



ELSEVIER

Contents lists available at ScienceDirect

## Aerospace Science and Technology

journal homepage: [www.elsevier.com/locate/aescte](http://www.elsevier.com/locate/aescte)

## Review

## Improving aircraft performance using machine learning: A review

Soledad Le Clainche<sup>a,\*</sup>, Esteban Ferrer<sup>a,b</sup>, Sam Gibson<sup>c</sup>, Elisabeth Cross<sup>c</sup>,  
Alessandro Parente<sup>d</sup>, Ricardo Vinuesa<sup>e</sup>

<sup>a</sup> ETSIAE-UPM-School of Aeronautics, Universidad Politécnica de Madrid, Plaza Cardenal Cisneros 3, E-28040 Madrid, Spain<sup>b</sup> Center for Computational Simulation, Universidad Politécnica de Madrid, Campus de Montegancedo, Boadilla del Monte, 28660 Madrid, Spain<sup>c</sup> Dynamics Research Group, University of Sheffield, UK<sup>d</sup> Brussels Faculty of Engineering, Université Libre de Bruxelles, Avenue Franklin D. Roosevelt 50, 1050, Brussels, Belgium<sup>e</sup> FLOW, Engineering Mechanics, KTH Royal Institute of Technology, Stockholm, Sweden

## ARTICLE INFO

## Article history:

Received 4 November 2022

Received in revised form 23 March 2023

Accepted 17 April 2023

Available online 23 April 2023

Communicated by Jérôme Morio

## ABSTRACT

This review covers the new developments in machine learning (ML) that are impacting the multi-disciplinary area of aerospace engineering, including fundamental fluid dynamics (experimental and numerical), aerodynamics, acoustics, combustion and structural health monitoring. We review the state of the art, gathering the advantages and challenges of ML methods across different aerospace disciplines and provide our view on future opportunities. The basic concepts and the most relevant strategies for ML are presented together with the most relevant applications in aerospace engineering, revealing that ML is improving aircraft performance and that these techniques will have a large impact in the near future.

© 2023 The Author(s). Published by Elsevier Masson SAS. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Contents

1. Introduction . . . . .	2
2. Machine learning methodology: a general overview . . . . .	4
2.1. Neural networks . . . . .	4
2.2. Regression and classification . . . . .	5
2.3. Semi-supervised learning . . . . .	6
2.4. Unsupervised learning: clustering and dimensionality reduction . . . . .	7
2.4.1. Clustering . . . . .	7
2.4.2. Dimensionality reduction . . . . .	7
2.4.3. Principal component analysis . . . . .	7
2.4.4. Local dimensionality reduction . . . . .	8
2.4.5. Kernel PCA . . . . .	8
2.4.6. Autoencoders . . . . .	8
3. Machine learning in fluid mechanics . . . . .	9
3.1. Computational fluid dynamics . . . . .	9
3.2. Reduced-order models . . . . .	10
3.3. Experiments . . . . .	11
4. Machine learning in aerodynamics . . . . .	11
4.1. Aerodynamic coefficients estimation . . . . .	12
4.2. Aeroelasticity . . . . .	13
4.3. Design optimization . . . . .	14
5. Machine learning in aeroacoustics . . . . .	14
6. Machine learning in combustion . . . . .	15
6.1. Data analysis and feature extraction . . . . .	16
6.2. Dimensionality reduction, classification and adaptive chemistry . . . . .	17

\* Corresponding author.

E-mail addresses: [soledad.leclainche@upm.es](mailto:soledad.leclainche@upm.es) (S. Le Clainche), [esteban.ferrer@upm.es](mailto:esteban.ferrer@upm.es) (E. Ferrer), [sjgibson1@sheffield.ac.uk](mailto:sjgibson1@sheffield.ac.uk) (S. Gibson), [e.j.cross@sheffield.ac.uk](mailto:e.j.cross@sheffield.ac.uk) (E. Cross), [alparent@ulb.ac.be](mailto:alparent@ulb.ac.be) (A. Parente), [rvinuesa@mech.kth.se](mailto:rvinuesa@mech.kth.se) (R. Vinuesa).

<https://doi.org/10.1016/j.ast.2023.108354>

1270-9638/© 2023 The Author(s). Published by Elsevier Masson SAS. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

6.3. Combustion closures .....	17
6.4. Reduced-order models for realistic combustion systems .....	18
7. Machine learning in structural assessment .....	18
8. Conclusions and further discussion: new technologies based on machine learning .....	19
Declaration of competing interest .....	21
Data availability .....	21
Acknowledgements .....	21
References .....	21

## 1. Introduction

Climate change and increasing resource scarcity are challenges that Europe needs to face in the coming decades. All this has a direct impact on air transport, which is struggling to maintain its performance and competitiveness while ensuring a development focused on sustainable mobility. Research and innovation are essential to maintain the capabilities of the aviation industry, driven by the rise of new markets and new competitors as a result of globalization. A new long-term vision for the aeronautics sector is essential to ensure its successful advancement. In this line, new requirements for the future aviation industry have been defined by the ACARE Flightpath 2050 [1], a Group of Recognized Personalities in the aeronautic sector, including stakeholders from the aeronautics industry, air traffic management, airports, airlines, energy providers and the research community. Aeronautics and air transport comprises both: air vehicle and system technology. The future of aviation should focus on improving design, reducing manufacturing time and cost (including certification and upgrade processes), and also improving the parts forming the overall air travel system (general aviation, aircraft, airlines, airports, air traffic management and maintenance, repair and overhaul).

The ACARE Flightpath 2050 has defined 5 goals that should be achieved by 2050 to guarantee the path through sustainable mobility:

- Compared to the capabilities of a typical aircraft in the year 2000, by 2050 new technologies should allow 90% reduction in NO<sub>x</sub> emissions, 75% reduction in CO<sub>2</sub> emissions per passenger/kilometre and 65% reduction in noise emission of flying aircraft.
- When taxiing, aircraft movements should be emission-free.
- Novel strategies to design and manufacture aerial vehicles should be developed to make them recyclable.
- Sustainable alternative fuels should be developed to position Europe as the centre of excellence in the field, and sustained by a strong European energy policy.
- By 2025, Europe should take the lead to establish global environmental standards, formulating and prioritizing an environmental action plan, and being at the forefront of atmospheric research.

At the same time, ensuring safety and security is also a major priority, with the aim at reducing by 80% the number of accidents by 2050 compared to 2000, taking into account the rising traffic.

To achieve these goals, it is extremely important to find newer eco-friendly alternatives suitable for the industry to reduce the aviation net carbon emissions and noise. To this aim, the aerospace industry is gathering efforts towards developing new aerodynamic designs, more efficient, reducing the oil consumption whilst maintaining the safety in the flight performance. Moreover, finding new alternatives to fossil fuels, improving the energy efficiency in combustion systems, or finding optimal routes for air traffic management (ATM) are also some of the key points where the aerospace industry should advance to minimize the environmental footprint. However, to achieve these objectives it is not enough to improve

the 'standard' configurations. The aerospace engineering industry is aware that to go beyond the state-of-the-art it is necessary to develop novel ground-breaking disruptive technologies.

Fluid and solid mechanics need to be advanced with applications in aerodynamics, acoustics, and combustion, to develop new technologies, resulting in novel aircraft designs with reduced environmental impact (see Fig. 1). Researchers in collaboration with the aeronautical industry should explore: (i) new aircraft configurations able to reduce noise and pollution emissions, (ii) cruise drag reduction by manipulating turbulent flow structures close to the aircraft surface (*i.e.*, delaying the boundary transition from laminar to turbulent flow), using novel friendly low-risk practices, (iii) novel strategies for flow control (rising the benefits achieved by only changing the external shape) to enhance the aerodynamic performance reducing drag, noise or flow transition, rising lift or controlling unsteadiness or flow separation, and (iv) reduce the system complexity with novel aircraft materials and lighter designs (which directly results in less fuel consumption), and reduced aircraft maintenance and life cost cycle [2].

High-fidelity numerical simulations and advanced experimental techniques (*i.e.*, wind tunnel experiments or open-air experiments as in the case of flight test, et cetera) allow collecting a large variety of data, containing relevant information about physical principles connected to the aerodynamic performance of the aircraft, the efficiency of the combustion system or the main instabilities driving the flow dynamics and the possibility of attenuating or boosting such instabilities using active or passive flow control techniques [3]. Additionally, experiments (*i.e.*, ultrasounds, non-intrusive testing, et cetera) and simulations provide information connected to the fundamentals of solid mechanics, the presence of noise and the structural health of the aircraft, allowing for noise control and early failure detection.

However, the economic and computational cost, related to the performance of experiments and simulations, encourages researchers to look for new alternatives, which allow to advance in the field *i.e.*, developing relevant technologies for the aerospace industry, while avoiding delays in the manufacturing and time-to-market process. Aircraft development, manufacturing, maintenance and support are four critical levels that must be accurate and reliable to ensure the success of the aerospace industry.

Artificial intelligence (AI) and machine learning (ML) have been introduced in the aerospace industry for various applications connected to the reduction of aircraft's environmental impact, including data interpretation [4], system management, customer service or aircraft modelling and to generate new high-fidelity databases at a reduced (economic and CPU) cost [5], solving problems of optimization, flow control, or even providing optimal sensors distributions for solid mechanics or aeroelasticity applications. In the recent review article by Brunton et al. [6], the authors summarize the new trends and perspective of ML in the aerospace industry, including its application for smart manufacturing, and in the development (and aircraft design), production and product support phases (aircraft design, manufacturing, verification and validation). The authors reveal the possibilities of ML to process data in light computations increasing the production rate, based on the idea of

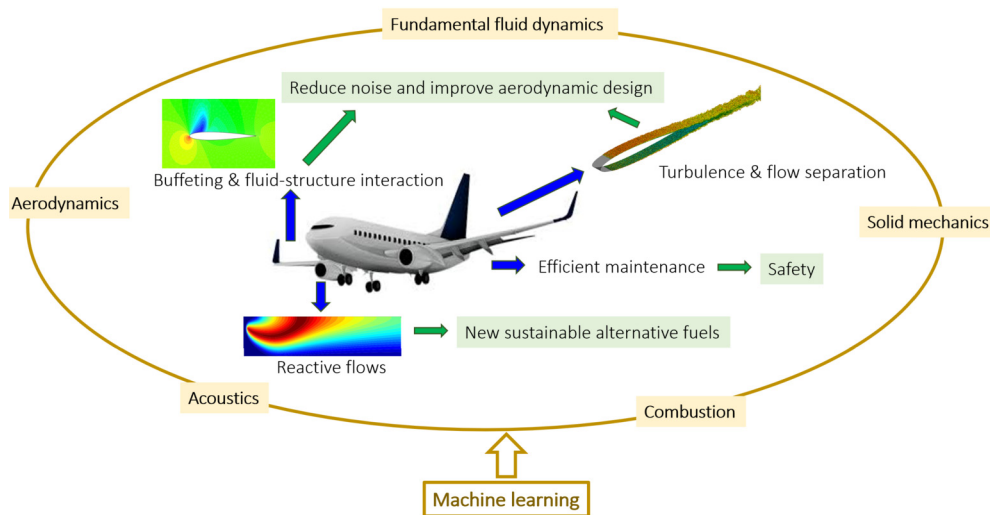


Fig. 1. Towards sustainable aviation using machine learning.

**Table 1**  
Summary of machine learning applications to improve aircraft performance.

DISCIPLINE	MACHINE LEARNING APPLICATION
Aerodynamics	Optimizing the design and operation of the aircraft. Wing design. Flight control. Aircraft performance optimization.
Aeroacoustics	Noise reduction. Aeroacoustic optimization. Flight path optimization. Noise prediction.
Combustion	Optimizing the combustion process in aircraft engines. Combustion optimization. Emission prediction. Fuel optimization. Predictive maintenance.
Structural health monitoring	Enabling early detection and diagnosis of structural damage or degradation. Predictive maintenance. Damage detection. Structural health prognosis. Structural design optimization.

the use of ML techniques that are measurable, interpretable and certifiable.

ML is generally understood as a branch of AI, although there are nuances in the definition: ML aims at improving systems performance using self-learning algorithms, while AI tries to mimic natural intelligence solving complex problems and enabling decision making (although not maximizing the system efficiency) [7]. Both AI and ML are connected to Big Data, a term that linked to the enormous volume of data that floods the aeronautical sector every day or that is generated from CFD simulations or experimental measurements and connected to aircraft aerodynamic performance [8,9]. Combining Big Data with ML techniques, it is possible to develop reduced dimensional systems, such as reduced order models (ROMs) [10,11], capable to accurately predict the evolution of the flow dynamics [3,12,13] (i.e., flow control, reduce cruise drag [14], boundary layer transition, etc.), or surrogate models, capable to predict the aerodynamic forces and moments acting on the aircraft as a function of some parameters (i.e., Reynolds number, Mach number, geometry shape, etc.).

This review provides a state-of-the-art of AI and ML applications in the aerospace engineering field. The basic concepts and most relevant strategies for ML and AI are brought together to explain the similarities found in the nomenclature of similar techniques used in different fields, also shedding light on new applications of these algorithms, quite extended in other fields but not known to the aerospace industry. For example, the review details the use of machine learning for reduced order modelling, which can be used to accelerate numerical simulations, or for temporal and spatial forecast (including non-intrusive sensing). Additionally, we include relevant applications of the field, including flow control, acoustics, combustion, flight test and structural health monitoring.

This article intends to explore the possibilities of ML, an emerging field for the aerospace industry, identifying new research lines of potential interest and bringing new ideas to developing the technologies of the future, and founded on a primary goal: to fight climate change. This article reviews the main disciplines connected to the aerodynamic performance of the aircraft (see Fig. 1): fundamental fluid dynamics, aerodynamics, acoustics, combustion, and general solid mechanics; and based on the idea of finding novel efficient designs, capable to reduce noise and pollutant emissions, while at the same time, ensuring safety and security. It is worth mentioning that there are more than 1000 related papers in the field, but this reviewer only details a few of them. These articles have been selected to represent the most popular applications in machine learning with the idea of providing a general overview about what is the state-of-the-art in the field. Nevertheless, we recommend the reader not only to read the paper mentioned, but also to explore other possibilities.

In fluid dynamics, machine learning approaches can be used to improve aircraft performance by optimizing the aerodynamic design and reducing drag, improving fuel efficiency, and increasing speed. Some of the most relevant applications of ML to improve aircraft efficiency are summarized in Table 1. Details about the methodologies, state-of-the-art and needs to advance in the field are presented along this review article.

The article is organized as follows. Section 2 introduces the machine learning methodologies, and the literature review of the main applications in aerospace engineering is presented in Section 3 for fundamental fluid dynamics, Section 4 for aerodynamics, Section 5 for aeroacoustics, Section 6 for combustion and Section 7 for solid mechanics. The main conclusions are presented in Section 8.

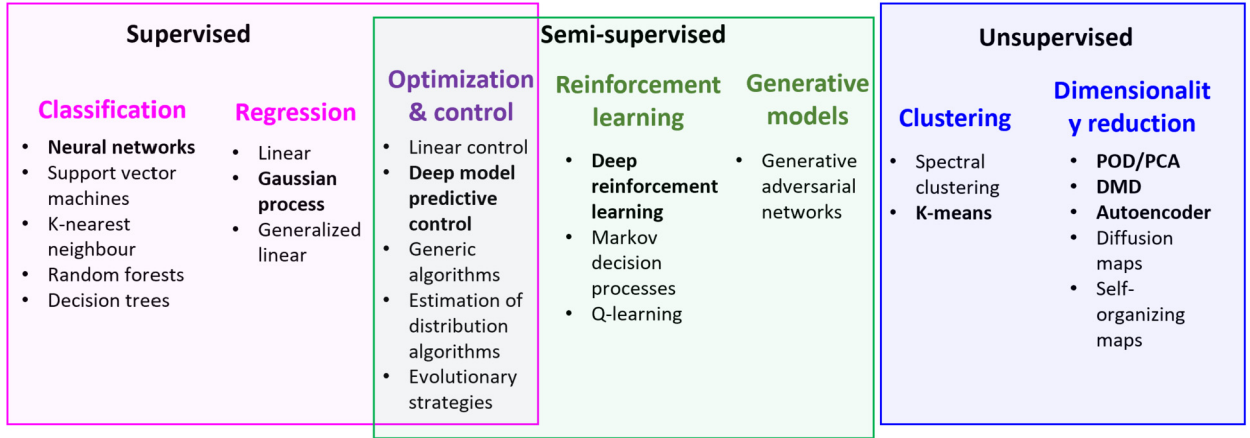


Fig. 2. Machine learning methods: a general overview. Classification extracted from Ref. [15]. In bold, the most popular techniques in the field of aerospace engineering.

## 2. Machine learning methodology: a general overview

Current advances in computer science are strongly related to the increasing amount of data generated and stored in the different disciplines conforming aerospace engineering. The valuable information contained in these databases encourages researchers to develop and test sophisticated algorithms to exploit such information, to gain insight and knowledge from the data and to subsequently propose and develop new commercial strategies aligned with the ideas behind the concept of sustainable aviation: developing new cleaner and safer aircraft designs. ML is a fast-growing science in the field of aerospace engineering due to its good capabilities to extract information from complex databases, which it later used to develop models, such as ROMs or surrogate models. Based on available data and the type of training carried out within the analysis, ML algorithms can be classified into unsupervised, semi-supervised or supervised learning, as presented in Fig. 2. This section briefly introduces the basic idea behind some of these ML algorithms, which have been used for different applications in the field of aerospace engineering.

### 2.1. Neural networks

ML uses artificial neural networks (ANNs), also called as neural networks (NNs), to process and extract information from databases. The name of this computing system is inspired by the biological neural networks of the human brain. ML uses NNs to solve an optimization problem. More specifically, using back propagation and stochastic gradient descent algorithms, ML optimizes the following compound function

$$\arg \min_{A_j} (f_P(A_P, \dots, f_2(A_2, f_1(A_1, X))) + \lambda g(A_j)), \quad (1)$$

which represents a NN formed by  $P$  layers. The weights connecting years  $k$  and  $k+1$  are given in matrix  $A_k$ . To accurately represent the data, which is the main goal of ML and NNs, it is necessary to regularize this system, which is massively undetermined, with the function bias  $g(A_j)$ . Using different regularization strategies, it is possible to prevent *overfitting*, allowing the NNs to generalize the solution obtained from the training to different data sequences.

To define the architecture of a NN, it is necessary to set the dimension and number of layers and the type of connections and mappings (linear/non-linear) between the different layers, in order to get the best system performance to represent data. Depending on all these parameters and on the way the optimization problem is defined, it is possible to define the different type of ML archi-

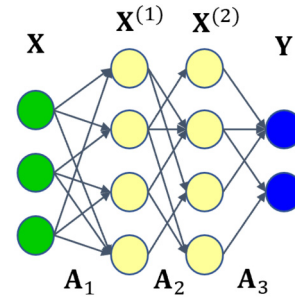


Fig. 3. Sketch representing a NN architecture with three layers. Structure extracted from Ref. [4].

tectures previously presented in Fig. 2. The generic structure of a multi-layer NN is presented in Fig. 3.

In this example, the input database  $X = [x_1 \ x_2 \ x_3] \in \mathbb{R}^3$ , is map  $Y = [y_1 \ y_2]$  in the output layer, with a different space dimension,  $\mathbb{R}^2$ . The relationship between the different layers in this NN is defined in eq. (2) as

$$\begin{aligned} X^{(1)} &= f_1(A_1, X) \\ X^{(2)} &= f_2(A_2, X^{(1)}) \\ Y &= f_3(A_3, X^{(2)}). \end{aligned} \quad (2)$$

This expression is also defined by the following compound functions  $f_j$ , defined for  $P$  layers as

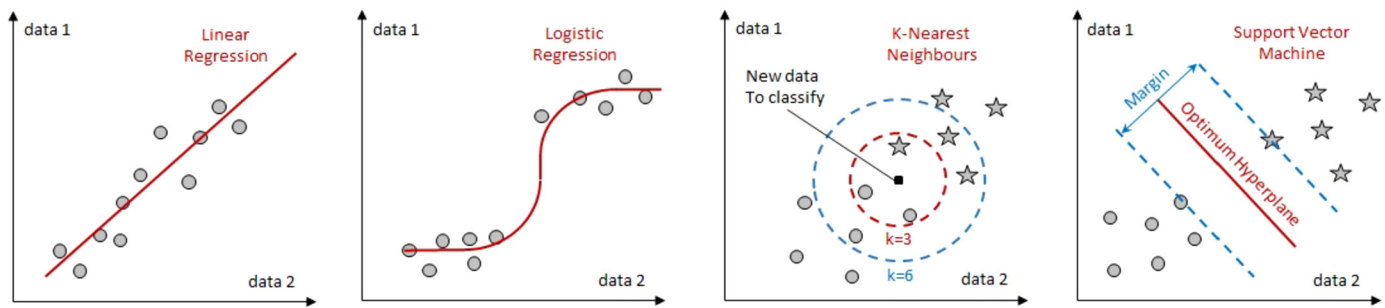
$$Y = f_P(A_P, \dots, f_2(A_2, f_1(A_1, X)) \dots), \quad (3)$$

which defines the general optimization problem presented in eq. (1). This expression defines a system of equations that is highly under-determined, which requires additional constraints to be solved, where the  $P$  matrices generate the best possible mapping.

The activation function  $f_j$  (for the layer  $j$ ), can be both linear or non-linear (*i.e.*, sigmoid, tanh, Rectified Linear Unit -ReLU, ...). In aerodynamics, radial basis function neural networks (RBFNNs) have been used since the past, due to the good generalization capabilities of the RBF as activation function [16]. This NNs model, proposed by Broomhead & Lowe [17] for the first time, generally uses a non-linear activation function  $f_j$  defined by a Gaussian basis function as

$$f_j(x_k) = \exp\left(-\frac{\|x_k - v_j^2\|}{2\sigma_j^2}\right), \quad (4)$$





**Fig. 4.** Summary of methods for regression and classification in machine learning. From left to right: linear regression, logistic regression, K-nearest neighbours and support vector machines.

where the two hyper-parameters  $\mathbf{v}$  and  $\sigma$  are known as the centre vector and the width parameter, respectively. The value set for these two parameter strongly affects the NN performance [18,19].

Depending on the type of connection between the layers and the neuron of the NNs, it is possible to identify different architectures. For instance, Recurrent Neural Networks (RNNs) process sequence of data using recurrent connections between the different layers. Among the different architectures of RNNs it is possible to distinguish Long Short-Term Memory (LSTM) networks, which in sequence prediction problems are capable of learning order dependence. Another type of architecture widely used for ML applications in fluid dynamics is Convolutional Neural Networks (CNNs). This type of networks transfers information locally between the different neurons and layers. Based on a threshold, each neuron may transfer or not transfer information to the next layer of neurons. More details describing deep learning basic concepts can be found in the book by Brunton & Kutz [20], and some applications combining NNs linear mappings with non-linear modal interactions for turbulent flows are presented by Le Clainche et al. [4].

## 2.2. Regression and classification

Regression and classification belong to the category of supervised learning. In supervised learning, labelled data sets are used, to guide the machine during the training process. The algorithm learns by minimizing the error with respect to the labelled target. Regression and classification are widely popular and very effective techniques, but can lead to erroneous results if data (or categories) are missing, which leads to problems during training and/or classification. The missing data or category can relate to a problem misconception leading to a category missing in the data, to data removal because of large deviations, to equipment malfunctioning, etc. [21]. To avoid problems in these supervised-learning algorithms, a good a-priori knowledge (and analysis) of the problem at hand is advisable to ensure that the labelled training data is representative of the overall dataset. In this section, we focus on the most common regression and classification algorithms, including: regression and logistic regression, decision trees, random forest, K-nearest neighbours, support-vector machines and Naive-Bayes classification.

A common task in supervised learning is regression, which is used to identify patterns and relationships within a dataset. In this task, the algorithm approximates a function for a continuous output (e.g., a real value). Classic examples are linear (or quadratic) regressions, where linear (or quadratic) curves are fit to data. This type of regressions can be understood from a purely algebraic viewpoint, through least-square approximations (i.e., Ridge regression using the L2 norm) but generalizes to ML using NNs.

In linear regression with one variable, the output  $y$  will fulfil the equation of a straight line:  $y = mx + c$ , where  $m$  is the slope,  $c$  the intercept and  $x$  the input. The calculation of ( $m$ ) and ( $c$ ) will be based on the minimization of a cost function, i.e., the

minimizing the error between the training data and the linear model. An example is shown in Fig. 4. The regression model can be used to interpolate missing data or to extrapolate (forecast) outside the training-data range. When considering regression with multiple dependent variables, it is necessary to analyze the influence of each variable on the overall trend of the dependent variable (system output). Multiple linear regression is a useful technique for more than one independent variable. This technique can achieve better fits than when using simple linear regression if multiple independent variables are involved.

Classification also belongs to the supervised-learning family. In this problem, the output will take a discrete set of values or categories. The algorithm will use pre-set labelled data to classify the data into multiple categories. According to the type of variables, we can distinguish categorical classification from numerical classification. Logistic regression is the simplest classification algorithm that performs a binary classification (between only two categories) and can help to determine the occurrence (or not) of an event. The algorithm starts with data previously labelled in the two possible categories. These two categories are joined through a sigmoid curve, which will determine the probability that new data falls into one category or the other, see Fig. 4. The sigmoid curve

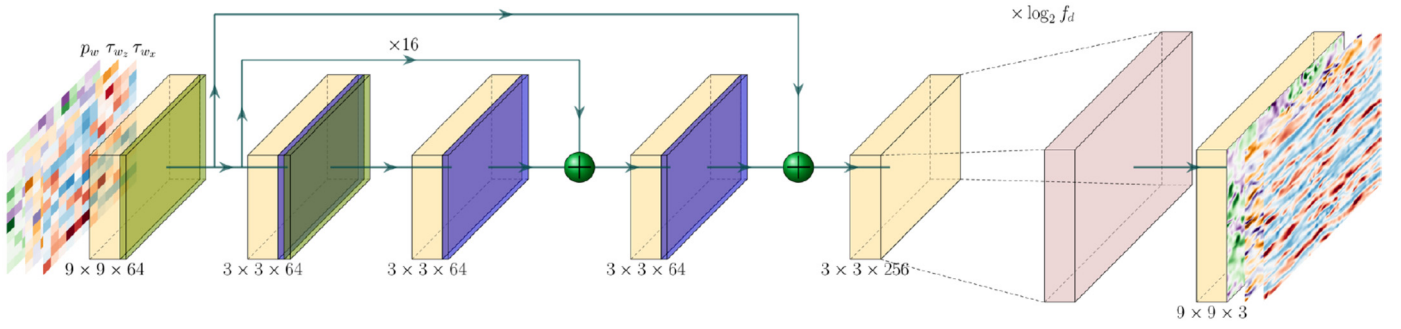
$$P(t) = \frac{1}{1 + \exp(-t)}$$

helps classify the probability of success of an event; starting with a null probability (first category) and ending at a certain event (second category). The curve shows a change of curvature when the probability of success is 50%. This point, called threshold, will differentiate the data classified into both categories, as depicted in Fig. 4.

The decision tree is a conceptually simple algorithm, useful for classifying data. The algorithm uses a tree, in which each of the branches implies a decision leading to a classification. To make these decisions and determine which characteristics to discern, it is necessary to train it with data. The construction of the trees, as well as the determination of the order of the different branches alludes to the reduction of dispersion (or entropy) in the data in each step providing an optimal tree with minimal dispersion.

Random forest [22] is an algorithm for both classification and regression tasks and is particularly useful for data sets in which there can be loss of information. Its development starts from the construction of multiple decision trees (hence the concept of “forest”). Each decision tree performs a classification task leading to the final result of the random forest, which is the best classification of all trees.

K-Nearest Neighbours (KNN) are one of the simplest algorithms used for supervised classification. It classifies the input data based on the closeness to already classified data. The algorithm generates “neighbours” of data and measures the distance (or “similarity”) of new data with respect to the classified neighbouring data.



**Fig. 5.** Schematic representation of a GANS architecture used to increase the resolution of the quantities measured at the wall in a turbulent channel flow. The colour coding for each layer is 2D convolution (beige), parametric-ReLU activation (dark green), batch normalization (blue), sub-pixel convolution (pink) and ReLU activation (light green). Note that the kernel size and the number of filters are shown at the bottom of each of the layers. Figure adapted from Ref. [27] with permission of the publisher. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

The KNN algorithm is effective if the data set is not excessively large (slow learning algorithm) and the data is free of noise. The KNN concept can be understood by drawing a circle centred on the new data point, which increases its radius until there are  $K$  elements (belonging to a pre-set category) within the circle.  $K$  represents the number of elements needed within the circumference to classify the model input; this gives the name to the algorithm K-nearest neighbours, as shown in Fig. 4. The choice of the parameter  $K$  is essential to achieve the desired resolution of the problem, and needs to be large enough to avoid the influence of noisy data, but small enough to limit the computation time. A possible value is the square root of the labelled data ( $n$ ), i.e.,  $K = \sqrt{n}$ .

Support-vector machines (SVMs) are similar to K-Nearest Neighbours and can be used to perform classification and regression tasks where the transformed feature space is very large. However, for large datasets, SVMs suffer from high computational cost. SVMs create a hyperplane to delimit the neighbours or regions in the data set, see Fig. 4. To determine the position of the hyperplane, the algorithm finds support vectors, defining the distance to the neighbours. This distance must be maximum for the classification to be successful. Note that the hyperplanes may have various shapes: linear, quadratic, cubic, etc. SVMs can deal with a wide variety of classification problems including high dimensions and non-linearities. A drawback is that SVMs require calibration of the parameters to attain good results.

Finally, the Naive-Bayes classifier has its mathematical basis in the conditional probability proposed by Bayes' theorem, whose formulation is:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}.$$

The logical operation is similar to the error back-propagation algorithm. The final probability will be based on the total probability taking into account the influence of the individual events on the following events (conditional probability). The total probability theorem, used to give the final output of the model is  $P(B) = P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \dots$ . The process requires to calculate the probability of each event individually and analyze its influence on the final response. This probabilistic analysis will finally lead to a decision. The algorithm will predict events, based on present and past data, knowing the probability of occurrence of the predicted events. Some advantages of Naive-Bayes classifiers are outlined in [21] and include: (i) they remain smooth to small changes in the network, (ii) they are flexible, allowing for transfer learning between classification and regression, (iii) they can handle missing data.

Gaussian-process (GP) regression is an extension of Bayesian-regression method that is used widely within the field of structural health monitoring. Rather than predicting a single deterministic

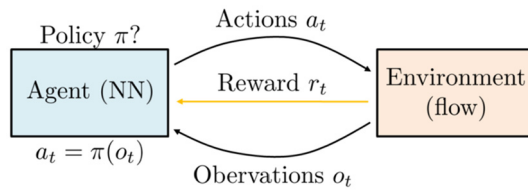
function, the predictive posterior is instead a distribution of functions that are consistent with the data that the model has been conditioned by. At any finite number of points in the output space, the distribution of these functions is Gaussian. Like SVMs, GP regression uses a *kernel*, or *covariance* function to estimate the similarity between two inputs and it is the selection of this kernel that defines the form that the posterior functions take, in combination with the optimization of the model hyperparameters. The GP can be defined by its mean and this covariance; often a zero mean prior is assumed, however, knowledge of the process can be incorporated into the model relatively simply using a *mean function*. A commonly-used covariance function is the squared exponential:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2\ell^2} \|\mathbf{x} - \mathbf{x}'\|^2\right),$$

where  $\mathbf{x}$  and  $\mathbf{x}'$  are the inputs to the model,  $\sigma_f^2$  is the signal variance hyperparameter and  $\ell$  is the lengthscale hyperparameter. The reader is referred to Ref. [23] for an in-depth review of GP regression theory.

### 2.3. Semi-supervised learning

In this section we will discuss two types of semi-supervised algorithms: generative adversarial networks (GANs) and deep reinforcement learning (DRL). GANs [24] are a type of generative model, which comprises two different parts: the generator ( $G$ ) and the discriminator ( $D$ ). These two parts of the network are assigned competing tasks, and they are trained against each other using game theory. A widely-used application of GANs is that of super-resolution, i.e., starting from a low-resolution input image  $L_R$ , the GAN architecture needs to produce a high-resolution version  $\hat{H}_R$  which should be consistent with the statistical features of the original high-resolution image  $H_R$ . In this setup, the generator will be tasked with creating the  $\hat{H}_R$  images given the  $L_R$  input and the reference  $H_R$ . This process is illustrated for a turbulent flow in Fig. 5. Note that this super-resolution task can also be conducted in an unsupervised manner by imposing physical properties on the data, as documented by Kim et al. [25]. The discriminator is thus trained to differentiate  $\hat{H}_R$  and  $H_R$ , so that the generator gets progressively better at producing realistic images, and the discriminator improves its performance when identifying the ones that have been artificially generated by  $G$ . It is common to use a convolutional neural network (CNN) as generator, and the resolution increase is usually carried out at the end of the network by means of sub-pixel convolution layers [26]. A common architecture for the discriminator in this type of task [27] is to also use a CNN, followed by fully-connected layers, and using a sigmoid activation in the last one it is possible to obtain a probability value between



**Fig. 6.** Schematic of the deep-reinforcement-learning process, where the aim is to find the optimal decision policy ensuring that the cumulative reward is maximized. Figure extracted from Ref. [31] with permission of the publisher.

0 and 1 to discern whether the high-resolution image is real or not.

When it comes to DRL, in this framework an agent (which can be a neural network) performs actions on an environment (which would be the flow) in a closed loop. If we consider instant  $t$ , at which the environment has a certain state  $s_t$ , the agent receives a partial observation of that state denoted  $o_t$ . Based on this observation, the agent decides to apply a certain action  $a_t$ , which will have an impact on the system, modifying its state to  $s_{t+1}$ , and so forth. The quality of those actions under a certain norm is measured by the so-called reward  $r_t$ , which is provided periodically to the agent. Generally in reinforcement learning, the framework is aimed at developing an optimal decision policy  $a_t = \pi(o_t)$  with the goal of obtaining the maximum cumulative reward over a certain time horizon. This process is summarized in Fig. 6. One widely-used implementation is the proximal-policy-optimization (PPO) method [28], which has the advantage of being faster than e.g., the trust-region policy optimization (TRPO) [28]. There are other suitable options for DRL, such as deep-Q network (DQN) learning [29], which is very efficient mainly with non-continuous control. In this sense, it is important to thoroughly assess the dynamics of the problem at hand to choose the most suitable DRL implementation, see for instance the variations of DQN learning reported by Gu *et al.* [30].

## 2.4. Unsupervised learning: clustering and dimensionality reduction

Compared to classification and regression approaches, unsupervised learning deals with unlabelled data with the objective of “learning” the data structure without assistance. Unsupervised methods allow processing large amounts of data, leading to the determination of the most relevant features that can be used for a lower-dimensional representation of the original data. As such, unsupervised learning techniques are highly relevant in aerospace applications.

### 2.4.1. Clustering

Clustering (e.g., K-means, mean shift, Gaussian mixture models) [32], is an unsupervised-learning method that does not require pre-sampled data to cluster regions and can automatically discover grouping within the data. An overview of clustering methods and comparisons for simple problems can be found in Ref. [33]. Here, we review the popular K-means, mean shift and Gaussian mixture models.

K-means is the simplest and faster clustering technique. The user inputs the number of clusters “k” and the algorithm initializes “k” centroids, assigning each point in the dataset to the closest centroid. Using these temporary clusters, new centroids are calculated employing the mean distance from all the data points. The process is repeated iteratively until the centroids do not move in the subsequent iterations.

The mean-shift clustering method is a variant of K-means where instead of the centroid, a sliding circular window of radius “Kernel” are considered. The goal is to slide the windows until the location that maximizes the density of points inside the window is

found. Density-based spatial clustering of applications with noise (DBSCAN) is a variant of mean-shift that considers noisy unclassified data, and can provide enhanced robustness.

Gaussian-mixture models assume that the data follows a Gaussian distribution with an associated mean and standard deviation. The algorithm estimates the mean and variance of each normal distribution iteratively using the expectation-maximization method [34] to provide the optimal estimation of these parameters.

There is some recent interesting in aerodynamics and CFD, which uses clustering to detect flow regions. For Reynolds Averaged Navier–Stokes (RANS) simulations, Saetta & Tognaccini [35] and Lanzetta *et al.* [36] showed that various viscous sensors could be used as input to a clustering ML framework. The aim was to detect the boundary layer and wake regions for turbulent flow past aeronautical applications (airfoils, aircraft). The results showed that the ML clustering outperformed the classic viscous sensors used to identify the flow regions in RANS. In the context of Large Eddy Simulations (LES), similar ideas have been proposed using Gaussian mixture to identify laminar and turbulent regions [37,38], which can also be employed to perform local mesh adaptation to increase the accuracy of simulations. Furthermore, Callaham *et al.* [39] used unsupervised-learning techniques to identify the dominant physical processes for different flow scenarios and in Ref. [40] the authors trained a neural network to classify different types of vortex wakes.

### 2.4.2. Dimensionality reduction

One of the most common tasks in unsupervised learning is dimensionality reduction. Autoencoders use NNs (with linear or non-linear activation functions) to reduce the data dimensionality and extract the main data features [41]. In fluid dynamic problems, a very rich literature exists concerning the application of proper orthogonal decomposition (POD) [42]. POD is also known with a variety of names, depending on the application field, i.e., Karhunen–Loève decomposition (KLD) [43] in image and signal processing, Principal Component Analysis (PCA) [44] in the statistics literature, Empirical Orthogonal Functions [45] in oceanography and meteorology, etc. A detailed discussion on the relation between POD, PCA and KLD, and their connection with Singular Value Decomposition (SVD) is presented in [46].

### 2.4.3. Principal component analysis

Dimensionality reduction techniques rely on the hypothesis that a reduced and optimal basis exists to represent the data. PCA thus finds a new, orthogonal basis to represent the data set, which is approximated by retaining only a subset of fewer directions, namely the ones accounting for most of the data variance in a transformed data set. Prior to performing PCA (as in all ML approaches) the raw data set is generally transformed (centred and scaled) to allow dealing with variables of different units and sizes, and to focus on fluctuations. Many options exist for data pre-processing and optimal choices cannot be defined universally but rather depending on the problem at hand [47]. The eigenvalue decomposition of the covariance matrix yields the eigenvectors, denoted as the principal components (PCs), ordered by the magnitude of their corresponding eigenvalues. Truncating the PCs to a desired level of variance provides a low-dimensional representation of the original data (Fig. 7), which is the best linear predictor of the data matrix  $\mathbf{X}$  in terms of squared prediction error [44].

PCs are determined to maximize variance, not interpretability. The complex structure of the PCs and the presence of non-negligible weights for most of the original variables on an individual PC might complicate the interpretation and exploitation of the PCs identified without supervision. To alleviate that, the eigenvector matrix can be rotated to aid physical interpretation of the scores, i.e., the projection of the data onto the new basis. Many

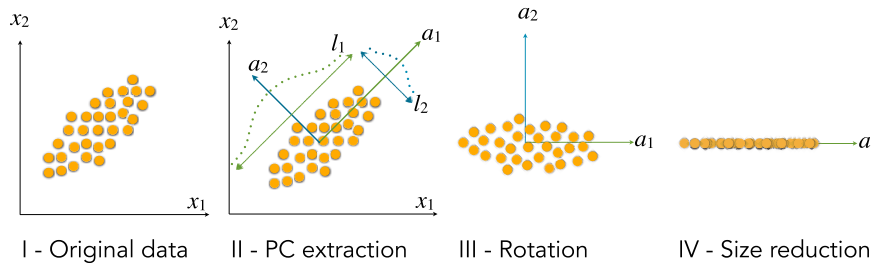


Fig. 7. Steps in the dimensionality reduction process using PCA.

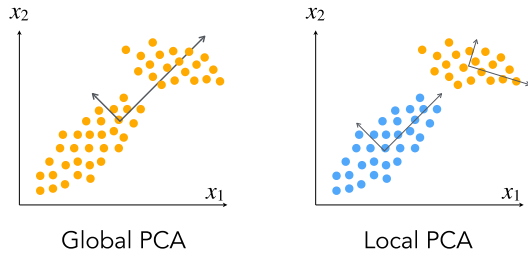


Fig. 8. Schematic representation of global and local PCA.

rotation methods exist, they can either be orthogonal or oblique. Orthogonal transformations, such as Varimax rotation [48], rotate the PCs rigidly while maximizing the sum of the variances of the squared coefficients within each eigenvector. The rotation generally leads to rotated PCs with few significant coefficients, thus aiding interpretation. After rotation, the amount of variance accounted for by a set of PCs is conserved but the proportion is redistributed amongst the components.

#### 2.4.4. Local dimensionality reduction

Typical flow data sets relevant to aerospace applications display non-linear relationships among variables. PCA is a multi-linear technique, *i.e.*, additional PCs are needed to approximate the original data set with low reconstruction error to compensate for the data non-linearity. To overcome the limitations of global PCA, an alternative is to identify local principal curves [49] and locally linear embeddings (LLE) [50], and to perform local modal decompositions, *i.e.*, applying PCA locally (LPCA).

Local PCA (LPCA) can be exploited, partitioning the data into clusters by either using an *a priori* chosen feature, or using a Vector Quantization (VQ) technique coupled to the local application of PCA within clusters. The resulting approach, denoted as VQPCA [51,52], assumes that the non-linear manifold can be locally approximated by a linear one as the data manifold will not curve too much over the extent of the local region. In the first case a supervised clustering is carried out, while the latter is fully unsupervised. In VQPCA, the data are assigned to the different clusters by minimizing the reconstruction error of each point to a given set of clusters (Fig. 8). The algorithm is iterative and convergence is achieved when the centroids' change from one iteration to the next one is below a prescribed threshold.

#### 2.4.5. Kernel PCA

Kernel PCA (KPCA) is a non-linear dimensionality reduction technique that makes use of kernel methods. In KPCA [53], the original data-set is transformed into an arbitrarily high-dimensional feature space where the relationship amongst the variables is considered to be linear and PCA can be effectively carried out. A non-trivial, arbitrary function is chosen but never calculated explicitly, which allows using high-dimensional feature spaces. The non-linearity of the KPCA method comes from

the use of a non-linear kernel function that populates the covariance matrix (kernel trick). Thus, the eigenvectors and eigenvalues of the covariance matrix in the feature space are never actually solved. The choice of the kernel function modelling the covariance between two different points in the feature space is up to the designer. An example is the choice of a Gaussian kernel. Since no operations are directly carried out in the feature space, KPCA does not compute the PCs themselves, but the projections of our data onto those components, namely the scores.

One of the main drawbacks of kernel PCA is that PCA is performed on a covariance matrix that scales with the number of observations rather than the number of variables, thus increasing the computational burden of the method. In addition, no straightforward operation is available to reconstruct the data from the KPCA scores, since the kernel is never computed explicitly. This is in contrast to linear PCA, where a direct mapping exists. To reconstruct the data in Kernel PCA, a minimization problem must be solved, making the process computationally more expensive compared to PCA.

#### 2.4.6. Autoencoders

An autoencoder (AE) [54] is an unsupervised neural network (NN) aiming at learning the reduced representation (encoding) of a dataset. Along with the encoder, which projects the data into a non-linear manifold (encodes), a decoder reconstructs the data from the so-called latent space, minimizing the reconstruction error. The simplest form of an AE is a feed-forward, non-recurrent neural network having an input layer, an output layer and one or more hidden layers connecting them, where the output layer has the same number of neurons as the input layer. This has the purpose of reconstructing the input data (minimizing the difference between input and output) via a loss function that can accommodate regularization and sparsity terms. It is worth noting that a shallow AE with linear activation functions corresponds to PCA.

The variational autoencoder (VAE) [55,56] is a particular type of probabilistic generative architecture which can also be used for modal decomposition of high-dimensional data. These models combine statistical descriptions of the data with deep neural networks, and enable introducing stochasticity to the latent space, where the low-dimensional representation of the data is obtained. In this approach, new samples can be generated through the decoder from the latent space, a characteristic that has attracted the interest of the community [41,57,58]. An interesting property of the latent representation is the disentanglement, *i.e.* being able to produce statistically-independent variables in the latent space [59]. By regularizing the loss function of the VAE it is possible to encourage learning such statistically-independent variables  $r_i$  while penalizing the size of the latent space. This was proposed by Higgins et al. [60] with the  $\beta$ -variational autoencoder ( $\beta$ VAE). This can be achieved by minimizing the distance the probabilistic distribution  $p(r)$  and the product of its marginals [61–63].



### 3. Machine learning in fluid mechanics

Fluid mechanics is closely connected with the aerodynamic performance of airplanes, and it is therefore essential to properly model and predict the properties of the flow around the aircraft in order to carry out optimization and design tasks. Very close to the airplane surface a thin layer, called boundary layer [64], develops. This region plays a critical role in aircraft design, and it is extremely challenging to model, particularly at high Reynolds numbers, *i.e.*, when the boundary layer is turbulent [65].

In fluid mechanics there is great potential for ML methods due to the high dimensionality of the associated problems and the inherent nonlinearities present in turbulent flows. In particular, novel deep-learning architectures are enabling harnessing the vast amounts of data available in the literature to produce more accurate and robust models, outperforming traditional methods for prediction, modelling and control. Future developments will focus on finding novel strategies for modelling and control, leveraging the great potential of deep reinforcement learning (DRL), see for instance Refs. [66,67].

However, there are also limitations in the usage of ML for fluid-dynamics problems. One potential limitation is the need for large amounts of training data, which can be challenging and computationally expensive to acquire. Additionally, machine learning models may not always provide physical insights into the underlying fluid dynamics phenomena, which can be important for understanding and designing practical applications. ML methods have the potential to significantly improve our understanding of fluid dynamics and enable more efficient and accurate fluid flow simulations, but careful consideration of their limitations and appropriate implementations is required.

#### 3.1. Computational fluid dynamics

The numerous success stories in ML documented for a wide range of areas [68] over the past years motivate the possibility of using ML methods for simulating fluid flows. An overview of how ML can help computational fluid dynamics (CFD) is provided by Vinuesa and Brunton [69], who essentially identified potential in many areas including accelerating CFD and improving turbulence modelling. It is important to note that ML methods will not constitute a replacement for traditional CFD methods [70,71], which are based on discretizing the governing equations and integrating them numerically. Instead, it is important to identify concrete areas within CFD that can benefit from ML. For instance, interesting work has been developed by Kochkov et al. [72] on improving the accuracy of coarse simulations by means of deep learning in simplified flows. Other areas of potential are the efficient resolution of the Poisson problem [73], which is encountered in operator-splitting methods, or approaches to reduce the size of the computational domain through efficient inflow [74,75] and far-field [76] conditions. Manrique de Lara and Ferrer [77,78] have used NNs to accelerate high-order simulations by including a corrective forcing that achieves high accuracy when running a low-order case, providing a framework for accelerated simulations. When it comes to turbulence modelling, Duraisamy et al. [79] provided a review of the potential of ML to improve RANS models. Note that these models, which are based on Reynolds averaging and typically rely on an eddy-viscosity approach [80] to represent the Reynolds stresses (some are based on more sophisticated nonlinear constitutive relations [81]), are not necessarily satisfactory when simply using high-fidelity data to address the closure problem [82]. The anisotropy tensor, which is an important component of RANS models, was predicted by Ling et al. [83] using a deep-learning architecture which embedded Galilean invariance, an approach leading to excellent results when comparing the flow pre-

dictions with the real solution. On the other hand, Wu et al. [84] proposed combining the physical knowledge from the flow with deep-learning architectures, producing a framework that allowed predicting both linear and nonlinear parts of the Reynolds-stress tensor separately. In addition to improved interpretability, this approach led to very good results in a number of configurations. The interpretability [85] of the deep-learning models for RANS simulations was one of the main features of the framework proposed by Jiang et al. [86]. When it comes to interpretability, we would like to highlight that deep-learning models are generally able to provide accurate predictions based on the input data, but it is typically not possible to establish the concrete relationships relating input and output. An interpretable model allows to establish such as connection, and there are a number of research directions [87] starting to enable this property in deep-learning models. A good example of this, which relies on genetic programming, is the work by Cranmer et al. [88]. It is also worth noting that Weatheritt and Sandberg [89] developed a methodology for RANS modelling based on gene-expression programming [90] which essentially produces models that are interpretable since they rely on combinations of basic symbolic expressions. Their models exhibit excellent results in a number of complex geometries where RANS models typically fail, including cases with secondary flows and significant three-dimensionality. Along these lines, a generalization of the constitutive relations used to model the Reynolds stresses can also help to properly predict complex flows [81], and in this sense data-driven model discovery is a promising research direction. An essential aspect to consider when developing ML-based RANS models is their associated uncertainty, an aspect currently being investigated in several studies [91,92].

Another simulation area where ML has shown potential is that of large-eddy simulation (LES). As opposed to RANS, where all the turbulent scales are modelled through the Reynolds stresses, in LES a certain range of (the larger) scales is simulated, and low-pass filtering is employed. The scales with higher wavenumbers (*i.e.*, with smaller wavelengths) are then modelled through a so-called subgrid-scale model (SGS). There are several deep-learning-based methods to develop such SGS models, *e.g.* the approach based on CNNs by Beck et al. [93]. CNNs are widely used in computer vision, and they can exploit the spatial correlations present in the data to enhance their predictions. Thus, they are suitable for making accurate predictions in the context of turbulent flows. Another CNN-based approach was proposed by Lapeyre et al. [94], in their case applied to reactive flows. Despite the potential of these approaches, in principle they rely on data of higher fidelity (*e.g.*, direct numerical simulations, DNSs) to train the deep-learning models. The higher-fidelity data is used to develop an SGS model able to reproduce the behaviour of the non-modelled simulation in the LES case. The limitation of this approach relies on the need of such high-fidelity data, which limits the applicability of these models in a wide range of cases beyond those considered in the training stage. An alternative to this approach was proposed by Novati et al. [95], who used reinforcement learning (RL) to develop an SGS model. The advantage of this method compared with the one above is that it is unsupervised, *i.e.*, it does not require the high-fidelity data as a reference. Furthermore, it is able to develop SGS models that do not necessarily emulate the behaviour of a high-fidelity simulation, but rather establish a condition consistent with the coarse case being simulated. This in principle can favour the generalizability of such LESs. Another area of potential interest within ML is that of wall-model development. Wall models are employed at very-high-Reynolds-number applications, where the whole near-wall region is replaced by a suitable boundary condition [96]. Note that at the moment there is not a clear approach to accomplish this for general cases, given the fact that the flow in the outer region (for example streamwise pressure gra-

dients) significantly determines many of the flow features closer to the wall [97,98]. Another difficulty is the fact that in wall-bounded turbulence over roughness the outer region has no way of knowing whether the wall is smooth or rough, and therefore there is uncertainty regarding the best approach to set these boundary conditions [99]. Despite these limitations, important progress has been made based on data-driven methods relying on physical arguments [100,101]. Deep learning is also providing strategies to be able to map the flow in the outer region and at the wall-normal location where the boundary condition would be set, either through direct prediction [102] or by determining a virtual velocity [103]. Another promising approach relying on ML is that developed by Bae and Koumoutsakos [104], which is based on reinforcement learning to determine the correct boundary information close to the wall.

To summarize, there are a number of promising directions for ML-enabled turbulence modelling for RANS. High-fidelity data can be leveraged, together with modern deep-learning architectures, to produce more sophisticated closures. The main challenge in this is the need of high-fidelity data, which introduces limitations in the generalization of the obtained models. In this sense, exploiting transfer learning [105], so as to improve the range of application of the models, can be one of the key solutions to this. When it comes to LES, although several computer-vision-based solutions have been proposed to supplement the coarse simulation with the unresolved scales in a supervised manner, this approach exhibits problems when performing a-posteriori tests [93]. A more promising approach relies on ensuring that the simulated flow satisfies the required physical constraints, a condition that can be achieved by dynamically adapting the SGS term via deep reinforcement learning (DRL) [66].

### 3.2. Reduced-order models

Fluid-mechanics systems involving turbulence are high dimensional and chaotic, thus it is convenient to develop frameworks to reduce their dimensionality and simplify their analysis. The so-called ROMs are not only relevant when it comes to shedding light on the physics of these systems, but also regarding other applications such as flow control or optimization [106]. Classical methods for ROM development rely on linear algebra, where a widely-used approach is the so-called proper-orthogonal decomposition (POD) [107], which relies on the singular-value decomposition (SVD) as discussed above. The POD framework, which is essentially data driven, is applied to  $N + 1$  three-dimensional snapshots of the flow field  $\mathbf{u}(\mathbf{x}, t)$ , where boldface denotes tensors,  $\mathbf{u}$  is a vector containing the three velocity components at instant  $t$  and  $\mathbf{x}$  is a vector containing the spatial coordinates. The POD of the spatio-temporal signal  $\mathbf{u}(\mathbf{x}, t)$  would yield:

$$\mathbf{u}(\mathbf{x}, t) = \bar{\mathbf{u}}(\mathbf{x}) + \sum_{i=1}^N a_i(t) \phi_i(\mathbf{x}), \quad (5)$$

where  $\bar{\mathbf{u}}(\mathbf{x})$  is the mean flow,  $a_i(t)$  are the temporal coefficients and  $\phi_i(\mathbf{x})$  are the spatial modes. Consequently, POD separates spatial and temporal information in the flow, and because the spatial modes are ranked by their energy contribution to the original signal [107], it is possible to truncate the sum in eq. (5) at a given number of modes, defining a certain level of energy reconstruction. Another interesting related technique is the so-called dynamic-mode decomposition (DMD) [108], which also identifies a low-dimensional model of the original system, with modes ranked by frequency (instead of energy). These methods are linear, and since turbulent flows are inherently non linear, it may be interesting to explore methods which include and exploit such non linearity. An

example of this is the so-called higher-order dynamic-mode decomposition (HODMD) [109], which embeds non-linearity in the temporal dependencies among the snapshots used to develop the ROM. In this sense, ML has great potential to also help develop suitable ROMs able to exploit the massive amounts of data available and the non-linearities characterizing these flows. An early example of this is the approach by Milano and Koumoutsakos [110], who employed deep neural networks to learn the temporal dynamics of ROMs of the near-wall region of turbulent channels. In fact, they established meaningful connections between neural networks and POD, namely the fact that when nonlinearities are removed from neural networks one recovers POD when developing ROMs. Other types of ROMs purely based on HODMD (a tool generally used in fluid dynamics to identify the main flow physical patterns) have been used to predict the temporal evolution of numerical databases in transitional [11,111] and turbulent flows [12]. The advantage of these models is that they were based on physical principles (driving the main flow dynamics), although the type of models used were purely linear, limiting their application in some case. Recently, Le Clainche et al. [4] combined HODMD with NNs to develop a model capable to predict the wall shear stress in a channel with a porous wall leading to drag reducing/increasing effects (anisotropic turbulence). The model used HODMD to identify the main instabilities leading the flow dynamics. The non-linear interaction of the HODMD (also known as DMD) modes was also considered in the model to then combining this information (related to the flow physics) with simple architectures of NNs to predict the wall shear stress for long time intervals. Based on the same idea of developing predictive ROMs containing information modelling the flow physics, Abadía-Heredia et al. [3] combined POD modes with NNs using CNNs or LSTM architectures to predict the temporal evolution of the flow using data from the transitory of CFD simulations. In all these examples, using data dimensionality techniques based on the physical principles describing the flow dynamics, allowed to use one-dimensional architectures for the NNs. López-Martín et al. [5] also show that using both CNNs or LSTM it is also possible to accurately predict the temporal flow evolution from transient CFD databases, but using three-dimensional NNs architectures instead.

The low-order model of the near-wall cycle of turbulence proposed by Moehlis et al. [112], which exhibits all the relevant phenomena present in the high-dimensional data, has been used as a platform to assess the feasibility of ML techniques to predict the temporal dynamics of the flow. For instance, LSTM networks were used by Srinivasan et al. [14] to model the system, obtaining excellent predictions of turbulence statistics, Poincaré maps and Lyapunov exponents (following system perturbation). Other successful approaches to model the temporal dynamics in Ref. [112] include the Koopman-based framework (with non-linear forcing) by Eivazi et al. [113] (which can provide in some cases better short-term predictions than the LSTM at a reduced cost) and the technique based on physics-constrained reservoir computing by Doan et al. [114] (which is able to accurately model extreme events). Borrelli et al. [115] employed deep learning to predict the temporal dynamics of a minimal turbulent channel, achieving very good results, in line with those achieved for low-order models of turbulent flows. Finally, AEs are another very promising method for modal decomposition, because they enable developing non-linear ROMs. CNNs have been employed for AE-based modal decomposition [116,117], and Eivazi et al. [41] have shown their applicability to complex turbulent flows. In fact, they showed significant reductions of the number of modes necessary to reconstruct the original system compared with POD, while retaining orthogonality of the modes.

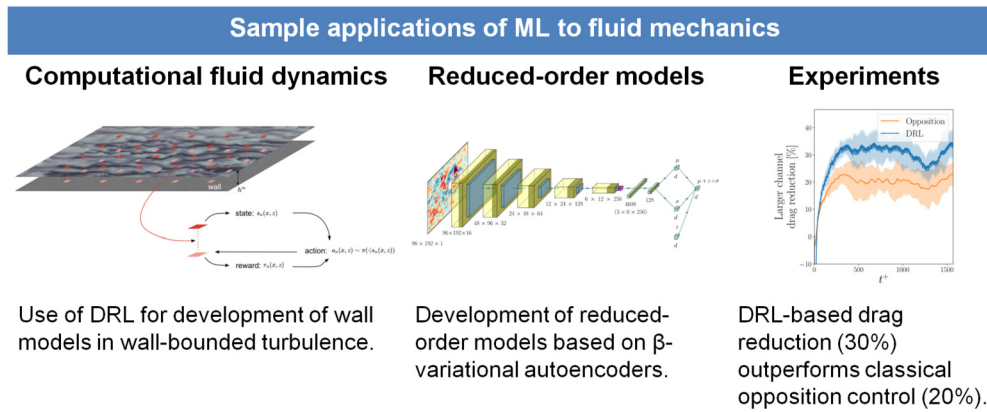


Fig. 9. Sample applications of ML to fluid mechanics, focusing on computational fluid dynamics, reduced-order models and experiments. Figures extracted from Refs. [41,67,104], with permission from the publishers.

### 3.3. Experiments

Experiments can also benefit from the numerous developments in the ML field. In general, exploiting the vast amount of data available in numerical simulations to improve experimental procedures is a very promising endeavour. Kim and Changhoon [118] have shown that the measured wall-shear stress can be used to predict the heat flux in a wall-bounded flow by means of deep neural networks, thus paving the way to augmented experimental data when no thermal quantities are measured. Another important area within experiments is that of non-intrusive sensing, i.e., the possibility to measure at the wall and predict the flow behaviour farther from the wall. This also has important applications in the context of flow control, as discussed below. Guastoni et al. [105] showed that fully-convolutional neural networks (FCNs) can be used to predict the velocity fluctuations above the wall in a turbulent channel, using as input the wall-shear-stress components and the wall pressure. Note that in FCNs all the layers throughout the network are based on convolutions, whereas this restriction does not apply in the more general CNNs. This is a very important result, because the wall quantities are easier to measure without disrupting the flow. Note that they obtained excellent predictions in the near-wall region, although their quality deteriorates farther away. Furthermore, they showed that deep neural networks are more suitable than traditional prediction methods such as linear stochastic estimation (LSE) [119] or extended POD (EPOD) [120] to predict turbulent flows. Güemes et al. [121] proposed an extension to the FCN approach by combining it with POD, such that the neural network is used to predict the temporal coefficients of the POD modes of the flow on a certain plane. Later, Güemes et al. [27] extended this work to the implementation of generative adversarial networks (GANs), which are able to perform super-resolution tasks. The GANs-based approach enabled performing accurate and robust predictions from few sparse measurements at the wall, as opposed to the previous studies, which required finely-resolved wall inputs. Another important framework for experiments is that of physics-informed neural networks (PINNs) [122]. This is a framework which exploits the machinery around the training process of neural networks (mainly the automatic differentiation) to formulate and solve partial differential equations (PDEs). Note that PINNs are currently being used to augment experimental measurements [123] and to improve the quality of the measurements in fluid-mechanics problems [124,125]. It is important to note that experiments in fluid mechanics rely on a number of empirical corrections, where data-driven methods can help to develop more general correction schemes [126,127]. And as mentioned above, ML is also helping to develop more robust flow-control strategies. Non-intrusive sensing, particularly combined with opposition

control [128], is an excellent tool to enhance the experimental implementation of flow-control algorithms. Other data-driven methods for flow control include Bayesian regression based on Gaussian processes [23], as done in the context of boundary-layer control by Mahfoze et al. [129], and genetic programming, which has been recently used to control several complex flows [130,131]. Perhaps the most promising approach for control of fluid flows is DRL, which is a data-driven technique capable to discovering novel and more effective control solutions, as proved by Rabault et al. [132] in the case of the two-dimensional flow around a cylinder. An application of DRL to control turbulent channel flow has been documented by Guastoni et al. [67], and an overview of DRL applications to turbulent flows, including flow control, has been provided by Vinuesa et al. [31].

In Fig. 9 we show several examples of applications of ML-based methods to fluid mechanics, all discussed in this section.

## 4. Machine learning in aerodynamics

Improving aerodynamic design in aircraft is one of the main goals of the aerospace industry. Aircraft wings, airfoil and turbine engine blades are some of the aircraft parts that need to be studied in detail to provide more efficient designs reducing drag (minimizing pollutant emissions) and noise (diminishing airframe and engine noise). The aerospace industry is already concerned about some design concepts that could result in more efficient designs reducing noise and drag. For instance, high aspect ratio wings with adaptive sections are connected to reduced vortex and wave drag in the aircraft, while to reduce friction and pressure drag it is necessary to develop flow control strategies to promote laminar flow, reducing the drag of the turbulent skin friction and controlling the unsteady flow separation and turbulent noise source. To reduce critical loads in the aircraft, in addition to control flow separation, it is important to control the presence of shock waves and unsteady loads for instance delaying the onset of buffeting, preventing flutter [8], which is induced by linear aerodynamics, or controlling other aeroelastic responses leading to limit cycle oscillations (LCO), which are induced by aerodynamic non-linearities [2]. Optimization, flow control and aeroelasticity are challenging topics that the current aerospace engineering industry need to enface during the next upcoming years to be able to develop more efficient, less pollutant and more silent aircraft designs.

CFD numerical simulations and wind tunnel experiments are generally carried out to extract information connected to the aircraft performance, with the aim at using such knowledge to improve aerodynamic design. Aerodynamic data gathered from numerical and experimental sources can be divided in two groups: (i) distributed data, which includes skin friction and surface pressure



distribution, and (ii) integral quantities, which include drag, lift and moment coefficients [133]. However, the high cost associated to the performance of numerical simulations (large computational times and memory resources) and experiments (elevated monetary cost), encourages researchers to use aerodynamic models to solve design and multiphysics problems.

The lack of accuracy in classic aerodynamic models, which rely on empirical laws and classical theory (potential flow), combined with the rapid advance in data science and the current possibilities to generate and store large amount of high-fidelity numerical and experimental databases, motivate researchers to develop novel strategies to develop data-driven aerodynamic models. These models are more reliable and accurate than theoretical models and capable to provide relevant information connected to the physical insight of the problem under study, which opens new possibilities for applying the data-driven models to new flow control, optimization, aeroelasticity, flight dynamics and flow reconstruction applications [134].

Aerodynamic data-driven models can be divided into three groups: (i) semi-empirical models, which obtain the drag, lift and moment coefficients of airfoils as the result of linear or non-linear differential equations in charge of describing the flow dynamics, (ii) surrogate models, which are generally used in steady applications to find unknown variables using as general input geometrical parameters or flow conditions (i.e., Mach number, Reynolds number, ...), and (iii) ROMs, which include models for feature extraction (i.e., POD [135], DMD [136]) that are generally used in large dimensionality flow datasets, and models for system identification based on input-output aerodynamic data (i.e., neural networks, Kriging [137,138], Volterra series [139,140], other linear and non-linear models such as Eigensystem Realization Algorithm -ERA, Auto-Regressive with Exogenous input -ARX, Non-linear ARX, Auto Regressive Moving Average -ARMA, Non-linear ARMA with exogenous input -NARMAX, et cetera).

During the last years, AI, ML and more specifically, NNs, have emerged as highly potential tools to improve and transform engineering design and analysis, offering flexible, robust and versatile tools to develop ROMs and surrogate models, both suitable for finding optimal solutions in aerodynamic optimization [141,142]. ML applications to improve aerodynamic design solving optimization problems are multiple. For instance, NNs have been used to develop models (i) capable to accelerate CFD simulations, which increases the optimization efficiency since it is possible to calculate the integral quantities and distributed data of several design geometries in a short time simplifying the possibilities of finding the optimal designs [143,144], (ii) also these models can be used to reconstruct high-fidelity data from sensor measurement [145] or (iii) to provide a general idea about the proper combination of aircraft weight, wing area and engine thrust helping in the early stages of aircraft design to take decisions [146]. ROMs can be combined with other analytical tools, can predict the aerodynamic loads and can represent aerodynamic systems, allowing several types of analysis at a reduced computational cost [147].

Even though ML is an emerging science in most of the subject connected to aerospace engineering, NNs have been used since the 90s to model aerodynamic loads for flight dynamics [148]. In the recent extensive and detailed review of Kou & Zhang [134], the authors present several data-driven modelling methods applied to improve aerodynamic performance, including several references related to NNs applications since the 90s. The authors conclude as a general observation for the field of data-driven aerodynamic modelling that the future trends in this field should focus on developing novel data fusion techniques, allowing to combine knowledge from numerical simulations with different fidelities and experiments, maximizing in this way the capabilities of the models developed. The authors also reveal the need of developing new mod-

els less dependent on data and adding information about the flow physics, which includes interpretability on the results. The need of solving more complex applications, such as turbulent flows, complex geometries, industrial problems, three-dimensional and multi-scale flows (i.e., massive flow separation), etc., encourages researchers to find novel alternatives to develop non-linear accurate models to improve aircraft designs. Evaluating the real applicability of these models by means of using uncertainty quantification tools, is critical for the industry of the future. These conclusions, generalized for data-driven modelling, can also be particularized to extend, and improve ML and NNs applications to solve aerodynamic problems.

To summarize, ML methods are suitable to improve aircraft aerodynamic performance because they allow for the optimization of complex systems using large amounts of data. Traditional methods for aerodynamic design and optimization are limited by the computational power and the costly models that can be used. Machine learning methods can handle large amounts of data and are able to identify complex relationships that may not be apparent using traditional methods. ML methods can help solve several issues related to aircraft aerodynamic performance. For instance, (i) ML algorithms can optimize complex systems with a large number of variables, allowing for the design of more efficient and effective aerodynamic components and systems; (ii) ML algorithms can analyze large amounts of data from sensors and other sources to identify patterns and relationships that can be used to optimize aerodynamic performance; (iii) ML algorithms can reduce the time and cost required for the design and optimization of aerodynamic components and systems by automating the process and reducing the need for physical prototypes. Fig. 10 summarizes the benefits of using ML methods for aerodynamic design optimization, solving the limitations of traditional methods.

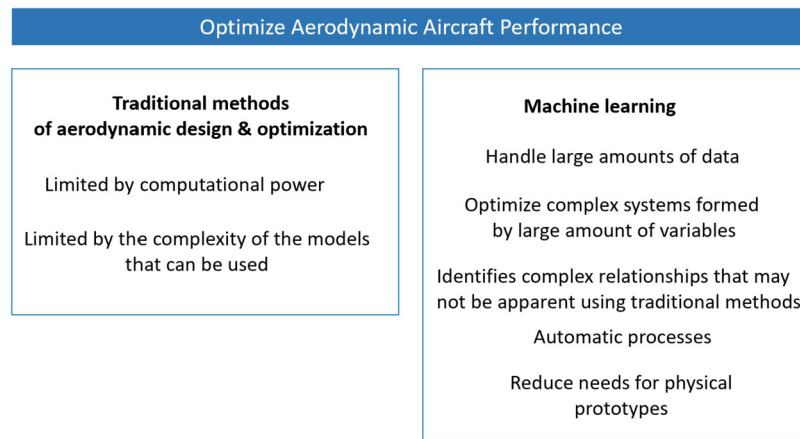
However, there are also limitations to machine learning methods in the context of aircraft aerodynamic performance improvement. Some limitations include: (i) lack of interpretability: ML algorithms can be difficult to interpret, making it challenging to understand how they arrived at a particular solution or optimization. This can limit the ability to understand the trade-offs involved in a particular design or optimization decision; (ii) lack of transparency: ML algorithms can be opaque, making it difficult to understand how they are making decisions. This can be particularly concerning in safety-critical systems such as aircraft; (iii) data quality issues: ML algorithms rely on high-quality data to be effective. If the data used to train the algorithm is incomplete or inaccurate, it can lead to sub-optimal designs or optimizations.

Overall, while ML methods offer a powerful tool for improving aircraft aerodynamic performance, it is important to be aware of their limitations and to use them appropriately in the design and optimization process. Details of some relevant ML applications in aerodynamics are presented in the following sections.

#### 4.1. Aerodynamic coefficients estimation

The performance of the ML model depends on the type of identification method, the model structure and the type of data analyzed. Flight testing, experiments and numerical simulations are different data sources to obtain unsteady aerodynamic forces. Using ML tools and considering data fusion techniques, which allows to combine data from different fidelities and types (experimental and numerical), would contribute to maintain the proper balance between cost of data generation and model accuracy [134]. However, most of the models, only consider data coming from a single source. In this line, it is possible to mention the representative work of the following authors, ordered in chronological order since the 90s, who applied NNs for the identification of aerodynamic forces (lift and drag). More specifically, Linse & Stengel [148] used





**Fig. 10.** Capabilities of machine learning methods to solve the limitations found in traditional methods for optimal aircraft design optimization.

NNs to estimate the aerodynamic coefficients, presenting a novel methodology that was based on extracting additional training information about the first partial derivatives of the aerodynamic coefficients. The method was successfully tested in flight test data generated numerically, solving a non-linear problem of twin-jet transport, and used for non-linear control of aircraft. Suresh et al. [149] used RNNs to identify the dynamic stall effect of rotor blade on an OA212 airfoil. The authors predicted the lift coefficient in a case at high angle of attack using experimental databases and intended to show the generalization capabilities of the method presenting a way to generalize the error to model training. More recently, Secco & Mattos [150] developed a surrogate model using NNs to predict aerodynamic coefficients of transport airplanes with high accuracy. The NN-based model used data from about 100,000 cases calculated with CFD numerical simulations and the flight condition, airfoil geometry and wing planform were used as an input. The computational time to estimate the aerodynamic coefficients was  $\sim 4000$  times smaller than the time required by the CFD solver. This model was also successfully used for optimization design. Recently, Hou et al. [151] used NNs to study the effect of aerodynamic disturbances (due to manoeuvres or gusts) on the wing surface using pressure measurements. The database was generated by an inviscid vortex method in a two-dimensional flat plate. NNs are used to estimate the angle of attack and the leading-edge suction parameter. The authors first combine CNNs and RNNs providing an accurate estimation these variables. In a second approach, a novel and more accurate approach is presented, where ML is integrated in a dynamical system framework to further correct the prediction using a smaller number of parameters. The authors also show that adding random noise to the database it is possible to prevent overfitting. In the same year, to overcome the limitations in adapting input variables with different orders of ROMs, Wang et al. [152] proposed using a fuzzy scalar radial basis function NN, which combines fuzzy rules and radial basis function neural networks (RBFNNs), enhancing in this way, the generalization capabilities of the model to complex dynamical systems. This model was successfully tested to estimate the unsteady aerodynamics in transonic flow airfoils. RBFNNs were previously used by Ghoteyshi et al. [153] to compute the unsteady aerodynamic loads in an airfoil with pitching and plunging motion.

#### 4.2. Aeroelasticity

Studying in detail aerodynamic linear and non-linear responses, for instance producing flutter (linear) or leading to LCO (non-linear), is extremely important to reduce critical loads in the aircraft. Non-linear dynamical modelling is a challenging task, since

instabilities can be present in both training and simulation, additionally, purely non-linear models sometimes are unable to identify the main dynamics in systems with dominant linear behaviour. The possibilities of using NNs to combine both, linear and non-linear effects to develop models, have motivated several authors to focus their work on developing ML-based models to estimate aeroelastic responses in the aircraft. To name a few, Marques & Anderson [154] used a finite memory NN to model the nonlinear transonic aerodynamic aircraft response, motivated by the modelling difficulties generally found to control aeroelastic responses. The NN model used a supervised training process with multiple input-output sets from numerical data solving Euler equations. The finite memory NN architecture allowed to identify the memory in the hidden layer of the finite impulse response. Zhang et al. [155] developed a non-linear unsteady aerodynamic model using RBFNNs employing the aeroelastic system self-excited vibration signal as input. The non-linear aerodynamic fluid-structure interaction (FSI) model was coupled with the structural equations of motion, accurately estimating the LCO change with velocity in a transonic flow with large shock motions. Mazhar et al. [156] used NNs to successfully interpolate the pressure loads on unmanned aerial vehicles (UAVs). The aim was to present an accurate model suitable for UAV structural design, which was based on one-way FSI. The performance of NNs was superior compared to standard interpolation techniques using high order polynomials. Mannarino & Mantegazza [157] used recursive NNs to develop a model using high-fidelity numerical data solving a non-linear aerodynamic model. The NN model was successfully used to determine aeroelastic limit cycles using two different approaches: (i) through time marching simulations and (ii) through direct time collocation. The model was successfully tested in the Benchmark Active Control Technology (BACT) wing [158], which consist o a rectangular wing pitching around its mid chord axis and with a NACA0012 airfoil section. To model non-linear unsteady aerodynamics, both at constant and varying flow conditions, Kou & Zhang [18] successfully tested the performance of multi-kernel NNs, which are able to improve the generalization capabilities of the model. They also succeeded in using recursive NNs to predict non-linear aeroelastic behaviours and aerodynamic forces in a NACA 64A010 airfoil for multiple flow conditions in transonic flow. More recently, Li et al. [159] used NNs, more specifically, LSTM networks to develop a model capable to predict non-linear unsteady aerodynamics, which was successfully tested for various Mach numbers on a NACA 64A010 airfoil that was pitching and plunging in a transonic flow. The model accurately captured the aerodynamic and aeroelastic capabilities of the dynamical system, showing the generalization capabilities of this NN architecture. LSTM architectures

are perfect to capture the time-delayed effects of unsteady aerodynamics without the need of including any additional information about the orders selection (in unsteady aerodynamics it is necessary to consider the time-delayed effects from the input-output dynamical system).

#### 4.3. Design optimization

Design optimization is also a highly important discipline in aerospace engineering, where machine learning has recently gain strong relevance in the field. Aerodynamic design is related to airplane wing, airfoil geometry, UAVs, turbine engine blades, flight conditions, etc. In the recent review article by Li et al. [142] the authors show the application of machine learning for design optimization, including design space, objective modelling, constraint modelling and optimization convergence rate. Optimization should be smart, efficient, and capable to easily balance design space (*i.e.*, shape, morphed geometry, etc.) and modelling efforts. It is notorious that design optimization strongly influences the performance of the airplane in the fields of aerodynamics, combustion, or aeroacoustics. In the recent review article by Sun & Wang [141] the authors summarize the main contribution of the research community in the use of NNs for surrogate modelling applied to design optimization. The article discusses the effect in optimization of the type of NN selected, the type of data used for the training and if special data treatment is needed, and it reveals the potential of using NNs for optimization purposes. For instance, the authors show that multi-level surrogate modelling, where various levels of information are used, is highly effective to obtain the optimal design area with a small computational cost. NNs can be combined with several optimization algorithms and can also make use of large databases, extracting relevant information suitable for optimization purposes. In this review paper, the authors reveal the relevance of deep learning in aerodynamic design, which reduces the manual interaction in the optimization process. In the line of using NNs for design optimization, it is worth mentioning the work carried out by the following authors during the last 10 years. Chen et al. [160] used NNs, used a supervised algorithm combining Self-Organizing Map networks, for selecting referenced airfoils, with Back-Propagation, to learn the relationship between the aerodynamic performance and airfoil geometry, to estimate airfoil aerodynamic characteristics for aerodynamic design. Sun et al. [161] introduced an inverse design method using NNs in trained wing and airfoil databases. Kozel & Leifsson [162] developed a methodology for solving transonic airfoil aerodynamic shape optimization problems. The authors combined a low-fidelity CFD model (coarse mesh and relaxed convergence criteria) with NNs, in charge of correcting the low-fidelity simulations with space mapping. The method was able to maximize lift and minimize drag in a two-dimensional transonic flow. Oktay et al. [163] combined NNs with the algorithm of simultaneous perturbation stochastic approximation (SPSA) to estimate the drag coefficient optimum value of a fuselage. By this combination, the authors used SPSA without any objective function, supported by the NNs capabilities, which were trained using wind tunnel data. The NN was generalized to estimate the drag coefficient of different geometry fuselages (different fuselage shape parameters). More recently, Xu et al. [164] used NNs to develop surrogate models, later employed in a genetic algorithm to select optimized airfoil designs for transonic supercritical conditions, with the idea of minimizing time-averaged drag and reducing buffet effect (to prevent structural damages).

The identification of aerodynamic forces, the estimation of linear and non-linear responses in the aircraft (aeroelasticity and FSI) and design optimization, are three main topics where researchers have put effort since the past to advance in the field of aerospace engineering and aircraft optimal design and manufacturing. Using

ML in the field, allows to develop accurate ROMs and surrogate models, suitable for a wide range of problems, from incompressible to transonic compressible flows, at various flight conditions. Nevertheless, the lack of generalization in the methodologies proposed, as well as the possibility of merging different types of data (numerical and experimental), with different types of fidelities, encourages researchers to continue advancing in the field, combining ML with data fusion techniques that will also consider the physical insight of the problem under study. Developing highly accurate predictive model based on machine learning, which will be applicable to solve different problems and requiring small manual interaction, will definitely advance the field of aerospace engineering and the achievement of sustainable aviation.

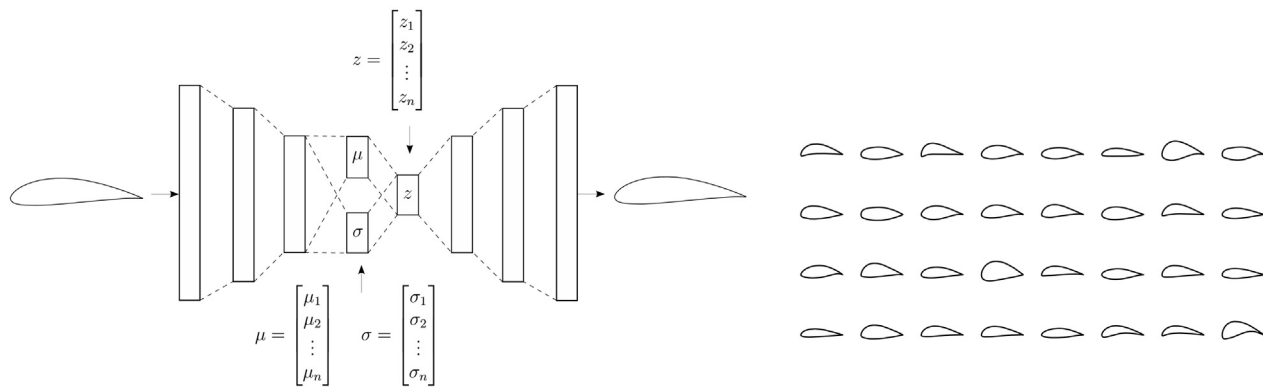
#### 5. Machine learning in aeroacoustics

The recent review by Bianco et al. [165] summarizes the advantages of using ML for modelling pure acoustic phenomena. In the review, the authors cover a variety of topics including source localization in speech processing, source localization in ocean acoustics, bioacoustics and environmental sounds. The main ideas in this review include localizing noise sources and identifying acoustic features or patterns using ML. The references included in Ref. [165] show that ML-based methods can provide enhanced performance when compared with conventional signal-processing methods. However, ML-based techniques are limited by the existing data, since large amounts of data are typically necessary for training.

More relevant to aeronautical applications is the field of aeroacoustics and computational aero-acoustics (CAA). Aeroacoustics problems are generally more complex than pure acoustics once since the coupling between the turbulent flow interacting with the geometries, the associated acoustic generation and subsequent propagation need to be considered. The most important sources of noise in aeronautical applications include aerodynamic noise, jet noise and rotating machinery noise, see Moreau [166] for the more recent advances on these fields. To the authors' knowledge there is no review covering the use of ML to model aeroacoustics or to accelerate CAA simulations.

Aeroacoustics and CAA exhibit two distinct problems: noise generation (near field) and noise propagation (far field). Although direct simulation of acoustics (near and far fields are solved together) is possible, and only for small cases to date, the hybrid approach is generally preferred to lower the cost in large complex simulations. Acoustic analogies revolutionized the field of aeroacoustics in the 1960s (see Lighthill [167,168]), and have enabled a certain degree of decoupling between the generation and propagation of noise. Following these analogies, equivalent acoustic sources can be extracted from the well-resolved near-field region. These can be subsequently modelled and inputted into a new simulation to predict their propagation. Simulations of aero-acoustics can use this decoupling to enable faster simulations, since the near field that requires highly-accurate and costly flow computations can be performed separately from the acoustic propagation (wider region but coarser mesh). The field has evolved considerably due to the development of faster computers that enable larger computations, see Lele & Nichols [169] or Moreau [166].

Following the acoustic analogies, models or correlations can be applied to either the generation or the propagation regions, or even to both regions at the same time. The model (or correlation, regression) can use NNs. When considering only propagation within the aeroacoustic context, we find only a few works with interesting ideas. For example, Alguacil et al. [170] predicted the propagation of acoustic waves using deep CNNs and developed deep-learning surrogates to characterizes the transfer function for the propagation and scattering of acoustic waves in quiescent flows



**Fig. 11.** Variational autoencoder for airfoil parametrization with  $n$ -dimensional latent variables used for aeroacoustic optimization (left).  $\mu$  and  $\sigma$  denote the mean and standard deviation of the latent variables  $z$ . Smooth airfoils generated with a variational autoencoder (right). Figures adapted from [178].

[171]. Tang et al. [172] studied dynamic interactive sound propagation by means of neural-network-based-learned scattered fields coupled with ray tracing to generate specular, diffuse, diffraction, and occlusion effects. Kužnar et al. [173] proposed to use linear-regression models to improve vehicle acoustics.

When considering generation and propagation of aeroacoustics, artificial neural networks are being deployed to model/correlate a variety of input parameters to sound pressure levels. The use of ML in acoustics is scarce and has been restricted to the automotive industry or simple flows.

Beigmoradi et al. [174] used Taguchi methods and NNs to reduce the computational complexity for the aeroacoustics of a rear end simplified car model. The model was used for optimization using genetic algorithms. Rüttgers et al. [175] trained a deep NN using a Lattice-Boltzman method to predict the aeroacoustic far field of a 2D square domain that include randomly distributed rectangular and circular objects, and monopole sources. Moeen Uddin et al. [176] used acoustic analogies based on RANS and Ffowcs Williams and Hawkings, to predict the aero-acoustic tire noise at near-field and far-field receivers around the tire. Artificial neural networks-based regression were used to study the highly non-linear relationships in the system: between A-weighted sound pressure level and tire parameters (Groove depth, Groove width, Temperature and velocity). Abreu et al. [177] have used spectral POD to characterize the wave packets originated at the trailing edge of turbulent airfoils. Note that these structures have important implications in aircraft noise generation, and ROMs aimed at predicting their behaviour can help to design more-silent aerodynamic solutions. Finally, Kou et al. [178] used autoencoders to improve the optimization of airfoils where a multidisciplinary objective function combined aerodynamic and aeroacoustic targets. Fig. 11 shows the variational autoencoder used for airfoil parametrization, which challenged a classic CST parameterization method. It was observed that variational autoencoders can represent a wide variety of geometries, with only four encoded variables, leading to efficient optimizations, which resulted in more silent optimal shapes.

The main conclusions of this section are that there is huge potential in using ML methods for aeroacoustics. This field is at its infancy and has not yet been deployed to solve aeronautical applications. The main difficulty is the that the aeroacoustics transfer functions (linking flow/geometrical parameters and far field acoustics) are not always smooth and care is needed when performing tasks involving derivatives or such functions (e.g., gradient or adjoint based optimization). NNs do not always perform well when modelling non-smooth functions [179], and therefore their use in this field needs caution. A possible way to alleviate the lack of smoothness is to decouple noise generation and propagation and use NNs separately for each part. Alternatively one may regularize

the transfer function using for example smooth flow models (e.g. based on RANS equations).

To summarize, ML methods are needed to reduce aircraft acoustic footprint because they can help optimize aircraft designs and operational procedures to minimize noise. Noise pollution is a significant concern for both the aviation industry and the communities that live near airports, and reducing aircraft acoustic footprint can improve the quality of life for those communities. ML methods can help solve several issues related to reducing aircraft acoustic footprint. For instance: (i) ML algorithms can be used to optimize the design of aircraft components to minimize noise. For example, they can be used to optimize the shape and size of engine components or the design of the aircraft's wings to reduce noise; (ii) ML algorithms can be used to optimize operational procedures such as takeoff and landing procedures to reduce noise. For example, they can be used to identify the most efficient and quietest takeoff and landing procedures for a given aircraft; (iii) ML algorithms can be used to predict the noise levels of different aircraft configurations or operational procedures. This can help designers and operators optimize aircraft designs and procedures to minimize noise. However, there are also limitations to ML in the context of reducing aircraft acoustic footprint. These limitations consider the following aspects: (i) ML algorithms can be difficult to interpret, making it challenging to understand how they arrived at a particular solution or optimization. This can limit the ability to understand the trade-offs involved in a particular design or operational decision; (ii) ML algorithms rely on high-quality data to be effective. If the data used to train the algorithm is incomplete or inaccurate, it can lead to sub-optimal designs or optimizations; (iii) reducing aircraft acoustic footprint is a complex problem that involves many variables, such as the aircraft design, operational procedures, and environmental factors. ML algorithms may not be able to capture all of the nuances of the problem, leading to sub-optimal solutions. Overall, while machine learning methods offer a powerful tool for reducing aircraft acoustic footprint, it is important to be aware of their limitations and to use them appropriately in the design and optimization process. Other approaches, such as regulations and community engagement, may also be necessary to address noise pollution from aviation.

## 6. Machine learning in combustion

The ultimate goal of combustion research is to develop accurate, generalizable and predictive models to describe the phenomena occurring in combustion devices, which primarily involve turbulent flows. The recent years have witnessed significant advances in the fields of DNS and LES, as well as in the development and reduction of chemical kinetics mechanisms for hydrocarbon fuels. The

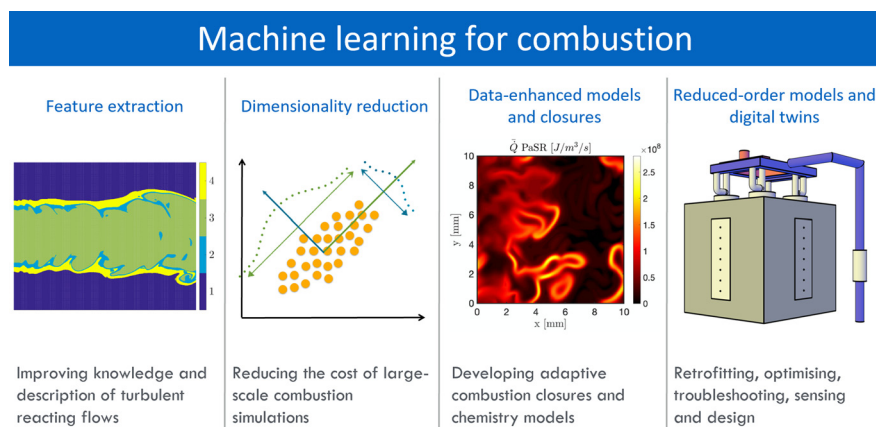


Fig. 12. Summary of the main applications and impact perspectives of ML in combustion.

currently available computing power allows to perform parametric DNS studies closer to the conditions of practical interest. LES is becoming more and more present in industry, although RANS remains the preferred option for its affordable computation cost. Despite the significant progress, existing modelling approaches generally lack generality and predictivity [180]. The main challenges are associated with the number of species involved in combustion processes, the small scales and the non-linear turbulence-chemistry interactions [181–183]. ML methods can contribute to combustion science and treat some of the previously unmet challenges, providing interpretable feature extraction techniques, delivering generally applicable approaches to locally adapt comprehensive chemical mechanisms and sub-grid models, designing new closures to parametrize the unresolved fluctuations, and developing ROMs for fast and yet-accurate system evaluations [184].

In the following, the current state and perspective associated to the use of ML in combustion physics are reviewed and discussed. A summary of the main applications and perspectives is shown in Fig. 12.

### 6.1. Data analysis and feature extraction

The field of combustion science has traditionally dealt with vast amounts of data from experiments and large-scale numerical simulations. In fact, combustion has been dealing with big data for almost thirty years [185] thanks to advances in petascale computing, which allows for detailed simulations of laboratory-scale turbulent flames with detailed chemistry. While exascale computing will enable closer-to-reality conditions, the study of direct numerical simulation (DNS) of turbulent combustion is already a crucial research area for developing and validating models [186–188]. There have been many techniques developed over the past 50 years to handle and process combustion data from experiments and simulations, which have relied heavily on domain expertise and heuristic algorithms. However, these analyses can also be viewed as early applications of machine learning [189]. Several early studies focused on the use of dimensionality reduction techniques such as PCA [44], with the objective of identifying empirical low-dimensional manifolds in combustion systems [47,52,190–197]. The analysis focused both on experimental [198–203] and numerical simulation data [204–211]. The increasing availability of DNS data of complex configurations is shifting the interest towards computational data, for the possibility of mining large sets and accessing variables not available from experiments (e.g., source terms and turbulence-related variables). The dimensionality reduction offered by PCA can be limited for combustion data. Indeed, PCA is a multi-linear technique and typical reacting flow data exhibit strong non-linearities. This implies that additional components are

identified by PCA to make up for the data non-linearity, leading to an overestimation of the true problem dimensionality. To overcome the limitations of classic PCA with non-linear systems, the application of non-linear methods has been proposed in the community, including non-linear and KPCA [196,209,212], isometric mapping (IsoMap) [213], T-distributed Stochastic Neighbor Embedding (T-SNE) [214,215] and AEs [216,217]. Recently, data-driven approaches based on CNNs were proposed for regime identification [218,219]. Moreover, non-linear regression based on Neural Networks [220] was employed to identify the main features of Moderate or Intense Low-oxygen combustion (MILD)<sup>1</sup> and predict the co-existence of different combustion modes, paving the way to the adaptive selection of closures during numerical simulations. Non-linear methods greatly improve the dimensionality reduction potential. However, they become computationally intractable for large data sets (KPCA and IsoMap) or involve many hyper-parameters to optimize and cross validate (t-SNE, NNs and CNNs). Finally, non-linear algorithms extract features that can be more challenging to interpret and that cannot be easily converted into predictive modelling approaches. An alternative to non-linear methods is to identify local principal curves [221] and locally-linear embeddings (LLE) [50], as well as to perform local modal decomposition, i.e., applying PCA locally (LPCA) [51]. Algorithms based on the definition of an *a priori* prescribed (supervised) conditioning variable (such as mixture fraction), and on the iterative (unsupervised) minimization of the reconstruction error have been investigated in the context of combustion data analysis [52]. The latter has proven particularly effective and even competitive with more sophisticated approaches such as AEs for dimensionality reduction [216], as well as for clustering tasks [222,223]. Coupling vector quantization with dimensionality reduction using PCA (VQPCA) [51,52] provides robust classification algorithms, less prone to overfitting and applicable to conditions different than those met during training [224]. Indeed, VQPCA requires the determination of only two hyper-parameters, the number of eigenvalues, controlled by the amount of desired variance, and clusters, based on quantitative metrics [225–228] as well as the statistical representativeness of the clusters [222]. The classification properties of LPCA have been exploited to identify the leading principal variables (PV) and processes in different regions of the state-space [229–232], and to develop adaptive chemistry models [222,223].

<sup>1</sup> MILD combustion is a combustion technology able to ensure very high combustion efficiencies with no soot, and very low nitrogen oxides and noise emissions, compared to conventional combustion regimes, due to the reduced temperature peaks and macro-scale homogeneity. In addition, it is a fuel-flexible technology.



## 6.2. Dimensionality reduction, classification and adaptive chemistry

The large dimensionality of combustion systems has driven the development of dimensionality reduction approaches. Considering the impact of large kinetic mechanisms on the computational time of detailed numerical simulations [182], many different strategies have been developed in the last two decades for the systematic generation of reduced mechanisms, including Sensitivity Analysis [233], PCA [233,234], Path Flux analysis [191], graph-search methods such as Directed Relation Graph (DRG) [235], DRG-aided sensitivity analysis [236], DRG with error propagation (DRGEP) [237], lumping [238], quasi-steady state approximation (QSS) using Computational Singular Perturbation (CSP) [239], Tangential Stretch Rate [240], Level of Importance (LOI) [241], via production/consumption analysis [242], and using error estimation [243]. An effective strategy to reduce the burden of detailed chemistry is to use adaptive chemistry approaches, which adjust the mechanisms to the local flow conditions [244–249], performing the reduction on the fly. Approaches relying on pre-tabulated libraries of reduced mechanisms [222–224,250–252] have been recently proposed, thanks to the development of efficient classification algorithms. This opens the way to the use of reduction methods characterized by higher overhead and unsuited for use at runtime, *i.e.*, error-controlled approaches [237,253,254] and methods relying on the dynamical description of the system [239,240].

In order to cope with the large number of uncertain parameters in comprehensive chemical mechanisms, uncertainty quantification and optimization have been adopted in the process of chemical mechanism development [255–258]. In particular, several methodologies have been designed to solve the so called “inverse problem” [259,260], and improve predictions [261] based on increasingly available experimental data. Notably, the use of genetic algorithms [262,263] has gained momentum to deal with the highly structured objective functions, typical of mechanism optimization, often characterized by the presence of multiple local minima/maxima [259]. Considering the increasing amount of kinetic parameters from theoretical *ab initio* calculations [258,264], one challenge in optimization approaches is the estimation and/or determination of plausible uncertainty ranges based on the level of theory adopted throughout the calculation protocols for electronic structures, potential energy surfaces and phenomenological reaction rate constants [263,265,266]. The availability of highly-efficient optimization algorithms and tools [267] is pushing research in the development of highly reduced chemical mechanisms for complex fuels, to allow high-fidelity simulations with realistic chemistry rather than global mechanisms [268,269]. In this line of research, new frameworks combining dimensionality reduction and non-linear regression (*e.g.*, using b-spline interpolants and NNs) have been proposed to replace the tabulation of reaction rates in flamelet/progress variable [270–275] as well as in finite-rate chemistry approaches [276,277]. In particular, neural frameworks were recently proposed for the efficient handling of ordinary differential equations in chemical kinetic solvers [278]. The potential of machine learning in reaction pathway discovering was also demonstrated [279,280], opening the way to the automatic generation and tailoring of automatic, mechanisms to specific conditions. Along the same line, machine learning techniques were used to simplify the chemistry of complex hydrocarbons combining data-based models for representative pyrolysis species and foundational chemical mechanisms for the remaining ones [281]. Besides rate-based methods that identify and eliminate redundant species and reactions, approaches based on the re-parameterization of the chemical state-space [282–286] have been further developed and combined to data-driven approaches. In particular, PCA was used to develop reduced-order models for combustion simulations based on the resolution of transport equa-

tions for the principal components [194]. To reduce the number of transported components and improve the reconstruction of the state-space as well as the associated chemical source terms, PCA was coupled to non-linear regression techniques, replacing the (multi-linear) PCA mapping between the reduced PC space and the original one. Different regressions have been used to this purpose including Multi Adaptive Regression Splines (MARS) [195,197], NNs [196,287] and Gaussian Process Regression (GPR) [288]. This modelling approach has been demonstrated for simple reactors [196,288,289] One Dimensional Turbulence (ODT) simulations [196,290], premixed and non-premixed DNS simulations [209,210,291,292], as well as in the context of RANS simulations using kernel density estimation (KDE) [293,294]. Recently, the method combining PCA and GPR (PC-GPR) was applied, for the first time, to the three-dimensional LES simulation of flames D-F [295] and of the Cabra flame [296], showing excellent predictive capabilities. The ability of the method to generalize the formulation of flamelet/progress variables approaches and facilitate the formulation of sub-grid closures (being the selected scalars uncorrelated) appears particularly promising. Recent research indicates that the quality and regressibility of low-dimensional manifolds has importance in building ROMs [296,297]. Recently developed tools to characterize topologies of low-dimensional manifolds from the perspective of non-uniqueness and feature sizes can help address manifold challenges in the future [298,299].

## 6.3. Combustion closures

In turbulent flows, temperature fluctuations can be as high as several hundred Kelvins. Considering that Arrhenius reaction rates are highly non-linear functions of temperature, accurate statistical closures for filtered approaches cannot be based on an expansion about mean properties [182,300]. The task of a combustion model is to provide a description of the unresolved scales based on the information available during a simulation. Combustion models have been often classified into two categories, the flamelet-like<sup>2</sup> [282–286] and PDF-like approaches<sup>3</sup> [301–303]. Besides them, we can also mention reactor-based<sup>4</sup> [304–307] and conditional-moment approaches (CMC)<sup>5</sup> [308]. Data-driven methods are significantly impacting the design of improved sub-grid models for LES and RANS simulations, beside turbulence closures [309,310]. Recently, Deep CNNs have been used by different authors, to improve the prediction of the unresolved flame surface wrinkling [94] in the context of the Thickened Flame model [311], as well as to predict the filtered progress variable source term and the unresolved scalar transport terms in LES [312–314]. NNs and Deep Learning (DL) were employed to develop presumed probability-density-function (PDF) models for conventional [315] and MILD combustion [316]. Data-driven approaches were recently used to assess the validity of existing reactor-based closures in MILD combustion [317,318], to obtain conditional statistics [319,320] in the context of CMC, and for the dynamic assignment of combustion closures [321].

<sup>2</sup> Flamelet-like models are based on the re-parametrization of the thermochemical state using a reduced number of variables identifying a low-dimensional manifold in the composition space on which the evolution of the system is constrained.

<sup>3</sup> PDF-like approaches do not make the assumption of a low-dimensional manifolds and treat the mean/filtered reaction rates exactly, solving the, one-point one-time, joint PDF of fluid composition. Mixing at the molecular level requires a closure in PDF approaches.

<sup>4</sup> Reactor-based models assume that a fraction or the entire computational cell behaves like an idealized reactor (perfectly-stirred reactor or plug-flow reactor), allowing an affordable treatment of finite-rate chemistry.

<sup>5</sup> CMC relies on the concept of conditional manifolds and equations are derived and modelled for the conditional moments of the reactive scalars.

#### 6.4. Reduced-order models for realistic combustion systems

In some applications, a very fast evaluation of a system response is required (for control, optimization, ...), thus limiting the use of time-consuming simulations. In this context, ROMs are used to approximate the underlying hidden relationship between inputs and outputs, using available observations to estimate the system response for unexplored conditions. Black-box approaches have been widely used in combustion to create static input-output maps [322] and for system identification [323], to predict macroscopic quantities such as exhaust gas emissions and temperature and to detect oscillatory patterns such as thermoacoustic instabilities, respectively. In combustion, NNs have been used to evaluate and tabulate reaction rates as discussed above [270–272,324], as well as to estimate modelling errors in LES [186,325]. Black-box approaches can be very powerful although not equipped with the guarantees of physics-based approaches. To cope with that, nonlinear system identification techniques were proposed to enlighten black box approaches by discovering the underlying physics [326,327]. An interesting approach to generate combustion ROMs is the use of reactor network models [328–330], which fall in the category of grey-box models, as they combine a theoretical structure, *i.e.*, the canonical reactors, with data to generate the network. The approach can be also regarded as an example of multi-fidelity methods, relying on high-fidelity tools (*i.e.*, CFD) to construct the reactor network and the lower-fidelity network model 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 the interest for data-driven approaches in this area of research. Projection-based reduced-order models (ROMs), such as principal component analysis (PCA) and other similar techniques, have been applied to both experimental and numerical combustion data in various studies [47,52,190–197]. However, direct evolution of modal coefficients in computational fluid dynamics (CFD) codes, known as intrusive ROMs, has been limited in non-linear problems such as combustion due to the high cost of the associated Galerkin projection and the complexity of integrating them with legacy and commercial CFD codes [331]. On the other hand, non-intrusive projection-based ROMs have been successfully employed, in combination with non-linear regression approaches such as Gaussian process regression (GPR), neural networks (NNs), and polynomial chaos expansion (PCE) [332,333]. In cases where the number of variables of interest is high, training numerous ROMs can be costly and any correlation between variables can be lost. PCA and similar techniques can reduce the number of ROMs required, encoding the original variables into a set of fewer uncorrelated scalars that explain most of the system variance. In the combustion community, dimensionality reduction with PCA and non-linear regression with GPR have been combined to develop ROMs for uncertainty quantification [334] and simulation-based digital twins of industrial systems [335]. The prediction of a new system state based on continuous incoming data from sensors is crucial for combining CFD simulations with real-time data, and this is only possible if the prediction becomes instantaneous [335,336].

### 7. Machine learning in structural assessment

Aerospace structures are subjected to a wide spectrum of loads, operating in harsh environments [337] and phenomena such as high-velocity debris have the ability to severely impact the integrity of components [338]. Furthermore, while developments in composite materials present large advantages in terms of strength and weight, failure mechanisms are more difficult to predict as a

result of an increased number of failure modes compared to metals [338].

Structural health monitoring (SHM) is the implementation of a damage-detection strategy for engineering structures [339]. Applied across many industries, the integration of SHM into the aerospace sector is particularly beneficial as a result of the difficulty of regular inspection and the very high cost of failure. Integrating SHM into aerospace structures has the capacity to reduce the downtime, operating and maintenance costs of aircraft, while also improving the safety and reliability simultaneously [340].

While vibration-based techniques are the most commonly employed [339], acoustic emission and electro-mechanical impedance (EMI) are examples of sensing techniques that are often discussed in the context of damage detection within SHM for aerospace structures. The former uses sensors to detect high-frequency stress waves that are generated during crack propagation and, while aerospace applications have been considered since the 1970s [341], significantly enhanced accuracy of crack-location prediction has been achieved for aerospace applications using ML methods such as Gaussian process regression [342] and NNs [343] in recent years. Further examples of the use of ML methods being used for damage detection classification within the aerospace industry include probabilistic neural networks [344], support-vector machines [345] and logistic regression [346].

One problem facing damage detection using SHM methods is the requirement of complex and heavy wiring requirements of the sensor networks which can be prohibitive for aerospace structures [347]. As a result, *low power* [347] or *no power* equipment [348] has been developed to enable wireless sensor networks. Such approaches require efficient methods of communicating data and, thus, new methods of interpreting the data have been developed. In Ref. [349], low-rank-approximation is used to estimate sparse data from a wireless sensor network and a K-nearest-neighbour algorithm is employed to classify this imputed data. Furthermore, sensor and wiring reliability can be considered to be one of the 'weak-links' of an effective SHM system, particularly in the harsh environments aerospace structures are subjected to, and classification techniques like KNN can be used to detect sensors faults [350].

Digital twins are widely discussed across a broad range of engineering domains, where in SHM applications, they are developed with an aim of providing an accessible insight into the health of a structure or inform predictive maintenance. This is achieved by having a digital model of the structure in which measured detail can be incorporated [351]. A common example of this within the aerospace sector is performing updated finite-element modelling; however, this can be computationally costly. ML can, therefore, be used to perform surrogate modelling using methods such as Gaussian processes [352,353] to significantly reduce the computational cost.

Many aerospace components such as landing gear operate a safe-life methodology, in which there is zero-crack tolerance [354]. Therefore, damage-detection methods, based *e.g.* on Acoustic emission, are not always relevant (although their inclusion is advocated, nonetheless). A sub-category of SHM is *structural loads monitoring* or *virtual sensing*. This involves predicting the loads, as well as the subsequent stress and strain, that a structure is subjected to using incomplete available measurements. This enables damage prognosis, which can be considered the natural next step of structural health monitoring with applications in all forms of defence and commercial aviation [355]. Fatigue analysis is of particular interest here, as more than 80% of structures fail due to fatigue [356], which occurs after the structure is subjected to a large number of stress cycles.

Within the helicopter health and usage monitoring systems (HUMS) domain, there are numerous ML-based approaches to

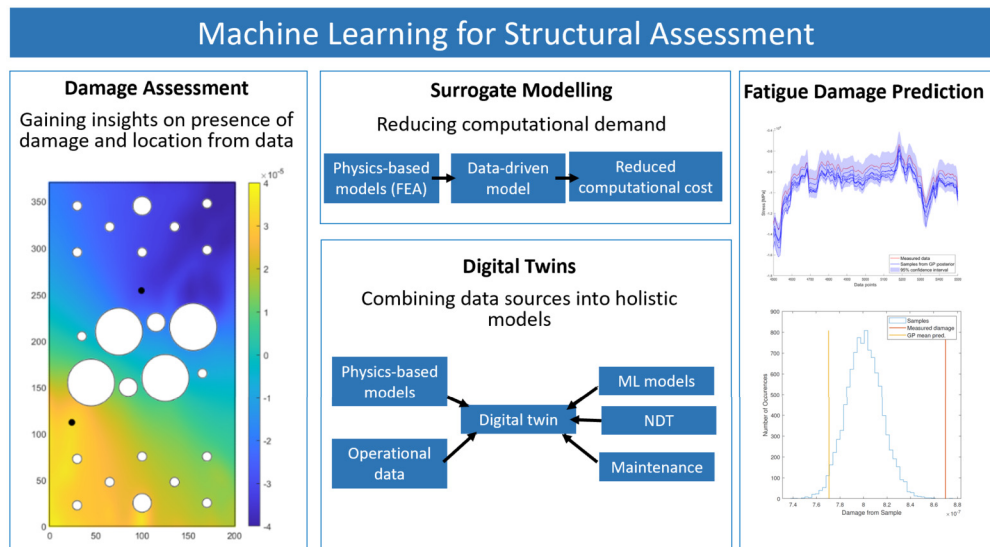


Fig. 13. Machine learning for structural assessment.

loads monitoring and fatigue analysis. In Ref. [357], the mechanical loads are predicted using a multilayer perceptron (MLP) NN. In Ref. [358], an extreme learning machine (a subclass of feed-forward NN) is used for load prediction and the results are fed through the *stress-life* fatigue analysis procedure to predict damage accumulation.

*Grey-box modelling* is the integration of our knowledge of the physics of the structure into the model. The most common way of doing this, and something the authors advocate for whenever possible, is via *semi-physical modelling* or *input augmentation*, involving manipulating inputs to the model to indicate physical processes [359–361]. An example of this is squaring the measured airspeed before using it as an input to the model (in this case a NN) to give an indication of dynamic pressure [359,362]. Similarly, in Ref. [361], a *mathematical network* is used for fatigue monitoring, in which physics-inspired merging functions are used as the first layer to the network. An alternative way of integrating physics into models is via *residual modelling* in which ML methods such as Gaussian process regression [363,364] are used to predict the discrepancy between a physics-based or empirical prediction and the measured data. An introduction to grey-box modelling for SHM applications can be found in Ref. [365].

A data-driven model is fundamentally an estimate of the similarity of the test case compared to the training data that the model has seen. Practically, it is difficult to include all possible loading phenomena of aerospace structures in the training dataset due to magnitude of the loading spectrum and computational limits. Therefore, probabilistic methods are becoming increasingly popular. Gaussian process regression is one of the most popular ML methods for virtual sensing problems, in part, due to its inherently probabilistic nature and subsequent capacity to indicate uncertainty. In the case of Gaussian process regression, the estimate of the similarity of the training and the test data is based on the *kernel*, or *covariance function*. Based on prior knowledge of the data, an appropriate kernel can be chosen and combining kernels using addition and multiplication enable a rich language that the functional form of the model can take [366]. A novel way of inputting physical insight into Gaussian process regression models is to integrate the knowledge into the kernel itself, which has the added benefit of providing a physical interpretation to the samples drawn from the model posterior [365].

One example of Gaussian process regression used for virtual loads monitoring can be found in Ref. [367], where the loads on

the landing gear of an aircraft are predicted with a high degree of accuracy. In Ref. [360], the loads on an aircraft wing are predicted and fatigue analysis is carried out using the predicted strain values. However, the probabilistic aspect of the model is often not fully utilized. In Ref. [368], Gaussian process regression is again used to predict the stress on the wing of an aircraft. By taking a large number of draws from the posterior covariance of the model, the model uncertainty can be propagated through the *stress-life* fatigue analysis procedure and, thereby, a probabilistic estimate for fatigue damage accumulation is developed.

Such approaches can fit into probabilistic risk-based decision frameworks for structural health monitoring [369], which can be used to make risk-informed decisions for safety-critical components like those found in the aerospace industry. Adopting a probabilistic approach to fatigue in aerospace components has the potential to better control the risk level and plan maintenance more effectively [370,371].

Data collection alone does not constitute structural health monitoring. It is only by using this data to make informed decisions that we are performing SHM [372] and it is the use of machine learning that is enabling insights to be gathered from the increasingly widespread collection of data across aerospace engineering. However, one of the main limitations of the currently employed methods is the fact that data-driven approaches require representative training data of the respective phenomena, which can often be difficult to acquire for structural assessment. It is for this reason, among others, that the authors suggest the use of ML as a compliment to, rather than a replacement of, traditional physics-based modelling for structural assessment [365] (see Fig. 13).

## 8. Conclusions and further discussion: new technologies based on machine learning

The continued growth of machine learning (ML) is leading to a progressively larger impact on a wide range of scientific areas, including aerospace engineering. A number of emerging ML-based technologies are already impacting every aspect of this area, e.g. in terms of simulation capabilities and enhanced physical insight. The main goal pursued by the aerospace industry is to use ML to develop several applications connected to the reduction of aircraft's environmental impact, also considering system management, customer service, data interpretation and even developing new methods based on ML-tools capable to generate new high-



fidelity databases at a reduced cost, which will be later used to develop new designs, more efficient, in a fast and efficient way. These applications involve acquiring knowledge and developing new ML tools suitable to solve problems of fundamental fluid dynamics (experimental and numerical), aerodynamics, acoustics, combustion and structural health monitoring.

The type of identification method, the type of data analyzed, and the model structure, influences the performance of the ML model. To maintain the proper balance between the model accuracy and the cost of data generation, a suitable option would be using ML tools considering data fusion techniques, which combine data from different types (experimental and numerical) and different fidelities. The development of models that are generalizable within the different applications, would suppose and advance in the field, reducing in this way the complex process of calibration behind selecting the proper architecture composing the neural networks composing the model and extracting relevant information from the different data sources available, which is highly dependant on the type of application: aerodynamics, combustion, acoustics, solid mechanics, etc. However, most of the models currently used by the industry, only consider data coming from a single source, leaving new research opportunities to advance in the field. Also, considering models grounded in physics, which combine modern neural network architectures with other dimensionality reduction techniques based on the identification of physical patterns, would suppose and advance in the field. Some researchers are currently developing their work following these ideas, showing the good performance of these models grounded in physics: a few examples are found in the fields of fundamental fluid mechanics, aerodynamics combustion and structural health monitoring, nevertheless there are still many open questions that should be addressed, and also the lack of robust and generalizable schemes still leaves it as an open topic that should be addressed by researchers and the aerospace industry in the near future.

When it comes to the application of ML to CFD, there is great potential in several areas related to improving the efficiency of numerical simulations of fluid flows. The future development of ML should not be focused on replacing CFD, but rather improving the efficiency and robustness of subproblems within CFD solvers. For instance, regarding acceleration of CFD (e.g., by accelerating the Poisson problem or improving coarse simulations), enhancing modelling (for LES and RANS) and improving ROM development. In the latter, AEs are progressively becoming a more useful technique since they enable leveraging the non-linearity while, through recent developments [41], retaining the orthogonality of the AE modes, which are ranked by their contribution to the reconstruction as in POD. One point of future improvement for AEs is the possibility to effectively integrate the temporal dynamics into the predictions, a step that currently relies on using the decoder and the latent vectors. When it comes to wind-tunnel experiments, non-intrusive sensing has greatly benefited from recent progress in deep-learning architectures for computer vision. Another area holding great potential is that of flow control, where a number of methods (e.g., Gaussian-process regression, genetic programming and deep reinforcement learning) have the potential of discovering very effective strategies. From the perspective of fluid mechanics, ML-based methods can help to significantly improve numerical simulations, experimental techniques and the development of reduced-order models. These will lead to better a characterization of the flow, and potentially to novel strategies to control it. Ultimately, these will produce a positive impact on the aerodynamic design, leading to configurations with higher aerodynamic efficiency.

The critical technology for aerodynamic aircraft performance improvement using ML approach is the integration of ML algorithms into the design and optimization of aircraft components

and systems. This involves using data-driven models and simulations to predict how changes to the design of the aircraft's aerodynamic components will impact its performance, such as its lift, drag, and fuel efficiency. Some specific areas where machine learning is being used to improve aerodynamic performance include: (i) wing design optimization: ML algorithms can be used to optimize the shape, size, and placement of the wings to reduce drag and increase lift, improving overall aerodynamic performance; (ii) engine performance optimization: ML algorithms can be used to optimize the operation of the aircraft's engines, such as adjusting the thrust, fuel mixture, and other parameters to improve fuel efficiency and reduce emissions; (iii) flight control systems: ML algorithms can be used to optimize the aircraft's flight control systems, such as the autopilot, to ensure that the aircraft is operating at peak performance in all conditions; (iv) structural optimization: ML algorithms can be used to optimize the structure of the aircraft, such as reducing the weight of components without compromising strength, to improve overall aerodynamic performance. By using machine learning algorithms to optimize the design and operation of aircraft components and systems, aerodynamic performance can be improved, resulting in greater fuel efficiency, reduced emissions, and better overall performance.

Also, to reduce fuel consumption, electrical (or hybrid) propulsion systems are combined with new aerodynamic concepts such as distributed propulsion systems, boundary layer ingestion, embedded propulsion or truss-braced wings to improve aircraft efficiency [373]. The aeroacoustic footprints associated to these new concepts are still unclear and complex and costly aeroacoustic analyses will be necessary to avoid acoustic sources producing passenger discomfort. ML will help to minimize costly experiments and simulations by helping to find correlations (between aerodynamics and acoustics), produce fast reduced order models and enhance optimization tools for design. In addition, ML helps to classify and understand physics behind data and hence can bring light into the massive amount data, which is produced by aircraft manufacturers when conceiving new designs.

The development of soft-sensing strategies for real-time engine performance monitoring is a key area where machine learning can contribute to enhancing aircraft performance by improving the combustion process. Traditional engine monitoring systems rely on pre-set thresholds or rules to determine maintenance needs based on sensor measurements of parameters such as fuel flow rate, temperature, pressure, and exhaust emissions. The combination of soft-sensing strategies based on machine learning algorithms represents a critical technology to enable more efficient and effective engine performance monitoring. These technologies help reduce maintenance costs, increase reliability and safety, and improve aircraft performance.

The use of machine learning is crucial in the development of structural health monitoring. Widespread collection of data is commonplace across all engineering sectors, but it is often the use of machine learning that enables this data to provide information of the structural safety and inform decisions on maintenance. In doing so, we are able to reduce downtime while improving reliability and safety. Physics-informed machine learning and population-based SHM (PBSHM) are two research areas currently attracting a lot of interest. In physics-informed machine learning (also known grey-box modelling), the aim is to integrate knowledge of physics into machine learning models, enabling significant improvements in model performance. In PBSHM, the aim is to use transfer learning to use data from one structure in a fleet to inform knowledge of the health of a different structure. Overall, the use of machine learning will be critical for developing SHM systems for aircraft. This will enable a more efficient and effective monitoring of structural health and help to ensure the safety and reliability of aircraft structures.



This article demonstrates the potential of ML to enhance aircraft performance. It includes a comprehensive review of key applications of ML in relevant aerospace engineering disciplines, such as fluid dynamics, aerodynamics, aeroacoustics, combustion, and structural health monitoring, providing an overview of the current state-of-the-art in the field. However, while ML methods are powerful tools that can advance research, their limitations must always be considered and they should be applied judiciously.

### 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.

### Data availability

No data was used for the research described in the article.

### Acknowledgements

E.F. and S.L.C. would like to thank the support of the Comunidad de Madrid through the call Research Grants for Young Investigators from the Universidad Politécnica de Madrid. S.L.C. also acknowledges the grant PID2020-114173RB-I00 funded by MCIN/AEI/10.13039/501100011033. E.F. also acknowledges the Spanish Ministry MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR for the grant “Europa Investigación 2020” EIN2020-112255. RV acknowledges the financial support from ERC grant no. 443 ‘2021-CoG-101043998, DEEPCONTROL’. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

### References

- [1] Directorate-General for Mobility European Commission, Directorate-General for Research Transport, and Innovation. Flightpath 2050: Europe's vision for aviation: maintaining global leadership and serving society's needs, Publications Office, 2011.
- [2] A. Abbas, J. de Vicente, E. Valero, Aerodynamic technologies to improve aircraft performance, *Aerosp. Sci. Technol.* 28 (2013) 100–132.
- [3] R. Abadía-Heredia, M. López-Martín, B. Carro, J.I. Arribas, J.M. Pérez, S. Le Clainche, A predictive hybrid reduced order model based on proper orthogonal decomposition combined with deep learning architectures, *Expert Syst. Appl.* 187 (2022) 115910.
- [4] S. Le Clainche, M. Rosti, L. Brandt, A data-driven model based on modal decomposition: application to the turbulent channel flow over an anisotropic porous wall, *J. Fluid Mech.* 939 (2022) A5.
- [5] M. Lopez-Martin, S. Le Clainche, B. Carro, Model-free short-term fluid dynamics estimator with a deep 3d-convolutional neural network, *Expert Syst. Appl.* 177 (2021) 114924.
- [6] S.L. Brunton, N. Kutz, K. Manohar, A.Y. Aravkin, K. Morgansen, Data-driven aerospace engineering: reframing the industry with machine learning, *AIAA J.* 59 (8) (2021) 2820–2847.
- [7] S. Pandey, J. Schumacher, K.R. Sreenivasan, A perspective on machine learning in turbulent flows, *J. Turbul.* 21 (2020) 567–584.
- [8] C. Méndez, S. Le Clainche, R. Moreno-Ramos, J.M. Vega, A new automatic, very efficient method for the analysis of flight flutter testing data, *Aerosp. Sci. Technol.* 114 (2021) 106749.
- [9] A. Corrochano, A.F. Neves, B. Khanal, S. Le Clainche, N.J. Lawson, Des of a slingsby firefly aircraft: unsteady flow feature extraction using pod and hodmd, *J. Aerosp. Eng.* 35 (5) (2022) 04022063.
- [10] J.M. Vega, S. Le Clainche, Higher Order Dynamic Mode Decomposition and Its Applications, Elsevier, 2020.
- [11] S. Le Clainche, Prediction of the optimal vortex in synthetic jets, *Energies* 12 (9) (2018) 1635.
- [12] S. Le Clainche, E. Ferrer, A reduced order model to predict transient flows around straight bladed vertical axis wind turbines, *Energies* 11 (3) (2018) 566–576.
- [13] J. Garicano-Mena, B. Li, E. Ferrer, E. Valero, A composite dynamic mode decomposition analysis of turbulent channel flows, *Phys. Fluids* 31 (11) (2019) 115102.
- [14] P.A. Srinivasan, L. Guastoni, H. Azizpour, P. Schlatter, R. Vinuesa, Predictions of turbulent shear flows using deep neural networks, *Phys. Rev. Fluids* 4 (2019) 054603.
- [15] S.L. Brunton, B.R. Noack, P. Koumoutsakos, Machine learning for fluid mechanics, *Annu. Rev. Fluid Mech.* 52 (2019) 477–508.
- [16] S. Haykin, *Neural Networks and Learning Machines*, Prentice Hall, Upper Saddle River, 2009.
- [17] D.S. Broomhead, D. Lowe, Multivariable functional interpolation and adaptive networks, *Complex Syst.* 2 (3) (1988) A5.
- [18] J. Kou, W. Zhang, Multi-kernel neural networks for nonlinear unsteady aerodynamic reduced-order modeling, *Aerosp. Sci. Technol.* 67 (2017) 309–326.
- [19] J. Kou, W. Zhang, Reduced-order modeling for nonlinear aeroelasticity with varying Mach numbers, *Aerosp. Sci. Technol.* 31 (6) (2018) 04018105.
- [20] S. Brunton, J. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, Cambridge Univ. Press, Cambridge, 2019, A5.
- [21] Aized Amin Soofi, Arshad Awan, Classification techniques in machine learning: applications and issues, *J. Basic Appl. Sci.* 13 (Jan. 2017) 459–465.
- [22] Gilles Louppe, Understanding random forests: from theory to practice, 2014.
- [23] C.E. Rasmussen, Gaussian processes for machine learning, *J. Urol. Urogynäkol.* (2003).
- [24] I. Goodfellow, et al., Generative adversarial networks, *Commun. ACM* 63 (2020) 139–144.
- [25] H. Kim, J. Kim, S. Won, C. Lee, Unsupervised deep learning for super-resolution reconstruction of turbulence, *J. Fluid Mech.* 910 (2021) A29.
- [26] C. Ledig, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017*, pp. 105–114.
- [27] A. Güemes, S. Discetti, A. Ianiro, B. Sirmacek, H. Azizpour, R. Vinuesa, From coarse wall measurements to turbulent velocity fields through deep learning, *Phys. Fluids* 33 (2021) 075121.
- [28] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms, Preprint, arXiv:1707.06347, 2017.
- [29] V. Mnih, et al., Human-level control through deep reinforcement learning, *Nature* 518 (2015) 529–533.
- [30] S. Gu, T. Lillicrap, U. Sutskever, S. Levine, Continuous deep Q-learning with model-based acceleration, in: *33rd International Conference on Machine Learning, ICML, vol. 6, 2016*, pp. 4135–4148.
- [31] R. Vinuesa, O. Lehmkuhl, A. Lozano-Durán, J. Rabault, Flow control in wings and discovery of novel approaches via deep reinforcement learning, *Fluids* 865 (2019) 281–302.
- [32] Christopher M. Bishop, *Pattern recognition and machine learning*, in: *Information Science and Statistics*, vol. 1613-9011, 1 edition, Springer, New York, NY, 2006.
- [33] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay Scikit-learn, *Machine learning in python, J. Mach. Learn. Res.* 12 (Nov. 2011) 2825–2830.
- [34] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the em algorithm, *J. R. Stat. Soc., Ser. B, Methodol.* 39 (1) (1977) 1–22.
- [35] Ettore Saetta, Renato Tognaccini, Identification of flow field regions by machine learning, in: *AIAA SCITECH 2022 Forum*, 2022.
- [36] Marco Lanzetta, Benedetto Mele, Renato Tognaccini, Advances in aerodynamic drag extraction by far-field methods, *J. Aircr.* 52 (6) (2015) 1873–1886.
- [37] Kheir-Eddine Otmani, Gerasimos Ntoukas, Oscar A. Mariño, Esteban Ferrer, Toward a robust detection of viscous and turbulent flow regions using unsupervised machine learning, 2023.
- [38] Kenza Tlales, Kheir-Eddine Otmani, Gerasimos Ntoukas, Gonzalo Rubio, Esteban Ferrer, Machine learning adaptation for laminar and turbulent flows: applications to high order discontinuous Galerkin solvers, 2022.
- [39] Jared Callahan, James Koch, Bingni Brunton, J. Kutz, Steven Brunton, Learning dominant physical processes with data-driven balance models, *Nat. Commun.* 12 (02.2021).
- [40] Brendan Colvert, Mohamad Alsaman, Eva Kanso, Classifying vortex wakes using neural networks, *Bioinspir. Biomim.* 13 (09.2017).
- [41] H. Eivazi, S. Le Clainche, S. Hoyas, R. Vinuesa, Towards extraction of orthogonal and parsimonious non-linear modes from turbulent flows, *Expert Syst. Appl.* 202 (2022) 117038.
- [42] G. Berkooz, P. Holmes, J.L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, *Annu. Rev. Fluid Mech.* 25 (1) (1993) 539–575.
- [43] K. Karhunen, Über lineare Methoden in der Wahrscheinlichkeitsrechnung, *Ann. Acad. Sci. Fenn., Ser. A 1* (1947).
- [44] I.T. Jolliffe, *Principal Component Analysis*, Springer, New York, 1986.

- [45] E.N. Lorenz, Statistical Forecasting Project (Massachusetts Institute of Technology). Empirical Orthogonal Functions and Statistical Weather Prediction, Scientific report, Massachusetts Institute of Technology, Department of Meteorology, 1956.
- [46] C.G. Wu, Y.C. Liang, W.Z. Lin, H.P. Lee, S.P. Lim, A note on equivalence of proper orthogonal decomposition methods, *J. Sound Vib.* 265 (5) (2003) 1103–1110.
- [47] Alessandro Parente, James C. Sutherland, Principal component analysis of turbulent combustion data: data pre-processing and manifold sensitivity, *Combust. Flame* 160 (2.2013) 340–350.
- [48] Henry F. Kaiser, The varimax criterion for analytic rotation in factor analysis, *Psychometrika* 23 (3) (1958) 187–200.
- [49] J. Einbeck, G. Tutz, L. Evers, Local principal curves, *Stat. Comput.* 15 (4) (October 2005) 301–313.
- [50] Sam T. Roweis, Lawrence K. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science* 290 (12.2000) 2323–2326.
- [51] Kambhatla Nanda, Todd K. Leen, Dimension reduction by local principal component analysis, *Neural Comput.* 9 (7) (1997) 1493–1516.
- [52] A. Parente, J.C. Sutherland, B.B. Dally, L. Tognotti, P.J. Smith, Investigation of the mild combustion regime via principal component analysis, *Proc. Combust. Inst.* 33 (2011) 3333–3341.
- [53] Bernhard Schölkopf, Alexander Smola, Klaus Robert Müller, Kernel principal component analysis, in: Wulfram Gerstner, Alain Germond, Martin Hasler, Jean-Daniel Nicoud (Eds.), *Artificial Neural Networks - ICANN 1997 - 7th International Conference, Proceedings, 08-10-1997 through 10-10-1997*, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 1997, pp. 583–588. Publisher Copyright: ©.
- [54] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science* 313 (2006) 504–507, <https://doi.org/10.1126/science.1127647>, <http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed&uid=16873662&cmd=showdetailview&indexed=google>.
- [55] Diederik P. Kingma, Max Welling, Auto-encoding variational Bayes, Preprint, arXiv:1312.6114, 2022.
- [56] Danilo Jimenez Rezende, Shakir Mohamed, Daan Wierstra, Stochastic back-propagation and approximate inference in deep generative models, Preprint, arXiv:1401.4082, 2014.
- [57] Raban Iten, Tony Metger, Henrik Wilming, Lidia del Rio, Renato Renner, Discovering physical concepts with neural networks, *Phys. Rev. Lett.* 124 (2020) 010508.
- [58] Romit Maulik, Kai Fukami, Nesar Ramachandra, Koji Fukagata, Kunihiro Taira, Probabilistic neural networks for fluid flow surrogate modeling and data recovery, *Phys. Rev. Fluids* 5 (2020) 104401.
- [59] Yoshua Bengio, Aaron Courville, Pascal Vincent, Representation learning: a review and new perspectives, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (2013) 1798–1828.
- [60] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, Alexander Lerchner, Beta-VAE: learning basic visual concepts with a constrained variational framework, in: *International Conference on Learning Representations*, 2017.
- [61] Alessandro Achille, Stefano Soatto, Information dropout: learning optimal representations through noisy computation, *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (12) (2018) 2897–2905.
- [62] Christopher P. Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins, Alexander Lerchner, Understanding disentangling in  $\beta$ -vae, Preprint, arXiv:1804.03599, 2018.
- [63] Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Rätsch, Sylvain Gelly, Bernhard Schölkopf, Olivier Bachem, Challenging common assumptions in the unsupervised learning of disentangled representations, Preprint, arXiv:1811.12359, 2019.
- [64] F.H. Clauser, The turbulent boundary layer, *Adv. Appl. Mech.* 4 (1956) 1–51.
- [65] I. Marusic, B.J. McKeon, P.A. Monkewitz, H.M. Nagib, A.J. Smits, K.R. Sreenivasan, Wall-bounded turbulent flows at high Reynolds numbers: recent advances and key issues, *Phys. Fluids* 22 (2010) 065103.
- [66] M. Kurz, P. Offenhäuser, A. Beck, Deep reinforcement learning for turbulence modeling in large eddy simulations, *Int. J. Heat Fluid Flow* 99 (2023) 109094.
- [67] L. Guastoni, J. Rabault, P. Schlatter, H. Azizpour, R. Vinuesa, Deep reinforcement learning for turbulent drag reduction in channel flows, Preprint, arXiv:2301.09889, 2023.
- [68] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Feländer, S.D. Langhans, M. Tegmark, F. Fuso Nerini, The role of artificial intelligence in achieving the sustainable development goals, *Nat. Commun.* 11 (2020) 233.
- [69] R. Vinuesa, S.L. Brunton, Enhancing computational fluid dynamics with machine learning, *Nat. Comput. Sci.* 2 (2022) 358–366.
- [70] S. Godunov, I. Bohachevsky, Finite difference method for numerical computation of discontinuous solutions of the equations of fluid dynamics, *Mat. Sb.* 47 (3) (1959) 271–306.
- [71] C. Canuto, M.Y. Hussaini, A. Quarteroni, A. Thomas Jr., *Spectral Methods in Fluid Dynamics*, Springer Science & Business Media, 2012.
- [72] D. Kochkov, J.A. Smith, A. Alieva, Q. Wang, M.P. Brenner, S. Hoyer, Machine learning-accelerated computational fluid dynamics, *Proc. Natl. Acad. Sci. USA* 118 (2021) e2101784118.
- [73] E. Ajuria, A. Alguacil, M. Bauerheim, A. Misdariis, B. Cuenot, E. Benazera, Towards a hybrid computational strategy based on deep learning for incompressible flows, in: *AIAA AVIATION Forum*, June 15–19, 2020, pp. 1–17.
- [74] K. Fukami, Y. Nabae, K. Kawai, K. Fukagata, Synthetic turbulent inflow generator using machine learning, *Phys. Rev. Fluids* 4 (2019) 064603.
- [75] M. Yousif, L. Yu, H. Lim, Physics-guided deep learning for generating turbulent inflow conditions, *J. Fluid Mech.* 936 (2022) A21.
- [76] Y. Morita, S. Rezaeiravesh, N. Tabatabaei, R. Vinuesa, K. Fukagata, P. Schlatter, Applying Bayesian optimization with Gaussian-process regression to computational fluid dynamics problems, *J. Comput. Phys.* 449 (2022) 110788.
- [77] Fernando Manrique de Lara, Esteban Ferrer, Accelerating high order discontinuous Galerkin solvers using neural networks: 1D Burgers' equation, *Comput. Fluids* 235 (2022) 105274.
- [78] Fernando Manrique de Lara, Esteban Ferrer, Accelerating high order discontinuous Galerkin solvers using neural networks: 3D compressible Navier-Stokes equations, 2022.
- [79] K. Duraisamy, G. Iaccarino, H. Xiao, Turbulence modeling in the age of data, *Annu. Rev. Fluid Mech.* 51 (2019) 357–377.
- [80] J.V. Boussinesq, *Théorie analytique de la chaleur: mise en harmonie avec la thermodynamique et avec la théorie mécanique de la lumière* T. 2, Refroidissement et échauffement par rayonnement conductibilité des tiges, lames et masses cristallines courants de convection théorie mécanique de la lumière, Gauthier-Villars, 1923.
- [81] P.R. Spalart, Strategies for turbulence modelling and simulations, *Int. J. Heat Fluid Flow* 21 (2000) 252–263.
- [82] J. Wu, H. Xiao, R. Sun, Q. Wang, Reynolds-averaged Navier–Stokes equations with explicit data-driven Reynolds stress closure can be ill-conditioned, *J. Fluid Mech.* 869 (2019) 553–586.
- [83] J. Ling, A. Kurzawski, J. Templeton, Reynolds averaged turbulence modelling using deep neural networks with embedded invariance, *J. Fluid Mech.* 807 (2016) 155–166.
- [84] J.-L. Wu, H. Xiao, E. Paterson, Physics-informed machine learning approach for augmenting turbulence models: a comprehensive framework, *Phys. Rev. Fluids* 3 (2018) 074602.
- [85] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, *Nat. Mach. Intell.* 1 (2019) 206–215.
- [86] C. Jiang, R. Vinuesa, R. Chen, J. Mi, S. Laima, H. Li, An interpretable framework of data-driven turbulence modeling using deep neural networks, *Phys. Fluids* 33 (2021) 055133.
- [87] R. Vinuesa, B. Sirmacek, Interpretable deep-learning models to help achieve the sustainable development goals, *Nat. Mach. Intell.* 3 (2021) 926.
- [88] M. Cranmer, A. Sanchez-Gonzalez, P. Battaglia, R. Xu, K. Cranmer, D. Spergel, S. Ho, Discovering symbolic models from deep learning with inductive biases, in: *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, Vancouver, Canada, 2020, arXiv:2006.11287.
- [89] J. Weatheritt, R.D. Sandberg, A novel evolutionary algorithm applied to algebraic modifications of the RANS stress-strain relationship, *J. Comput. Phys.* 325 (2016) 22–37.
- [90] J.R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, 1992.
- [91] M. Emory, J. Larsson, G. Iaccarino, Modeling of structural uncertainties in Reynolds-averaged Navier–Stokes closures, *Phys. Fluids* 25 (2013) 110822.
- [92] A.A. Mishra, G. Iaccarino, Uncertainty estimation for Reynolds-averaged Navier–Stokes predictions of high-speed aircraft nozzle jets, *AIAA J.* 55 (2017) 3999–4004.
- [93] A.D. Beck, D.G. Flad, C.-D. Munz, Deep neural networks for data-driven LES closure models, *J. Comput. Phys.* 398 (2019) 108910.
- [94] C.J. Lapeyre, A. Misdariis, N. Cazard, D. Veynante, T. Poinot, Training convolutional neural networks to estimate turbulent sub-grid scale reaction rates, *Combust. Flame* 203 (2019) 255.
- [95] G. Novati, H.L. de Laroussilhe, P. Koumoutsakos, Automating turbulence modelling by multi-agent reinforcement learning, *Nat. Mach. Intell.* 3 (2021) 87–96.
- [96] C. Moeng, A large-eddy-simulation model for the study of planetary boundary-layer turbulence, *J. Atmos. Sci.* 13 (1984) 2052–2062.
- [97] R. Vinuesa, P.S. Negi, M. Atzori, A. Hanifi, D.S. Henningson, P. Schlatter, Turbulent boundary layers around wing sections up to  $Re_c = 1,000,000$ , *Int. J. Heat Fluid Flow* 72 (2018) 86–99.
- [98] R. Pozuelo, Q. Li, P. Schlatter, R. Vinuesa, An adverse-pressure-gradient turbulent boundary layer with nearly constant  $\beta \approx 1.4$  up to  $Re_\theta \approx 8700$ , *J. Fluid Mech.* 939 (2022) A34.
- [99] J. Jiménez, Turbulent flows over rough walls, *Annu. Rev. Fluid Mech.* 36 (2004) 173–196.
- [100] Y. Mizuno, J. Jiménez, Wall turbulence without walls, *J. Fluid Mech.* 723 (2013) 429–455.

- [101] K. Sasaki, R. Vinuesa, A.V.G. Cavalieri, P. Schlatter, D.S. Henningson, Transfer functions for flow predictions in wall-bounded turbulence, *J. Fluid Mech.* 864 (2019) 708–745.
- [102] G.B. Arivazhagan, L. Guastoni, A. Güemes, A. Ianiro, S. Discetti, P. Schlatter, H. Azizpour, R. Vinuesa, Predicting the near-wall region of turbulence through convolutional neural networks, in: *Proc. 13th ERCOFTAC Symp. on Engineering Turbulence Modelling and Measurements (ETMM13)*, Rhodes, Greece, September 16–17, 2021, arXiv:2107.07340.
- [103] N. Moriya, K. Fukami, Y. Nabae, M. Morimoto, T. Nakamura, K. Fukagata, Inserting machine-learned virtual wall velocity for large-eddy simulation of turbulent channel flows, Preprint, arXiv:2106.09271, 2021.
- [104] H.J. Bae, P. Koumoutsakos, Scientific multi-agent reinforcement learning for wall-models of turbulent flows, *Nat. Commun.* 13 (2022) 1443.
- [105] L. Guastoni, A. Güemes, A. Ianiro, S. Discetti, P. Schlatter, H. Azizpour, R. Vinuesa, Convolutional-network models to predict wall-bounded turbulence from wall quantities, *J. Fluid Mech.* 928 (2021) A27.
- [106] Kunihiko Taira, Steven L. Brunton, Scott Dawson, Clarence W. Rowley, Tim Colonius, Beverley J. McKeon, Oliver T. Schmidt, Stanislav Gordeyev, Vassilios Theofilis, Lawrence S. Ukeiley, Modal analysis of fluid flows: an overview, *AIAA J.* 55 (12) (2017) 4013–4041.
- [107] J.L. Lumley, The structure of inhomogeneous turbulence, in: A.M. Yaglom, V.I. Tatarski (Eds.), *Atmospheric Turbulence and Wave Propagation*, Nauka, Moscow, 1967, pp. 166–178.
- [108] P.J. Schmid, Dynamic mode decomposition of numerical and experimental data, *J. Fluid Mech.* 656 (August 2010) 5–28.
- [109] S. Le Clainche, J.M. Vega, Higher order dynamic mode decomposition, *SIAM J. Appl. Dyn. Syst.* 16 (2) (2017) 882–925.
- [110] M. Milano, P. Koumoutsakos, Neural network modeling for near wall turbulent flow, *J. Comput. Phys.* 182 (2002) 1–26.
- [111] Soledad Le Clainche, José M. Vega, Higher order dynamic mode decomposition to identify and extrapolate flow patterns, *Phys. Fluids* 29 (8) (August 2017) 084102, <https://doi.org/10.1063/1.4997206>.
- [112] J. Moehlis, H. Faisst, B. Eckhardt, A low-dimensional model for turbulent shear flows, *New J. Phys.* 6 (2004) 56.
- [113] H. Eivazi, L. Guastoni, P. Schlatter, H. Azizpour, R. Vinuesa, Recurrent neural networks and Koopman-based frameworks for temporal predictions in a low-order model of turbulence, *Int. J. Heat Fluid Flow* 90 (2021) 108816.
- [114] N.A.K. Doan, W. Polifke, L. Magri, Short- and long-term predictions of chaotic flows and extreme events: a physics-constrained reservoir computing approach, *Proc. R. Soc. A* 477 (2021) 20210135.
- [115] G. Borrelli, L. Guastoni, H. Eivazi, P. Schlatter, R. Vinuesa, Predicting the temporal dynamics of turbulent channels through deep learning, *Int. J. Heat Fluid Flow* 96 (2022) 109010.
- [116] T. Murata, K. Fukami, K. Fukagata, Nonlinear mode decomposition with convolutional neural networks for fluid dynamics, *J. Fluid Mech.* 882 (2020) A13.
- [117] K. Fukami, T. Nakamura, K. Fukagata, Convolutional neural network based hierarchical autoencoder for nonlinear mode decomposition of fluid field data, *Phys. Fluids* 32 (2020) 095110.
- [118] J. Kim, C. Lee, Prediction of turbulent heat transfer using convolutional neural networks, *J. Fluid Mech.* 882 (2020) A18.
- [119] M.P. Encinar, J. Jiménez, Logarithmic-layer turbulence: a view from the wall, *Phys. Rev. Fluids* 4 (2019) 114603.
- [120] J. Borée, Extended proper orthogonal decomposition: a tool to analyse correlated events in turbulent flows, *Exp. Fluids* 35 (2003) 188–192.
- [121] A. Güemes, S. Discetti, A. Ianiro, Sensing the turbulent large-scale motions with their wall signature, *Phys. Fluids* 31 (2019) 125112.
- [122] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.* 378 (2019) 686–707.
- [123] Maziar Raissi, Alireza Yazdani, George Em Karniadakis, Hidden fluid mechanics: learning velocity and pressure fields from flow visualizations, *Science* 367 (6481) (2020) 1026–1030.
- [124] H. Eivazi, R. Vinuesa, Physics-informed deep-learning applications to experimental fluid mechanics, Preprint, arXiv:2203.15402, 2022.
- [125] H. Eivazi, M. Tahani, P. Schlatter, R. Vinuesa, Physics-informed neural networks for solving Reynolds-averaged Navier–Stokes equations, *Phys. Fluids* 34 (2022) 075117.
- [126] R. Vinuesa, H.M. Nagib, Enhancing the accuracy of measurement techniques in high Reynolds number turbulent boundary layers for more representative comparison to their canonical representations, *Eur. J. Mech. B, Fluids* 55 (2016) 300–312.
- [127] R. Vinuesa, P. Schlatter, H.M. Nagib, Role of data uncertainties in identifying the logarithmic region of turbulent boundary layers, *Exp. Fluids* 55 (2014) 1751.
- [128] H. Choi, P. Moin, J. Kim, Active turbulence control for drag reduction in wall-bounded flows, *J. Fluid Mech.* 262 (6481) (1994) 75–110.
- [129] O.A. Mahfoze, A. Moody, A. Wynn, R.D. Whalley, S. Laizet, Reducing the skin-friction drag of a turbulent boundary-layer flow with low-amplitude wall-normal blowing within a Bayesian optimization framework, *Phys. Rev. Fluids* 3 (2019) 094601.
- [130] R. Li, B.R. Noack, L. Cordier, J. Borée, F. Harambat, Drag reduction of a car model by linear genetic programming control, *Exp. Fluids* 58 (2017) 103.
- [131] G. Minelli, T. Dong, B.R. Noack, S. Krajnović, Upstream actuation for bluff-body wake control driven by a genetically inspired optimization, *J. Fluid Mech.* 893 (2020) A1.
- [132] J. Rabault, M. Kuchta, A. Jensen, U. Réglade, N. Cerardi, Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control, *J. Fluid Mech.* 7 (2022) 62.
- [133] K. Li, J. Kou, W. Zhang, Deep learning for multi-fidelity aerodynamic distribution modelling from experimental and simulation data, arXiv:2019.12966v2, 2021.
- [134] J. Kou, W. Zhang, Data-driven modelling for unsteady aerodynamics and aeroelasticity, *Prog. Aerosp. Sci.* 125 (2021) 100725.
- [135] L. Sirovich, Turbulence and the dynamics of coherent structures. Parts i – iii, *Q. Appl. Math.* 45 (3) (1987) 561–571.
- [136] P. Schmid, Dynamic mode decomposition of numerical and experimental data, *J. Fluid Mech.* 656 (2010) 5–28.
- [137] B. Glaz, L. Liu, P.P. Friedmann, Reduced-order nonlinear unsteady aerodynamic modeling using a surrogate-based recurrence framework, *AIAA J.* 48 (10) (2010) 2418–2429.
- [138] A. Da Ronch, M. Ghoreyshi, K. Badcock, On the generation of flight dynamics aerodynamic tables by computational fluid dynamics, *Prog. Aerosp. Sci.* 47 (8) (2011) 597–620.
- [139] D.E. Raveh, Reduced-order models for nonlinear unsteady aerodynamics, *AIAA J.* 39 (8) (2001).
- [140] P. Marzocca, W.A. Silva, L. Librescu, Nonlinear open-/closed-loop aeroelastic analysis of airfoils via Volterra series, *AIAA J.* 42 (4) (2004) 673–686.
- [141] G. Sun, S. Wang, A review of the artificial neural network surrogate modeling in aerodynamic design, *Proc. Inst. Mech. Eng., G J. Aerosp. Eng.* 233 (16) (2019) 5863–5872.
- [142] Jichao Li, Xiaosong Du, Joaquim R.R.A. Martins, Machine learning in aerodynamic shape optimization, *Prog. Aerosp. Sci.* 134 (2022) 100849.
- [143] E. Iuliano, E.A. Pérez, Application of Surrogate-Based Global Optimization to Aerodynamic Design, Springer, New York, 2016.
- [144] N. Van Nguyen, J. Lee, M. Tyan, et al., Repetitively enhanced neural networks method for complex engineering design optimisation problems, *Aeronaut. J.* 119 (2015) 1253–1270.
- [145] J.N. Kani, A.H. Elsheikh, Dr-rnn: a deep residual recurrent neural network for model reduction, arXiv:1709.00939 [cs.CE], 2017.
- [146] M.A.A. Oroumieh, S.M.B. Malaek, M. Ashrafzaadeh, et al., Dr-rnn: a deep residual recurrent neural network for model reduction, *Aerosp. Sci. Technol.* 26 (2013) 244–258.
- [147] J. Kou, W. Zhang, An approach to enhance the generalization capability of nonlinear aerodynamic reduced-order models, *Aerosp. Sci. Technol.* 49 (2016) 197–208.
- [148] D.J. Linse, R.F. Stengel, Identification of aerodynamic coefficients using computational neural networks, *J. Guid. Control Dyn.* 16 (6) (1993) 1018–1025.
- [149] S. Suresh, S.N. Omkar, V. Mani, T.N. Guru Prakash, Lift coefficient prediction at high angle of attack using recurrent neural network, *Aerosp. Sci. Technol.* 7 (8) (2003) 595–602.
- [150] N.R. Secco, B.S.d. Mattos, Artificial neural networks to predict aerodynamic coefficients of transport airplanes, *Aircr. Eng. Aerosp. Technol.* 89 (2017) 211–230.
- [151] W. Hou, D. Darakananda, J.D. Eldredge, Machine-learning-based detection of aerodynamic disturbances using surface pressure measurements, *AIAA J.* 57 (12) (2019) 5079–5093.
- [152] X. Wang, J. Kou, W. Zhang, Unsteady aerodynamic modeling based on fuzzy scalar radial basis function neural networks, *Proc. Inst. Mech. Eng., G J. Aerosp. Eng.* 233 (14) (2019) 5107–5121.
- [153] M. Ghoreyshi, A. Jirasek, R.M. Cummings, Computational approximation of nonlinear unsteady aerodynamics using an aerodynamic model hierarchy, *Aerosp. Sci. Technol.* 28 (1) (2013) 133–144.
- [154] F.D. Marques, J. Anderson, Identification and prediction of unsteady transonic aerodynamic loads by multi-layer functionals, *J. Fluids Struct.* 15 (1) (2001) 83–106.
- [155] W. Zhang, B. Wang, Z. Ye, J. Quan, Dr-rnn: a deep residual recurrent neural network for model reduction, *AIAA J.* 50 (5) (2012) 1019–1028.
- [156] F. Mazhar, A.M. Khan, I.A. Chaudhry, et al., On using neural networks in uav structural design for cfd data fitting and classification, *Aerosp. Sci. Technol.* 30 (2013) 210–225.
- [157] A. Mannarino, P. Mantegazza, Nonlinear aeroelastic reduced order modeling by recurrent neural networks, *J. Fluids Struct.* 48 (2014) 103–121.
- [158] J. Rivers, B. Dansberry, R. Bennet, M. Durham, W. Silva, Naca 0012 benchmark model experimental flutter results with unsteady pressure distributions, in: *33rd Structures, Structural Dynamics and Material Conference*, 1992, 92–2396.
- [159] K. Li, J. Kou, W. Zhang, Deep neural network for unsteady aerodynamic and aeroelastic modeling across multiple Mach numbers, *Nonlinear Dyn.* 96 (3) (2019) 2157–2177.
- [160] J. Chen, G. Sun, X. Jin, Intelligent aerodynamic design for airfoil based on artificial neural network method, in: *2nd International Conference on Computer and Automation Engineering*, vol. 42, ICCAE, Singapore, 2010, pp. 428–436.



- [161] G. Sun, Y. Sun, S. Wang, Artificial neural network based inverse design airfoils and wings, *Aerosp. Sci. Technol.* 42 (2015) 415–428.
- [162] S. Koziel, L. Leifsson, Knowledge-based airfoil shape optimization using space mapping, in: 30th AIAA Applied Aerodynamics Conference, New Orleans, Louisiana, vol. 34, 2012, p. 3016.
- [163] T.C. Oktay, H. Elik, M. Uzun, A novel learning algorithm to estimate the optimum fuselage drag coefficient, *Sakarya Univ. J. Sci.* 21 (2017) 63–68.
- [164] Z. Xu, J.H. Saleh, V. Yang, Optimization of supercritical airfoil design with buffet effect, *AIAA J.* (2018).
- [165] Michael J. Bianco, Peter Gerstoft, James Traer, Emma Ozanich, Marie A. Roch, Sharon Gannot, Charles-Alban Deledalle, Machine learning in acoustics: theory and applications, *J. Acoust. Soc. Am.* 146 (5) (2019) 3590–3628.
- [166] Stéphane Moreau, The third golden age of aeroacoustics, *Phys. Fluids* 34 (3) (2022) 031301.
- [167] Michael James Lighthill, Maxwell Herman Alexander Newman, On sound generated aerodynamically i. General theory, *Proc. R. Soc. Lond. Ser. A, Math. Phys. Sci.* 211 (1107) (1952) 564–587.
- [168] Michael James Lighthill, On sound generated aerodynamically ii. Turbulence as a source of sound, *Proc. R. Soc. Lond. Ser. A, Math. Phys. Sci.* 222 (1148) (1954) 1–32.
- [169] Sanjiva K. Lele, Joseph W. Nichols, A second golden age of aeroacoustics?, *Philos. Trans. - Royal Soc., Math. Phys. Eng. Sci.* 372 (2022) (2014) 20130321.
- [170] Antonio Alguacil, Michaël Bauerheim, Marc C. Jacob, Stéphane Moreau, Predicting the propagation of acoustic waves using deep convolutional neural networks, *J. Sound Vib.* 512 (2021) 116285.
- [171] Antonio Alguacil, Michaël Bauerheim, Marc C. Jacob, Stéphane Moreau, A generic deep learning framework for propagation and scattering of acoustic waves in quiescent flows.
- [172] Zhenyu Tang, Hsien-Yu Meng, Dinesh Manocha, Learning acoustic scattering fields for dynamic interactive sound propagation, in: 2021 IEEE Virtual Reality and 3D User Interfaces (VR), 2021, pp. 835–844.
- [173] Damjan Kužnar, Martin Možina, Marina Giordanino, Ivan Bratko, Improving vehicle aeroacoustics using machine learning, *Eng. Appl. Artif. Intell.* 25 (5) (2012) 1053–1061.
- [174] Sajjad Beigmoradi, Hassan Hajabdollahi, Asghar Ramezani, Multi-objective aero acoustic optimization of rear end in a simplified car model by using hybrid robust parameter design, artificial neural networks and genetic algorithm methods, *Comput. Fluids* 90 (2014) 123–132.
- [175] Mario Rüttgers, Seong-Ryong Koh, Jenia Jitsev, Wolfgang Schröder, Andreas Lintermann, Prediction of acoustic fields using a lattice-Boltzmann method and deep learning, in: Heike Jagode, Hartwig Anzt, Guido Juckeland, Hatem Ltaief (Eds.), *High Performance Computing*, Springer International Publishing, Cham, 2020, pp. 81–101.
- [176] Ghulam Moeen Uddin, Sajawal Gul Niazi, Syed Muhammad Arafat, Muhammad Sajid Kamran, Muhammad Farooq, Nasir Hayat, Sher Afghan Malik, Abe Zeid, Sagar Kamarthi, Sania Saqib, Ijaz Ahmad Chaudhry, Neural networks assisted computational aero-acoustic analysis of an isolated tire, *Proc. Inst. Mech. Eng., Part D, J. Automob. Eng.* 234 (10–11) (2020) 2561–2577.
- [177] Leandra I. Abreu, Alvaro Tanarro, André V.G. Cavalieri, Philipp Schlatter, Ricardo Vinuesa, Ardeshtir Hanifi, Dan S. Henningson, Spanwise-coherent hydrodynamic waves around flat plates and airfoils, *J. Fluid Mech.* 927 (2021) A1.
- [178] Jiaqing Kou, Laura Botero-Bolívar, Román Ballano, Oscar Marino, Leandro de Santana, Eusebio Valero, Esteban Ferrer, Aeroacoustic airfoil shape optimization enhanced by autoencoders, 2023.
- [179] Masaaki Imaizumi, Kenji Fukumizu, Deep neural networks learn non-smooth functions effectively, in: *AISTATS*, 2019.
- [180] Steven L. Brunton, Bernd R. Noack, Petros Koumoutsakos, Machine learning for fluid mechanics, *Annu. Rev. Fluid Mech.* 52 (2020) 477–508.
- [181] R.W. Bilger, The role of combustion technology in the 21st century, 2011.
- [182] Stephen B. Pope, Small scales, many species and the manifold challenges of turbulent combustion, *Proc. Combust. Inst.* 34 (2013) 1–31.
- [183] A. Giusti, E. Mastorakos, Turbulent combustion modelling and experiments: recent trends and developments, *Flow Turbul. Combust.* 103 ( ) (2019) 847–869.
- [184] Kamila Zdybał, Mohammad Rafi Malik, Axel Coussement, James C. Sutherland, Alessandro Parente, Reduced-order modeling of reactive flows using data-driven approaches, in: N. Swaminathan, A. Parente (Eds.), *Lecture Notes in Energy: Machine Learning and Its Application to Reacting Flows*, Springer, 2022, Chapter 9.
- [185] Jacqueline H. Chen, Petascale direct numerical simulation of turbulent combustion - fundamental insights towards predictive models, *Proc. Combust. Inst.* 33 (2011) 99–123.
- [186] P. Trisjono, H. Pitsch, Systematic analysis strategies for the development of combustion models from dns: a review, *Flow Turbul. Combust.* 95 (10.2015) 231–259.
- [187] Achim Wick, Antonio Attili, Fabrizio Bisetti, Heinz Pitsch, Dns-driven analysis of the flamelet/progress variable model assumptions on soot inception, growth, and oxidation in turbulent flames, *Combust. Flame* 214 (4.2020) 437–449.
- [188] H. Pitsch, A. Attili, *Data Analysis for Direct Numerical Simulations of Turbulent Combustion: From Equation-Based Analysis to Machine Learning*, Springer International Publishing, 2020.
- [189] Matthias Ihme, Wai Tong Chung, Aashwin Ananda Mishra, Combustion machine learning: principles, progress and prospects, *Prog. Energy Combust. Sci.* 91 (2022) 101010.
- [190] Ulrich Maas, Dominique Thévenin, Correlation analysis of direct numerical simulation data of turbulent non-premixed flames, in: *Symposium (International) on Combustion*, vol. 27, 1998, pp. 1183–1189.
- [191] C.E. Frouzakis, Y.G. Kevrekidis, J. Lee, K. Boulouchos, A.A. Alonso, Proper orthogonal decomposition of direct numerical simulation data: data reduction and observer construction, *Proc. Combust. Inst.* 28 (1.2000) 75–81.
- [192] Sean J. Danby, Tarek Echekki, Proper orthogonal decomposition analysis of autoignition simulation data of nonhomogeneous hydrogen-air mixtures, *Combust. Flame* 144 (1.2006) 126–138.
- [193] A. Parente, J.C. Sutherland, L. Tognotti, P.J. Smith, Identification of low-dimensional manifolds in turbulent flames, *Proc. Combust. Inst.* 32 (1) (2009) 1579–1586.
- [194] James C. Sutherland, Alessandro Parente, Combustion modeling using principal component analysis, *Proc. Combust. Inst.* 32 (1) (2009) 1563–1570.
- [195] Amir Biglari, James C. Sutherland, A filter-independent model identification technique for turbulent combustion modeling, *Combust. Flame* 159 (5.2012) 1960–1970.
- [196] Hessam Mirgolbabaei, Tarek Echekki, A novel principal component analysis-based acceleration scheme for les-odt: an a priori study, *Combust. Flame* 160 (5.2013) 898–908.
- [197] Yue Yang, Stephen B. Pope, Jacqueline H. Chen, Empirical low-dimensional manifolds in composition space, *Combust. Flame* 160 (10.2013) 1967–1980.
- [198] R.S. Barlow, J.H. Frank, Effects of turbulence on species mass fractions in methane/air jet flames, 1998.
- [199] R.S. Barlow, G.J. Fiechtner, C.D. Carter, J.-Y. Chen, Experiments on the scalar structure of turbulent co/h 2 /n 2 jet flames, 2000.
- [200] R.S. Barlow, J.H. Frank, A.N. Karpetis, J.Y. Chen, Piloted methane/air jet flames: transport effects and aspects of scalar structure, *Combust. Flame* 143 (12.2005) 433–449.
- [201] B.B. Dally, A.N. Karpetis, R.S. Barlow, Structure of turbulent non-premixed jet flames in a diluted hot coflow, *Proc. Combust. Inst.* 29 (1) (2002) 1147–1154.
- [202] Ch Schneider, A. Dreizler, J. Janicka, E.P. Hassel, Flow field measurements of stable and locally extinguishing hydrocarbon-fuelled jet flames, *Combust. Flame* 135 (10.2003) 185–190.
- [203] Paul R. Medwell, Peter A.M. Kalt, Bassam B. Dally, Simultaneous imaging of oh, formaldehyde, and temperature of turbulent nonpremixed jet flames in a heated and diluted coflow, *Combust. Flame* 148 (1.2007) 48–61.
- [204] Tarek Echekki, Jacqueline H. Chen, Direct numerical simulation of autoignition in non-homogeneous hydrogen-air mixtures, *Combust. Flame* 134 (8.2003) 169–191.
- [205] Evatt R. Hawkes, Ramanan Sankaran, James C. Sutherland, Jacqueline H. Chen, Scalar mixing in direct numerical simulations of temporally evolving plane jet flames with skeletal co/h2 kinetics, *Proc. Combust. Inst.* 31 (1) (2007) 1633–1640.
- [206] Evatt R. Hawkes, Ramanan Sankaran, Jacqueline H. Chen, Sebastian A. Kaiser, Jonathan H. Frank, An analysis of lower-dimensional approximations to the scalar dissipation rate using direct numerical simulations of plane jet flames, *Proc. Combust. Inst.* 32 (1) (2009) 1455–1463.
- [207] Chun Sang Yoo, Edward S. Richardson, Ramanan Sankaran, Jacqueline H. Chen, A dns study on the stabilization mechanism of a turbulent lifted ethylene jet flame in highly-heated coflow, *Proc. Combust. Inst.* 33 (2011) 1619–1627.
- [208] Naveen Punati, James C. Sutherland, Alan R. Kerstein, Evatt R. Hawkes, Jacqueline H. Chen, An evaluation of the one-dimensional turbulence model: comparison with direct numerical simulations of co/h2 jets with extinction and reignition, *Proc. Combust. Inst.* 33 (2011) 1515–1522.
- [209] Axel Coussement, Olivier Gicquel, Alessandro Parente, Kernel density weighted principal component analysis of combustion processes, *Combust. Flame* 159 (9.2012) 2844–2855.
- [210] Axel Coussement, Olivier Gicquel, Alessandro Parente, Mg-local-pca method for reduced order combustion modeling, *Proc. Combust. Inst.* 34 (2013) 1117–1123.
- [211] Gaurav Bansal, Ajith Mascarenhas, Jacqueline H. Chen, Direct numerical simulations of autoignition in stratified dimethyl-ether (dme)/air turbulent mixtures, *Combust. Flame* 162 (3.2015) 688–702.
- [212] Hessam Mirgolbabaei, Tarek Echekki, Nejib Smaoui, A nonlinear principal component analysis approach for turbulent combustion composition space, *Int. J. Hydrog. Energy* 39 (9) (2014) 4622–4633.
- [213] A.A. Mascarenhas, G. Bansal, J.H. Chen, Identification of intrinsic low dimensional manifolds in turbulent combustion using an isomap based technique, Technical report, Sandia National Lab. (SNL-CA), Livermore, California, United States, 2011.
- [214] Ehsan Fooladgar, Christophe Duwig, A new post-processing technique for analyzing high-dimensional combustion data, *Combust. Flame* 191 (5.2018) 226–238.



- [215] Ehsan Fooladgar, Pál Tóth, Christophe Duwig, Characterization of flameless combustion in a model gas turbine combustor using a novel post-processing tool, *Combust. Flame* 204 (6.2019) 356–367.
- [216] Kamila Zdybał, Giuseppe D'Alessio, Gianmarco Aversano, Mohammad Rafi Malik, Axel Coussement, James C. Sutherland, Alessandro Parente, Advancing reactive flow simulations with data-driven models, in: Miguel A. Mendez, Andrea Ianiro, Bernd R. Noack, Steven L. Brunton (Eds.), *Data-Driven Fluid Mechanics: Combining First Principles and Machine Learning*, Cambridge University Press, 2022. Chapter 15.
- [217] Tryambak Gangopadhyay, Vikram Ramanan, Adedotun Akintayo, Paige K. Boor, Soumalya Sarkar, Satyanarayanan R. Chakravarthy, Soumik Sarkar, 3d convolutional selective autoencoder for instability detection in combustion systems, *Energy AI* 4 (2021) 100067.
- [218] Xiaobin Zhu, Zun Cai, Jianjun Wu, Yuqiang Cheng, Qiang Huang, Convolutional neural network based combustion mode classification for condition monitoring in the supersonic combustor, *Acta Astronaut.* 159 (6.2019) 349–357.
- [219] Kaidi Wan, Sandra Hartl, Luc Vervisch, Pascale Domingo, Robert S. Barlow, Christian Hasse, Combustion regime identification from machine learning trained by Raman/Rayleigh line measurements, *Combust. Flame* 219 (9.2020) 268–274.
- [220] Kherlen Jigjid, Chitoshi Tamaoki, Yuki Minamoto, Ryota Nakazawa, Nakamasa Inoue, Mamoru Tanahashi, Data driven analysis and prediction of mild combustion mode, *Combust. Flame* 223 (1.2021) 474–485.
- [221] Jochen Einbeck, Gerhard Tutz, Ludger Evers, Local principal curves, 2005.
- [222] Giuseppe D'Alessio, Alessandro Parente, Alessandro Stagni, Alberto Cuoci, Adaptive chemistry via pre-partitioning of composition space and mechanism reduction, *Combust. Flame* 211 (2020) 68–82.
- [223] Giuseppe D'Alessio, Alberto Cuoci, Gianmarco Aversano, Mauro Bracconi, Alessandro Stagni, Alessandro Parente, Impact of the partitioning method on multidimensional adaptive-chemistry simulations, *Energies* 13 (10) (2020).
- [224] Giuseppe D'Alessio, Alberto Cuoci, Alessandro Parente, Feature extraction and artificial neural networks for the on-the-fly classification of high-dimensional thermochemical spaces in adaptive-chemistry simulations, *Data-Centric Eng.* 2 (4.2021).
- [225] S. Fu, S.Y. Lu, David L. Davies, Donald W. Boulidin, The string-to-string correction problem, 1977.
- [226] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, *Commun. Stat.* 3 (1) (1974) 1–27.
- [227] Peter Rousseeuw, P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.* 20 (11.1987) 53–65.
- [228] Robert Tibshirani, Guenther Walther, Trevor Hastie, Estimating the number of clusters in a data set via the gap statistic, *J. R. Stat. Soc., Ser. B, Stat. Methodol.* 63 (2) (2001) 411–423.
- [229] Zhiyi Li, Stefanie Tomasch, Zhi X. Chen, Alessandro Parente, Ivar S. Ertesvåg, Nedunchezian Swaminathan, Study of mild combustion using les and advanced analysis tools, *Proc. Combust. Inst.* 38 (4) (2021) 5423–5432.
- [230] Giuseppe D'Alessio, Antonio Attili, Alberto Cuoci, Heinz Pitsch, Alessandro Parente, Unsupervised data analysis of direct numerical simulation of a turbulent flame via local principal component analysis and procustes analysis, in: Álvaro Herrero, Carlos Cambra, Daniel Urda, Javier Sedano, Héctor Quintián, Emilio Corchado (Eds.), *15th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2020)*, Springer International Publishing, Cham, 2021, pp. 460–469.
- [231] M. Valorani, F. Creta, P.P. Ciottoli, R. Malpica Galassi, D.A. Goussis, H.N. Najm, S. Paolucci, H.G. Im, E.-A. Tingas, D.M. Manias, A. Parente, Z. Li, T. Grenga, Computational singular perturbation method and tangential stretching rate analysis of large scale simulations of reactive flows: feature tracking, time scale characterization, and cause/effect identification. Part 2, analyses of ignition systems, laminar and turbulent flames, Springer International Publishing, Cham, 2020, pp. 65–88.
- [232] Kamila Zdybał, Giuseppe D'Alessio, Antonio Attili, Axel Coussement, James C. Sutherland, Alessandro Parente, Local manifold learning and its link to domain-based physics knowledge, *Appl. Energy Combust. Sci.* 14 (2023) 100131, <https://doi.org/10.1016/j.jaecs.2023.100131>.
- [233] Tam Turanyi, Sensitivity analysis of complex kinetic systems. Tools and applications, 1990.
- [234] J.F. Griffiths, Reduced kinetic models and their application to practical combustion systems, 1995.
- [235] Tianfeng Lu, Chung K. Law, A directed relation graph method for mechanism reduction, *Proc. Combust. Inst.* 30 (2005) 1333–1341.
- [236] Ramanan Sankaran, Evatt R. Hawkes, Jacqueline H. Chen, Tianfeng Lu, Chung K. Law, Structure of a spatially developing turbulent lean methane-air bunsen flame, *Proc. Combust. Inst.* 31 (1) (2007) 1291–1298.
- [237] P. Pepiot-Desjardins, H. Pitsch, An efficient error-propagation-based reduction method for large chemical kinetic mechanisms, *Combust. Flame* 154 (1) (2008) 67–81.
- [238] H. Huang, M. Fairweather, J.F. Griffiths, A.S. Tomlin, R.B. Brad, A systematic lumping approach for the reduction of comprehensive kinetic models, *Proc. Combust. Inst.* 30 (2005) 1309–1316.
- [239] Mauro Valorani, Francesco Creta, Dimitris A. Goussis, Jeremiah C. Lee, Habib N. Najm, An automatic procedure for the simplification of chemical kinetic mechanisms based on csp, *Combust. Flame* 146 (7.2006) 29–51.
- [240] Riccardo Malpica Galassi, Pietro P. Ciottoli, S. Mani Sarathy, Hong G. Im, Samuel Paolucci, Mauro Valorani, Automated chemical kinetic mechanism simplification with minimal user expertise, *Combust. Flame* 197 (2018) 439–448.
- [241] Automatic reduction procedure for chemical mechanisms applied to premixed methane/air flames, *Proc. Combust. Inst.* 28 (2) (2000) 1809–1815.
- [242] A.C. Zambon, H.K. Chelliah, Explicit reduced reaction models for ignition, flame propagation, and extinction of  $\text{C}_2\text{H}_4/\text{CH}_4/\text{H}_2$  and air systems, *Combust. Flame* 150 (7.2007) 71–91.
- [243] Alison S. Tomlin, Tamás Turányi, Michael J. Pilling, Chapter 4 mathematical tools for the construction, investigation and reduction of combustion mechanisms, in: M.J. Pilling (Ed.), *Low-Temperature Combustion and Autoignition*, in: *Comprehensive Chemical Kinetics*, vol. 35, Elsevier, 1997, pp. 293–437.
- [244] Long Liang, John G. Stevens, Sumathy Raman, John T. Farrell, The use of dynamic adaptive chemistry in combustion simulation of gasoline surrogate fuels, *Combust. Flame* 156 (7.2009) 1493–1502.
- [245] Francesco Contino, Hervé Jeanmart, Tommaso Lucchini, Gianluca D'Errico, Coupling of in situ adaptive tabulation and dynamic adaptive chemistry: an effective method for solving combustion in engine simulations, *Proc. Combust. Inst.* 33 (2011) 3057–3064.
- [246] Lei Zhou, Haiqiao Wei, Chemistry acceleration with tabulated dynamic adaptive chemistry in a realistic engine with a primary reference fuel, *Fuel* 171 (5.2016) 186–194.
- [247] Zhuyin Ren, Yufeng Liu, Tianfeng Lu, Liuyan Lu, Oluwayemisi O. Oluwole, Graham M. Goldin, The use of dynamic adaptive chemistry and tabulation in reactive flow simulations, *Combust. Flame* 161 (1.2014) 127–137.
- [248] Zhuyin Ren, Chao Xu, Tianfeng Lu, Michael A. Singer, Dynamic adaptive chemistry with operator splitting schemes for reactive flow simulations, *J. Comput. Phys.* 263 (4.2014) 19–36.
- [249] Yu Shi, Long Liang, Hai Wen Ge, Rolf D. Reitz, Acceleration of the chemistry solver for modeling di engine combustion using dynamic adaptive chemistry (dac) schemes, *Combust. Theory Model.* 14 (2.2010) 69–89.
- [250] Youwen Liang, Stephen B. Pope, Perrine Pepiot, A pre-partitioned adaptive chemistry methodology for the efficient implementation of combustion chemistry in particle pdf methods, *Combust. Flame* 162 (9.2015) 3236–3253.
- [251] Ashish S. Newale, Youwen Liang, Stephen B. Pope, Perrine Pepiot, A combined ppac-rcce-isat methodology for efficient implementation of combustion chemistry, *Combust. Theory Model.* 23 (11.2019) 1021–1053.
- [252] A.S. Newale, S.B. Pope, P. Pepiot, Computationally-efficient and accurate particle pdf simulations of turbulent combustion using coupled pre-partitioned adaptive chemistry and tabulation, *Proc. Combust. Inst.* 38 (1.2021) 2721–2729.
- [253] A. Stagni, A. Frassoldati, A. Cuoci, T. Faravelli, E. Ranzi, Skeletal mechanism reduction through species-targeted sensitivity analysis, *Combust. Flame* 163 (1.2016) 382–393.
- [254] Kyle E. Niemeyer, Chih Jen Sung, Mandhapati P. Raju, Skeletal mechanism generation for surrogate fuels using directed relation graph with error propagation and sensitivity analysis, *Combust. Flame* 157 (9.2010) 1760–1770.
- [255] Michael Frenklach, Transforming data into knowledge-process informatics for combustion chemistry, *Proc. Combust. Inst.* 31 (1) (2007) 125–140.
- [256] Alison S. Tomlin, The role of sensitivity and uncertainty analysis in combustion modelling, *Proc. Combust. Inst.* 34 (2013) 159–176.
- [257] Wang Hai, David A. Sheen, Combustion kinetic model uncertainty quantification, propagation and minimization, *Prog. Energy Combust. Sci.* 47 (2015) 1–31.
- [258] James A. Miller, Raghu Sivaramakrishnan, Yujie Tao, C. Franklin Goldsmith, Michael P. Burke, Ahren W. Jasper, Nils Hansen, Nicole J. Labbe, Peter Glarborg, Judit Zádor, Combustion chemistry in the twenty-first century: developing theory-informed chemical kinetics models, *Prog. Energy Combust. Sci.* 83 (2021) 100886.
- [259] Michael Frenklach, Hai Wang, Martin J. Rabinowitz, Optimization and analysis of large chemical kinetic mechanisms using the solution mapping method—combustion of methane, *Prog. Energy Combust. Sci.* 18 (1) (1992) 47–73.
- [260] David A. Sheen, Hai Wang, The method of uncertainty quantification and minimization using polynomial chaos expansions, *Combust. Flame* 158 (12) (2011) 2358–2374.
- [261] T. Varga, T. Nagy, C. Olm, I.Gy. Zsely, R. Palvolgyi, E. Valko, G. Vincze, M. Cserhati, H.J. Curran, T. Turanyi, Optimization of a hydrogen combustion mechanism using both direct and indirect measurements, *Proc. Combust. Inst.* 35 (1) (2015) 589–596.
- [262] L. Elliott, D.B. Ingham, A.G. Kyne, N.S. Mera, M. Pourkashanian, C.W. Wilson, Genetic algorithms for optimisation of chemical kinetics reaction mechanisms, *Prog. Energy Combust. Sci.* 30 (3) (2004) 297–328.
- [263] A. Bertolino, M. Fürst, A. Stagni, A. Frassoldati, M. Pelucchi, C. Cavallotti, T. Faravelli, A. Parente, An evolutionary, data-driven approach for mechanism optimization: theory and application to ammonia combustion, *Combust. Flame* 229 (7.2021).

- [264] Henry J. Curran, Developing detailed chemical kinetic mechanisms for fuel combustion, *Proc. Combust. Inst.* 37 (2019) 57–81.
- [265] R.J. Shannon, A.S. Tomlin, S.H. Robertson, M.A. Blitz, M.J. Pilling, P.W. Seakins, Global uncertainty propagation and sensitivity analysis in the  $\text{CH}_3\text{OCH}_2 + \text{O}_2$  system: combining experiment and theory to constrain key rate coefficients in  $\text{dme}$  combustion, *J. Phys. Chem. A* 119 (7.2015) 7430–7438.
- [266] Stephen J. Klippenstein, From theoretical reaction dynamics to chemical modeling of combustion, *Proc. Combust. Inst.* 36 (2017) 77–111.
- [267] Magnus Fürst, Andrea Bertolino, Alberto Cuoci, Tiziano Faravelli, Alessio Frassoldati, Alessandro Parente, Optismoke++: a toolbox for optimization of chemical kinetic mechanisms, *Comput. Phys. Commun.* 264 (2021) 107940.
- [268] Nicolas Jaouen, Luc Vervisch, Pascale Domingo, Auto-thermal reforming (atr) of natural gas: an automated derivation of optimised reduced chemical schemes, *Proc. Combust. Inst.* 36 (2017) 3321–3330.
- [269] Nicolas Jaouen, Luc Vervisch, Pascale Domingo, Guillaume Ribert, Automatic reduction and optimisation of chemistry for turbulent combustion modelling: impact of the canonical problem, in: *Special Issue in Honor of Norbert Peters*, *Combust. Flame* 175 (2017) 60–79.
- [270] F.C. Christo, A.R. Masri, E.M. Nebot, S.B. Pope, An integrated pdf/neural network approach for simulating turbulent reacting systems, 1996.
- [271] J.A. Blasco, N. Fueyo, C. Dopazo, J. Ballester, Modelling the temporal evolution of a reduced combustion chemical system with an artificial neural network, 1998.
- [272] Matthias Ihme, Christoph Schmitt, Heinz Pitsch, Optimal artificial neural networks and tabulation methods for chemistry representation in les of a bluff-body swirl-stabilized flame, *Proc. Combust. Inst.* 32 (1) (2009) 1527–1535.
- [273] Opeoluwa Owoyele, Prithwish Kundu, Pinaki Pal, Efficient bifurcation and tabulation of multi-dimensional combustion manifolds using deep mixture of experts: an a priori study, *Proc. Combust. Inst.* 38 (4) (2021) 5889–5896.
- [274] Takafumi Honzawa, Reo Kai, Kotaro Hori, Makoto Seino, Takayuki Nishiie, Ryoichi Kurose, Experimental and numerical study of water sprayed turbulent combustion: proposal of a neural network modeling for five-dimensional flamelet approach, *Energy AI* 5 (2021) 100076.
- [275] Elizabeth Armstrong, Michael A. Hansen, Robert C. Knaus, Nathaniel A. Trask, John C. Hewson, James C. Sutherland, Accurate compression of tabulated chemistry models with partition of unity networks, *Combust. Sci. Technol.* (2022) 1–18.
- [276] A.K. Chatzopoulos, Stelios Rigopoulos, A chemistry tabulation approach via rate-controlled constrained equilibrium (rcce) and artificial neural networks (anns), with application to turbulent non-premixed  $\text{CH}_4/\text{H}_2/\text{N}_2$  flames, *Proc. Combust. Inst.* 34 (2013) 1465–1473.
- [277] Lucas L.C. Franke, Athanasios K. Chatzopoulos, Stelios Rigopoulos, Tabulation of combustion chemistry via artificial neural networks (anns): methodology and application to les-pdf simulation of Sydney flame I, *Combust. Flame* 185 (2017) 245–260.
- [278] Opeoluwa Owoyele, Pinaki Pal Chemnode, A neural ordinary differential equations framework for efficient chemical kinetic solvers, *Energy AI* 7 (2022) 100118.
- [279] Jinzhe Zeng, Liqun Cao, Mingyuan Xu, Tong Zhu, John Z.H. Zhang, Complex reaction processes in combustion unraveled by neural network-based molecular dynamics simulation, *Nat. Commun.* 11 (1) (2020) 5713.
- [280] Weiqi Ji, Sili Deng, Autonomous discovery of unknown reaction pathways from data by chemical reaction neural network, *J. Phys. Chem. A* 125 (4) (2021) 1082–1092.
- [281] Alqahtani Sultan, Tarek Echekki, A data-based hybrid model for complex fuel chemistry acceleration at high temperatures, *Combust. Flame* 223 (2021) 142–152.
- [282] N. Peters, Laminar flamelet concepts in turbulent combustion, *Symp., Int., Combust.* 21 (1) (1988) 1231–1250. Twenty-First Symposium (International on Combustion).
- [283] J.A. van Oijen, L.P.H. de Goeij, Modelling of premixed laminar flames using flamelet-generated manifolds, *Combust. Sci. Technol.* 161 (1) (2000) 113–137.
- [284] Olivier Gicquel, Nasser Darabiha, Dominique Thé Venin, Laminar premixed hydrogen/air counterflow flame simulations using flame prolongation of ildm with differential diffusion, 2000.
- [285] B. Fiorina, R. Baron, O. Gicquel, D. Thevenin, S. Carpentier, N. Darabiha, Modelling non-adiabatic partially premixed flames using flame-prolongation of ildm, *Combust. Theory Model.* 7 (9.2003) 449–470.
- [286] Charles D. Pierce, Parviz Moin, Progress-variable approach for large-eddy simulation of non-premixed turbulent combustion, *J. Fluid Mech.* 504 (4.2004) 73–97.
- [287] Deepak K. Dalakoti, Bruno Savard, Evatt R. Hawkes, Armin Wehrfritz, Haiou Wang, Marc S. Day, John B. Bell, Direct numerical simulation of a spatially developing n-dodecane jet flame under spray a thermochemical conditions: flame structure and stabilisation mechanism, *Combust. Flame* 217 (7.2020) 57–76.
- [288] Benjamin J. Isaac, Jeremy N. Thornock, James Sutherland, Philip J. Smith, Alessandro Parente, Advanced regression methods for combustion modelling using principal components, *Combust. Flame* 162 (6.2015) 2592–2601.
- [289] Mohammad Rafi Malik, Benjamin J. Isaac, Axel Coussement, Philip J. Smith, Alessandro Parente, Principal component analysis coupled with nonlinear regression for chemistry reduction, *Combust. Flame* 187 (2018) 30–41.
- [290] Amir Biglari, James C. Sutherland, An a-posteriori evaluation of principal component analysis-based models for turbulent combustion simulations, *Combust. Flame* 162 (10.2015) 4025–4035.
- [291] Axel Coussement, Benjamin J. Isaac, Olivier Gicquel, Alessandro Parente, Assessment of different chemistry reduction methods based on principal component analysis: comparison of the mg-pca and score-pca approaches, *Combust. Flame* 168 (6.2016) 83–97.
- [292] Opeoluwa Owoyele, Tarek Echekki, Toward computationally efficient combustion dns with complex fuels via principal component transport, *Combust. Theory Model.* 21 (4) (2017) 770–798.
- [293] Rishikesh Ranade, Tarek Echekki, A framework for data-based turbulent combustion closure: a priori validation, *Combust. Flame* 206 (8.2019) 490–505.
- [294] Rishikesh Ranade, Tarek Echekki, A framework for data-based turbulent combustion closure: a posteriori validation, *Combust. Flame* 210 (2019) 279–291.
- [295] Mohammad Rafi Malik, Pedro Obando Vega, Axel Coussement, Alessandro Parente, Combustion modeling using principal component analysis: a posteriori validation on sandia flames d, e and f, *Proc. Combust. Inst.* 38 (1.2021) 2635–2643.
- [296] Mohammad Rafi Malik, Axel Coussement, Tarek Echekki, Alessandro Parente, 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 (2022) 112134.
- [297] Kamila Zdybał, James C. Sutherland, Alessandro Parente, Manifold-informed state vector subset for reduced-order modeling, *Proc. Combust. Inst.* 39 (2022) 1–10.
- [298] Kamila Zdybał, Elizabeth Armstrong, Alessandro Parente, James C. Sutherland, PCAfold: python software to generate, analyze and improve PCA-derived low-dimensional manifolds, *SoftwareX* 12 (2020) 100630.
- [299] Kamila Zdybał, Elizabeth Armstrong, James C. Sutherland, Alessandro Parente, Cost function for low-dimensional manifold topology assessment, *Sci. Rep.* 12 (2022) 14496.
- [300] Thierry Poinot, Denis Veynante, Theoretical and Numerical Combustion, R.T. Edwards Inc., 2005.
- [301] Stephen B. Pope, Turbulent Flows, Cambridge University Press, 2000.
- [302] Rodney O. Fox, Computational Models for Turbulent Reacting Flows, Cambridge Series in Chemical Engineering, Cambridge University Press, 2003.
- [303] D.C. Haworth, S.B. Pope, Transported Probability Density Function Methods for Reynolds-Averaged and Large-Eddy Simulations, Springer, Netherlands, Dordrecht, 2011, pp. 119–142.
- [304] B.F. Magnussen, B.H. Hjertager, On mathematical modeling of turbulent combustion with special emphasis on soot formation and combustion, *Symp., Int., Combust.* 16 (1) (1977) 719–729.
- [305] J. Chomiak, Combustion a study in theory, fact and application, 1.1990.
- [306] Marco Ferrarotti, Zhiyi Li, Alessandro Parente, On the role of mixing models in the simulation of mild combustion using finite-rate chemistry combustion models, *Proc. Combust. Inst.* 37 (2019) 4531–4538.
- [307] Zhiyi Li, Alberto Cuoci, Alessandro Parente, Large eddy simulation of mild combustion using finite rate chemistry: effect of combustion sub-grid closure, *Proc. Combust. Inst.* 37 (2019) 4519–4529.
- [308] A.Y. Klimenko, R.W. Bilger, Conditional moment closure for turbulent combustion, *Prog. Energy Combust. Sci.* 25 (6) (1999) 595–687.
- [309] Mathis Bode, Michael Gauding, Zeyu Lian, Dominik Denker, Marco Davidovic, Konstantin Kleinheinz, Jenia Jitsev, Heinz Pitsch, Using physics-informed enhanced super-resolution generative adversarial networks for subfilter modeling in turbulent reactive flows, *Proc. Combust. Inst.* 38 (1.2021) 2617–2625.
- [310] Bo Liu, Jiupeng Tang, Haibo Huang, Xi Yun Lu, Deep learning methods for super-resolution reconstruction of turbulent flows, *Phys. Fluids* 32 (2.2020).
- [311] O. Colin, F. Ducros, D. Veynante, T. Poinot, A thickened flame model for large eddy simulations of turbulent premixed combustion, *Phys. Fluids* 12 (2000) 1843–1863.
- [312] Z.M. Nikolaou, C. Chrysostomou, L. Vervisch, S. Cant, Progress variable variance and filtered rate modelling using convolutional neural networks and flamelet methods, *Flow Turbul. Combust.* 103 (8.2019) 485–501.
- [313] Andrea Seltz, Pascale Domingo, Luc Vervisch, Zacharias M. Nikolaou, Direct mapping from les resolved scales to filtered-flame generated manifolds using convolutional neural networks, *Combust. Flame* 210 (12.2019) 71–82.
- [314] Z.M. Nikolaou, C. Chrysostomou, Y. Minamoto, L. Vervisch, Evaluation of a neural network-based closure for the unresolved stresses in turbulent premixed v-flames, *Flow Turbul. Combust.* 106 (2.2021) 331–356.
- [315] Marc T. Henry de Frahan, Shashank Yellapantula, Ryan King, Marc S. Day, Ray W. Grout, Deep learning for presumed probability density function models, *Combust. Flame* 208 (2019) 436–450.
- [316] Zhi X. Chen, Salvatore Iavarone, Golnoush Ghiasi, Veeraraghavan Kannan, Giuseppe D'Alessio, Alessandro Parente, Nedunchezhan Swaminathan, Application of machine learning for filtered density function closure in mild combustion, *Combust. Flame* 225 (3.2021) 160–179.
- [317] Salvatore Iavarone, Arthur Péquin, Zhi X. Chen, Nguyen Anh Khoa Doan, Nedunchezhan Swaminathan, Alessandro Parente, An a priori assessment of the

- partially stirred reactor (pasr) model for mild combustion, vol. 38, Elsevier Ltd, 2021, pp. 5403–5414.
- [318] Arthur Péquin, Salvatore Iavarone, Riccardo Malpica Galassi, Alessandro Parente, The partially stirred reactor model for combustion closure in large eddy simulations: physical principles, sub-models for the cell reacting fraction, and open challenges, *Phys. Fluids* 34 (5) (2022) 055122.
- [319] S. Yao, B. Wang, A. Kronenburg, O.T. Stein, Modeling of sub-grid conditional mixing statistics in turbulent sprays using machine learning methods, *Phys. Fluids* 32 (11.2020).
- [320] S. Yao, B. Wang, A. Kronenburg, O.T. Stein, Conditional scalar dissipation rate modeling for turbulent spray flames using artificial neural networks, *Proc. Combust. Inst.* 38 (1.2021) 3371–3378.
- [321] Wai Tong Chung, Aashwin Ananda Mishra, Nikolaos Perakis, Matthias Ihme, Data-assisted combustion simulations with dynamic submodel assignment using random forests, *Combust. Flame* 227 (5.2021) 172–185.
- [322] Thomas Palmé, Magnus Fast, Marcus Thern, Gas turbine sensor validation through classification with artificial neural networks, *Appl. Energy* 88 (11) (2011) 3898–3904.
- [323] Soteris A. Kalogirou, Artificial intelligence for the modeling and control of combustion processes: a review, *Prog. Energy Combust. Sci.* 29 (6) (2003) 515–566.
- [324] Mathis Bode, Nathan Collier, Fabrizio Bisetti, Heinz Pitsch, Adaptive chemistry lookup tables for combustion simulations using optimal b-spline interpolants, *Combust. Theory Model.* 23 (7.2019) 674–699.
- [325] Lukas Berger, Konstantin Kleinheinz, Antonio Attili, Fabrizio Bisetti, Heinz Pitsch, Michael E. Mueller, Numerically accurate computational techniques for optimal estimator analyses of multi-parameter models, *Combust. Theory Model.* 22 (5.2018) 480–504.
- [326] Steven L. Brunton, Joshua L. Proctor, J. Nathan Kutz, William Bialek, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl. Acad. Sci. USA* 113 (2016) 3932–3937.
- [327] Kathleen Champion, Bethany Lusch, J. Nathan Kutz, Steven L. Brunton, Data-driven discovery of coordinates and governing equations, *Proc. Natl. Acad. Sci. USA* 116 (11.2019) 22445–22451.
- [328] J. Swithenbank, I. Poll, M.W. Vincent, D.D. Wright, Combustion design fundamentals, *Symp., Int., Combust.* 14 (1) (1973) 627–638, Fourteenth Symposium (International) on Combustion.
- [329] M. Falcitelli, L. Tognotti, S. Pasini, An algorithm for extracting chemical reactor network models from cfd simulation of industrial combustion systems, *Combust. Sci. Technol.* 174 (2002) 27–42.
- [330] S. Lyra, R.S. Cant, Analysis of high pressure premixed flames using equivalent reactor networks for predicting nox emissions, *Fuel* 107 (2013) 261–268.
- [331] Shane A. McQuarrie, Cheng Huang, Karen E. Willcox, Data-driven reduced-order models via regularised operator inference for a single-injector combustion process, *J. R. Soc. N.Z.* 51 (2021) 194–211.
- [332] Gianmarco Aversano, John Camilo Parra-Alvarez, Benjamin J. Isaac, Sean T. Smith, Axel Coussement, Olivier Gicquel, Alessandro Parente, Pca and Kriging for the efficient exploration of consistency regions in uncertainty quantification, *Proc. Combust. Inst.* 37 (2019) 4461–4469.
- [333] Gianmarco Aversano, Giuseppe D'Alessio, Axel Coussement, Francesco Contino, Alessandro Parente, Combination of polynomial chaos and Kriging for reduced-order model of reacting flow applications, *Results Eng.* 10 (6.2021).
- [334] Gianmarco Aversano, Aurélie Bellemans, Zhiyi Li, Axel Coussement, Olivier Gicquel, Alessandro Parente, Application of reduced-order models based on pca & Kriging for the development of digital twins of reacting flow applications, *Comput. Chem. Eng.* 121 (2.2019) 422–441.
- [335] Gianmarco Aversano, Marco Ferrarotti, Alessandro Parente, Digital twin of a combustion furnace operating in flameless conditions: reduced-order model development from cfd simulations, *Proc. Combust. Inst.* 38 (2021) 5373–5381.
- [336] Joern Kraft, Stefan Kuntzagk, Engine fleet-management: the use of digital twins from a mro perspective, 2017.
- [337] R.W. Ross, Integrated vehicle health management in aerospace structures, in: Fuh-Gwo Yuan (Ed.), *Structural Health Monitoring (SHM) in Aerospace Structures*, Elsevier Science & Technology, 2016, pp. 3–66. Chapter 1.
- [338] Helena Rocha, Christopher Sempimoshnig, João P. Nunes, Sensors for process and structural health monitoring of aerospace composites: a review, *Eng. Struct.* 237 (2021).
- [339] Charles R. Farrar, Keith Worden, *Structural Health Monitoring: A Machine Learning Perspective*, John Wiley and Sons, Nov. 2012.
- [340] F.G. Yuan, *Structural Health Monitoring (SHM) in Aerospace Structures*, Elsevier Science & Technology, 2016.
- [341] D.O. Harris, A review of aerospace applications of acoustic emission, in: *Advances in Test Measurement*, vol. 13, January 1976, pp. 225–233.
- [342] J. Hensman, R. Mills, S.G. Pierce, K. Worden, M. Eaton, Locating acoustic emission sources in complex structures using Gaussian processes, *Mech. Syst. Signal Process.* 24 (1) (2010) 211–223.
- [343] Chandrashekhara Bhat, M.R. Bhat, C.R.L. Murthy, Acoustic emission characterization of failure modes in composites with ANN, *Compos. Struct.* 61 (3) (2003) 213–220.
- [344] Victor Giurgiutiu, Andrei Zagrai, Damage detection in thin plates and aerospace structures with the electro-mechanical impedance method, *Struct. Health Monit.* 4 (2) (2005) 99–118.
- [345] T.H. Loutas, A. Panopoulou, D. Roulas, V. Kostopoulos, Intelligent health monitoring of aerospace composite structures based on dynamic strain measurements, *Expert Syst. Appl.* 39 (9) (2012) 8412–8422.
- [346] Fengyang Jiang, Zhidong Guan, Zengshan Li, Xiaodong Wang, A method of predicting visual detectability of low-velocity impact damage in composite structures based on logistic regression model, *Chin. J. Aeronaut.* 34 (1) (2021) 296–308.
- [347] Carmine Gianni, Marco Balsi, Salvatore Esposito, Francesco Ciampa, low-power global navigation satellite system-enabled wireless sensor network for acoustic emission localisation in aerospace components, *Struct. Control Health Monit.* 27 (6) (2020) 1–13.
- [348] Ondrej Rubes, Jan Chalupa, Filip Ksica, Zdenek Hadas, Development and experimental validation of self-powered wireless vibration sensor node using vibration energy harvester, *Mech. Syst. Signal Process.* 160 (2021) 107890.
- [349] Hadi Salehi, Saptarshi Das, Shantanu Chakrabarty, Subir Biswas, Rigoberto Burgueno, Damage identification in aircraft structures with self-powered sensing technology: a machine learning approach, *Struct. Control Health Monit.* 25 (12) (2018) 1–3.
- [350] Thomas Melia, Alan Cooke, Siobhan Grayson, Machine learning techniques for automatic sensor fault detection in HUMS systems, in: 17th Australian International Aerospace Congress: AIAC 2017, Melbourne, Engineers Australia, 2017.
- [351] Eric J. Tuegel, Anthony R. Ingraffea, Thomas G. Eason, S. Michael Spottswood, Reengineering aircraft structural life prediction using a digital twin, *Int. J. Aerosp. Eng.* (2011) 2011.
- [352] Chenzhao Li, Sankaran Mahadevan, You Ling, Sergio Choe, Liping Wang, Dynamic Bayesian network for aircraft wing health monitoring digital twin, *AIAA J.* 55 (3) (2017) 930–941.
- [353] S. Chakraborty, S. Adhikari, R. Ganguli, The role of surrogate models in the development of digital twins of dynamic systems, *Appl. Math. Model.* 90 (2021) 662–681.
- [354] Joshua Hoole, Pia Sartor, Jonathan Cooper, Safe-life fatigue and sensitivity analysis: a pathway towards embracing uncertainty, in: 5th Aircraft Structural Design Conference, Royal Aeronautical Society, 2016.
- [355] Charles R. Farrar, Nick A.J. Lieven, Damage prognosis: the future of structural health monitoring, *Philos. Trans. - Royal Soc., Math. Phys. Eng. Sci.* 365 (1851) (2007) 623–632.
- [356] P.H. Wirsching, Probabilistic fatigue analysis, in: *Probabilistic Structural Mechanics Handbook*, Springer, 1995, pp. 146–165.
- [357] Michael T. Manry, Cheng Hsiung Hsieh, Hema Chandrasekaran, Near-optimal flight load synthesis using neural nets, in: *Neural Networks for Signal Processing - Proceedings of the IEEE Workshop*, 1999, pp. 535–544.
- [358] Julio J. Valdes, Catherine Cheung, Alejandro Lehman Rubio, Extreme learning machines to approximate low dimensional spaces for helicopter load signal and fatigue life estimation, in: 2017 International Joint Conference on Neural Networks (IJCNN), 2017-May, IEEE, 2017.
- [359] S. Reed, A parametric-based empennage fatigue monitoring system using artificial neural networks, in: *Proceedings of the 23rd Symposium of the International Committee on Aeronautical Fatigue, DGLR-Breicht Hamburg*, 2005, pp. 693–704.
- [360] Ramon Fuentes, Elizabeth Cross, Andrew Halfpenny, Keith Worden, Robert J. Barthorpe, Aircraft parametric structural load monitoring using Gaussian process regression, in: *EWSHM-7th European Workshop on Structural Health Monitoring*, 2014.
- [361] H. Azzam, A practical approach for the indirect prediction of structural fatigue from measured flight parameters, *Proc. Inst. Mech. Eng., G J. Aerosp. Eng.* 211 (1) (1997) 29–38.
- [362] S.C. Reed, Development of a parametric-based indirect aircraft structural usage monitoring system using artificial neural networks, *Aeronaut. J.* 111 (1118) (2007) 209–230.
- [363] Daniel J. Pitchforth, Timothy J. Rogers, Ulf T. Tygesen, Elizabeth J. Cross, Grey-box models for wave loading prediction, *Mech. Syst. Signal Process.* 159 (2021) 107741.
- [364] Hua-Ping Wan, Wei-Xin Ren, A residual-based Gaussian process model framework for finite element model updating, *Comput. Struct.* 156 (2015) 149–159.
- [365] Elizabeth J. Cross, S.J. Gibson, M.R. Jones, D.J. Pitchforth, S. Zhang, T.J. Rogers, Physics-informed machine learning for structural health monitoring, in: *Structural Health Monitoring Based on Data Science Techniques*, Springer, 2022, pp. 347–367.
- [366] David Duvenaud, Automatic model construction with Gaussian processes, PhD thesis, University of Cambridge, 2014.
- [367] Geoffrey Holmes, Pia Sartor, Stephen Reed, Paul Southern, Keith Worden, Elizabeth Cross, Prediction of landing gear loads using machine learning techniques, *Struct. Health Monit.* 15 (5) (2016) 568–582.
- [368] S.J. Gibson, E.J. Cross, T.J. Rogers, Data-driven strain prediction models and fatigue damage accumulation, in: *Proceedings of ISMA2020 International Conference on Noise and Vibration Engineering*, 2020, pp. 3067–3075.

- [369] Aidan J. Hughes, Robert J. Barthorpe, N. Dervilis, Charles R. Farrar, Keith Worden, A probabilistic risk-based decision framework for structural health monitoring, *Mech. Syst. Signal Process.* 150 (2021) 107339.
- [370] Giorgio Cavallini, Roberta Lazzeri, A probabilistic approach to fatigue risk assessment in aerospace components, *Eng. Fract. Mech.* 74 (18) (2007) 2964–2970.
- [371] Joshua Hoole, Pia Sartor, Julian Booker, Jonathan Cooper, Xenofon V. Gogouvitis, Amine Ghouali, R. Kyle Schmidt, Probabilistic Fatigue Methodology for Safe-Life Design and Analysis, Technical report, DSTL, 2021.
- [372] Elizabeth Cross, On structural health monitoring in changing environmental and operational conditions, PhD thesis, University of Sheffield, 2012.
- [373] Peter F. Pelz, Philipp Leise, Marvin Meck, Sustainable aircraft design — a review on optimization methods for electric propulsion with derived optimal number of propulsors, *Prog. Aerosp. Sci.* 123 (2021) 100714.