Proceedings of the Combustion Institute 42 (2026) 106664

Contents lists available at ScienceDirect

Proceedings of the Combustion Institute

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

Uncovering flame physics with machine learning: Application to the reaction rate in hydrogen flames

Antonio Attili [a,*], Ludovico Nista [b], Tommaso Baffetti [b], Geveen Arumapperuma [a], Sofiane Al Kassar [a], Lukas Berger [c,d], Christoph D.K. Schumann [e], Temistocle Grenga [f], Heinz Pitsch [b]

[a] School of Engineering, Institute for Multiscale Thermofluids, University of Edinburgh, Edinburgh, EH9 3FD, United Kingdom
[b] Institute for Combustion Technology, RWTH Aachen University, Aachen, 52056, Germany
[c] Thermo and Fluid Dynamics (FLOW), Faculty of Engineering, Vrije Universiteit Brussel, Brussels 1050, Belgium
[d] Brussels Institute for Thermal-Fluid Systems and Clean Energy (BRITE), VUB-ULB, Brussels 1050, Belgium
[e] Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, United Kingdom
[f] Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, SO17 1BJ, United Kingdom

ARTICLE INFO

Keywords:
Hydrogen flames
Convolutional Neural Networks
Source-term parametrisation
Machine learning

ABSTRACT

Convolutional Neural Networks (CNNs) are used as analytical tools to investigate the relationship between the progress variable field and the local chemical source term in lean premixed hydrogen flames. Rather than employing machine learning for modelling, CNNs are leveraged to analyse the physical information contained in spatially resolved fields from direct numerical simulations. CNNs are particularly well suited for this task because they exploit spatial correlations and multi-scale structures, allowing an assessment of how spatial organisation influences the source term. The analysis demonstrates that the progress variable based on water, C_{H_2O}, contains all the information required to accurately parametrise the chemical source term whereas C_{H_2} alone does not. The inclusion of the mixture fraction Z improves the accuracy of the latter but provides no significant additional information to the CNN when the spatial field of C_{H2O} is used as input. The same behaviour is observed in both a laminar thermodiffusively unstable flame and a turbulent slot-jet hydrogen flame, indicating that C_{H_2O} is a robust single variable for parametrisation. A complementary scale analysis in the laminar case shows that spatial features extending over at least two laminar flame thicknesses are required to reconstruct the source term accurately, thereby identifying the characteristic scale in the progress variable field that carries this information. These results demonstrate how machine learning can uncover physical dependencies that remain hidden to classical statistical analyses.

Novelty and significance statement: This work introduces Convolutional Neural Networks (CNNs) as analysis tools to investigate the physical information embedded in flame fields, rather than as modelling instruments. Exploiting their ability to extract spatial correlations from high-fidelity data, CNNs are used to test whether the progress variable field alone contains sufficient information to describe the local chemical source term in lean premixed hydrogen flames. This provides a novel machine-learning framework for analysing combustion physics and source-term parametrisation. The results show that the progress variable based on water contains all the information required to accurately parametrise the source term in both laminar and turbulent flames, whereas the hydrogen-based progress variable does not. In addition, a minimum spatial extent of about two flame thicknesses is identified as the relevant scale for accurate parametrisation. These findings highlight machine learning as a powerful diagnostic tool complementary to classical statistical analyses.

1. Introduction

Lean premixed hydrogen flames are highly susceptible to thermodiffusive effects. Under laminar conditions, they exhibit complex spatial patterns and can experience several-fold increases in flame speed [1–3], while in turbulent regimes these effects interact synergistically with the flow field, producing even stronger departures from classical premixed-flame behaviour [4–6]. A direct consequence is that the local source term and local burning rate become strongly non-uniform, with

* Corresponding author.
E-mail address: antonio.attili@ed.ac.uk (A. Attili).

https://doi.org/10.1016/j.proci.2026.106664
Received 15 March 2026; Accepted 12 June 2026
Available online 16 October 2026
1540-7489/© 2026 The Authors. Published by Elsevier Inc. on behalf of The Combustion Institute. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

pronounced variations along the flame surface. This has major implications for modelling, since a parametrisation of the chemical source term is often required in combustion models. Understanding the spatial variability of the chemical source term and identifying the variables that control it are therefore important for both physical interpretation and the development of reliable reduced-order models. It is well known that parametrising the source term in hydrogen flames in terms of local thermochemical quantities requires at least two variables [7–9]. This contrasts with methane flames [10,11], for which the source term can often be described accurately as a function of a single progress variable C. Several choices for the second control variable have been proposed, with slightly different levels of accuracy. A widely used and effective option is the mixture fraction Z [7], while a second progress variable also provides good performance [12].

An alternative possibility is to replace the second variable with spatial information derived directly from the progress variable field itself [9]. This idea is motivated by the known correlation between chemical source term and local progress-variable curvature [13], since curvature contains partial information about the local spatial organisation of the flame. Although this approach improves the parametrisation, its accuracy remains limited [2,9], suggesting that the relevant spatial features extend beyond those captured by purely local geometric quantities. The objectives of this paper are therefore:

(i) to determine whether the progress variable field alone contains sufficient information to describe the source term
(ii) to compare different progress-variable definitions, such as those based on H_2 and H_2O
(iii) to identify the minimum spatial extent required for accurate parametrisation.

Several data-analysis techniques can be used to extract information from reacting-flow datasets. Proper orthogonal decomposition (POD) [14], for example, identifies dominant energetic modes and provides a low-dimensional representation of the flow field, while correlation-based approaches and conditional statistics can be used to quantify dependencies between selected variables. These methods have provided substantial insight into turbulent and reacting flows. However, they generally require either a linear decomposition, a predefined set of conditioning variables, or an explicitly chosen form of the correlation to be analysed. In principle, conditional statistics could also be extended to include non-local information by conditioning on the values of a scalar field at several neighbouring points. In practice, however, this would lead to a very high-dimensional conditional space, requiring prohibitively large amounts of data to obtain converged statistics. As a consequence, these approaches are less suited to determining whether a spatially distributed scalar field contains sufficient information to reconstruct a target quantity through nonlinear and non-local dependencies.

To address these objectives, machine learning is employed in this work. Differently from its more common use as a modelling tool, it is used here as a means of analysing the underlying flame physics. This represents a novel use of deep learning, where neural networks are used to reveal and quantify physical relationships rather than to construct closure models. In the present case, the questions of interest cannot be readily formulated in terms of simple statistical quantities, while a direct assessment through conventional statistical analyses would require prohibitively large datasets. Machine learning offers a practical alternative, enabling the exploration of complex high-dimensional relationships and providing direct access to the information content embedded in the spatial structure of the fields.

2. Datasets used in the analysis

Data from two direct numerical simulations (DNS) of laminar and turbulent lean premixed hydrogen flames are analysed in this work.

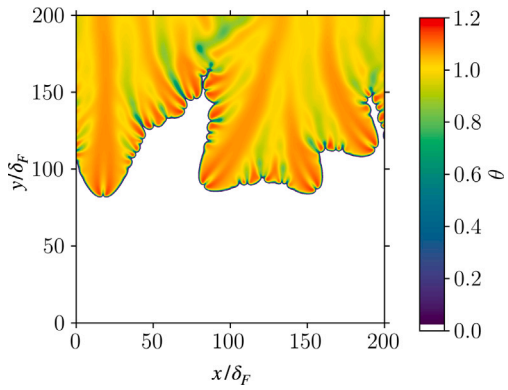


Fig. 1. Normalised temperature field in the laminar unstable hydrogen flame.

For both cases, the species, energy, and Navier–Stokes equations in the low-Mach number limit are solved using a semi-implicit finite-difference method [15], which has been extensively applied in a variety of reacting-flow configurations [4,11,16,17]. Spatial derivatives are discretised using second-order finite differences for the momentum equation and scalar diffusive terms, while a weighted essentially non-oscillatory (WENO) scheme [18] is employed for the convective terms in the scalar equations. The chemical source terms are treated through operator splitting and integrated using the stiff ordinary differential equation solver CVODE [19].

The first configuration is a two-dimensional laminar flame freely propagating in a square domain sufficiently large to capture the fully developed nonlinear regime of the thermodiffusive instability [2]. The equivalence ratio is $\phi = 0.4$, the pressure is atmospheric, and the unburned-gas temperature is $T_u = 298$ K. In this configuration, the thermodiffusive instability develops naturally from the initial perturbations, altering the flame shape and propagation speed and producing strong source-term variations along the flame surface. This configuration has been used extensively in previous studies [2,20–22] and is recognised as a convenient canonical case for investigating the fundamental physics of thermodiffusively unstable hydrogen flames and testing modelling hypotheses. The domain is discretised with a uniform mesh of resolution Δ, such that $\delta_f/\Delta \approx 10$, where δ_f is the thermal flame thickness, defined as $\delta_f = (T_b - T_u)/\|\nabla T\|_{\max}$, with T_b and T_u the burned- and unburned-gas temperatures. The normalised temperature field for this flame is shown in Fig. 1.

The second configuration is a turbulent jet premixed flame surrounded by a coflow of burned gases [4,23]. The jet consists of a hydrogen-air mixture with equivalence ratio $\phi = 0.4$ at ambient conditions ($p = 1$ atm, $T_u = 298$ K), with Reynolds number $Re = UH/\nu = 11\,200$. Based on the inlet velocity, jet width, and laminar flame speed, the nominal Karlovitz number is in the range $Ka = 20$–40, where $Ka = \delta_f^2/\eta^2$ and $\eta = (\bar{\nu}^3/\tilde{\varepsilon})^{1/4}$ is the Kolmogorov scale, with $\tilde{\varepsilon}$ the Favre-averaged energy dissipation and $\bar{\nu} = \overline{\mu/\rho}$ the ensemble-averaged viscosity. The domain is periodic in the spanwise direction (z), with open outlet conditions in the streamwise direction (x) and slip conditions imposed at the lateral boundaries (y). Turbulent inlet fluctuations are generated from an auxiliary simulation of fully developed channel flow. A uniform mesh is used in all three directions, with $\Delta/\eta \approx 1$ and $\delta_f/\Delta \approx 10$. The total number of grid points is approximately 1.4 billion. A two-dimensional slice of the vorticity field for the slot-burner configuration is shown in Fig. 2.

3. Local parametrisation of chemical source term in premixed hydrogen flames

The need for two local quantities to parametrise the source term in hydrogen flames [7–9] is illustrated in Fig. 3, which reports data



Fig. 2. Two-dimensional slice of the vorticity field in the turbulent hydrogen flame. The red line shows an isocontour of the hydrogen mass fraction, identifying the flame surface.

from the turbulent flame DNS introduced in Section 2. The figure shows local parametrisations based on both progress-variable definitions considered in this work, C_{H2} and C_{H2O}, either alone or in combination with the mixture fraction Z. The scatter plots in the space of progress variable C, mixture fraction Z, and source term reveal well-defined two-dimensional manifolds, indicating that the parametrisation in terms of (C, Z) is appropriate for both progress-variable definitions. However, the source term varies significantly at fixed C as Z changes, demonstrating that a parametrisation based solely on local values of the progress variable is inadequate.

A quantitative assessment is provided by the irreducible error associated with parametrisations based on C alone and on (C, Z) [8]. The irreducible error measures the residual variance of the target quantity around its conditional mean for a given set of parameters and therefore represents the minimum error achievable with that parametrisation [2, 24]. As shown in Fig. 3, the irreducible error remains large when only C_{H2} or only C_{H2O} is used, whereas it is strongly reduced when Z is included. This confirms that, for both progress-variable definitions, the local value of C alone is not sufficient to accurately parametrise the source term, and that mixture fraction is needed when a purely local parametrisation is considered.

One of the central questions addressed in the present work is whether this limitation of local parametrisations can be overcome by exploiting the spatial structure of the progress variable field. The underlying idea is that spatial features may contain additional information about the local flame state, allowing the source term to be reconstructed without explicitly introducing a second thermochemical variable.

4. CNNs as analysis tools

Convolutional Neural Networks (CNNs) are employed here as a data-driven framework to examine the relationship between the spatial structure of the progress variable field and the local chemical source term in hydrogen flames. CNNs are particularly well suited for analysing spatially structured data, as they exploit local correlations through convolutional filters that progressively extract multi-scale features from one or more input fields. Among the available architectures, the U-Net structure [25] is especially effective for field-to-field regression tasks and has already demonstrated excellent performance in combustion applications [26–28] and it will be used in the analysis presented. Its encoder–decoder structure, combined with skip connections, allows local and global information to be processed simultaneously: as the receptive field increases through the network, progressively more complex features are learned while preserving spatial detail. The U-Net employed here follows the same architecture and training strategy as in Ref. [28]. It is a fully convolutional encoder–decoder network composed of downsampling and upsampling blocks connected by skip connections. Each downsampling block applies convolutional operations, batch normalisation, and nonlinear activation, while the upsampling path reconstructs the output field at the original resolution. The network is trained in a supervised patch-to-patch manner: spatial blocks are extracted from the DNS fields, the selected input quantities are provided to the CNN, and the corresponding DNS source-term fields are used as targets. The trainable parameters are optimised by minimising a mean-squared-error loss, while an independent validation dataset is used during training to monitor convergence and possible overfitting. The final accuracy is then evaluated on test fields that are not used during the optimisation procedure.

The network's performance, quantified through the reconstruction error between predicted and DNS source terms, is interpreted as a quantitative indicator of the physical information contained in the input fields. A low reconstruction error indicates that the selected inputs contain sufficient information to determine the source term, whereas larger errors reveal missing information in the parametrisation. To assess this systematically, the CNN is trained and evaluated using different combinations of the input fields shown in Fig. 4. First, only the progress variable field is used for the CNN training, in order to test whether its spatial structure alone is sufficient to reconstruct the source term. The mixture fraction Z is then added as a second input, and different progress-variable definitions are considered. Finally, the network is trained using progressively larger sub-domains of the progress variable field, making it possible to quantify how spatial extent affects predictive accuracy and to identify the minimum characteristic length scale below which the parametrisation deteriorates. Other combinations of inputs could be considered, for example two different progress variables or additional species fields. However, the main objective of the present analysis is to determine whether a single suitably chosen progress variable can provide sufficient information to infer the chemical source term. From this perspective, testing many additional combinations would not provide significant additional insight into the central question addressed in the paper.

For selected cases, the analysis was also repeated using a different neural-network architecture, a Generative Adversarial Network (GAN) [29,30], in order to verify that the observed trends were not specific to the chosen machine-learning framework. In this approach, the generator network employs the same architecture as the CNN described above (a U-Net structure), while the discriminator is a deconvolutional network [29,30].

5. Flame analysis leveraging CNNs

5.1. Parametrisation of source term

The first step of the analysis consists in assessing how accurately different input fields allow the CNN to reconstruct the local source term, with the objective of identifying which quantities contain the information required for an effective parametrisation. The analysis is first carried out for the laminar flame introduced in Section 2. Four input combinations are considered (see Fig. 4):

- C_{H_2O}
- C_{H_2O}, Z
- C_{H_2}
- C_{H_2}, Z

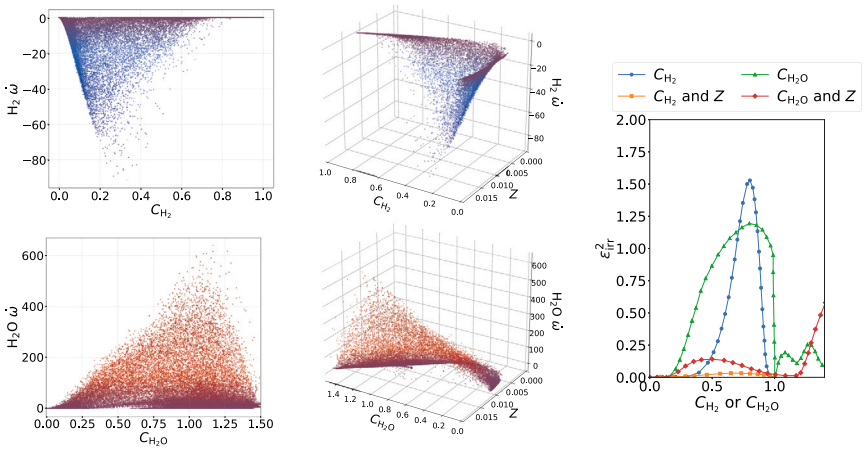


Fig. 3. Classical parametrisation of chemical source term in a lean, premixed, turbulent hydrogen flames using local variables only. Left: Manifolds in the space of progress variable (based on the hydrogen and water mass fraction) and source term. Centre: Manifolds in the space of progress variable, mixture fraction (based on the Bilger definition [2]), and source term. Right: Irreducible error analysis.

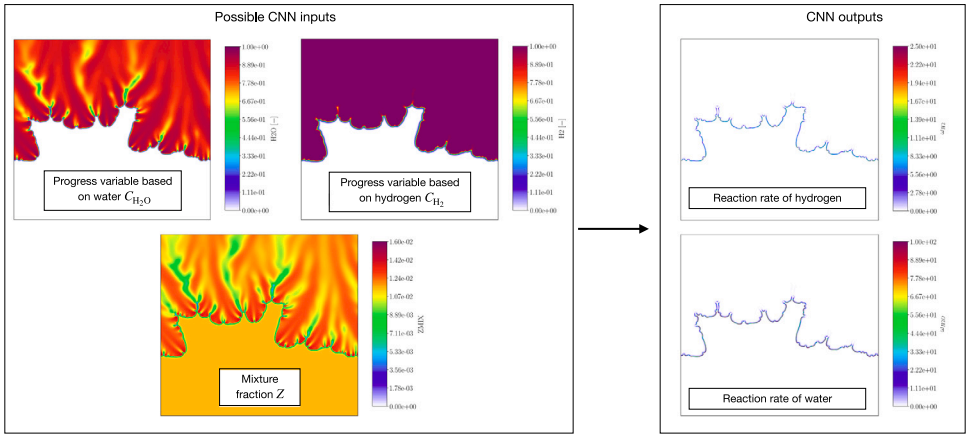


Fig. 4. Input and output fields considered for the different CNN trainings and performance evaluations in the unstable laminar hydrogen flame. The reaction rates are normalised with the peak values in the corresponding one-dimensional unstretched laminar flame.

For each case, an independent network is trained using the same architecture and optimisation procedure, and performance is quantified through the reconstruction error between the predicted and DNS source-term fields.

It is useful to note that the two progress-variable definitions have different monotonicity and boundedness properties. The progress variable based on hydrogen, C_{H2}, is monotonic and remains bounded between 0 and 1. This property is one of the reasons why H_2-based definitions are commonly used in classical flamelet models [8], where the progress variable must provide a well-posed coordinate for tabulation. Conversely, the progress variable based on water, C_{H2O}, is not strictly monotonic and may locally exceed unity. These features would be problematic in a purely local flamelet parametrisation, both numerically and because they may lead to a non-unique or ill-posed mapping between the progress variable and the thermochemical state. In the present work, however, the CNN is used primarily as an analysis tool rather than as a combustion model, and monotonicity and strict boundedness are therefore not required. The CNN does not rely on the input field as a monotonic and bounded coordinate of a tabulated manifold, but instead infers the reaction rate from the local spatial distribution of the input field. For this reason, the non-monotonicity of C_{H2O} and its local excursions above unity do not prevent the CNN from extracting the information needed to predict the source term. Moreover, even if CNNs were used as combustion models, the lack of monotonicity or strict boundedness of the input field would not represent the same type of constraint as in classical flamelet models. The CNN learns a mapping from the spatial distribution of the input field to the reaction-rate output. Therefore, the network can exploit the spatial structure of the H_2O field, even if the value of C_{H2O} is not monotonic and is not everywhere bounded by unity.

Fig. 5 reports the results in the form of joint probability density functions (jPDFs) between CNN predictions and corresponding DNS values for the four input combinations. The closer the probability density is concentrated along the diagonal, the more accurate the reconstruction of the source-term field. Clear differences emerge among the different inputs. When the progress variable based on water, C_{H_2O}, is used as the sole input, the agreement is nearly perfect, with only negligible deviations from the diagonal. Adding the mixture fraction Z does not produce any visible improvement, indicating that the spatial field of C_{H2O} already contains the dominant information required to reconstruct the source term.

A markedly different behaviour is observed when the progress variable based on hydrogen, C_{H_2}, is used. In this case, the jPDF exhibits a broader spread around the diagonal, showing that the source term cannot be reconstructed with the same level of accuracy. This indicates that C_{H_2} does not fully encode the information needed for parametrisation of its source term when considered alone. However, when the mixture fraction Z is added as a second input, the reconstruction improves substantially and the agreement returns to a level comparable to that obtained with C_{H_2O}.

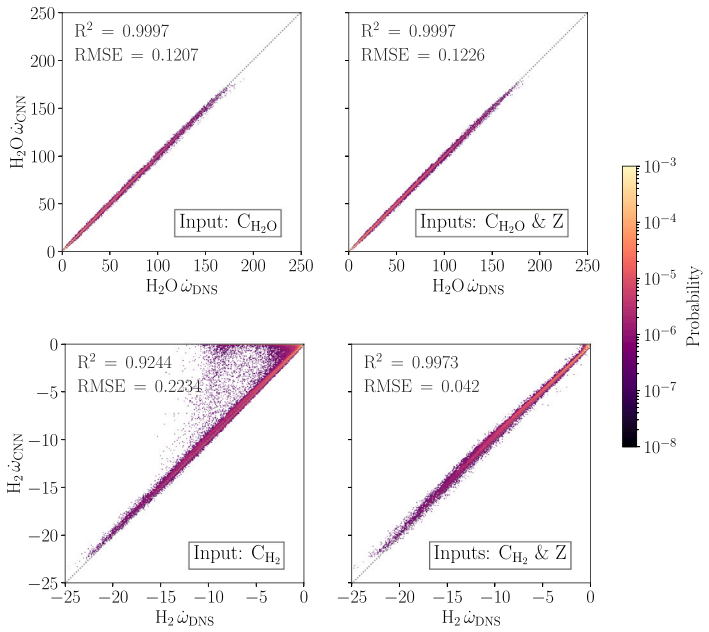


Fig. 5. Parametrisation of chemical source term in a laminar unstable flame using CNNs and different combinations of inputs. jPDFs of CNN predictions and ground truth DNS values for different combinations of inputs. Correlation coefficient and root mean square errors are also shown in each plot.

The superior performance of C_{H_2O} can be understood by considering its spatial structure within the flame. As shown in Fig. 4, the C_{H_2O} field preserves a richer representation of flame features. In addition, C_{H_2O} exhibits a stronger correlation with Z, meaning that part of the information normally provided by Z is already embedded in the spatial distribution of C_{H_2O}. By contrast, C_{H_2} misses part of this information, making the explicit inclusion of Z necessary to recover an accurate parametrisation.

It is important to emphasise that this result does not imply a simple local correlation of the form $\dot{\omega} = f(C_{H2O})$. The input to the CNN is not only the pointwise value of the progress variable, but its spatial distribution over a finite region. The correlation learned by the network is therefore non-local and high-dimensional, involving the value of C_{H2O} at the target point as well as in the surrounding flame structure. This distinction resolves the apparent contrast with the local parametrisation discussed in Section 3. In a purely local description, the source term is expressed as a function of a small number of pointwise variables, and the mixture fraction is required because the local value of the progress variable alone does not uniquely determine the thermochemical state. In the CNN analysis, instead, the surrounding spatial field provides additional information on the flame structure, transport history, and mixing state.

It is also worth noting that reproducing this analysis with conditional statistics would require conditioning on the values of C_{H2O} at the target point and at many neighbouring points, leading to an extremely high-dimensional conditional space that is not practically accessible. A comparison with a strictly local machine-learning model, such as a multilayer perceptron or a 1×1 CNN, would provide another way of testing the role of spatial information. However, such models use only pointwise input values and are therefore subject to the same information limitation as the local parametrisations shown in Fig. 3. Their best achievable performance is bounded by the irreducible error associated with the corresponding local conditional mean [24]. Since this error remains large when only the local value of C_{H2} or C_{H2O} is used, the improved performance obtained by the CNN with C_{H2O} must originate from the spatial information contained in the surrounding.

Additional tests were performed to verify that these conclusions are not restricted to the prediction of the source term associated with the species used to define the progress variable. These results are not shown here because they do not provide additional insight beyond the trends discussed above. In particular, using C_{H2O} as input to predict the H_2 source term leads to the same conclusions: the prediction remains accurate and the addition of mixture fraction does not significantly improve the result. Conversely, when the H_2-based progress variable is used to parametrise other quantities, the prediction quality deteriorates unless mixture fraction is also provided as an input.

To verify that these conclusions are not specific to the CNN architecture, selected laminar cases were also analysed using a Generative Adversarial Network (GAN) [29,30]. GANs differ from CNN training as they involve the joint training of a generator and a discriminator network. The resulting adversarial optimisation introduces learning dynamics that differ from standard regression based solely on predefined loss functions, and can lead to different reconstruction characteristics. Fig. 6 shows the corresponding jPDFs for the cases where C_{H_2O}, (C_{H_2O}, Z), and C_{H_2} are used as input. The GAN reproduces the same trends observed with the CNN: an almost perfect reconstruction is obtained when C_{H_2O} is used, the addition of Z does not produce any noticeable improvement, and a significantly lower accuracy is observed when only C_{H_2} is provided. The level of discrepancy relative to the DNS data remains essentially unchanged, indicating that both the superior performance of C_{H_2O} and the negligible contribution of Z are not influenced by the specific network or learning strategy used. This provides additional confidence that the observed hierarchy among the different parametrisations reflects intrinsic physical relationships between the input fields and the source term, rather than artefacts introduced by the machine-learning architecture employed.

The same analysis is then extended to the turbulent slot-jet hydrogen flame introduced in Section 2 [4]. This extension is important because, unlike the laminar case, turbulence introduces a broad range of interacting spatial scales, additional flame wrinkling, and enhanced fluctuations of the local thermochemical state, all of which could in principle modify the information required for an accurate parametrisation of the source term. Fig. 7 compares the jPDFs between CNN predictions and DNS source-term data obtained using C_{H_2O} alone and using (C_{H_2O}, Z) as input. In both cases, the agreement remains very good, with the probability density strongly concentrated along the diagonal. The reconstruction is slightly less accurate than in the laminar flame, which likely reflects the increased complexity of the local flame dynamics under turbulent strain. However, the two jPDFs remain essentially indistinguishable, indicating that the inclusion of Z does

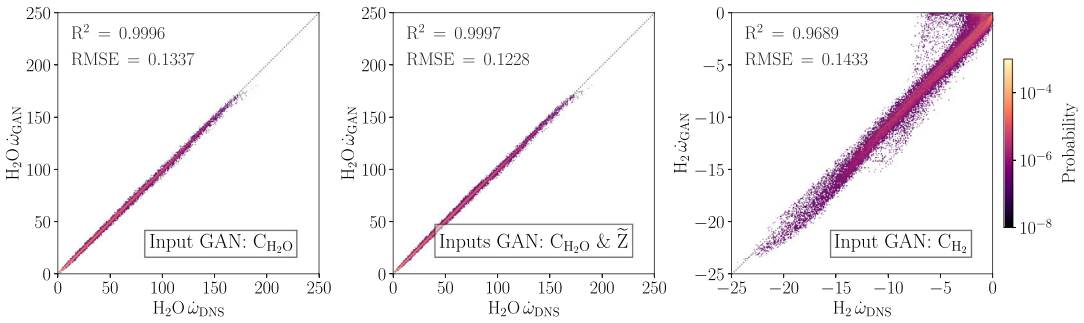


Fig. 6. Results obtained using a Generative Adversarial Network (GAN) [29,30] for source-term reconstruction in the laminar flame, included to assess the robustness of the conclusions with respect to the machine-learning architecture employed (CNN versus GAN).

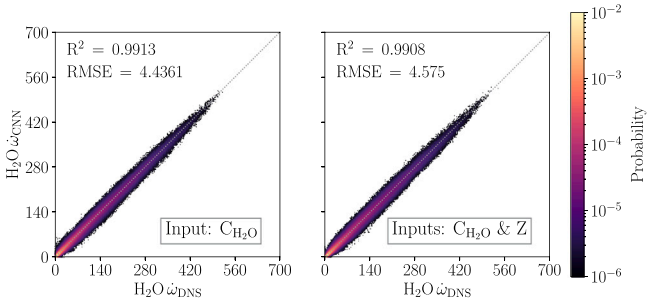


Fig. 7. Parametrisation of chemical source term in a turbulent flame using CNNs with progress variable only (left) and with the combination of progress variable and mixture fraction (right).

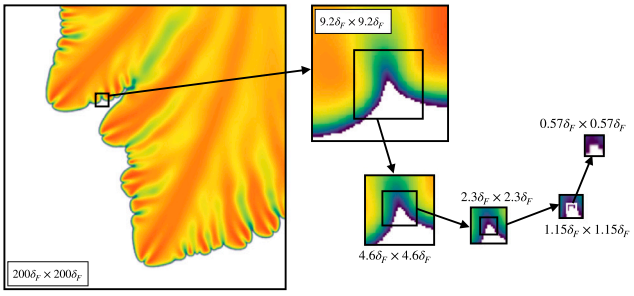


Fig. 8. Scale analysis of the chemical source term. Different box sizes used to train the different CNN models.

not provide any significant improvement in predictive accuracy. This demonstrates that the information embedded in the spatial structure of C_{H_2O} remains sufficient to reconstruct the local source term even under fully turbulent conditions. Although turbulence introduces additional variability, the information carried by C_{H_2O} alone is equivalent to that contained in (C_{H_2O}, Z), showing that the dominant physical relationship identified in the laminar flame is preserved. This confirms that C_{H_2O} acts as a robust single variable for source-term parametrisation across markedly different flame regimes.

These observations highlight the unique capability of machine-learning-based analysis to extract physical relationships directly from data, revealing dependencies that are difficult to access through classical statistical approaches, such as the complete information content of C_{H_2O} for describing the local source term.

5.2. Scale analysis of the chemical source term in the laminar flame

Having established, for the laminar flame, that C_{H_2O} contains all the information required to describe the source term, the next step is to identify the characteristic length scale of the spatial features that encode this information. A series of trainings was then performed for the same CNN architecture using input data in sub-domains ("boxes") of different sizes (SB), ranging from approximately 0.2 to $10\,\delta_F$, where δ_F is the laminar flame thermal thickness. Fig. 8 illustrates examples of these boxes superimposed on the flame field. By analysing the performance of the network for different sizes of the input field, it is possible to identify the minimum size required for a good parametrisation.

Fig. 9 reports selected jPDFs obtained for different box sizes. A clear transition is observed: for box sizes larger than approximately $2\,\delta_F$, the predictions remain highly accurate, with probability density concentrated close to the diagonal, whereas for smaller boxes the reconstruction quality deteriorates rapidly. This indicates that the source term cannot be determined from purely local information and that spatial features extending over at least two flame thicknesses are required to recover the relevant physical content embedded in the progress-variable field.

A more quantitative view is provided in Fig. 10, which shows the reconstruction error as a function of box size. The error remains low for large boxes but increases sharply below approximately $2\,\delta_F$, confirming that this scale represents a lower bound for accurate parametrisation. This characteristic scale is consistent with the spatial extent over which thermodiffusive structures develop along the flame front [17].

To ensure that this behaviour is not affected by numerical resolution, the analysis was repeated using data interpolated onto progressively finer meshes before training. This procedure guarantees that even the smallest boxes contain a sufficient number of grid points and avoids artefacts associated with very limited spatial sampling. The consistency of the resulting curves confirms that the observed transition is robust. For the finest interpolated dataset, a degradation of prediction accuracy is observed for the largest box sizes. This behaviour is attributed to convergence difficulties during CNN training, a known limitation when the number of input points becomes very large [31]. Since this effect is absent in the other cases, it does not affect the overall interpretation of the results.

Overall, the analysis shows that the progress-variable field must be sampled over a spatial extent of at least two laminar flame thicknesses to contain sufficient information for accurate reconstruction of the local source term. Extending the same scale analysis to turbulent flames would require a substantially larger database including different Reynolds and Karlovitz numbers, in order to separate the influence of turbulence from that of thermochemical structure. This lies beyond the scope of the present work and is left for future investigation.

6. Conclusions

This work employed a CNN-based architecture, the U-Net framework, as analytical tools to investigate the information content of progress-variable fields in lean premixed hydrogen flames and their relation to the local source term. Unlike conventional applications of machine learning aimed at prediction or reduced-order modelling, the networks were used here as diagnostic instruments to quantify

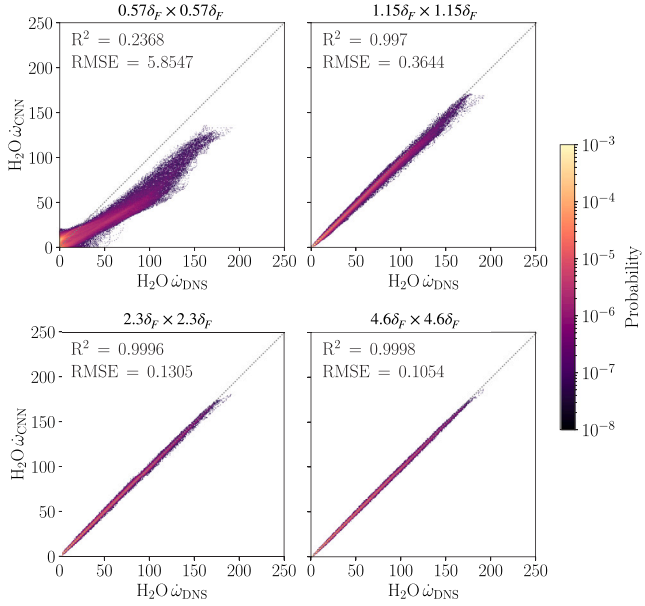


Fig. 9. Scale analysis of the source term. jPDF of CNN predictions and ground truth DNS for various box sizes.

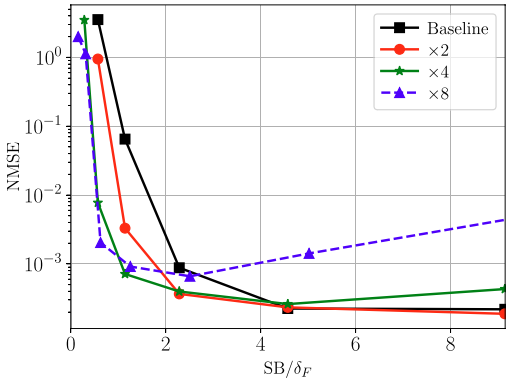


Fig. 10. Scale analysis of the source term. Error in the parametrisation for different box sizes. Different lines are for CNNs trained with data interpolated on increasingly finer meshes before training.

whether selected input fields contain sufficient physical information to reconstruct the source-term distribution.

The analysis showed that the progress variable based on water, C_{H_2O}, contains all the information required to accurately parametrise the source term in both laminar thermodiffusively unstable and turbulent slot-jet hydrogen flames. By contrast, the progress variable based on hydrogen, C_{H_2}, does not provide sufficient information when used alone, and requires the addition of the mixture fraction Z to recover comparable accuracy. These results indicate that the spatial structure of C_{H_2O} inherently embeds the information usually supplied by a second thermochemical variable. A complementary scale analysis performed for the laminar flame demonstrated that spatial features extending over at least two laminar flame thicknesses are necessary for accurate reconstruction of the source term. This identifies a characteristic minimum length scale over which the relevant physical information is distributed, showing that the source term cannot be determined from purely local values of the progress variable even when the optimal definition is used.

Overall, the study demonstrates that machine-learning-based analysis can reveal physically meaningful relationships directly from high-fidelity combustion data, uncovering dependencies that are difficult to access through classical statistical approaches. The methodology introduced here offers a general framework that can be extended to other combustion problems where the information content of scalar fields and the identification of relevant spatial scales remain open questions.

CRediT authorship contribution statement

Antonio Attili: Designed research, Performed analysis, Writing – original draft, Supervision. **Ludovico Nista:** Performed analysis, Discussion, Writing – original draft, Code implementation. **Tommaso Baffetti:** Performed analysis, Discussion, Code implementation. **Geveen Arumapperuma:** Performed analysis, Discussion, Writing – review & editing. **Sofiane Al Kassar:** Discussion, Writing – review & editing. **Lukas Berger:** Discussion, Writing – review & editing. **Christoph D.K. Schumann:** Discussion, Writing – review & editing. **Temistocle Grenga:** Supervision, Discussion, Writing – review & editing. **Heinz Pitsch:** Supervision, Discussion, Writing – review & editing.

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.

The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for this journal and was not involved in the editorial review or the decision to publish this article.

Acknowledgements

L.N. and H.P acknowledge the German Federal Ministry of Education and Research (BMBF) and the state of North Rhine-Westphalia for supporting this work as part of the NHR funding and the European Union under the European Research Council Advanced Grant HYDROGENATE, Grant Agreement No. 101054894. This work used the ARCHER2 UK National Supercomputing Service via the UK Turbulence Consortium (EP/X035484/1).

References

[1] L. Berger, A. Attili, H. Pitsch, Intrinsic instabilities in premixed hydrogen flames: Parametric variation of pressure, equivalence ratio, and temperature. part 1-dispersion relations in the linear regime, Combust. Flame 240 (2022) 111935.
[2] L. Berger, A. Attili, H. Pitsch, Intrinsic instabilities in premixed hydrogen flames: parametric variation of pressure, equivalence ratio, and temperature. Part 2–non-linear regime and flame speed enhancement, Combust. Flame 240 (2022) 111936.
[3] S. Al Kassar, L. Berger, P.E. Lapenna, F. Creta, H. Pitsch, A. Attili, Efficient and accurate calculation of dispersion relations for intrinsically unstable premixed flames, CF 269 (2024) 113640.
[4] L. Berger, A. Attili, H. Pitsch, Synergistic interactions of thermodiffusive instabilities and turbulence in lean hydrogen flames, Combust. Flame 244 (2022) 112254.
[5] T. Howarth, E. Hunt, A. Aspden, Thermodiffusively-unstable lean premixed hydrogen flames: Phenomenology, empirical modelling, and thermal leading points, CF 253 (2023) 112811.
[6] L. Berger, A. Attili, M. Gauding, H. Pitsch, Effects of karlovitz number variations on thermodiffusive instabilities in lean turbulent hydrogen jet flames, Proc. Combust. Inst. 40 (1–4) (2024) 105219.
[7] J.D. Regele, E. Knudsen, H. Pitsch, G. Blanquart, A two-equation model for non-unity lewis number differential diffusion in lean premixed laminar flames, Combust. Flame 160 (2) (2013) 240–250.
[8] L. Berger, A. Attili, M. Gauding, H. Pitsch, LES combustion model for premixed turbulent hydrogen flames with thermodiffusive instabilities: a priori and a posteriori analysis, J. Fluid Mech. 1003 (2025) A33.
[9] H. Böttler, D. Kaddar, T.J.P. Karpowski, F. Ferraro, A. Scholtissek, H. Nicolai, C. Hasse, Can flamelet manifolds capture the interactions of thermo-diffusive instabilities and turbulence in lean hydrogen flames?—an a-priori analysis, Int. J. Hydrog. Energy 56 (2024) 1397–1407.
[10] H. Pitsch, Large-eddy simulation of turbulent combustion, Annu. Rev. Fluid Mech. 38 (2006) 453–482.

[11] A. Attili, S. Luca, D. Denker, F. Bisetti, H. Pitsch, Turbulent flame speed and reaction layer thickening in premixed jet flames at constant karlovitz and increasing Reynolds numbers, Proc. Combust. Inst. 38 (2) (2021) 2939–2947.
[12] A. Remiddi, P.E. Lapenna, D. Cavalieri, D. Schintu, G. Indelicato, A. Attili, L. Berger, H. Pitsch, F. Creta, Data-driven modeling of resolved and filtered thermo-diffusively unstable hydrogen–air flames, Proc. Combust. Inst. 40 (1–4) (2024) 105713.
[13] H. Olguin, P. Domingo, L. Vervisch, C. Hasse, A. Scholtissek, On the closure of curvature in 2D flamelet theory, Combust. Flame 267 (2024) 113599.
[14] K. Bizon, G. Continillo, E. Mancaruso, S. Merola, B. Vaglieco, POD-based analysis of combustion images in optically accessible engines, Combust. Flame 157 (4) (2010) 632–640.
[15] O. Desjardins, G. Blanquart, G. Balarac, H. Pitsch, High order conservative finite difference scheme for variable density low mach number turbulent flows, JCP 227 (2008) 7125–7159.
[16] A. Attili, F. Bisetti, M.E. Mueller, H. Pitsch, Effects of non-unity lewis number of gas-phase species in turbulent nonpremixed sooting flames, 166, 2016, pp. 192–202.
[17] L. Berger, K. Kleinheinz, A. Attili, H. Pitsch, Characteristic patterns of thermodiffusively unstable premixed lean hydrogen flames, PCI 37 (2) (2019) 1879–1886.
[18] X. Liu, S. Osher, T. Chan, Weighted essentially non-oscillatory schemes, JCP 115 (1) (1994) 200–212.
[19] A. Hindmarsh, P. Brown, K. Grant, S. Lee, R. Serban, D. Shumaker, C. Woodward, SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers, ACM Trans. Math. Software 31 (3) (2005) 363–396.
[20] M. Schneider, H. Nicolai, V. Schuh, M. Steinhausen, C. Hasse, Flame-wall interaction of thermodiffusively unstable hydrogen/air flames, part II: Parametric variations of equivalence ratio, temperature, and pressure, Combust. Flame 279 (2025) 114319.
[21] F. Creta, P.E. Lapenna, R. Lamioni, N. Fogla, M. Matalon, Propagation of premixed flames in the presence of darrieus–Landau and thermal diffusive instabilities, Combust. Flame 216 (2020) 256–270.
[22] T. Howarth, A. Aspden, An empirical characteristic scaling model for freely-propagating lean premixed hydrogen flames, Combust. Flame 237 (2022) 111805.
[23] S.A. Kassar, S. Cantagalli, W. Lauder, G. Arumapperuma, A. Attili, Turbulent hydrogen premixed flames at high pressure and high temperature, 2026, arXiv preprint arXiv:2601.15848.
[24] L. Berger, K. Kleinheinz, A. Attili, F. Bisetti, H. Pitsch, M.E. Mueller, Numerically accurate computational techniques for optimal estimator analyses of multi-parameter models, Combust. Theory Model. 22 (3) (2018) 480–504.
[25] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241.
[26] C.J. Lapeyre, A. Misdariis, N. Cazard, D. Veynante, T. Poinsot, Training convolutional neural networks to estimate turbulent sub-grid scale reaction rates, Combust. Flame 203 (2019) 255–264.
[27] A. Seltz, P. Domingo, L. Vervisch, Z.M. Nikolaou, Direct mapping from LES resolved scales to filtered-flame generated manifolds using convolutional neural networks, Combust. Flame 210 (2019) 71–82.
[28] G. Arumapperuma, N. Sorace, M. Jansen, O. Bladek, L. Nista, S. Sakhare, L. Berger, H. Pitsch, T. Grenga, A. Attili, Extrapolation performance of convolutional neural network-based combustion models for large-eddy simulation: Influence of Reynolds number, filter kernel and filter size, Flow, Turbul. Combust. (2025) 1–30.
[29] L. Nista, C.D. Schumann, P. Petkov, V. Pavlov, T. Grenga, J.F. MacArt, A. Attili, S. Markov, H. Pitsch, Parallel implementation and performance of super-resolution generative adversarial network turbulence models for large-eddy simulation, Comput. & Fluids 288 (2025) 106498.
[30] L. Nista, H. Pitsch, C.D.K. Schumann, M. Bode, T. Grenga, J.F. MacArt, A. Attili, Influence of adversarial training on super-resolution turbulence reconstruction, Phys. Review Fluids 9 (2024) 064601.
[31] Y. Bengio, Practical recommendations for gradient-based training of deep architectures, in: Neural Networks: Tricks of the Trade: Second Edition, Springer, 2012, pp. 437–478.