

## Perspective

## Data-driven models and digital twins for sustainable combustion technologies

Alessandro Parente<sup>1,2,3,\*</sup> and Nedunchezian Swaminathan<sup>4,\*</sup>

## SUMMARY

**We highlight the critical role of data in developing sustainable combustion technologies for industries requiring high-density and localized energy sources. Combustion systems are complex and difficult to predict, and high-fidelity simulations are out of reach for practical systems because of computational cost. Data-driven approaches and artificial intelligence offer promising solutions, enabling renewable synthetic fuels to meet decarbonization goals. We discuss open challenges associated with the availability and fidelity of data, physics-based numerical simulations, and machine learning, focusing on developing digital twins capable of mirroring the behavior of industrial combustion systems and continuously updating based on newly available information.**

## INTRODUCTION

The race for carbon neutrality by 2050 is underway.<sup>1</sup> The current paradigm of considering the energy consumption in various sectors such as transport, industry, domestic application, etc. individually will impede reaching climate neutrality.<sup>1</sup> A holistic view of an interconnected system linking different sectors with appropriate energy carriers and shared infrastructures needs to be developed. The core of such an approach is energy efficiency and electrification, which relies massively on renewable energies.<sup>2</sup> Whether this is feasible or not is a complex question with significant uncertainties.<sup>3</sup>

About 50% of energy needs in hard-to-abate industries can be met through electrification.<sup>4</sup> They pose significant challenges for reducing their carbon emissions. Heavy industries such as steel, cement, and aluminum rely on high-temperature heat (e.g., >1,300 K)<sup>5</sup> that currently requires combustion to be cost-effective. In addition, producing some chemicals and fertilizers, such as ammonia, relies on non-energy sources (i.e., natural gas) as feedstock and emits greenhouse gases. The complexity and integration of these industrial processes make it difficult to find carbon-free alternatives, and the significant investments for these alternatives can challenge their implementation for industries. In sectors such as heavy-duty transportation, electrification is not straightforward because of the requirement of high-density, localized energy sources that fuels can only provide. For example, aviation, shipping, and trucking require fuels that cannot be substituted with solar panels or batteries due to the high energy density required. The World Energy Outlook from International Energy Agency (IEA)<sup>6</sup> recently noted, "... there are end uses for which it is not feasible or cost-effective to electrify directly with the technologies currently available on the market or likely to be commercialized in coming decades. Around half of energy demand in energy-intensive industries is for high grade process heat (above 400°C), which is challenging to electrify with current technologies. Ships, planes, and heavy trucks make up around 85% of transport non-electricity demand in 2050; at present it is very difficult to see how most of the energy they use could be electrified," stressing this criticality.

Hard-to-abate sectors with challenges for electrification will need to implement a series of measures to meet the ambitious decarbonization goals. Top priorities<sup>7</sup> are represented by energy efficiency and demand side management, electrification of heat and processes up to the maximum achievable levels, carbon capture and utilization, and the use of carbon-neutral fuels, renewable synthetic fuels, or biofuels. In this context, hydrogen and hydrogen carriers like ammonia enjoy renewed and rapidly growing attention as the hard-to-abate industries emit nearly 30% of the global CO<sub>2</sub>.<sup>6</sup> Energy-intensive industries are exploring replacing fossil fuels with synthetic and renewable energy carriers entirely or in combination with conventional fuels (natural gas) and hybrid configurations involving other types of heating. Many fundamental questions and new challenges must be addressed while developing the critical approaches to achieve such technological advancements. Specifically, the impact of changing fuel/burner technology on the combustion processes in large industrial furnaces, the associated pollutants emission, the overall process efficiency, and the effects on the product characteristics still need to be fully understood. This knowledge is critical for implementing major technological changes to large industrial assets, for which experimental testing is neither comprehensive nor feasible at the scales of interest. Also, current computational tools do not offer the required fidelity and computational efficiency to evaluate potential scenarios to be explored for decarbonizing these industries.

<sup>1</sup>Aero-Thermo-Mechanics Department, École polytechnique de Bruxelles, Université libre de Bruxelles, Avenue Franklin D. Roosevelt 50, 1050 Brussels, Belgium

<sup>2</sup>WEL Research Institute, Avenue Pasteur 6, 1300 Wavre, Belgium

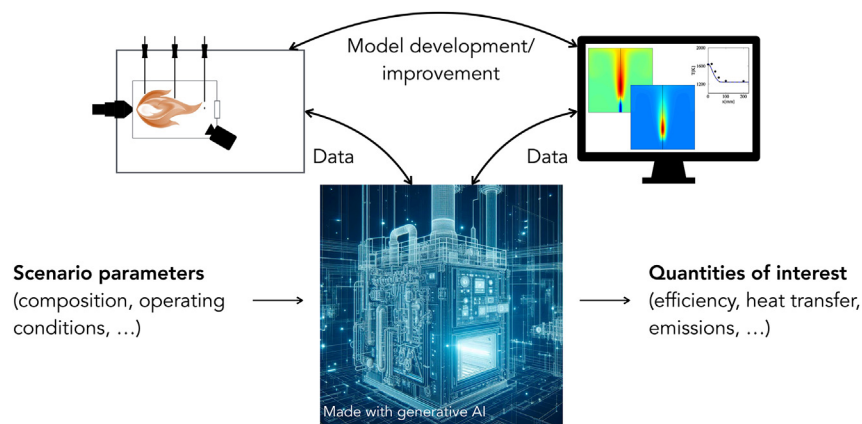
<sup>3</sup>Brussels Institute for Thermal-fluid systems and clean Energy (BRITE), Université libre de Bruxelles and Vrije Universiteit Brussel, 1050 Ixelles, Belgium

<sup>4</sup>Department of Engineering, Hopkinson Laboratory, Cambridge University, Cambridge CB2 1PZ, UK

\*Correspondence: [alessandro.parente@ulb.be](mailto:alessandro.parente@ulb.be) (A.P.), [ns341@cam.ac.uk](mailto:ns341@cam.ac.uk) (N.S.)

<https://doi.org/10.1016/j.isci.2024.109349>





**Figure 1. A schematic of a digital twin (DT) with data flow: cyber-physical infrastructure can drive decarbonization in hard-to-abate industries creating a link between physical assets and virtual models to assess reliably and affordably the impact of renewable synthetic fuels on the combustion processes, the associated pollutants emission, the overall process efficiency, and the effect on the characteristics of the products**

This perspective paper discusses the key developments in combustion science and its hybridization with other cutting-edge research topics required to develop a pervasive, physics-aware, data-intensive framework enabling sustainable combustion technologies and using renewable fuels for energy-intensive applications. The challenges and enabling technologies associated with deploying digital twins (DTs) for various stakeholders have been discussed by Rasheed et al.<sup>8</sup> This paper highlights some specific challenges associated with energy-intensive industries and key requirements from a modeling perspective. We focus on the opportunities offered by DTs and cyber-physical infrastructures by incorporating heterogeneous (both in type and accuracy) data streams into predictive models with well-defined confidence, highlighting the challenges and open research needs (Figure 1).

## THE ROLE OF DATA IN THE DEVELOPMENT OF SUSTAINABLE COMBUSTION TECHNOLOGIES

Combustion systems involve wide-ranging time and length scales, transitions, bifurcations, and multiple interacting physical phenomena that are highly complex to predict. While using turbulent combustion models has become common across industries, their current predictive capabilities do not yet meet the required standards for new design and regulation,<sup>9</sup> especially regarding pollutants and operational stability limits.<sup>10,11</sup> The advent of exascale computing has potentially unlocked the application of direct numerical simulations (DNSs) of combustion systems limited beyond specific aspects of turbulent combustion physics and building blocks. Still, not all relevant dimensionless numbers/conditions are numerically accessible. Moreover, exascale facilities are only available to a handful of researchers, and their availability for the industry and routine simulations is miles away.

DNSs produce vast data in some dimensions (spatial and temporal resolution) but very sparse in others (flow conditions, fuel composition, etc., to name a few); hence, generalizing the knowledge gained becomes challenging. Still, DNS data have rich information that can help decoding complex turbulence-chemistry interactions and guide the development of filtered and low-fidelity modeling approaches for quicker evaluations.<sup>12–14</sup> Indeed, combustion physics and its interaction with turbulence are small-scale phenomena that are expected to be quite universal, and, hence, the ML-enhanced models developed for these complex small-scale physics are expected to hold well across a wide range of conditions and problems.

Some recent attempts have been made to create databases and benchmark cases (e.g., the BLASTNet<sup>15</sup>). A recent review showed that more than 200 DNS studies on turbulent reacting flows over the last two decades were published during the last two decades.<sup>16</sup> However, very little of these data have become available to the broad combustion community. Challenges to be overcome are primarily related to standardization, both concerning the shared information and the codes used to generate the data. Efforts such as BLASTNet are great initiatives to define minimal requirements for datasets in terms of (meta)data required to associate them to a specific code and recreate specific labels for machine learning (ML) training. Handling code variability represents a different challenge and would require the inclusion of metrics to assess the quality of data. Finally, the typical size of DNS combustion data can impede data sharing, especially for combustion cases. This can be addressed by creating one or more centralized and maintained data storage facilities with easy access by any community member, ensuring that the data can be identified and searchable.

Similarly, high-fidelity measurements are available only for laboratory-scale systems, and the amount and accuracy of the measurements decrease going from small to large scales. Realistic combustion environments are typically very harsh, and sophisticated combustion diagnostics can hardly withstand them, limiting the diagnostic options and excluding the use of laser diagnostic techniques in most applications. Moreover, even with the most advanced diagnostic techniques, only a few variables of interest are accessible in industrial combustion systems, with limited spatial and temporal resolution. Differently to DNS data, the sharing practices for combustion experiments are well consolidated in the combustion community, and several workshops actively define benchmark cases for the simulation of different combustion

systems.<sup>17</sup> Velocity and scalar measurements are provided for the validation of numerical simulation approaches. However, while there is a tendency to define benchmark cases of increasing complexity and closer to applications of practical interests, available data remain fundamentally limited to laboratory-scale equipment. Recently, efforts have been made toward developing scale-bridging systems to fill the gap between laboratory units and real applications, particularly in the field of advanced combustion technologies enabling the combustion of renewable synthetic fuels.<sup>18–21</sup> While vital, these initiatives still require further development, especially concerning the standardization of the shared data and their attributes.

Designing a complex system is an iterative process requiring many model evaluations and extensive experimental testing. As such, it cannot rely solely on high-fidelity tools, admitting they are available, nor can it be carried out directly at the scale of interest. Indeed, neither experimental testing nor numerical simulations are self-sufficient, and a synergistic combination is required. A combination of approaches is required to bridge information from high- and low-fidelity models across scales and data types. Data-driven approaches and artificial intelligence have tremendous potential to enable a step change in combustion science and technology. A synergistic combination of experiments, simulations, and data-driven methods offers huge potential to address challenges of calibrating low-fidelity models, their parameterization, and the integration of heterogeneous data streams at different scales in calibration and model update processes. Specifically, research efforts to combine ML with physics-based knowledge are promising<sup>22</sup> to enable the pervasive use of data-driven methods and develop reliable and effective reduced-order models (ROMs) for capturing the underlying hidden relationship between inputs and quantities of interest.

### DTs OF INDUSTRIAL COMBUSTION SYSTEMS: WHERE DO WE STAND?

Combining simulations and measurements is key to developing DTs of large-scale combustion systems that can predict combustion evolution in real time and act as soft sensors. Indeed, extensive use of physical sensors in harsh environments like furnaces is infeasible, and deploying laser metrology for industrial combustion systems is difficult. Moreover, most sensors provide single-point information on a single parameter so that only a tiny portion of the system and a small set of physical/chemical parameters can be monitored. Gartner named DTs as one of the Top 10 Strategic Tech Trends 2018,<sup>23</sup> and they are now mainstream and considered an integral part of the Fourth Industrial Revolution. The quick adoption of DTs is mainly connected to the faster spread of enabling technologies: the cloud and internet of things (IoT).

Among many definitions, one can describe a DT as *an integrated multi-physics, multi-scale, probabilistic simulation of an as-built system enabled by a digital thread that uses the best available models, sensor information, and input data to mirror and predict activities/performance over the life of its corresponding physical twin*.<sup>24–26</sup> There are five components of a DT,<sup>27</sup> viz., (i) the ability to simulate the system functions over its life cycle, (ii) synchronizing the digital and physical systems, (iii) real-time data integration, (iv) modeling the entire physical state space, and (v) its service functionalities. Achieving these five components for industrial combustion systems is very challenging. These systems involve turbulence and combustion, which are still unsolved problems in classical physics and chemistry.<sup>9</sup> Their interactions take the challenges to a higher level, depending on the overlap between chemical and turbulent scales, resulting in many combustion regimes with fundamentally different reacting structures, morphologies, and topologies.<sup>3,10</sup>

Black-box approaches have been widely used in combustion to create static input-output maps<sup>28</sup> and system identification<sup>29</sup> to predict macroscopic quantities such as exhaust gas emissions and temperature and to detect oscillatory patterns such as thermoacoustic instabilities, respectively. In combustion, neural networks have been used to evaluate and tabulate reaction rates,<sup>30–33</sup> as well as to estimate modeling errors in large eddy simulations (LESs).<sup>12,34</sup> Black-box approaches can be very powerful, although not equipped with the guarantees of physics-based approaches. To cope with that, non-linear system identification techniques were proposed to enlighten black-box approaches by discovering the underlying physics.<sup>35</sup>

An interesting approach to generate combustion ROMs is the use of chemical reactor networks (CRNs),<sup>36–44</sup> which fall in the category of gray-box models, as they combine a theoretical structure, i.e., the canonical reactors, with data to generate the network. The approach can also be regarded as an example of multi-fidelity methods, being high-fidelity tools (i.e., computational fluid dynamics, CFD) used to construct the reactor network and the simplified network model used to evaluate pollutant emissions and other quantities, using detailed chemical mechanisms. These techniques are particularly useful for large and complex systems, such as furnaces or gas turbines. However, their current overall fidelity and generalizability are limited by the high-fidelity simulations required to generate the network structure, indicating an interest in data-driven approaches in this area of research.<sup>45</sup>

ROMs based on projection methods (such as principal-component analysis [PCA] and other techniques) have been employed for experimental and numerical combustion data.<sup>46–55</sup> The direct evolution of the modal coefficients in CFD codes (i.e., intrusive ROMs) has been quite limited for non-linear problems like combustion due to the high cost of the associated Galerkin projection and the complexity of coupling them to legacy and commercial CFD codes.<sup>56</sup> The application of non-intrusive projection-based ROMs has been, on the other hand, quite successful in combination with non-linear regression approaches such as Gaussian process regression (GPR),<sup>57</sup> neural networks, and polynomial chaos expansion (PCE).<sup>58</sup> Three-dimensional CFD simulations were combined with sensor data to develop the DT<sup>59,60</sup> of a semi-industrial furnace operating with methane-hydrogen mixtures and full hydrogen charges. The DT could predict scalar fields and, integral and point quantities with very high accuracy.

Other attempts at developing DTs of industrial applications involve the development of a simplified physics-based model with a set of coefficients tuned with an optimization method,<sup>61</sup> building ROMs for specific variables of interest (e.g., emissions, vibrational data, ...), and building estimators via regression algorithms for the system health states.<sup>62,63</sup>

## DTs OF INDUSTRIAL COMBUSTION SYSTEMS: WHAT DO WE NEED TO MOVE FORWARD?

The current state of the art highlights five key points for unlocking DT potentials for energy-intensive industries: the training data, data analysis and feature extraction, adaptive simulation frameworks and subgrid models, ROMs, and DTs with lifelong learning capabilities. Although these aspects are discussed for turbulent combustion systems, they are relevant for closely related multi-physics flow problems encountered in power generation, process industries, orbital re-entry, air pollution and atmospheric flows, and biological flows, to name a few.

### Training data

Collecting representative, accurate, and reliable training data is central for any data-driven approach. These datasets must cover a wide range of operating conditions and be representative of different potential applications. When looking at fields where ML prospers, such as advertising, image, or speech recognition, one realizes that the impact of each ML prediction is small and adds up to many similar decisions for which the training data are so vast that virtually all tasks may be regarded as interpolation. The situation for turbulent flows is very different. Indeed, the accuracy and physical realizability of reacting flow simulations are mandatory, thus requiring interpretable, explainable, and generalizable ML algorithms.<sup>64</sup>

Experimental data are required to validate models; they are our ultimate window into reality. However, data availability at the industrial scale is scarce due to the challenges of performing full-scale, high-fidelity experiments.<sup>11</sup> Configurations bridging the laboratory and industrial scales are required to acquire data whose quantity and quality are relevant to the application of interest. The concept of scale-bridging data is paramount for DTs of industrial systems and represents a key area of research. The notion of scale-bridging data is not new and can be related to the validation hierarchies introduced by Oberkampf and Trucano.<sup>65</sup> The challenge lies in identifying systems with a degree of physical interactions relevant to the final application of interest but allowing isolating and controlling quantities of interest.

Numerical simulations can offer a level of detail that is yet unattainable in experiments since physical sensors cannot be massively deployed because of harsh conditions and costs. Moreover, data must cover a wide range of operating conditions and be representative of different potential applications. High-fidelity simulations of turbulent reacting flows are still limited to small domains and archetypal configurations due to their complexity and cost, both computational and monetary. Moreover, they generate vast data in some dimensions (thanks to the fine spatial and temporal resolutions). Still, they are very sparse in others (flow conditions and fuel composition), posing challenges for generalization.

Ideally, one would rely on DNS data, but this is far from a viable solution. Massive datasets are available for very few operating conditions (regarding chemical composition, turbulence level, turbulence/chemistry interactions, etc.). The impact of each evaluation on the results is critical, and guarantees on performances and uncertainty in the predictions must be provided. This points out that one must be cautious while extending ML approaches from one case to another since these approaches are also prone to overfitting.

Generating an appropriate dataset for the optimal representation of a system is still an open question, not only for chemically reacting flows.<sup>66–68</sup> In tabulated chemistry approaches, it is common practice to rely on simple chemistry-based simulation attempting to replicate the ensemble of state spaces encountered in an actual three-dimensional numerical simulation.<sup>9</sup> This requires prior knowledge of the system since the system is constrained on a pre-computed manifold. Moreover, challenges can arise when multiple combustion regimes coexist, indicating that the state-space trajectories for one archetypal flame are insufficient. Merging datasets from different archetypal flames is also attempted using *ad hoc* and case-specific approximations.

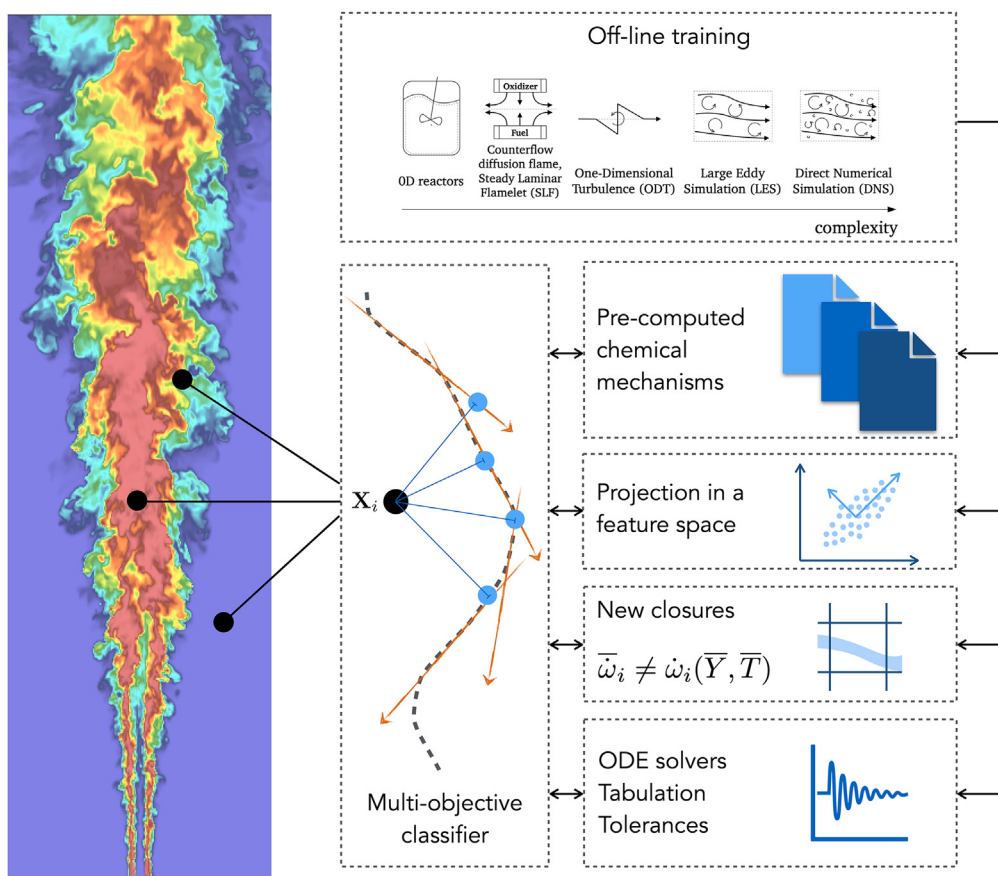
The opportunities offered by ML are extremely promising, such as developing unsupervised methodologies to “nest” different reactor and flame types into a single database for training and developing ROMs and tailoring modeling to specific local regions in numerical simulations.

### Data analysis and feature extraction

Experiments and high-fidelity simulations of combustion typically generate massive amounts of data. Hence, big data in combustion have been a reality for almost three decades.<sup>69</sup> This is also aided by the advent of petascale and exascale computing which allows DNS for close-to-reality conditions. These big data are typically used for model development and validation.<sup>12–14</sup> Many data-processing techniques for combustion have been developed in the past 50 years, and these techniques can be used for both physical and numerical experiments. While these analyses have strongly relied on domain expertise and heuristic algorithms, they can certainly be regarded as an early application of ML.

A key point is to link unsupervised feature extraction and domain knowledge. Reaction progress variable is, for instance, a key central concept in combustion theory and modeling. Nevertheless, no consensus exists in the literature about its definition and generalization.<sup>70</sup> Recent studies based on modal decomposition techniques such as PCA<sup>50,67,71,72</sup> have suggested that PCs can be associated with distinctive flame processes, linking feature extraction algorithms and domain knowledge. Other challenges in the analysis of combustion datasets are related to the representativity and sample density of different regions of the state space,<sup>73</sup> the state-space pre-processing,<sup>53,72,74–76</sup> the set of state variables to be selected<sup>54,72,76–82</sup> and the associated impact on the topology,<sup>83,84</sup> parameterization,<sup>76</sup> and (supervised) regression of the low-dimensional manifold.<sup>72,82</sup> To date, solutions to this problem are scarce and case specific, despite the significant potential impact on the performances of predictive combustion simulations.

A quantitative manifold-informed method to assess the manifold topology has been proposed,<sup>85</sup> opening the way to an optimal (from the topological perspective) parameterization of the low-dimensional manifold using both linear (e.g., using PCA) and non-linear (e.g., autoencoders) projections, and the adaptive mapping/regression of the variables of interest accounting for local manifold characteristics. Using linear/non-linear activation functions (for the encoding and decoding processes, respectively) combined with topology-aware loss functions



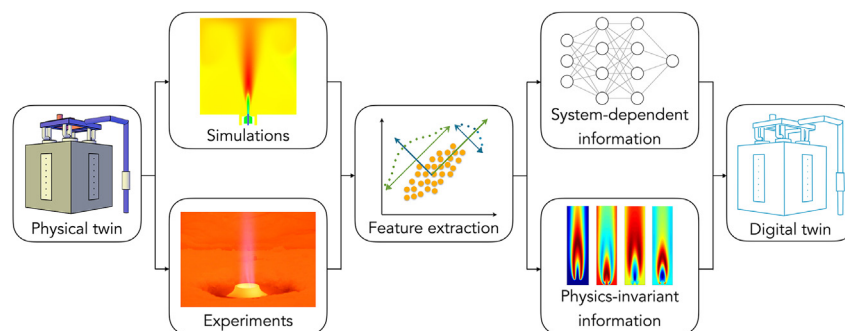
**Figure 2. Adaptive simulation framework**

On-the-fly classification for dynamic, turbulent combustion simulations, adapting the chemical mechanism, the number of chemical species, the combustion closure, and the numerical settings to the local flow conditions.

is particularly appealing. Indeed, optimized manifold topologies can be achieved by leveraging the effect of the back-propagation algorithm on the network weights while keeping a linear projection that has proven useful to derive physics-based transport equations of reacting scalars in the principal-component space.<sup>86</sup>

### Adaptive simulation frameworks and subgrid models

The development of simulation-based ROMs or DTs presents several challenges related to the computational power required to generate enough numerical simulations of different fidelity. Considering the size of detailed chemical mechanisms and the typical dimensions of combustion systems involved in energy-intensive industries, it is crucial to develop computational approaches aiming for local comprehensiveness. In connection with the discussion on **training data**, unsupervised clustering approaches appear particularly promising to open the way for fully adaptive models that adjust the chemical mechanism, the number of transported species, the closures, and the numerical solvers and tolerances in CFD simulations, depending on the local flow conditions. The coupling of clustering with dimensionality reduction using PCA has proven particularly effective in identifying regions of the state space with structurally different characteristics and distinct physical processes and in developing adaptive chemistry models.<sup>87,88</sup> This concept has been demonstrated for adapting the complexity of the chemical mechanism locally in numerical simulations but can be extended to the entire simulation workflow shown in Figure 2, with applications beyond the sole combustion field. Subfilter modeling for high-fidelity LES is an appealing area for data-driven models. In combustion problems, temperature fluctuations can be as high as several hundred Kelvins. Since Arrhenius reaction rates are highly non-linear functions of temperature, accurate statistical closures for filtered approaches cannot be based on an expansion around mean properties.<sup>9</sup> The task of a combustion model is to describe the unresolved scales based on the information available during a simulation. Data-driven methods significantly impact the design of improved subgrid combustion models.<sup>89–98</sup> While replacing the chemical source term with black-box approaches may be helpful for recurring numerical simulations, such an approach is limited in scope because of its intrinsic inability to generalize. More interesting options include embedding ML in physical models or incorporating physical knowledge into ML models.



**Figure 3. A digital twins development framework based on heterogeneous data streams, feature extraction/dimensionality reduction, and ML algorithms**

An example of the first approach is to define data-driven correction terms in the form of summative or multiplicative residuals (i.e., using ML for correction terms while preserving a physics-based approach) in the framework of different closure paradigms, flamelet-like to PDF-like approaches, from conditional closures to reactor-based approaches.<sup>99,100</sup> Incorporating physical knowledge into ML algorithms can be achieved by promoting the compliance of the results with the known laws of physics for the problem at hand. In neural networks, this can be accomplished by tailoring the loss function to include terms penalizing the violation of specific conservation principles or the entire residual of a governing equation.<sup>101</sup> In combustion, examples involve physics-informed neural networks applied to stiff chemical kinetics problems<sup>102</sup> and physics-informed enhanced super-resolution generative adversarial networks (PIERSGANs) for subfilter modeling.<sup>103</sup>

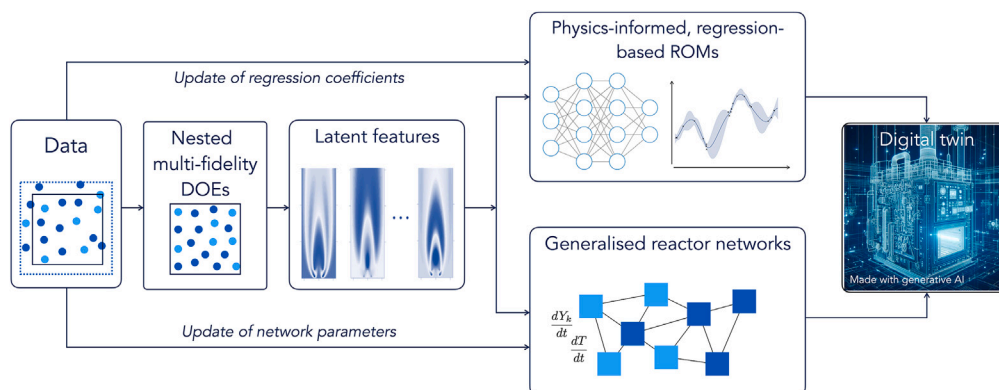
A final important remark concerns the actual use of ML-enhanced closures in turbulent combustion simulations. Many of the proposed approaches have been only demonstrated in *a priori* testing, but only a few have made it to a *posteriori* simulation.<sup>103</sup> Models showing good performances in an *a priori* framework could ultimately fail in an actual simulation, particularly those not accounting for the problem physics in any form, resulting in unphysical predictions and uncontrolled error propagation. Finally, there are challenges in posterior validation related to the numerical implementation and the coupling with the CFD codes, which are not directly relevant to the focus of this paper.

### ROMs

Complex systems design is an iterative process requiring several model evaluations and answers for “what if” scenarios. As such, it cannot rely solely on high-fidelity tools (such as CFD simulations). A combination of approaches is needed to bridge information from high- and low-fidelity models into a reduced representation of a complex system. Engineers have traditionally collected the results of experiments and numerical simulations through correlations to design and optimize systems. Today, the challenge is developing strategies to assimilate data of different natures and fidelity into a reduced representation of the system of interest. Approaches based on the combination of dimensionality reduction (using linear techniques such as PCA) and non-linear regression have been successful<sup>59,60</sup> in developing simulation-based DTs able to mimic the behavior of complex three-dimensional CFD numerical simulations (Figure 3). To further reduce prediction errors in the design space, non-linear dimensionality reduction (e.g., autoencoders-like) appear particularly well suited. Indeed, classic PCA tends to overestimate dimensionality, prompting the exploration of non-linear methods like kernel PCA,<sup>54,73</sup> T-distributed stochastic neighbor embedding (T-SNE),<sup>104</sup> and autoencoders.<sup>105</sup> While non-linear methods enhance dimensionality reduction, they are more challenging regarding physical interpretation and can be computationally costly since they involve numerous hyper-parameters to calibrate.<sup>106</sup>

Associating a level of confidence/uncertainty to the predictions also appears to be of paramount importance for ROMs, prompting the use of techniques such as GPR and Bayesian networks.

Finally, strategies need to be explored for the efficient fusion of data of different fidelity. The characterization of an industrial furnace may be, for instance, unfeasible, using LESs only. A common strategy in computer-aided design is to perform a decreasing number of simulations of increasing complexity: operating conditions are generally explored using lumped systems (e.g., 0D reactors) and one-dimensional models. Systems’ performances are then broadly investigated using reactor networks and RANS simulations, and few conditions of interest are analyzed in detail using LES. Moreover, when performing RANS or LES simulations, the computational effort can strongly vary depending on the level of fidelity employed to describe chemical kinetics (chemical schemes for methane combustion can, for instance, range from ~20 to more than 100 species), the number of chemical species impacting the computational cost more than quadratically.<sup>9</sup> There is an opportunity to exploit all the available data to generate and refine ROMs of complex combustion systems using heterogeneous data. Several multi-fidelity models, such as non-intrusive polynomial chaos,<sup>107</sup> CoKriging,<sup>108</sup> and radial basis function (RBF),<sup>109</sup> have been proposed in the literature. Techniques seeking to align the latent spaces spanned by the predictions of different fidelities<sup>58</sup> have been successfully applied in aerodynamics and other fields, demonstrating reduced computational costs and improved accuracy. Fusing low-fidelity and high-fidelity data in multi-fidelity ROM balances accuracy with computational budget constraints. This involves replacing some high-fidelity simulations with low-fidelity ones to minimize training costs while accepting a limited accuracy loss in the resulting ROM. The fundamental question is how many high-fidelity simulations can be replaced with low-fidelity ones in each design of experiment to train a ROM with satisfactory prediction accuracy.



**Figure 4. Schematic of the digital twin generation/adaptation process**

DTs can be updated using new data to modify the mapping between the feature and state spaces and between the input conditions and reactor network parameters. A new training process shall be carried out for new data significantly exceeding the original DOE.

### DTs with lifelong learning capabilities

Once built, ROMs can be used to replicate the numerical simulations. As such, they are affected not only by the error intrinsic in the ROM generation process but also by the limitations of the first-principle model (uncertainty in boundary conditions model parameters, numerical and experimental errors, ...). Strategies to update the DT (Figure 4) once in operation (lifelong learning) can profit from the vast literature in the field of data assimilation (DA) techniques. The latter have gained momentum in the combustion community to update numerical simulations based on high-fidelity measurements,<sup>110</sup> to calibrate physics-based ROMs,<sup>109,111</sup> and to update DTs based on few experimental measurements.<sup>112</sup> ROMs based on dimensionality reduction and regression methods are particularly appealing in this framework for the possibility of updating the regression coefficients (of a GPR regression or a neural network) based on the available data stream. ROMs based on CRNs can also be very appealing for their gray-box nature: data-driven algorithms can be designed to develop generalized CRNs consisting of combinations of ideal reactors for which the governing equations are known. The chemical network can be generated starting from simplified CFD simulations (thus relying on the multi-fidelity concept highlighted above) using unsupervised clustering algorithms modified to enforce the determination of spatially continuous clusters (using geometrical penalties or graph algorithms) to ensure spatial continuity of the identified clusters. The availability of multiple CFD simulations can potentially allow extraction of different CRNs, each tailored to precise simulation settings. To overcome the case-specific nature of the approach, a mapping between input conditions in the design space and network parameters (reactor volumes and mass exchanged between reactors) can be established using non-linear regression techniques. This approach can relate changes in internal mixing, induced by differences in operating conditions, to an adaptive set of network parameters, thus delivering a generalized CRN model that can represent the system under a wide range of operating conditions.

### CONCLUSIONS AND PROSPECTS

We emphasize the pivotal role of data in combustion technologies for energy-intensive industries. The inherent complexity of combustion systems necessitates innovative approaches, where data-driven methodologies and artificial intelligence emerge as transformative tools. The vision of creating robust and predictive DTs for large-scale combustion assets presents a promising avenue toward achieving sustainable combustion systems, enabling the potential of renewable synthetic fuels. However, some key challenges need to be addressed. These include obtaining representative training data, developing data analysis tools and adaptive simulation frameworks, creating accurate ROMs, and enabling DTs with lifelong learning capabilities. The need for a multi-disciplinary approach is highlighted to address these challenges and establish a comprehensive, physics-aware, data-intensive framework for sustainable combustion technologies in the realm of energy-intensive industries.

### LIMITATIONS OF THE STUDY

This paper reflects the authors' perspective on the role of data in advancing combustion technologies for energy-intensive industries. It reflects our interpretation and synthesis of the existing literature and trends and combines it with our vision of how the field will evolve. While we outline key challenges and potential solutions, the implementation and effectiveness of these strategies will vary depending on specific industrial contexts and technological advancements.

### ACKNOWLEDGMENTS

Some of the ideas presented here were developed during a collaborative project, ALCHEMY (Machine Learning for Complex Multiphysics problems), between Cambridge University and ULB funded by Fondation Wiener – Anspach, ULB, Brussels. The funding from this foundation is gratefully acknowledged.

A.P. received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement no. 714605. A.P. acknowledges the support of the Walloon Region as part of the FRFS-WEL-T strategic axis.

N.S. acknowledges the support of EPSRC through grant number EP/S025650/1.

## AUTHOR CONTRIBUTIONS

A.P.: conceptualization, writing – original draft preparation, writing – review and editing. N.S.: conceptualization, writing – review and editing.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

- Bruyn, S., Jongsma, C., Kampman, B., and Görlach, B. (2020). Thie J.E. Energy-intensive Industries – Challenges and Opportunities in Energy Transition, Study for the Committee on Industry, Research and Energy (ITRE) (Policy Department for Economic, Scientific and Quality of Life Policies. European Parliament).
- Powering a Climate-Neutral Economy: An EU Strategy for Energy System Integration, COM (2020), 299 final.
- Swaminathan, N., Bai, X., Brethouwer, G., and Haugen, N. (2022). Introduction. In *Advanced Turbulent Combustion Physics and Applications*, N. Swaminathan, X. Bai, N. Haugen, C. Fureby, and G. Brethouwer, eds. (Cambridge University Press), pp. 1–24.
- McKinsey & Company (2020). Plugging in: what electrification can do for industry. In *Energy Insights Practice*.
- Dreizler, A., Pitsch, H., Scherer, V., Schulz, C., and Janicka, J. (2021). The role of combustion science and technology in low and zero impact energy transformation processes. *AECs* 7, 100040.
- IEA (2022). *World Energy Outlook 2022* (OECD Publishing). <https://doi.org/10.1787/3a469970-en>.
- European Commission (2019). Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs, Masterplan for a Competitive Transformation of EU Energy-Intensive Industries Enabling a Climate-Neutral, Circular Economy by 2050 (Publications Office).
- Rasheed, A., San, O., and Kvamsdal, T. (2020). Digital Twin: Values, Challenges and Enablers From a Modeling Perspective. *IEEE Access* 8, 21980–22012.
- Pope, S.B. (2013). Small scales, many species and the manifold challenges of turbulent combustion. *Proc. Combust. Inst.* 34, 1–31.
- Bilger, R.W. (2011). The Role of Combustion Technology in the 21st Century. In *Turbulent Combustion Modeling. Fluid Mechanics and its Applications, vol 95*, T. Echekki and E. Mastorakos, eds. (Springer).
- Giusti, A., and Mastorakos, E. (2019). Turbulent Combustion Modelling and Experiments: Recent Trends and Developments. *Flow, Turbul. Combust.* 103, 847–869.
- Trisjono, P., and Pitsch, H. (2015). Systematic Analysis Strategies for the Development of Combustion Models from DNS: A Review. *Flow, Turbul. Combust.* 95, 231–259.
- Wick, A., Attili, A., Bisetti, F., and Pitsch, H. (2020). DNS-driven analysis of the Flamelet/Progress Variable model assumptions on soot inception, growth, and oxidation in turbulent flames. *Combust. Flame* 214, 437–449.
- H. Pitsch, and A. Attili, eds. (2020). *Data Analysis for Direct Numerical Simulations of Turbulent Combustion* (Springer).
- Chung, W.T., Jung, K.S., Chen, J.H., and Ihme, M. (2022). BLASTNet: A call for community-involved big data in combustion machine learning. *AECs* 12, 100087.
- Ihme, M., Chung, W.T., and Mishra, A.A. (2022). Combustion machine learning: Principles, progress and prospects. *PECS* 91, 101010.
- [www.combustioninstitute.org/resources/workshops-databases/](http://www.combustioninstitute.org/resources/workshops-databases/).
- Huang, X., Tummers, M.J., van Veen, E.H., and Roekaerts, D.J. (2022). Modelling of MILD combustion in a lab-scale furnace with an extended FGM model including turbulence–radiation interaction. *Combust. Flame* 237, 111634.
- Giuntini, L., Frascino, L., Ariemma, G.B., Galletti, C., Sorrentino, G., and Ragucci, R. (2023). Performance Assessment of Modeling Approaches for Moderate or Intense Low-Oxygen Dilution Combustion in a Scale-Bridging Burner. *Energy Fuels* 37, 9500–9513.
- Ferrarotti, M., De Paepe, W., and Parente, A. (2021). Reactive structures and NO<sub>x</sub> emissions of methane/hydrogen mixtures in flameless combustion. *Int. J. Hydrogen Energy* 46, 34018–34045.
- Jamshidiha, M., Kamal, M.M., Cafiero, M., Coussement, A., and Parente, A. (2024). Experimental and numerical characterization of hydrogen combustion in a reverse-flow micro gas turbine combustor. *Int. J. Hydrogen Energy* 55, 1299–1311.
- Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., and Kumar, V. (2017). Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Trans. Knowl. Data Eng.* 29, 2318–2331.
- Gartner. (2018). *Gartner Top 10 Strategic Technology Trends for 2018*.
- Grieves, M. (2014). *Digital Twin: Manufacturing Excellence through Virtual Factory Replication* (A White Paper; Michael Grieves, LLC).
- Shafto, M., Conroy, M., Doyle, R., Glaessgen, E., Kemp, C., LeMoigne, J., and Wang, L. (2010). *Draftmodeling Simulation Information Technology & Processing Roadmap* (Technology Area, 11).
- Kapteyn, M., Knezevic, D., Huynh, D., Tran, M., and Willcox, K. (2022). Data-driven physics-based digital twins via a library of component-based reduced-order models. *Int. J. Numer. Methods Eng.* 123, 2986–3003.
- Semeraro, C., Lezoche, M., Panetto, H., and Dassisti, M. (2021). Digital twin paradigm: A systematic literature review. *Comput. Ind.* 130, 103469.
- Palmé, M., Fast, M., and Thern, M. (2011). *Appl. Energy* 88, 3898–3904.
- Kalogirou, S.A. (2003). Artificial intelligence for the modeling and control of combustion processes: a review. *Prog. Energy Combust. Sci.* 29, 515–566.
- Christo, F., Masri, A., Nebot, E., and Pope, S.B. (1996). An integrated PDF/neural network approach for simulating turbulent reacting systems. *Symp (Int) Combust* 26, 43–48.
- Blasco, J.A., Fueyo, N., Dopazo, C., and Ballester, J. (1998). Modelling the Temporal Evolution of a Reduced Combustion Chemical System With an Artificial Neural Network. *Combust. Flame* 113, 38–52.
- Ihme, M., Schmitt, C., and Pitsch, H. (2009). Optimal artificial neural networks and tabulation methods for chemistry representation in LES of a bluff-body swirl-stabilized flame. *Proc. Combust. Inst.* 32, 1527–1535.
- Bode, M., Collier, N., Bisetti, F., and Pitsch, H. (2019). Adaptive chemistry lookup tables for combustion simulations using optimal B-spline interpolants. *Comb. Theor. Model.* 23, 674–699.
- Berger, L., Kleinheinz, K., Attili, A., Bisetti, F., Pitsch, H., and Mueller, M.E. (2018). Numerically accurate computational techniques for optimal estimator analyses of multi-parameter models. *Combust. Theory Model* 22, 480–504.
- Brunton, S.L., Proctor, J.L., and Kutz, J.N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* 113, 3932–3937.
- Swithenbank, J. (1970). *Combustion Fundamentals*, AFOSR 70-2110 TR.
- Faravelli, T., Bua, L., Frassoldati, A., Antifora, A., Tognotti, L., and Ranzi, E. (2001). A new procedure for predicting NO<sub>x</sub> emissions from furnaces. *Comput. Chem. Eng.* 25, 613–618.
- Falcitelli, M., Tognotti, L., and Pasini, S. (2002). An algorithm for extracting chemical reactor network models from cfd simulation of industrial combustion systems. *Combust. Sci. Technol.* 174, 27–42.
- Cuoci, A., Frassoldati, A., Buzzi Ferraris, G., Faravelli, T., and Ranzi, E. (2007). The ignition, combustion and flame structure of



- carbon monoxide/hydrogen mixtures. Note 2: Fluid dynamics and kinetic aspects of syngas combustion. *Int. J. Hydrogen Energy* 32, 3486–3500.
40. Lyra, S., and Cant, R. (2013). Analysis of high pressure premixed flames using Equivalent Reactor Networks for predicting NOx emissions. *Fuel* 107, 261–268.
  41. Stagni, A., Cuoci, A., Frassoldati, A., Faravelli, T., and Ranzi, E. (2014). A fully coupled, parallel approach for the post-processing of CFD data through reactor network analysis. *Comput. Chem. Eng.* 60, 197–212.
  42. Perpignan, A.V., Sampat, R., and Gangoli Rao, A. (2019). Modeling pollutant emissions of flameless combustion with a joint CFD and chemical reactor network approach. *Front. Mech. Eng.* 5, 63.
  43. Gupta, S., Malte, P., Brunton, S.L., and Novosselov, I. (2019). Prevention of lean flame blowout using a predictive chemical reactor network control. *Fuel* 236, 583–588.
  44. Savarese, M., Cuoci, A., De Paepe, W., and Parente, A. (2023). Machine learning clustering algorithms for the automatic generation of chemical reactor networks from CFD simulations. *Fuel* 343, 127945.
  45. Swaminathan, N., Parente, A. (eds) *Machine Learning and Its Application to Reacting Flows*. Lecture Notes in Energy, vol 44. Springer, Cham
  46. Maas, U., and Thévenin, D. (1998). Correlation analysis of direct numerical simulation data of turbulent non-premixed flames. *Proc. Combust. Inst.* 27, 1183–1189.
  47. Frouzakis, C.E., Kevrekidis, Y.G., Lee, J., Boulouchos, K., and Alonso, A.A. (2000). Proper orthogonal decomposition of direct numerical simulation data: Data reduction and observer construction. *Proc. Combust. Inst.* 28, 75–81.
  48. Danby, S.J., and Echehki, T. (2006). Proper orthogonal decomposition analysis of autoignition simulation data of nonhomogeneous hydrogen–air mixtures. *Combust. Flame* 144, 126–138.
  49. Parente, A., Sutherland, J.C., Tognotti, L., and Smith, P. (2009). Identification of low-dimensional manifolds in turbulent flames. *Proc. Combust. Inst.* 32, 1579–1586.
  50. Sutherland, J.C., and Parente, A. (2009). Combustion modeling using principal component analysis. *Proc. Combust. Inst.* 32, 1563–1570.
  51. Parente, A., Sutherland, J.C., Dally, B.B., Tognotti, L., and Smith, P.J. (2011). Investigation of the MILD combustion regime via Principal Component Analysis. *Proc. Combust. Inst.* 33, 3333–3341.
  52. Biglari, A., and Sutherland, J.C. (2012). A filter-independent model identification technique for turbulent combustion modeling. *Combust. Flame* 159, 1960–1970.
  53. Parente, A., and Sutherland, J.C. (2013). Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity. *Combust. Flame* 160, 340–350.
  54. Mirgolbabaei, H., and Echehki, T. (2013). A novel principal component analysis-based acceleration scheme for LES–ODT: An a priori study. *Combust. Flame* 160, 898–908.
  55. Yang, Y., Pope, S.B., and Chen, J.H. (2014). *Combust. Flame* 160, 1967–1980.
  56. McQuarrie, S.A., Huang, C., and Willcox, K.E. (2021). Data-driven reduced-order models via regularised Operator Inference for a single-injector combustion process. *J. R. Soc. N. Z.* 51, 194–211.
  57. Aversano, G., Bellemans, A., Li, Z., Coussement, A., Gicquel, O., and Parente, A. (2019). Application of reduced-order models based on PCA & Kriging for the development of digital twins of reacting flow applications. *Comput. Chem. Eng.* 121, 422–441.
  58. Aversano, G., D’Alessio, G., Coussement, A., Contino, F., Parente, A., and Parente, A. (2021). Combination of polynomial chaos and Kriging for reduced-order model of reacting flow applications. *RINENG* 10, 100223.
  59. Aversano, G., Ferrarotti, M., and Parente, A. (2020). *Proc. Comb. Inst.* 38, 5373–5381.
  60. Procacci, A., Amaduzzi, R., Coussement, A., and Parente, A. (2023). Adaptive digital twins of combustion systems using sparse sensing strategies. *Proc. Combust. Inst.* 39, 4257–4266.
  61. Wang, C., Wang, Z., Huang, K., Li, Y., and Yang, C. (2023). Digital twin for zinc roaster furnace based on knowledge-guided variable-mass thermodynamics: Modeling and application. *PSEP* 173, 39–50.
  62. Fu, X., Han, J., Castle, M., and Cao, K. (2023). Digital twin-driven vibration amplitude simulation for condition monitoring of axial blowers in blast furnace ironmaking. *Syst. Sci. Control. Eng.* 11, 2152400.
  63. Deon, B., Cotta, K., Silva, R., Batista, C., Justino, G., Freitas, G., Cordeiro, A., Barbosa, A., Loução, F., Simioni, T., et al. (2022). Digital twin and machine learning for decision support in thermal power plant with combustion engines. *Knowl. Based. Syst.* 253, 109578.
  64. Brunton, S.L., Noack, B.R., and Koumoutsakos, P. (2020). *Machine Learning for Fluid Mechanics*. *Annu. Rev. Fluid Mech.* 52, 477–508.
  65. Oberkampf, W.L., and Trucano, T.G. (2002). Verification and validation in computational fluid dynamics. *Prog. Aerosp. Sci.* 38, 209–272.
  66. D’Alessio, G., Cuoci, A., Aversano, G., Braconni, M., Stagni, A., and Parente, A. (2020). Impact of the partitioning method on multidimensional adaptive-chemistry simulations. *Energies* 13, 2567.
  67. Ranade, R., and Echehki, T. (2019). A framework for data-based turbulent combustion closure: A priori validation. *Combust. Flame* 206, 490–505.
  68. Renganathan, S.A., Harada, K., and Mavris, D.N. (2020). *Aerodynamic Data Fusion Toward the Digital Twin Paradigm*. *AIAA J.* 58, 3902–3918.
  69. Chen, J.H. (2011). Petascale direct numerical simulation of turbulent combustion—fundamental insights towards predictive models. *Proc. Combust. Inst.* 33, 99–123.
  70. Barlow, R.S., Magnotti, G., Cutcher, H.C., and Masri, A.R. (2017). On defining progress variable for Raman/Rayleigh experiments in partially-premixed methane flames. *Combust. Flame* 179, 117–129.
  71. Ranade, R., and Echehki, T. (2019). A framework for data-based turbulent combustion closure: A posteriori validation. *Combust. Flame* 210, 279–291.
  72. Malik, M.R., Obando Vega, P., Coussement, A., and Parente, A. (2021). Combustion modeling using Principal Component Analysis: A posteriori validation on Sandia flames D, E and F. *Proc. Combust. Inst.* 38, 2635–2643.
  73. Coussement, A., Gicquel, O., and Parente, A. (2012). Kernel density weighted principal component analysis of combustion processes. *Combust. Flame* 159, 2844–2855.
  74. Malik, M.R., Isaac, B.J., Coussement, A., Smith, P.J., and Parente, A. (2018). Principal component analysis coupled with nonlinear regression for chemistry reduction. *Combust. Flame* 187, 30–41.
  75. Biglari, A., and Sutherland, J.C. (2015). An a-posteriori evaluation of principal component analysis-based models for turbulent combustion simulations. *Combust. Flame* 162, 4025–4035.
  76. Isaac, B.J., Thornock, J.N., Sutherland, J., Smith, P.J., and Parente, A. (2015). Advanced regression methods for combustion modelling using principal components. *Combust. Flame* 162, 2592–2601.
  77. Mirgolbabaei, H., and Echehki, T. (2014). Nonlinear reduction of combustion composition space with kernel principal component analysis. *Combust. Flame* 161, 118–126.
  78. Echehki, T., and Mirgolbabaei, H. (2015). Principal component transport in turbulent combustion: A posteriori analysis. *Combust. Flame* 162, 1919–1933.
  79. Owoyele, O., and Echehki, T. (2017). Toward computationally efficient combustion DNS with complex fuels via principal component transport. *Combust. Theory Model.* 21, 770–798.
  80. Nguyen, H.-T., Domingo, P., Vervisch, L., and Nguyen, P.D. (2021). Machine learning for integrating combustion chemistry in numerical simulations. *Energy AI* 5, 100082.
  81. Gitushi, K.M., Ranade, R., and Echehki, T. (2022). Investigation of deep learning methods for efficient high-fidelity simulations in turbulent combustion. *Combust. Flame* 236, 111814.
  82. Malik, M.R., Coussement, A., Echehki, T., and Parente, A. (2022). Principal component analysis based combustion model in the context of a lifted methane/air flame: Sensitivity to the manifold parameters and subgrid closure. *Combust. Flame* 244, 112134.
  83. Armstrong, E., and Sutherland, J.C. (2021). A technique for characterising feature size and quality of manifolds. *Combust. Theory Model.* 25, 646–668.
  84. Zdybał, K., Sutherland, J.C., and Parente, A. (2023). Manifold-informed state vector subset for reduced-order modeling. *Proc. Combust. Inst.* 39, 5145–5154. <https://doi.org/10.1016/j.proci.2022.06.019>.
  85. Zdybał, K., Armstrong, E., Sutherland, J.C., and Parente, A. (2022). Cost function for low-dimensional manifold topology assessment. *Sci. Rep.* 12, 14496.
  86. Zdybał, K. (2023). PhD Dissertation. <https://github.com/kamilazdybal/phd-dissertation>.
  87. D’Alessio, G., Parente, A., Stagni, A., and Cuoci, A. (2020). Adaptive chemistry via pre-partitioning of composition space and mechanism reduction. *Combust. Flame* 211, 68–82.
  88. D’Alessio, G., Cuoci, A., and Parente, A. (2021). Feature extraction and artificial neural networks for the on-the-fly classification of high-dimensional thermochemical spaces in adaptive-chemistry simulations. *A. Parente Data-Centric Eng.* 2, e2.
  89. Bode, M., Gauding, M., Lian, Z., Denker, D., Davidovic, M., Kleinheinz, K., Jitsev, J., and

- Pitsch, H. (2021). Using physics-informed enhanced super-resolution generative adversarial networks for subfilter modeling in turbulent reactive flows. *Proc. Combust. Inst.* 38, 2617–2625.
90. Liu, B., Tang, J., Huang, H., and Lu, X.-Y. (2020). Deep learning methods for super-resolution reconstruction of turbulent flows. *Phys. Fluids* 32, 025105.
  91. Lapeyre, C.J., Misdariis, A., Cazard, N., Veynante, D., and Poinso, T. (2019). Training convolutional neural networks to estimate turbulent sub-grid scale reaction rates. *Combust. Flame* 203, 255–264.
  92. Colin, O., Ducros, F., Veynante, D., and Poinso, T. (2000). A thickened flame model for large eddy simulations of turbulent premixed combustion. *Phys. Fluids* 12, 1843–1863.
  93. Nikolaou, Z.M., Chrysostomou, C., Vervisch, L., and Cant, S. (2019). Progress variable variance and filtered rate modelling using convolutional neural networks and flamelet methods. *Flow Turbul. Combust.* 103, 485–501.
  94. Seltz, A., Domingo, P., Vervisch, L., and Nikolaou, Z.M. (2019). Direct mapping from LES resolved scales to filtered-flame generated manifolds using convolutional neural networks. *Combust. Flame* 210, 71–82.
  95. Nikolaou, Z.M., Chrysostomou, C., Minamoto, Y., and Vervisch, L. (2021). Evaluation of a neural network-based closure for the unresolved stresses in turbulent premixed V-flames. *Flow Turbul. Combust.* 106, 331–356.
  96. Henry de Frahan, M.T., Yellapantula, S., King, R., Day, M.S., and Grout, R.W. (2019). Deep learning for presumed probability density function models. *Combust. Flame* 208, 436–450.
  97. Chen, Z.X., Iavarone, S., Ghiasi, G., Kannan, V., D'Alessio, G., Parente, A., and Swaminathan, N. (2021). Application of machine learning for filtered density function closure in MILD combustion. *Combust. Flame* 225, 160–179.
  98. Nista, L., Schumann, C.D.K., Grenga, T., Attili, A., and Pitsch, H. (2023). Investigation of the generalization capability of a generative adversarial network for large eddy simulation of turbulent premixed reacting flows. *Proc. Combust. Inst.* 39, 5279–5288.
  99. Jigjid, K., Minamoto, Y., Doan, N.A.K., and Tanahashi, M. (2023). SGS Reaction rate modelling for MILD combustion based on machine-learning combustion mode classification: Development and a priori study. *Proc. Combust. Inst.* 39, 4489–4499.
  100. Freitas, R.S., Péquin, A., Galassi, R.M., Attili, A., and Parente, A. (2023). Model identification in reactor-based combustion closures using sparse symbolic regression. *Combust. Flame* 255, 112925.
  101. Raissi, M., Perdikaris, P., and Karniadakis, G.E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* 378, 686–707.
  102. Ji, W., Qiu, W., Shi, Z., Pan, S., and Deng, S. (2021). Stiff-PINN: Physics-Informed Neural Network for Stiff Chemical Kinetics. *J. Phys. Chem. A* 125, 8098–8106.
  103. Bode, M., Gauding, M., Goeb, D., Falkenstein, T., and Pitsch, H. (2023). Applying physics-informed enhanced super-resolution generative adversarial networks to turbulent premixed combustion and engine-like flame kernel direct numerical simulation data. *Proc. Combust. Inst.* 39, 5289–5298.
  104. Fooladgar, E., and Duwig, C. (2018). A new post-processing technique for analyzing high-dimensional combustion data. *Combust. Flame* 191, 226–238.
  105. Zdybal, K., Parente, A., and Sutherland, J.C. (2023). Improving reduced-order models through nonlinear decoding of projection-dependent outputs. *Patterns* 4, 100859.
  106. Le Clainche, S., Ferrer, E., Gibson, S., Cross, E., Parente, A., and Vinuesa, R. (2023). Improving aircraft performance using machine learning: A review. *Aerosp. Sci. Technol.* 138, 108354.
  107. Wai-Tsun Ng, L., and Eldred, M. (2012). Multifidelity Uncertainty Quantification Using Non-Intrusive Polynomial Chaos and Stochastic Collocation. In *AIAA 2012-1852. 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*.
  108. de Baar, J., Roberts, S., Dwight, R., and Mallol, B. (2015). Uncertainty quantification for a sailing yacht hull, using multi-fidelity kriging. *Comput. Fluids* 123, 185–201.
  109. Pellegrini, R., Iemma, U., Leotardi, C., Campana, E.F., and Diez, M. (2016). Multi-fidelity Adaptive global metamodel of expensive computer simulations. In *2016 IEEE Congress on Evolutionary Computation (CEC)*, Vancouver, BC, Canada, pp. 4444–4451.
  110. Labahn, J.W., Wu, H., Coriton, B., Frank, J.H., and Ihme, M. (2021). Data assimilation using high-speed measurements and LES to examine local extinction events in turbulent flames. *Proc. Comb. Inst.* 38, 2259–2266.
  111. Yu, H., Juniper, M.P., and Magri, L. (2021). A data-driven kinematic model of a ducted premixed flame. *Proc. Combust. Inst.* 38, 6231–6239.
  112. Donato, L., Galletti, C., and Parente, A. (2024). *Applied Thermal Engineering. Appl. Therm. Eng.* 236, 121431.