Data-Centric Engineering (2026), 7: e28
doi:10.1017/dce.2026.10063

RESEARCH ARTICLE

Generative prediction of laser-induced rocket ignition with dynamic latent space representations

Tony Zahtila[1], Ettore Saetta[2], Murray Cutforth[1], Davy Brouzet[1], Diego Rossinelli[1] and Gianluca Iaccarino[3]

[1]Center for Turbulence Research, Stanford University, USA
[2]Industrial Engineering, University of Naples Federico II, Italy
[3]Department of Mechanical Engineering, Stanford University, USA
Corresponding author: Tony Zahtila; Email: tzahtila@stanford.edu

Received: 30 December 2025 UTC; **Revised:** 02 April 2026 UTC; **Accepted:** 29 May 2026 UTC

Keywords: autoencoder compression; fast surrogate model; latent-space trajectory modeling; multifidelity

Abstract

Accurate and predictive scale-resolving simulations of laser-ignited rocket engines are highly time-consuming because the problem includes turbulent fuel–oxidizer mixing dynamics, laser-induced energy deposition, and high-speed flame growth. This is conflated with the large design space primarily corresponding to the laser operating conditions and target location. To enable rapid exploration and uncertainty quantification, we propose a data-driven surrogate modeling approach that combines convolutional autoencoders (cAEs) with neural ordinary differential equations (neural ODEs). The present target application of an machine learning-based surrogate model to leading-edge multiphysics turbulence simulation is part of a paradigm shift in the deployment of surrogate models toward increasing real-world complexity. Sequentially, the cAE spatially compresses high-dimensional flow fields into a low-dimensional latent space, wherein the system's temporal dynamics are learned via neural ODEs. Once trained, the model generates fast spatiotemporal predictions from initial conditions and specified operating inputs. By learning a surrogate to replace the entirety of the time-evolving simulation, the cost of predicting an ignition trial is reduced by several orders of magnitude, allowing efficient exploration of the input parameter space. Further, as the current framework yields a spatiotemporal field prediction, appraisal of the model output's physical grounding is more tractable. This approach marks a significant step toward real-time digital twins for laser-ignited rocket combustors and represents surrogate modeling in a complex system context.

Impact statement

We present a dynamical autoencoder (DnAE) framework that couples cAEs with neural ordinary differential equations (ODEs) to predict ignition trajectories in laser-ignited rocket combustors. The problem representation first reduces the simulation data from high-dimensional fields of primitive variables to path-integrated quantities that mimic experimental data acquisition. The resulting quantities are compressed via cAEs to a low-dimensional manifold, and subsequently, temporal evolution is forecast with parameterized neural ODEs, extending the application of such models beyond existing simple dynamics emulators to a complex engineering problem of interest. The DnAE reduces ignition prediction cost by orders of magnitude relative to large-eddy simulation while accurately capturing bifurcating ignition success and failure dynamics, representing a step toward real-time digital twins for propulsion systems.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

1. Introduction

Surrogate models are commonly used to accelerate the search for promising designs by replacing expensive evaluations or simulations (Forrester et al., 2008) or to estimate variability in quantities of interest (QoI) for an engineering system (Sudret et al., 2017). In many engineering applications, the surrogate represents a mapping $f:\mathbb{R}^d \to \mathbb{R}$, where the input vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)$ encodes design parameters and variables describing the uncertainty present in the system, and the output is a scalar representing local or system-level performance measures. There is no shortage of applications; examples include polynomial chaos expansions applied to wake flows (Lee et al., 2025; Lu et al., 2025) and Gaussian process regression used to compare experiments and simulations of a prism bluff body (Duan et al., 2019). Prediction of field quantities is more challenging and does not always provide direct engineering insight. Nonetheless, vector-to-vector surrogates of the form $f:\mathbb{R}^d \to \mathbb{R}^m$ have been explored, where the input encodes parametric dependence or initial conditions, and the output is a reduced feature-space representation of the flow field. Examples include parametric extensions of proper orthogonal decomposition (POD) for compressible aerodynamics (Bui-Thanh et al., 2003), and temporal extensions of POD such as dynamic mode decomposition (DMD) applied in a predictive capacity to simple partial differential equations (PDEs) (Lu and Tartakovsky, 2020). More recently, for steady aerodynamic fields, nonlinear dimensionality reduction has proven to be capable of extreme compression (Saetta et al., 2024b), and the flexibility of neural network architectures has facilitated strategies for entire flow-field generation. Subsequent extensions that encompass time-series learning efforts have notably used long short-term memory networks, transformers (Solera-Rico et al., 2024), and neural ordinary differential equations (neural ordinary differential equations [ODEs]/NODEs) to learn latent spaces of simpler spectrum problems such as the 1D Kuramoto–Sivashinsky equation for evolution on an inertial manifold (Linot and Graham, 2022), or two-dimensional Kolmogorov flow (Chakraborty et al., 2024). Neural ODEs provide a simpler, dynamics-consistent framework whose behavior is easy to inspect, in contrast to the comparatively complicated transformer architectures, which are more akin to a black-box model.

In the present study, a surrogate framework is applied to a reliability study in the context of rocket ignition success. In particular, we wish to assess binary ignition sensitivities in the combustor, in response to a high-dimensional input parameter space. A secondary aim is that the surrogate model provides realistic time-resolved spatial dynamics, which rules out the use of standard supervised classification methods. A natural approach to surrogate modeling is to first build a database, consisting of precalculated samples obtained from classical simulation methods, which then trains the surrogate model across an input parameter vector $\boldsymbol{\xi}$, with components ξ_i sampled according to their input probability distributions, e.g., $\Pr(\xi_i) \sim \mathcal{N}(\mu, \sigma^2)$ or $\Pr(\xi_i) \sim \mathcal{U}(a, b)$, and in this study chosen to be experimentally informed where possible (Strelau et al., 2024). In a laser-ignited rocket combustor context, each time a laser deposits an energy kernel, significant variability in the laser operating conditions (Zahtila et al., 2023) and more generally in the combustor system occurs, meaning that (i) no two ignition trials are the same, and (ii) variations in the system input space $\boldsymbol{\xi}$ can lead to bifurcating outcomes in the system response. Whether a deposited laser kernel will lead to a kernel evolution sequence with a successful ignition future state depends on the spatiotemporal evolution of the system, motivating a vector-to-vector surrogate. To predict whether ignition will occur, scale-resolving large-eddy simulation (LES) provides a physically accurate and computationally tractable approach among classical simulation strategies, such as Reynolds-averaged Navier–Stokes, and direct numerical simulation (DNS), with predictive ability for the current problem validated against experiments (Brouzet et al., 2025). An LES computation of an ignition trial amounts to time integration of high-dimensional 3D spatial field and is therefore still a costly calculation for the purpose of a single realization of ignition success. In this work, we first carry out an ensemble of LES calculations as ignition trials (Zahtila et al., 2025) to characterize the output space, capturing both input system variability and the resulting multimodal outcomes.

Building on this ensemble of LES trials, a data matrix of spatiotemporal flow realizations is constructed as $\mathcal{S} = \bigcup_{j=1}^{N_c} \left\{ \mathbf{x}^{(j)}(t_0), \mathbf{x}^{(j)}(t_1), \ldots, \mathbf{x}^{(j)}(t_{N_t-1}) \right\}$, where $\mathbf{x}^{(j)}(t_i) \in \mathbb{R}^d$, with d the dimension of the flattened

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

data snapshot, $i = 0, \ldots, N_t - 1$ the discrete snapshot index, and $j = 1, \ldots, N_c$ the LES case index, each with a new input parameter sample $\boldsymbol{\xi}^{(j)}$. Each individual data snapshot $\mathbf{x}^{(j)}(t_i)$ comes from a 2D integrated representation of the 3D flow field (detailed later in the manuscript). The constructed LES dataset therefore captures the range of spatiotemporal dynamics induced by the sampled ensemble of input uncertainty vectors $\boldsymbol{\xi}$ and forms the starting point for reduced-order modeling. The choice and construction of 2D representation for the flow field is crucial because the present system represents an instance of the compressible reacting Navier–Stokes equations and therefore contains many flow features such as shock–shear layer interactions, vorticity in the shear layer with associated turbulence spectrum, and hydrodynamic ejection caused by laser-induced optical breakdown (Wang et al., 2020). The main idea is that the three-dimensional flow field can first be reduced to an informative two-dimensional representation, such as through integrated fields, that retains the physics relevant to ignition while omitting information that is not central to the prediction task. These representations can then be compressed into a low-dimensional latent space, whose coordinates are learnable as functions of the input parameter vector. The present framework therefore nonlinearly spatially compresses the dataset through autoencoders (AEs) and subsequently learns trajectories in the latent space through a neural ODE. These two components together are sufficient to build a spatiotemporal surrogate model.

1.1. Dimensionality reduction

Dimensionality reduction is a well-established technique in fluid mechanics and has been popular for extraction of flow features and flow modes (Brunton et al., 2020; Jaroslawski et al., 2025; Lu et al., 2025). The most popular technique is POD, in which a field is expanded as $u(\mathbf{s}, t) = \sum_{k=1}^{\infty} a_k(t)\phi_k(\mathbf{s})$, where $\mathbf{s}$ denotes spatial position, $\phi_k(\mathbf{s})$ are orthonormal spatial basis functions, or POD modes, that capture the dominant structures in the data, and $a_k(t)$ are the corresponding time-dependent modal coefficients. POD projects the dataset $\mathcal{S}$ onto an optimal linear basis, and this may be useful firstly in inspecting the POD modes to determine the flow structures ranked in terms of energy content, but second, a truncated representation can serve as a reduced basis for compression of the original dataset. Although POD is successfully able to identify dominant coherent structures, it often requires a large number of modes to capture multiscale turbulence (Alfonsi and Primavera, 2007; Muralidhar et al., 2019) and can become unreliable when strong nonlinear dynamics dominate, motivating alternatives such as DMD (Schmid, 2022). A machine learning counterpart to POD is the AE, which encodes the data matrix into a compact latent space and subsequently decodes it for reconstruction. For a shallow two-layer AE with linear activation functions, the learned basis recovers the POD modes (Milano and Koumoutsakos, 2002), so that choosing a latent dimension N_ℓ corresponds to retaining the first N_ℓ POD modes. The latent representation is given by the vector $\mathbf{V} = (V_0, \ldots, V_{N_\ell - 1}) \in \mathbb{R}^{N_\ell}$, where each component V_ℓ is a latent coordinate learned from the data. When nonlinear activation functions are introduced, the network instead learns a nonlinear manifold embedded in the full state space $\mathcal{M} \subset \mathbb{R}^d$, allowing more aggressive compression than linear POD (Zeng et al., 2022; Kim and Heiland, 2023). In this work, we opt for a deterministic AE rather than a variational (VAE) counterpart to achieve dimension reduction into a latent space, to allow the latent space to more clearly capture the bifurcations between ignition success and failure.

1.2. Temporal evolution methods

After spatial compression, we predict the temporal evolution of the compressed variables: in POD, this would correspond to forecasting the modal coefficients $a_k(t)$, while for the AE, it corresponds to predicting the dynamics of the latent vector $\mathbf{V}(t)$. A neural ODE models the continuous evolution of the latent space variables as a system of ODEs,

$$\frac{d\mathbf{V}(t)}{dt} = f(\mathbf{V}(t), t; \boldsymbol{\theta}), \tag{1.1}$$

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

where $\mathbf{V}(t)$ denotes the latent state at time t, and $\boldsymbol{\theta}$ denotes the learnable parameters of the network. The neural ODE can be viewed as a continuous-depth analog of a residual network, where discrete residual updates are replaced by continuous evolution under an ODE (He et al., 2016; Chen et al., 2018). It is worthwhile to point out the similarity with the classical semidiscretization procedure, in which the governing equations are discretized in space, thereby reducing the PDEs to a system of coupled ODEs in time (Ferziger and Perić, 2002). By construction, this NODE formulation is Markovian in the latent space: the future trajectory depends only on the present state $\mathbf{V}(t)$ (and time t) and does not explicitly encode long-range temporal memory as in recurrent or attention-based architectures. Once the AE and NODE architectures are trained, the resulting surrogate is deterministic, producing a single latent trajectory and reconstructed field sequence for a given initial latent state and model input. This contrasts with previous work on the evolution of ignition kernels that applied stochastic strategies based on Wiener processes (Chung et al., 2024). The NODE framework has for some time been established in its continuous layers design and application (Chen et al., 2018), but typically, usage for time-series prediction has been limited to simpler trajectories such as spirals, or in the case of freely decaying turbulence, input variance through a single parameter, corresponding to initialization of turbulent energy spectrum (Portwood et al., 2019).

1.3. State of the art and outline of paper

In this work, we opt for a machine learning strategy because of the bifurcating dynamics, which are highly nonlinear and the need for a framework that can handle parametric dependency as well as state evolution. Linear methods that approach this aim are unsuitable for a variety of reasons. For instance, there is no expectation to capture bifurcating dynamics through DMD, which fits one global linear operator (Schmid, 2010; Lu and Tartakovsky, 2020). The closest linear time-stepping comparison to the present study would be POD spatial compression followed by linear autoregression with exogenous input (ARX) but as noted in preceding work in modeling of a cylinder wake (Siegel et al., 2008), a linear model was insufficient, and a network architecture was required. To meet this requirement, we introduce the dynamical autoencoder (DnAE) framework for the forecast of laser-induced rocket ignition.

The present effort may be viewed as a system-forecasting problem in combustion, a field in which machine learning has already seen considerable uptake across a diverse range of approaches. Deep learning has proven useful for accelerating chemical calculations (e.g., Sharma et al., 2020; Wan et al., 2021). Of particular relevance to the present study, since the thermochemical evolution equation itself takes the form of an ODE, one approach has been to train a neural ODE as a low-cost surrogate for the solution of detailed chemical kinetics (Owoyele and Pal, 2022). An alternative to explicit model-form treatment is the super-resolution of reacting fields (Chung et al., 2023), where 3D super-resolution benchmarks on DNS-derived turbulent flow data demonstrated sensitivity to architecture and incorporating physics-based loss terms. A further direction is the data-driven classification and modeling of combustion regimes in detonation waves, where machine learning has been used to identify regime structure and inform reduced-order modeling (Barwey et al., 2021). Closely related to the present aim, there is a wide variety of machine learning approaches for the prediction of ignition delay in combustion processes (Molana et al., 2024). More broadly, reviews of machine learning in combustion highlight the growing role of data-driven methods in classification, regime identification, and time-sequence prediction (Zhou et al., 2022; Mohammadi et al., 2023; Wu et al., 2025).

This paper initially provides a comprehensive description of the input parameter vector in Section 2.1, including the distributional assumptions and parameterization that define system variability. It then outlines the construction and choice of the QoI in Section 2.2, with possible candidates of several computational path-traced imaging modalities inside the combustor. Following this, the methodology for constructing the surrogate model is detailed in Sections 2.3 and 2.4, encompassing both spatial dimensionality reduction via AEs and the subsequent latent space temporal evolution using neural ODEs. The associated training strategy, including optimization procedures and data handling protocols, is then described in Section 3. Model performance is then quantitatively assessed in Section 4. An overall schematic of the data-driven surrogate model construction is given in Figure 1. With the surrogate model

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

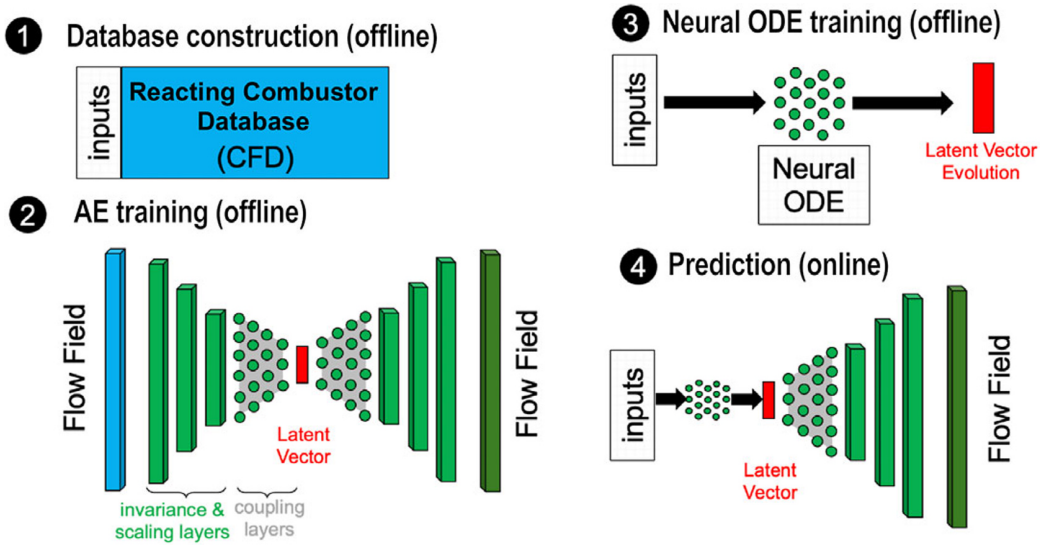


Figure 1. *Workflow describing the construction and deployment of the joint neural ODE and decoder as a generative model.*

ready for deployment, generative trajectories are obtained for a large Monte Carlo ensemble of ignition trials, with $N_{\mathrm{MC}} = 10^6$ surrogate evaluations. This enables a more detailed characterization of the ignition response across the input space, including decision boundaries for successful rocket ignition, and substantially improves upon the sparse and under-converged ignition probability maps obtained from the limited LES ensemble, as presented in Section 4.3.

2. Data and methodology

2.1. Sources of uncertainty and data representation

The ensemble of LES ignition trials described in Section 1 provides a data-driven view of how different operating conditions lead to ignition success or failure. To interpret and generalize these outcomes, we formalize the system variability within a structured uncertainty quantification (UQ) campaign (Cutforth et al., 2026). This framework identifies 15 input parameters (summarized in Table A1 in Appendix A) that encode variability (i.e., uncertainty) originating from the companion experimental campaign, the combustor system, and model-form assumptions. Specifically, the bounds on these parameters were obtained either from direct measurements in the reacting combustor experimental campaign (Strelau et al., 2024), or from detailed simulations of a laser spark matched to quiescent combustor experiments (Brouzet et al., 2025), from which the laser parameters were statistically inferred from the spark evolution. The remaining parameters were treated as model-form uncertainties and assigned narrow prior ranges to guard against overlooked sensitivities in the ignition response. These inputs form the parametric drivers for the surrogate model. Briefly, the uncertainty set includes epistemic variations in laser spark characteristics and focal accuracy (ξ_0–ξ_5). Deposition timing is influenced by two distinct mechanisms: (a) small-scale lags that introduce aleatoric uncertainty through turbulence-driven shifts (ξ_6) and (b) large-scale lags that capture epistemic uncertainty due to variability in the available methane system mass (ξ_7). Additional sources include modeling uncertainties in flame propagation (ξ_8–ξ_{10}), model-form uncertainty in the subgrid stress coefficient (ξ_{11}), epistemic uncertainties tied to fuel–oxidizer intake conditions (ξ_{12}–ξ_{13}), and geometric variability in combustor cross-section (ξ_{14}). The subsequent ignition trial computations were generated by Monte Carlo sampling of the 15-dimensional uncertain input space, with each parameter drawn from its prescribed marginal distribution and treated as statistically independent. Each

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

ignition trial is an LES simulation of a compressible reacting subscale rocket combustor with the Hypersonics Task-based Research (HTR) solver (Di Renzo et al., 2020), in which an O_2-CH_4 coflowing injector jet develops into a chamber. After development of the preignition turbulent state, a laser spark deposits an energy kernel whose subsequent evolution and interaction with the turbulent coflowing jet determine whether ignition succeeds or fails. The cost of one simulation is a wall-clock time of $\mathcal{O}(10)$ hours on 8 NVIDIA V100 GPU cards, which is a significant reduction relative to the high-fidelity simulation counterpart, achieved through reduced chemistry and strategic meshing together with tuning of deterministic features such as the jet inviscid core and spark evolution at common sample points. A detailed account of the simulations flow rates, timings, and combustor geometry is provided in the corresponding high-fidelity simulation campaign (Brouzet et al., 2026).

A central aim of the UQ campaign was to assess ignition sensitivities in the combustor, and each of the LES computations in the ensemble of ignition trials leads to a response of ignition failure or success. To this end, Figure 2 shows marginal distributions of the input parameters, separated into igniting and nonigniting cases. While some distributions differ only slightly (reflecting sampling convergence), others exhibit clear structural differences. To quantify these differences, the Kullback–Leibler (KL) divergence (Kullback and Leibler, 1951) was computed between the success and failure groups, with parameters exhibiting the largest divergence highlighted in Figure 2 with green check marks. To complement the KL-based marginal comparison, we also evaluated conditional Sobol sensitivities using a random-forest ignition classifier. In a preliminary analysis, ξ_7 (methane system mass) was found to be dominant, followed by the same leading parameters highlighted by the KL analysis; however, the strong influence of ξ_7 reduced the ability of the Sobol analysis to clearly resolve secondary interaction effects among the remaining inputs. Accordingly, ξ_7 was fixed and the Sobol sensitivities were recomputed over the remaining input space. In this conditional analysis, ξ_2 (lobe asymmetry ratio) and ξ_5 (energy deposited) remained the leading contributors, with first-order indices $S_1 = 0.54 \pm 0.06$ and 0.24 ± 0.05, respectively. The largest pairwise interaction was (ξ_2, ξ_5), with $S_2 = 0.06 \pm 0.08$, where the confidence bounds were estimated by bootstrap resampling of the surrogate model evaluations. These statistical trends have a clear physical interpretation. Increasing the deposited laser energy (ξ_5) produces a larger ignition kernel, making it more likely to overcome turbulent transport cooling and breakup (Passiatore et al., 2024). Increasing the methane system mass (ξ_7) increases the fuel available to the evolving kernel and improves the chance of sustained flame growth in locally favorable mixture regions. The ξ_2–ξ_5 interaction reflects the coupled role of spark structure and spark strength: ξ_5 sets the kernel energy, while ξ_2 influences the laser-spark asymmetry and hence its propagation toward favorable chemistry (Wang et al., 2020). However, ignition is fundamentally a Bernoulli random variable (success/failure), even if a continuous proxy is used, e.g., the maximum chamber pressure over time instead of a Boolean label. The underlying problem of predicting a sharp regime transition between ignition and failure is not removed by the proxy. Consequently, reliable probability estimates are slow to converge. For example, with only $N = 100$ independent trials and a nominal ignition probability of $p = 0.5$, the Wilson score interval (Wilson, 1927) still gives an uncertainty range of approximately [0.40,0.60], highlighting the substantial sampling uncertainty associated with limited LES ensembles. Given finite compute allocations and therefore limited sampling, the available LES ensemble is insufficient for robust convergence, and this motivates the construction of the DnAE surrogate model to efficiently augment ignition probability quantification. A comprehensive joint characterization of the most informative parameters is deferred; we return to it later using surrogate-enabled Monte Carlo.

To assemble a statistically useful LES dataset for reduced order modeling, we generated $N_c = 300$ ignition trials by first constructing tailored LES (Cutforth et al., 2026) built from large-scale experimentally validated calculations (Passiatore et al., 2024; Brouzet et al., 2025). The tailored LES computations differ from the increased scale-resolving calculations, for which $\mathcal{O}(1-10)$ are feasible, because they employ a significantly coarser mesh and feature reduced chemistry and are calibrated by employing an Ensemble Kalman filter (EnKF) for deterministic processes. This procedure matched the tailored LES to emulate the high-fidelity deterministic processes in the laser kernel dynamics and turbulence quantities in the coflowing jet. With this calibration in place, larger ensembles became feasible. The choice of $N_c = 300$ trials was

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

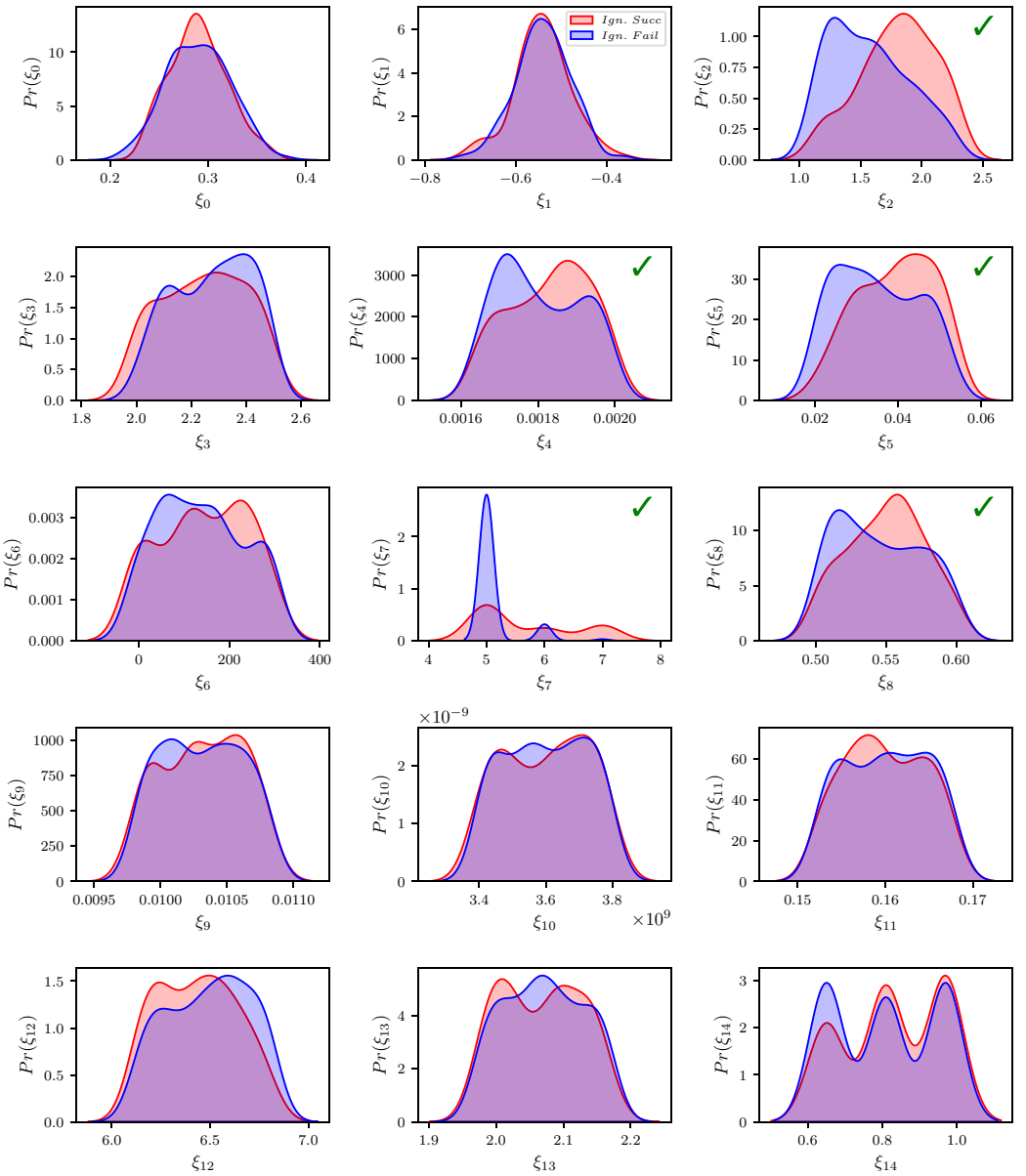


Figure 2. *Ignition probability density functions (PDFs) marginalized over the input parameters ξ_0 through ξ_{14}, comparing ignition success (red) and ignition failure (blue) through kernel density estimates; green check marks denote the five parameters with the largest KL divergence between igniting and nonigniting cases. Full description of parameters in Table A1 in Appendix A.*

guided by computational cost constraints and the convergence of the cumulative ignition-probability estimate, with the estimate $\hat{P}_{\text{ign}}$ stabilizing to within $\pm 5\%$ after $N_c = 300$ samples. The resulting dataset enabled the construction of ignition probability maps. Logistic fits of ignition probability as a function of radial focal position were consistent with the experimental probability maps in both the mean prediction and the 95% confidence interval, indicating statistical agreement between the computational ensemble data and the laboratory model rocket-combustor experiments (Strelau et al., 2024).

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

2.2. Quantity of interest: thermal imaging

Having generated a large and statistically useful tailored LES data ensemble, we proceed to select an appropriate representation of the combustor dynamics on which to train and develop the surrogate model. In this predictive rocket-ignition context, we therefore require a reduced data representation that remains physically meaningful yet tractable for learning. A further requirement for validation would be that the representation is relatable to the observables from experimental sensing modalities. Machine learning surrogate models trained on reduced or structure-promoting representations are preferable in turbulence due to the extremely high dimensionality, chaotic nature, and multiscale features of full 3D fields. Recent examples include, in the context of multiphase flows, investigation on the choice of interface treatment through diffuse, sharp, and level set approaches (Cutforth and Mirjalili, 2025), and in urban street canyons, 2D particle image velocimetry was preferable at a judiciously chosen plane that captured turbulence quadrant events (Jaroslawski et al., 2025). These reduced forms capture essential dynamics while significantly lowering computational cost and mitigating overfitting, making learning more tractable and generalizable. The essential features in the combustor include (1) coflowing (fuel and oxidizer) turbulent intake, (2) deposited energy kernel location and dynamics, and in the successful ignition scenario, (3) flame propagation and stable combustion.

Figure 3(a) displays rendered output from the computational fluid dynamics (CFD) solver (Di Renzo et al., 2020) and the associated flame visualization. To emulate the experimental data acquisition, our first

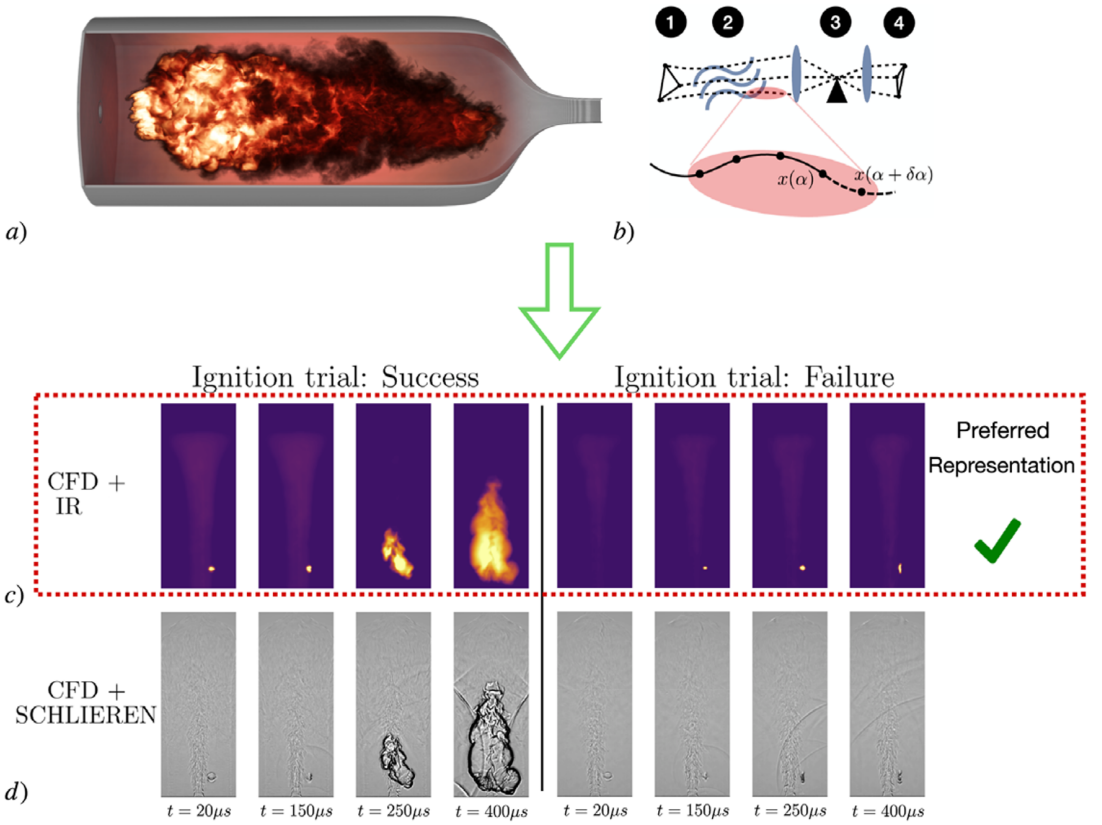


Figure 3. *Quantity of interest: spatiotemporal fields, (a) rendered simulation output of the successfully ignited combustor, (b) illustration of the integrated quantity acquisition, (1) light beam generation, (2) ray evolution according to the Eikonal equation, (3) Snell's law and occlusion, and (4) irradiance sampling, (c) successful and unsuccessful time series for the computational infrared representation, and (d) computational Schlieren.*

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

approach is to computationally emulate high-speed Schlieren by integrating light beams for a collimated light source passing through the variable density flow. The trajectories obey the governing equation derived from Fermat's principle:

$$\frac{d}{ds}\left(n\frac{d\mathbf{r}}{ds}\right) = \nabla n, \tag{2.1}$$

where $\mathbf{r}(s)$ is the light-beam trajectory parameterized by arc length s, and n is the refractive-index field calculated from the species mass fractions and Gladstone–Dale mixture relation, with full details in supplementary material of high-fidelity simulations (Brouzet et al., 2026), as part of a wider paradigm of digital enhancement of flow data (Rossinelli et al., 2026a,b). The path-tracing approach illustrated in Figure 3(b) mimics experimental line-of-sight integration to reduce the data to a 2D field quantity. While informative, this high-dimensional representation includes numerous flow and combustion features that are not necessarily predictive of ignition success.

The computational infrared (IR) imaging emulates the methodology used for chemiluminescence modeling (Brouzet et al., 2026) and instead of integrating refractive beam trajectories, this method relies on the orthographic projection of a volumetric emission field. While the chemiluminescence model utilizes species concentrations to define the source term, radiative emission is modeled to reproduce thermal imaging from the temperature field. Representative time series for both successful and failed ignition events are shown using IR imagery in Figure 3(c) and computational Schlieren imaging in Figure 3(d). Although the Schlieren images capture a wide range of turbulent spatial scales, we prioritized the CFD IR representation due to its more informative representation of spark location and subsequent dynamical evolution of the laser-deposited energy kernel. For the conducted $N_c = 300$ CFD cases, we generated $N_t = 20$ aligned postdeposition IR snapshots per case. This resolution captured the essential kernel dynamics and combustion evolution while balancing memory requirements to fit the resultant data matrix on a single graphical processing unit (GPU).

2.3. Autoencoder

Despite the spatial reduction from 3D to 2D through integrated quantities, the IR fields described in §2.2 must be further reduced in dimensionality in order to enable temporal modeling in the first place. We choose nonlinear convolutional autoencoders (cAEs) because they have been demonstrated to yield efficient compression while retaining essential physical structures (Saetta et al., 2024b); however, alternative approaches for learning latent manifolds are also available (Fan et al., 2025; Lu et al., 2026). In the present framework, we therefore employ a cAE that learns a low-dimensional latent representation using an architecture composed of three convolutional blocks that progressively extract multiscale spatial features while reducing resolution. These blocks utilize increasing channel counts and intervening pooling operations, augmented with residual connections and nonlinear activation functions. Subsequently, fully connected layers nonlinearly combine these multiresolution features to compress the $\mathbf{x} \in \mathbb{R}^{592\times 240}$ input IR fields to an eight-dimensional latent state $\mathbf{V}$. Finally, the decoder mirrors this configuration, employing symmetric upsampling and convolutional blocks to reconstruct the full high-resolution input space from $\mathbf{V}$, see Figure 4(a,c).

As a preprocessing step, each image matrix is min–max normalized to the range $[0,1]$ through affine rescaling, ensuring consistent numerical scale and improved conditioning during training (LeCun et al., 2012). Each normalized light–intensity field is passed to the AE as a single-channel 2D input, on which the network performs a nonlinear encoding–decoding operation:

$$\mathbf{V} = e(\mathbf{x}), \quad \hat{\mathbf{x}} = d(\mathbf{V}), \quad \hat{\mathbf{x}} = d(e(\mathbf{x})), \tag{2.2}$$

where $e : \mathbb{R}^{m\times n} \to \mathbb{R}^{N_\ell}$ is the encoder mapping for an input with dimensions $m \times n$, and $d : \mathbb{R}^{N_\ell} \to \mathbb{R}^{m\times n}$ the decoder mapping. Here, $\mathbf{x}$ denotes the input field, $\hat{\mathbf{x}}$ its reconstruction, and $\mathbf{V} \in \mathbb{R}^{N_\ell}$ the latent variables. Following Saetta et al. (2022), the optimization objective is a regularized reconstruction loss

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

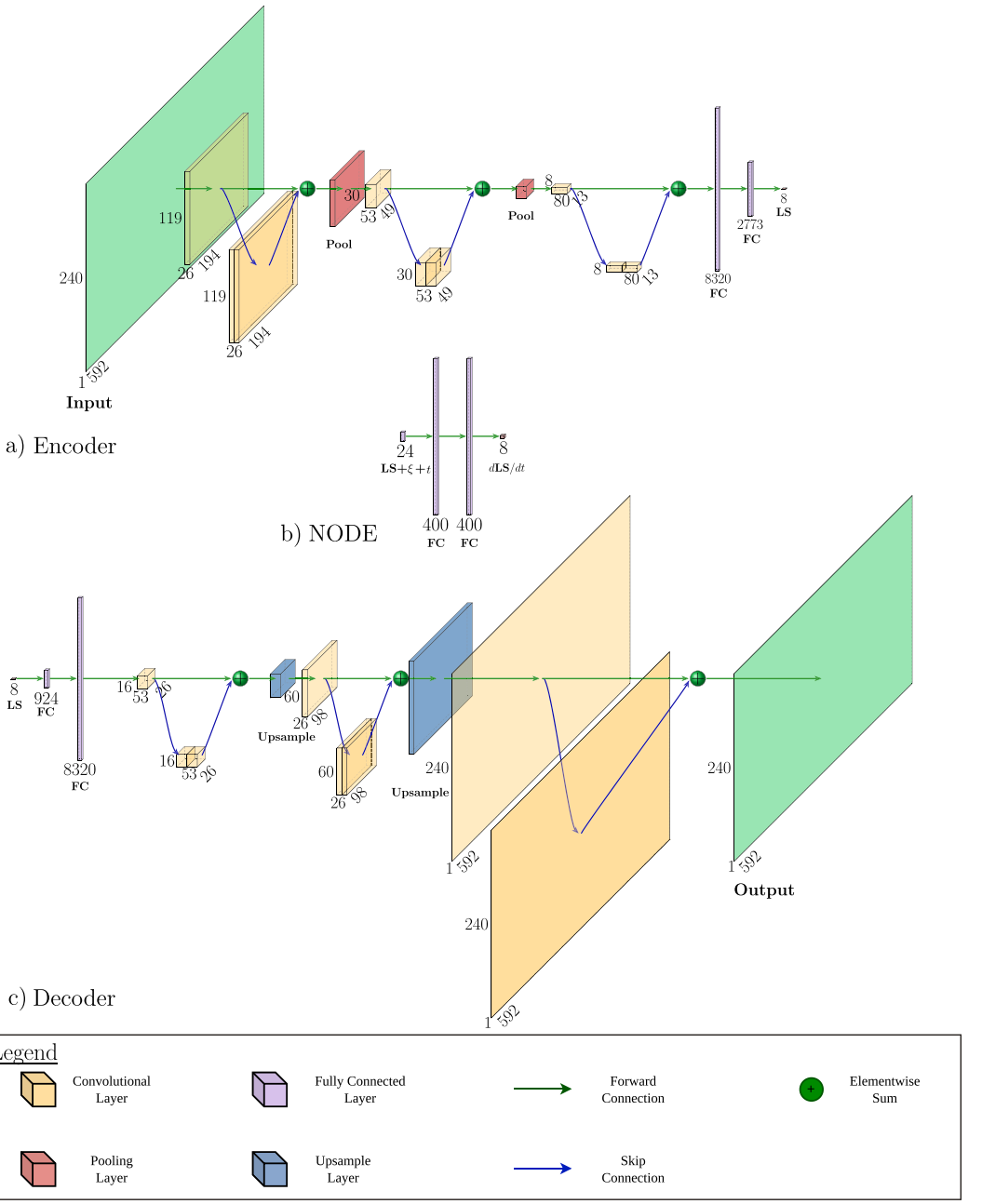


Figure 4. *Architecture schematic for the DnAE. (a) The encoder maps each single-channel IR field to an 8-dimensional latent state* $\mathbf{V}$ *using convolutional, pooling, residual, and fully connected layers. (b) The NODE represents the latent-space right-hand side: at each integration step, the augmented state* $(\mathbf{V},\boldsymbol{\xi},t)$ *is passed through two fully connected layers to evaluate the latent space derivative. Repeated evaluations within the time integrator advance the latent trajectory over the ignition trial. (c) The decoder maps* $\mathbf{V}$ *back to image space using fully connected, upsampling, and convolutional layers.*

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

$$\mathcal{L}_{\mathrm{AE}} = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2 + \lambda_{\mathrm{AE}} \frac{1}{N_e}\sum_{q=1}^{N_e} \left(l_q^e\right)^2, \tag{2.3}$$

where $\mathbf{x}_i$ is the ith AE training snapshot, sampled from any training ignition case and any available time snapshot, $\hat{\mathbf{x}}_i$ is its reconstruction, N is the total number of AE training snapshots, l_q^e denotes the qth encoder weight, N_e is the total number of encoder weights, and λ_{AE} is a user-defined regularization coefficient. In our implementation, we set $\lambda_{\mathrm{AE}} = 10^{-6}$ for L_2 autoencoder regularization, consistent with prior work on aerodynamic AEs. The latent space dimension N_ℓ is a critical hyperparameter that controls the latent-space manifold capturing the essential flow physics.

A key property of this construction is that the AE is deterministic: for any unseen input $\mathbf{x}$, the encoder yields a unique latent vector $\mathbf{V} = e(\mathbf{x})$, which the decoder deterministically maps back to the reconstructed field $\hat{\mathbf{x}} = d(\mathbf{V})$. In contrast to variational autoencoder (VAE) approaches that introduce stochasticity in the latent space (Fan et al., 2025), this deterministic structure ensures reproducibility of trajectories and avoids additional assumptions on the latent distribution (Saetta et al., 2024b). In steady-state settings, once the latent space is established, the decoder can be deployed as a generative model by interpolating within the latent manifold (Saetta et al., 2024b). For example, new realizations may be generated as $\mathbf{V}(\boldsymbol{\xi}) = \mathcal{I}\left(\boldsymbol{\xi}; \left\{\left(\boldsymbol{\xi}^{(j)}, \mathbf{V}^{(j)}\right)\right\}_{j=1}^{N_c}\right)$, where $\mathcal{I}$ denotes an interpolation operator acting on the latent ensemble and associated input parameters, yielding $\hat{\mathbf{x}}(\boldsymbol{\xi}) = d(\mathbf{V}(\boldsymbol{\xi}))$. For the present unsteady ignition problem, this idea is extended by learning the temporal evolution of the latent state with a parameterized neural ODE, rather than interpolating the full latent trajectory directly. Thus, the decoder remains the generative map from latent space to image space, while the NODE supplies the time-dependent latent trajectory.

2.4. Parameterized neural ODE

The temporal dynamics of the latent space representation $\mathbf{V}^{(j)}(t)$ for ignition trial j are advanced using a parameterized NODE framework to model the continuous evolution of the latent space variables, which forms a system of coupled ODEs,

$$\frac{d\mathbf{V}^{(j)}(t)}{dt} = f\left(\boldsymbol{\xi}^{(j)}, t, \mathbf{V}^{(j)}(t); \boldsymbol{\theta}\right), \tag{2.4}$$

where the present model form differs from the vanilla neural ODE shown in (1.1) by the incorporation of the input parameter vector $\boldsymbol{\xi}^{(j)}$, and as a reminder $\boldsymbol{\theta}$ denotes the learnable parameters of the network. The vanilla neural ODE is sufficient, for example, to emulate the latent space dynamics for flow past a cylinder with a single instance of the parameter space (Rojas et al., 2021), but the parameterized extension can account for variations in the input parameters (Lee and Parish, 2021). Instead of introducing uncertainty via a VAE in the preceding dimensionality reduction stage, we handle parametric uncertainty here by conditioning the latent NODE dynamics directly on the input vector $\boldsymbol{\xi}$.

The network in this study is a simple feed-forward network, but additional network features can be added, such as through convolutional layers (Shankar et al., 2020). Here, $f(\cdot)$ is a multilayer perceptron with two hidden layers of width w each with tanh activation, shown within the DnAE in Figure 4(b), mapping the input state $\left(\boldsymbol{\xi}^{(j)}, t, \mathbf{V}^{(j)}(t)\right)$ to the latent derivatives $\dot{\mathbf{V}}^{(j)}(t)$. Formally, the latent trajectory for ignition trial j can be written as an initial value problem (Chen et al., 2018):

$$\mathbf{V}^{(j)}(t) = \mathbf{V}^{(j)}(t_0) + \int_{t_0}^{t} f\left(\boldsymbol{\xi}^{(j)}, \tau, \mathbf{V}^{(j)}(\tau); \boldsymbol{\theta}\right) d\tau, \tag{2.5}$$

which expresses the time series as the initial condition plus the accumulated effect of the latent dynamics. Therefore, given an initial latent state $\mathbf{V}^{(j)}(t_0)$ from the AE compression for ignition trial j, the NODE-predicted trajectory $\hat{\mathbf{V}}^{(j)}(t)$ is propagated from this initial condition over discrete times $\{t_i\}_{i=0}^{N_t-1}$ by

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

$$\hat{\mathbf{V}}_{i+1}^{(j)} = \hat{\mathbf{V}}_i^{(j)} + \Delta_{\mathrm{RK4}}\left(f, t_i, \Delta t, \hat{\mathbf{V}}_i^{(j)}\right), \qquad i = 0, \ldots, N_t - 2, \tag{2.6}$$

where Δ_{RK4} denotes the fixed-step fourth-order Runge–Kutta increment applied to the parameterized vector field in (2.4), $\Delta t = t_{i+1} - t_i$, i indexes the discrete time snapshots, j indexes the ignition trial, and $\hat{\mathbf{V}}_i^{(j)} \equiv \hat{\mathbf{V}}^{(j)}(t_i)$. The loss function was specified to enforce trajectory reconstruction accuracy and consistency of temporal derivatives. Specifically,

$$\mathcal{L}_{\mathrm{NODE}} = \frac{1}{N_c(N_t-1)} \sum_{j=1}^{N_c} \sum_{i=1}^{N_t-1} \|\mathbf{V}_i^{(j)} - \hat{\mathbf{V}}_i^{(j)}\|_2^2 + \lambda_c \frac{1}{N_c(N_t-2)} \sum_{j=1}^{N_c} \sum_{i=1}^{N_t-2} \|\dot{\mathbf{V}}_i^{(j)} - \dot{\hat{\mathbf{V}}}_i^{(j)}\|_2^2, \tag{2.7}$$

where $\mathbf{V}_i^{(j)}$ is the true latent state for ignition trial j at time snapshot t_i, $\hat{\mathbf{V}}_i^{(j)}$ is the corresponding NODE prediction, and $\dot{\mathbf{V}}_i^{(j)}$ is a second-order finite-difference approximation of the latent velocity. The weighting hyperparameter $\lambda_c = 10^{-2}$ balances the overall contribution of the temporal-derivative penalty, which is the last term in (2.7). The inclusion of this temporal-derivative penalty is a soft constraint that encourages the neural ODE to learn directional changes, such as those associated with the latent space bifurcation that occurs in ignition.

Briefly, we remark on some present limitations. Separating time and space in our surrogate model's prediction for evolving temperature fields introduces limitations in their physical consistency. As evolution is handled fully by the neural ODE in a low-dimensional latent state, spatial interactions become effectively nonlocal through the learned AE representation, weakening the system causality. Second, because the latent space itself is not tied to physical conserved quantities, it is difficult to enforce constraints such as conservation laws directly; physics-informed NODE variants such as PINODE (Sholokhov et al., 2023) partially address this by embedding PDE structure into the dynamics, but they require explicit knowledge of the governing equations, which we do not have for our CFD IR fields.

3. Training

For the bifurcating physics and complex spectrum in rocket ignition and combustion, the effectiveness of the proposed framework relies critically on training methodology; we therefore describe the procedures adopted for the AE and neural ODE to robustly learn the compressed spatial fields and latent trajectories. For both AE and NODE training, the dataset was split by ignition trial into training and validation subsets, with 10% of the original cases held out for validation. The spatial and temporal components are trained separately to simplify the schedule and, more importantly, joint optimization incurs higher computational cost without necessarily yielding performance gains (Vlachas et al., 2022). For spatial compression, the encoder reduces the input data matrix $\mathcal{S}$ to a compact latent space, and the decoder mirrors this architecture for reconstruction. The immediate output of this stage is therefore a latent coordinate $\mathbf{V} = e(\mathbf{x})$ for each individual IR snapshot $\mathbf{x}$; temporal or case-dependent trajectory structure is only imposed subsequently by postprocessing these latent coordinates according to their ignition trial and snapshot index. Confidence in the network design stems from prior work on spatial compression of flow fields (Saetta et al., 2022) and preliminary studies therein evaluating the influence of architectural hyperparameters (e.g., number of layers, activation functions). The latent space dimension, N_ℓ, is varied, and corresponding training results are presented in Figure 5(a). The validation performance for $N_\ell = 4, 8$, and 16 shows a continuous, albeit small, improvement. We selected $N_\ell = 8$ as a tradeoff, offering good compression while retaining a relatively low latent space dimension. Visualizations of the reconstructed images show that increasing the latent space dimension of the AE first recovers the large-scale morphology of the flame, with finer-scale turbulent structure and sharper local features emerging only at higher N_ℓ. As an aside, when the computational Schlieren in Figure 3(d) is used for training, the scale-retaining features of this imaging modality required a significantly higher latent space dimension before meaningful compression was achieved.

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

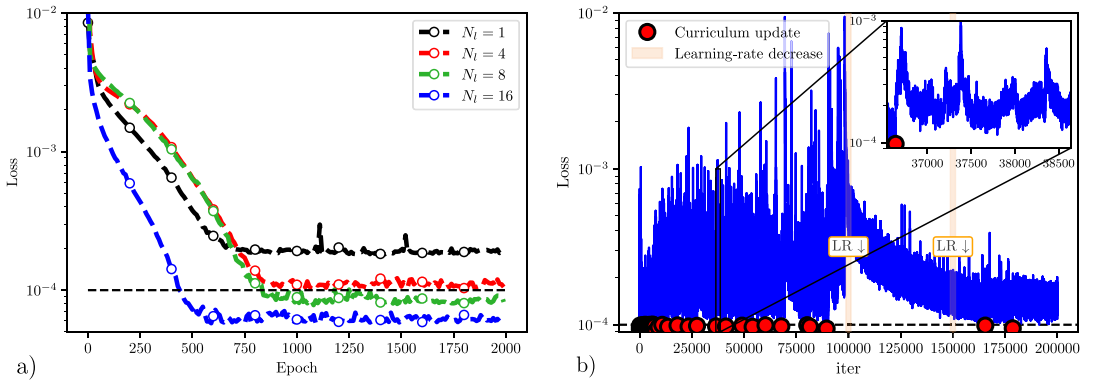


Figure 5. *(a) Evolution of the validation loss against epoch during training of the AE, for choices of latent space dimension hyperparameter $N_\ell = 1,4,8$ and 16. (b) Representative loss during training of the NODE over 200,000 iterations, for hidden-layer width hyperparameter $w = 400$. The horizontal black dashed line indicates the loss threshold for a curriculum increment. Red markers denote curriculum updates, triggered at specified threshold crossing, and the shaded orange bands indicate scheduled learning-rate decreases.*

Training NODEs on complex trajectories requires backpropagating through repeated evaluations of the neural network during time integration. Despite standard stabilizations, this integration depth often leads to exploding or vanishing gradients and reduced sensitivity (Chakraborty et al., 2024). This difficulty is amplified when the target trajectories contain sharp transients, oscillatory segments, or bifurcating branches, all of which increase the sensitivity of the learned dynamics to small integration errors. Several training strategies have been designed to overcome the reduced sensitivity for neural ODEs, such as shooting methods (Turan and Jäschke, 2021) to capture higher-frequency responses and the multistep penalty method (Chakraborty et al., 2024) that splits trajectories into multiple, nonoverlapping time windows and includes an optimization loss term to penalize the window discontinuities. In the present work, vanilla training of the parameterized NODE model resulted in a loss function that failed to decay, and the learned solution trajectories artificially suppressed oscillatory components rather than capturing the true underlying dynamics. To address this, a curriculum learning strategy was introduced, which shares conceptual features with the multistep penalty method. In general, curriculum learning structures the optimization process such that training begins with simpler tasks and gradually progresses to more complex ones, thereby improving both convergence and generalization (Bucci et al., 2023).

In the present application, task difficulty was defined by the temporal extent of the latent space trajectories. To clarify, all training samples of the input parameter space $\boldsymbol{\xi}$ were available for learning, but only a restricted time interval of each trajectory. This contrasts the curriculum regularization approach (Krishnapriyan et al., 2021) used in training of physics-informed neural networks that progressively samples different parts of the parameter space, leading to lower error rates. Curriculum learning was implemented as a staged optimization over progressively increasing trajectory lengths:

$$\text{Stage loss}: \mathcal{L}(T) = \frac{1}{N_c T} \sum_{j=1}^{N_c} \sum_{i=1}^{T} \|\mathbf{V}_i^{(j)} - \hat{\mathbf{V}}_i^{(j)}\|_2^2, \tag{3.1}$$

$$\text{Update rule}: T \mapsto T + \Delta T \quad \text{whenever } \mathcal{L}(T) < \epsilon. \tag{3.2}$$

Thus, training initially considers only the earliest part of each trajectory, and the target time interval for learning is extended by ΔT whenever the model reaches the prescribed loss tolerance. For NODE training,

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

the encoded trajectories were linearly up-sampled from the original $N_t = 20$ postdeposition snapshots to $N_t' = 100$ temporal samples. The full time series of the trajectory ensemble was divided into $N_L = 50$ folds, chosen as the maximum number that partitions the signal into integer-length segments, so that the curriculum increment was $\Delta T = N_t'/N_L$ samples. The curriculum update tolerance was set to $\epsilon = 1 \times 10^{-4}$, which by inspection corresponded to accurate capture of the characteristic curvature of the trajectories. Accordingly, the variation of the loss during NODE training is shown in Figure 5(b). The sharp increases in loss correspond to curriculum updates, where a new trajectory segment is exposed to the optimizer, while the subsequent decay indicates that the NODE progressively learns the newly introduced portion of the latent trajectory. This staged expansion of the learned trajectory segment is illustrated for representative latent components in Figure 6. As the training horizon is extended, the newly revealed portion of the trajectory initially exhibits larger error, after which optimization adjusts the latent dynamics to recover the correct direction of evolution in the added segment. In practice, these updates reduced the error in the newly introduced interval without noticeably degrading the already-learned earlier portion of the trajectory, although the curriculum update does not guarantee monotonic improvement on the previously exposed segment.

The present training strategy uses a single NVIDIA Tesla P100 (16 GB) GPU for 24 hours to efficiently train the AE to spatially compress the data matrix $\mathcal{S}$ into a learnable latent space, and subsequently, training the parameterized NODE takes 48 hours on a CPU. The key to making the present complex problem learnable in a realistic time frame was (i) a problem representation that retained the laser-spark kernel evolution dynamics, and combustion features, without being dominated by the multiscale complex spectrum of turbulence present in the numerical schlieren of a coflowing jet, and (ii) a curriculum learning strategy that allowed for complex trajectories with multiple modes and bifurcations to be learned.

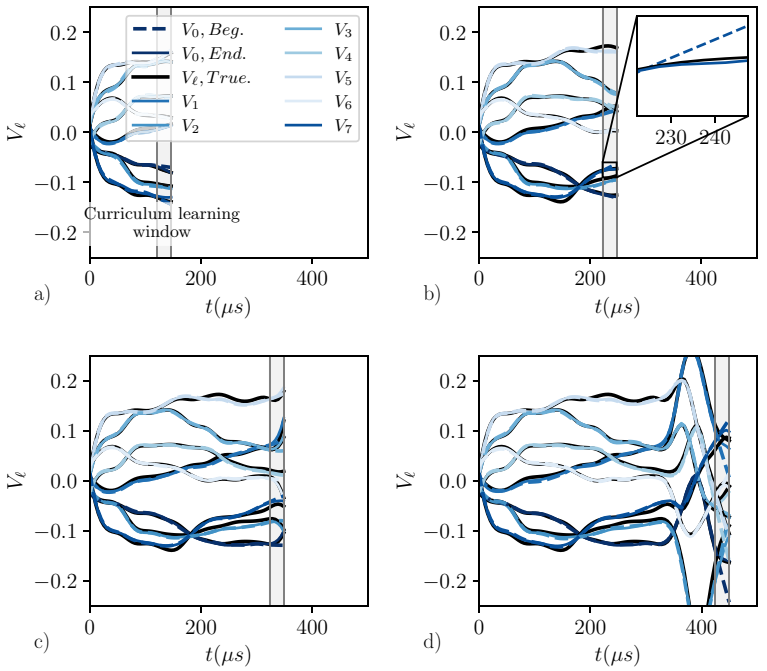


Figure 6. *Representative case of latent space trajectories V_ℓ illustrating the curriculum learning process. Panels (a)–(d) display the advancing training window (gray band), highlighting the progressive learning of newly revealed trajectory segments as the time horizon extends. Dashed lines indicate the beginning prediction at the start of the update, while solid lines show the final learned trajectory.*

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

4. Results

4.1. Latent space trajectories

To assess whether the trained model captures the essential dynamics of ignition, we first examine the latent space trajectories, i.e., the mapping of the CFD IR sequences from image space $\mathbf{x}^{(j)}(t_i) \in \mathbb{R}^{m \times n}$ $(m,n = 592 \times 240)$ by the AE into low-dimensional vectors $\mathbf{V}^{(j)}(t_i) \in \mathbb{R}^{N_\ell}$ that is required to extract meaningful latent representations. At the time of spatial compression, the AE is not temporally aware and simply encodes possible combustor states; as a postprocessing step for analysis, we order the latent vectors in time $\mathbf{V}^{(j)}(t_i)$ for each trial with unique sample of $\boldsymbol{\xi}^{(j)}$, yielding a family of trajectories. Figure 7 displays the latent components V_ℓ as time series across the ensemble; after linear up-sampling, each trial forms a smooth curve. At early times, the coordinates capture the initial stages of laser deposition and kernel formation, with variance across cases small and tightly clustered. Subsequently, the trajectories reflect interaction with the co-flowing fuel–oxidizer shear layer, and finally, when $t \gtrsim 250\mu s$ the bifurcating outcomes of either: (i) runaway combustion/ignition success or (ii) a dissipating kernel/ignition failure.

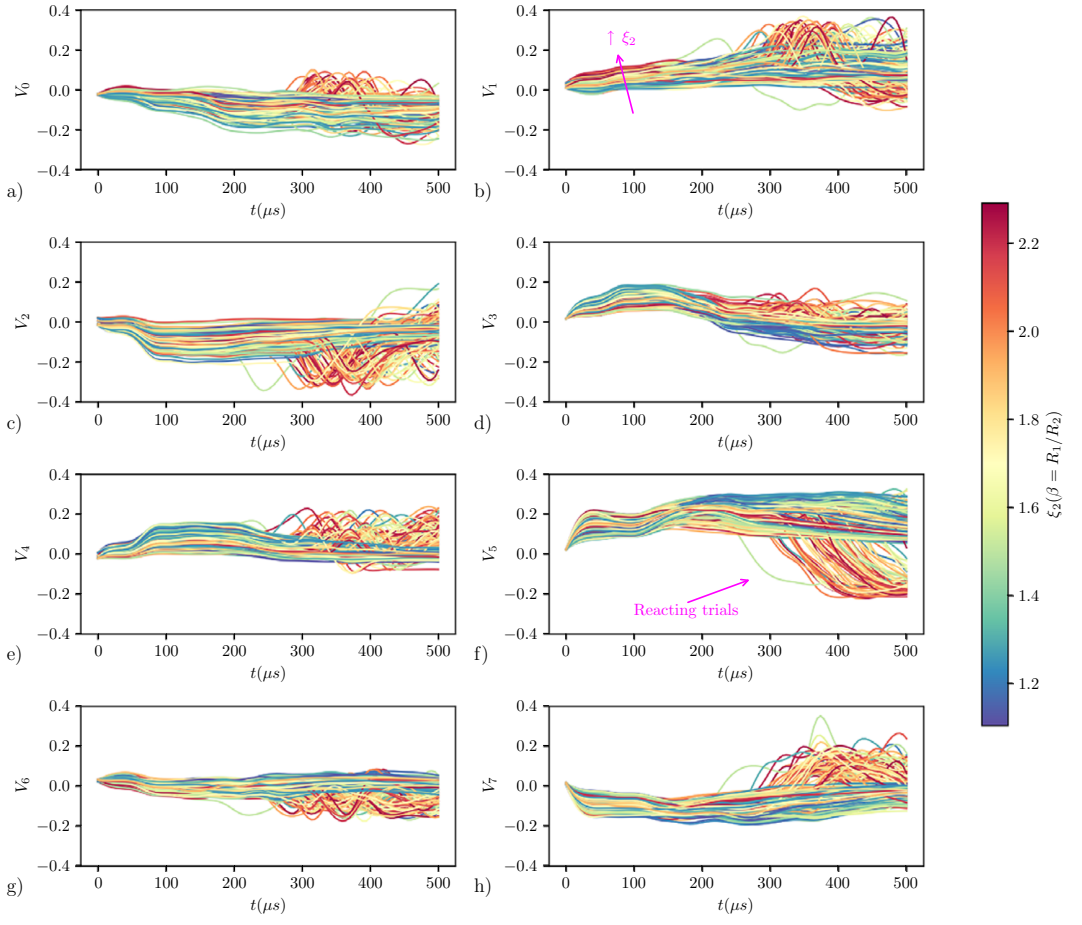


Figure 7. *Temporal evolution of the eight latent-space components for the reacting rocket-combustor LES ensemble. Each curve corresponds to one ignition trial, with color indicating the informative input parameter $\xi_2 = \beta = R_1/R_2$. Each ignition trial contributes one trajectory $\mathbf{V}^{(j)}(t)$ to the ensemble, and each subpanel shows a different component $V_\ell^{(j)}(t)$ of that latent vector.*

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

The latent space thus provides a reduced representation that preserves essential dynamics in compact form. For reference, in previous work on steady aerodynamics, systematic analysis of latent components yielded physical insight into the individual role of inviscid and viscous contributions (Saetta et al., 2024a, b). This was achieved by sequentially training and conditioning the latent space to separate physical processes. In the present application, the latent space could therefore be interrogated by systematically deactivating components of the latent space vector $\mathbf{V}$ to uncover what physics each component V_ℓ is capturing, analogous to modal analysis; however, analysis of the physical processes captured by each latent space variable is beyond the present scope of work. Instead, the present focus is to develop a predictive generative model for rocket ignition. In this setting, as the input vector $\boldsymbol{\xi}$ is sampled for the input conditions for a particular ignition trial, the key requirement for the model is to predict how the corresponding latent state evolves in time, i.e., we require a structured mapping $\boldsymbol{\xi} \mapsto \mathbf{V}(t;\boldsymbol{\xi})$, where $\mathbf{V}(t;\boldsymbol{\xi})$ denotes the latent trajectory generated over the ignition sequence for that choice of input parameters. Figure 7 illustrates this structured mapping by coloring each trajectory with an informative input parameter $\xi_2 = \beta = R_1/R_2$ (laser lobe asymmetry, see Table A1, Appendix A), highlighting the correlation between input variability and latent evolution.

The early stages ($t < 200\mu s$) of the laser spark dynamics are only weakly modulated by the presence of the coflowing shear layer and instead are governed by the baroclinic torque, which varies with laser input parameters (Wang et al., 2020). The corresponding representation in latent space should therefore be a smooth response to the input parameter vector $\boldsymbol{\xi}$. Accordingly, Figure 7(b) is annotated at $t = 100\mu s$ to highlight that increasing values of ξ_2 on average correspond to an increase in V_1. Further, a statistically clear mapping from ξ_2 to the latent space of V_3 in Figure 7(d) is also observed. Note, postprocessing or sensitivity analysis could be applied to inspect the degree of structured mapping from input parameter to latent space for every component of $\boldsymbol{\xi}$, and when $\xi_4 = l_{axial}$ (laser axial length) was chosen, clearer and higher degree of correlation was observed in early-time kernel reflecting the importance to the early laser deposition physics. Concerning early-time kernel dynamics, preliminary work on nonreacting sparks (Zahtila et al., 2024) showed that a high amount of variance occurred very early, with thereafter a period of relatively constant values, underscoring the importance of fast kernel dynamics.

Focusing on dynamics after the ignition time scale $t \gtrsim 250\mu$s, consistent trajectories emerge: successful ignitions oscillate and for some latent space components branch off while ignition failures persist with stable values and small variation. Oscillations appear in several components of the latent space $\mathbf{V}$ for igniting cases, but the degree of oscillation in each component varies; for instance, in Figure 7(a), the range of the oscillation values does not depart significantly from the nonignited cases, unlike in Figure 7(h), where positive values of V_7 are observed exclusively during the oscillatory phase. Strong separation in the latent space between successful and failed ignition is most visible in the clearly separated bifurcation branches of component V_5, visualized in Figure 7(f). Given the clear role of component V_5 in separating cases by ignition fate in the latent space, we project the latent trajectories onto two-dimensional component planes to provide interpretable slices of the manifold, in which clustering, variance, and separation are more readily discerned, as shown in Figure 8. Although some projection planes do not reveal strong separation by ignition fate and are less informative, in Figure 8(c), a very clear region of the latent space path is followed in the $V_5 - V_2$ plane for ignition success. Clusters of the terminating ignition trial trajectory points in the $V_5 - V_2$ plane reflect the bistable ignition bifurcation and the evolution path through latent space.

Previous surrogate-modeling efforts combining AEs with latent dynamics, including NODE-based approaches (Lazzara et al., 2023), have typically addressed systems with a simpler spectrum such as the viscous 1D Burgers equation with an expectation that they admit a coherent latent representation. In contrast, the present rocket–combustor system is a multiphysics problem involving turbulence, heat release, and bistable ignition outcomes, so it cannot be assumed a priori that the AE will produce a latent space with smooth and physically meaningful organization. We therefore first examined the latent space structure to determine whether nearby points corresponded to physically similar states and whether variations in the latent trajectories remained systematically related to the governing input parameters $\boldsymbol{\xi}$ and the associated ignition dynamics. This provided evidence that the encoded dynamics evolved on a

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

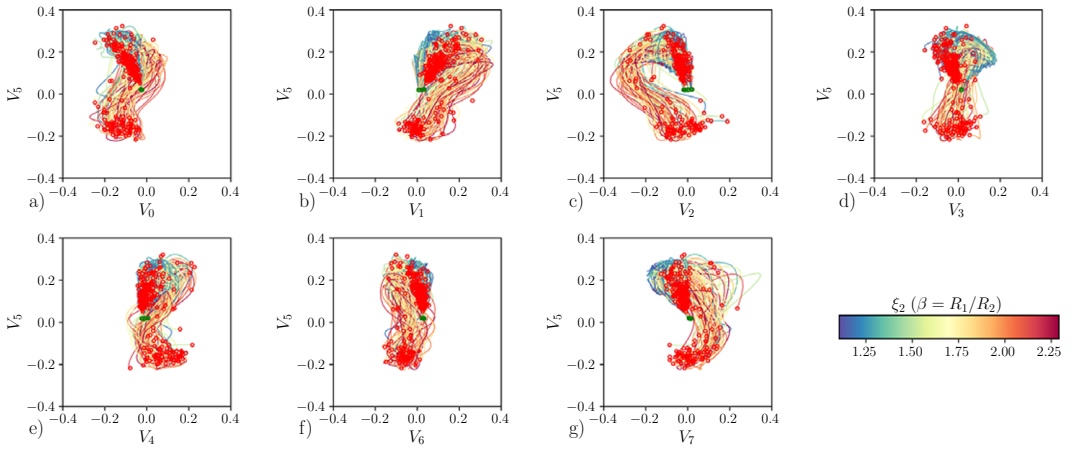


Figure 8. *Latent space projections of V_5 against all other latent variables for the reacting sparks simulation ensemble, colored by the laser lobe asymmetry input parameter $\xi_2 = \beta = R_1/R_2$. Each trajectory begins at a green hollow circle and terminates at a red hollow circle, illustrating consistent trajectories in the low-dimensional manifold and separation of igniting from nonigniting cases.*

coherent low-dimensional manifold whose organization retained dependence on $\boldsymbol{\xi}$, and hence that the latent trajectories were amenable to learning. Overall, Figures 7 and 8 together demonstrate that (i) ignition trials embed as smooth latent manifold trajectories; (ii) variance across trajectories reflects sensitivity to variability of input parameters; and (iii) bifurcations encode ignition fate, particularly through the role of V_5. Taken together, these observations confirm that the learned manifold provides a physically meaningful description of the ignition process, in which early-time clustering, mid-time shear-layer interaction, and ignition-time bifurcations are all preserved. This representation therefore establishes a suitable foundation for NODE-based temporal modeling, as the latent coordinates capture both common trajectories and their bistable outcomes in a consistent reduced-order form.

4.2. Performance of the DnAE

Following the accurate encoding of key physical processes captured by the separation and clustering of the AE latent space, we now investigate two components of the DnAE framework: (i) the predictive modeling capability of the NODE to temporally integrate an ODE within the reduced latent space, and (ii) the fidelity of the resulting spatiotemporal sequence output from the decoder. The only required inputs to yield a prediction from the trained DnAE are specification of the input parameter vector $\boldsymbol{\xi}$ and an initial latent state $\mathbf{V}_0$, which together define a unique ignition trial scenario. In this section, we first evaluate the NODE-integrated latent trajectories $\hat{\mathbf{V}}(t)$ relative to their target trajectories $\mathbf{V}(t)$ in the latent space, and subsequently, focus on the decoder output $\hat{\mathbf{x}}(t) = d\left[\hat{\mathbf{V}}(t)\right]$ to assess the framework's ability to generate physically realistic reconstructions and to predict ignition success across a diverse range of initial conditions.

Figure 9 presents the individual components of the AE latent space with overlaid trajectories predicted from the NODE with representative cases shown corresponding to the 5th, 50th, and 95th percentiles of prediction error, illustrating strong, median, and weak accuracy, respectively. The trajectories are disjointly arranged into training and validation runs. For the training runs, it becomes clear that the curriculum-learning trained neural ODE is able to reconstruct the full diversity of latent space trajectories and trajectory variance introduced by variation in the input parameter vector $\boldsymbol{\xi}$. This represents a considerable extension to the complexity of features learned by neural ODE trajectories when compared with previous work on springs (Garsdal et al., 2022), or exponentially growing, saturating trajectories (Owoyele and Pal, 2022). Based on the training trajectories, the NODE accurately captures the AE latent

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

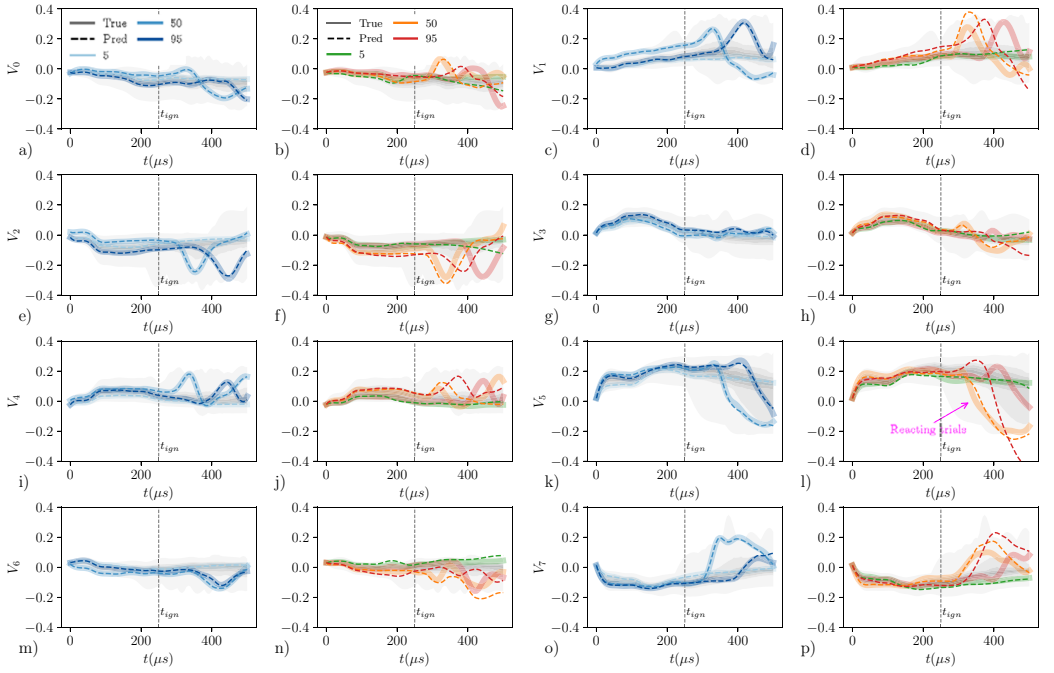


Figure 9. *Training (left columns) and validation (right columns) latent trajectories generated by the neural ODE, compared with the target latent-space trajectories. Representative cases are shown for the 5th, 50th, and 95th percentiles of RMSE, corresponding to strong, median, and weak prediction accuracy, respectively. The shaded region indicates the ensemble distribution of the full set of latent-space trajectories.*

space dynamics during the early stages of kernel evolution and reproduces the oscillatory behavior characteristic of reacting cases.

For the validation cases, the NODE-evolved latent trajectories $\hat{\mathbf{V}}(t)$ broadly followed the true dynamics $\mathbf{V}(t)$ during the early stages of kernel evolution and up until the ignition time scale but began to diverge once the reacting phase commenced. This behavior is consistent with the chaotic nature of the reacting system, where predictive accuracy typically degrades as trajectories evolve. To assess this quantitatively, we divided each trajectory into four equal-duration time intervals, denoted Q1–Q4 from earliest to latest. For each quarter, we computed both the Frechet distance (Sochopoulos et al., 2024) and L_2 norm between the predicted and true latent coordinates, and both metrics showed strong agreement in trends. As shown in Table 1, errors were relatively low across the first two segments

Table 1. *Error metric in NODE-predicted latent space trajectories across four temporal quarters (Q1–Q4) across training and validation data split*

	Train				Validation			
Metric	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Fréchet distance	0.009	0.013	0.018	0.035	**0.075**	0.100	0.190	**0.289**
L2 Norm	0.002	0.003	0.005	0.007	0.021	0.029	0.048	0.080

Note. Results reflect transition from early-time kernel dynamics to late-time divergence in the chaotic reacting phase.

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

(Q1–Q2), but increased substantially in later times (Q3–Q4), reflecting divergence on a chaotic attractor which arises from turbulent combustion. Despite this divergence, the NODE reliably captured key qualitative features—such as the collapse of trajectories for nonigniting cases and persistent oscillations for successful ignition. To clarify, the NODE is expected to track the target latent trajectories $\mathbf{V}(t)$ during the early-time kernel dynamics and to correctly trace the shear-layer interactions that bifurcate toward ignition success or failure. Thereafter, the NODE is only expected to capture the qualitative features of either combustion success or failure rather than one-to-one matching of the turbulent state. The remedy to long-time divergence would require state assimilation methods (e.g., EnKF; Özalp et al., 2025), which offer dramatically improved late-time accuracy; however, they require repeated observations and impose strict system sampling constraints, which on the time scales of rocket ignition are out of the question. The present framework, by contrast, requires no additional system observations beyond the initial condition.

At the conclusion of the neural ODE integration, the resultant latent trajectory is passed through the decoder $d(\cdot)$ to generate the reconstructed spatiotemporal sequence $\hat{\mathbf{x}}(t_i) = d\left[\hat{\mathbf{V}}(t_i)\right]$ at the sampled times which for the available dataset can be directly compared one-to-one with the corresponding IR post-processed CFD sequence $\mathbf{x}^{(j)}(t)$. For automated ignition classification, each reconstructed field $\hat{\mathbf{x}}^{(j)}(t)$ is segmented into iso-temperature contours, and ignition success is determined based on the area enclosed by a T_0-isotherm,

$$A_{T_0}^{(j)}(t) = \int_{\Omega_{T_0}^{(j)}(t)} \mathrm{d}A, \qquad \Omega_{T_0}^{(j)}(t) = \left\{ \boldsymbol{s} : \hat{T}^{(j)}(\boldsymbol{s}, t) \geq T_0 \right\}, \tag{4.1}$$

where T_0 is the user-prescribed threshold temperature, $\boldsymbol{s}$ denotes position in the projected two-dimensional image plane, and $\Omega_{T_0}^{(j)}(t)$ denotes the spatially evolving region bounded by the corresponding temperature isotherm. Ignition success is subsequently recorded when $A_{T_0}^{(j)}(t)$ exhibits sustained growth over the postdeposition sequence, with $T_0 = 1000\,\mathrm{K}$ in the present study. This criterion corresponds to sustained flame expansion and was found to be insensitive to variations in the chosen threshold temperature.

The DnAE performance in ignition success prediction is therefore summarized by the confusion matrices presented in Figure 10, with 10% of the original data held out for validation. The training confusion matrix in Figure 10(a) shows near-perfect classification performance, with a single ignition trial incorrectly labeled. For the ignition (positive) class, the precision and recall were 0.99 and 1.00. In general, classification

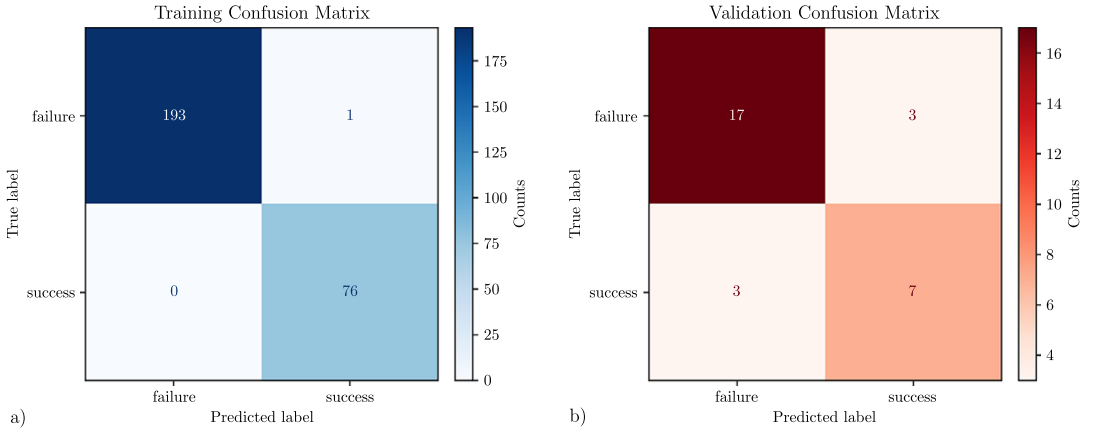


Figure 10. *Confusion matrices for the binary classification of ignition outcomes. (a) Training set results, with near-perfect separation between failure and success cases. (b) Validation set performance.*

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

difficulty is predicated on class imbalance and for the present rocket ignition reliability task, imbalance is quantified through class entropy, $H(p_{\text{ign}}) = -p_{\text{ign}}\log_2 p_{\text{ign}} - (1-p_{\text{ign}})\log_2(1-p_{\text{ign}}) = 0.85\text{bits}$, where for class entropy the maximum uncertainty, i.e., $p = 0.5$ results in 1.0 bits, therefore the present classification challenge is significant. The validation confusion matrix shown in Figure 10(b) indicates reduced but still reasonable predictive performance, with an overall ignition classification accuracy of 80%, with ignition prediction precision and recall both 0.70 indicating balanced false-positive and false-negative rates on this validation split. Rather than stemming from systematic class bias, misclassifications are attributed to the divergence of the predicted latent state from the ground truth, causing the trajectory to cross into a region of the latent space that evolves toward the incorrect ignition fate—a characteristic limitation when modeling chaotic dynamical systems (Solera-Rico et al., 2024). The validation performance on unseen cases indicated by the confusion matrix in Figure 10(b) is satisfactory although misclassifications suggest that broader sampling of the input space would improve generalization. The near-perfect training performance, together with the lower validation accuracy, indicates a gap between in-sample fit and out-of-sample generalization. This suggests that, given the present dataset size, coverage of the parameter space, and problem complexity, the model would benefit from broader sampling, but this is characteristic of the data scenario in machine learning for combustion, where CFD simulations are costly.

Visualizations of the reconstructed DnAE spatiotemporal sequence, $\hat{\mathbf{x}}(t) = d\left[\hat{\mathbf{V}}(t)\right]$, are shown in Figure 11 and compared with the one-to-one counterpart CFD IR imaging. A subset of cases was selected to illustrate ignition success and failure within the training dataset, as shown in Figure 11(a, b). These results show that the latent space of the AE is sufficiently smooth and continuous that time-evolved latent states remain physically accurate. Additionally, representative outcomes for each category in the validation confusion matrix are presented in Figure 11(c–f) to demonstrate both correctly and incorrectly classified ignition events. Starting with the training reconstruction of an igniting sequence shown in Figure 11(a), the DnAE correctly identifies the ignition onset timing relative to the reference CFD IR sequence. The reconstructed flame exhibits a comparable spatial extent to the CFD IR field, though with visibly reduced fine-scale structure. This loss of small-scale content is consistent with prior findings that accurate compression of the full energy spectrum of turbulent flow fields requires a very large latent dimension, typically orders of magnitude larger than present, to preserve small-scale information and avoid the masking of reconstruction errors (Vinograd and Di Leoni, 2025). Figure 11(c,f) illustrate correctly classified spatiotemporal sequences in which both the flame growth and ignition onset are well reproduced. Across the dataset, the DnAE typically achieves ignition onset predictions within an error of $\Delta t \approx 25\mu$s of the reference. Misclassifications, shown in Figure 11(d,e), nonetheless retain physically coherent structures, indicating that the latent representation remains dynamically consistent even when classification errors occur.

The present performance of the DnAE highlights that accurate integration across bistable time horizons remains inherently difficult; however, the model performance is satisfactory given the highdegree of variability in ignition process arising from the aleatoric uncertainty associated with instantaneous turbulence. Using the law of total variance, with ξ_6 (laser lag time) treated as the aleatoric uncertainty accounting for the instantaneous structure of turbulence and the remaining inputs treated as epistemic, the binary ignition fate was decomposed as $\text{Var}(Ign) = \mathbb{E}_{\boldsymbol{\xi}_{\text{epi}}}\left[\text{Var}\left(Ign|\boldsymbol{\xi}_{\text{epi}}\right)\right] + \text{Var}_{\boldsymbol{\xi}_{\text{epi}}}\left(\mathbb{E}\left[Ign|\boldsymbol{\xi}_{\text{epi}}\right]\right)$. This yielded run-to-run ignition fate variance fractions for the aleatoric and epistemic contributions of 0.39 and 0.61, respectively. A preceding physics-based study on the role of the instantaneous turbulent structure in ignition dynamics found that small changes in local flow features can drive bifurcating ignition outcomes (Wang et al., 2024). With the physically consistent reconstructions observed in Figures 11(d,e), we suggest that each trial is a physically realizable ignition trajectory, reflecting the stochasticity encoded within the latent space due to the variance introduced by the structure of instantaneous turbulence. The observed variability largely reflects the underlying randomness and sensitivity of the system rather than systematic model error.

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

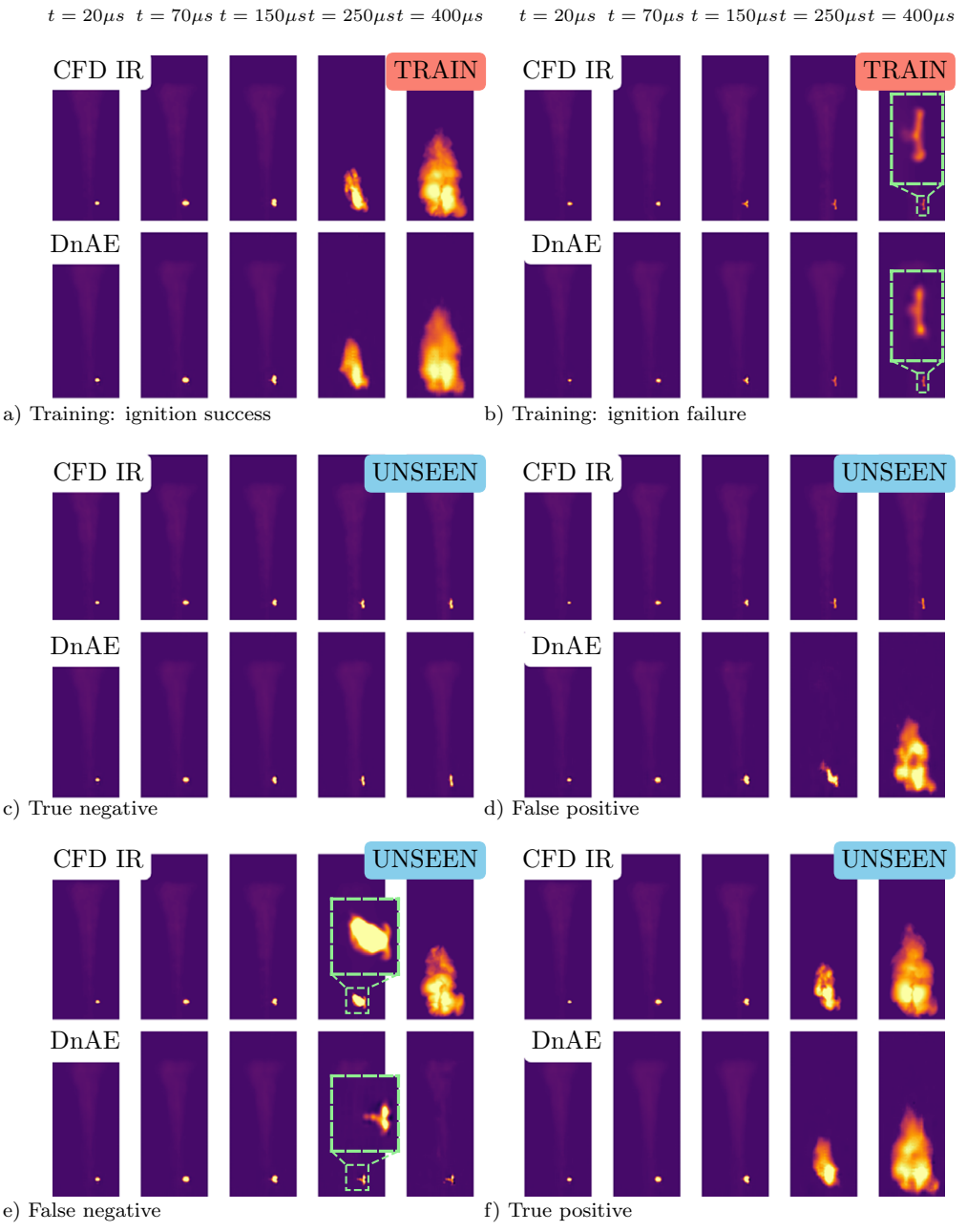


Figure 11. *One-to-one comparison between DnAE-reconstructed IR sequences and the corresponding CFD IR reference sequences for representative ignition outcomes. Panels (a, b) show training cases for ignition success and ignition failure, respectively. Panels (c–f) show validation cases corresponding to the confusion-matrix outcomes in Figure 10: (c) true negative, (d) false positive, (e) false negative, and (f) true positive.*

4.3. Large-scale sampling and generative features

The motivation for employing a generative model in UQ arises from the prohibitive cost of repeated CFD evaluations. Even in low-fidelity settings, a single deployment typically requires wall-clock times on the order of hours (Zahtila et al., 2023; Lee et al., 2025) on modern heterogeneous architectures.

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

Consequently, efficient characterization of system responses across the input space becomes computationally intractable, particularly in the presence of high-dimensional input parameter spaces. The solution of the forward CFD problem with machine learning has received significant attention (Vlachas et al., 2020) and is a current development topic for appropriate generalization and baselines (McGreivy and Hakim, 2024). In this work, we leverage the rapid deployment capabilities of the DnAE framework to propagate input uncertainties and approximate ignition response at a scale of $N_{\mathrm{MC}} = \mathcal{O}(10^6)$ realizations, and we accept as a tradeoff a reduction in physical fidelity that would arise from performing scale-resolving LES computations. Obtaining $N_{\mathrm{MC}} = \mathcal{O}(10^6)$ samples of the input space would be infeasible through classical numerical simulation but using the present framework is deployable on a single workstation. This accessibility does not preclude scaling; the framework could be transitioned to high-performance clusters where significantly higher dimensional latent spaces could be utilized, followed by decoding and analysis of the resulting spatiotemporal quantities.

The additional requirement, compared to the preceding section featuring one-to-one corresponding CFD IR model output, is the specification of an initial condition in the latent space, obtained here via radial basis function interpolation of the sampled initial conditions as a function of the input parameter vector $\boldsymbol{\xi}$, i.e., $\mathbf{V}_0(\boldsymbol{\xi}) = \mathrm{RBF}\left(\boldsymbol{\xi}; \left\{\boldsymbol{\xi}^{(k)}, \mathbf{V}_0^{(k)}\right\}_{k=1}^{N_c}\right)$, where $\boldsymbol{\xi}^{(k)}$ denotes the kth sampled parameter vector and $\mathbf{V}_0^{(k)}$ the corresponding initial condition from the LES ensemble set, an approximation error ϵ is introduced to the initial latent space state but we expect this error to be small because the initial image in the sequence is sampled just prior to laser spark deposition and therefore reflects only the aleatoric uncertainty in the uncontrollable instantaneous jet structure; moreover, there is little variance in the magnitude of the latent space vector at this initial stage. New samples of $\boldsymbol{\xi}$ are therefore drawn in accordance with their probability distributions documented in Appendix A Table A1. To ensure the generated samples remain within the geometric support of the training data, they would ideally reside within the convex hull of the uncertainty space spanned by $\boldsymbol{\xi}$ and would therefore be interpolative. However, because this space is high-dimensional and explicit convex-hull computation scales poorly with dimensionality, a simpler criterion is adopted: generative samples are instead constrained to lie within the component-wise bounds (minimum and maximum) of each uncertainty component ξ_i as determined from the training ensemble.

A restricted sample of the resulting large-scale generated set of trajectories is shown in Figure 12. For all latent space components V_ℓ, the variance during early-time intervals remains consistent with that of the training trajectories, indicating that the generative model follows the latent manifold dynamics. However, in the flame growth regime, oscillatory trajectories extend beyond the variance of the CFD IR set, but this is consistent with the behavior expected from more dense sampling from a chaotic turbulent attractor. Trajectories are classified into igniting and nonigniting on the basis of separation in component V_5, where departure from the steady solution branch into negative V_5 indicates ignition success, as noted in Section 4.1, introducing separation in latent space as a classifier. This is taken as an alternative to the previous contour-area growth criterion from the reconstructed spatiotemporal sequence $\hat{\mathbf{x}}(t) = d\left[\hat{\mathbf{V}}(t)\right]$ as the decoder introduces a significant additional cost required for the segmentation process. Although beyond the scope of the present focus, a high-accuracy spatiotemporal surrogate provides a pathway to examine detailed combustion behaviors, including flame anchoring, directly via the NODE evolution and quantitative comparison with reference physics (Brouzet et al., 2026).

Given the present reliability context of rocket ignition success, a key outcome of this framework is the ability to estimate ignition probability landscapes over the input parameter space at a resolution unattainable with CFD alone. By drawing $N_{\mathrm{MC}} = \mathcal{O}(10^6)$ Monte Carlo samples, we construct joint probability distributions of the statistically important controllable laser operating parameters identified in Section 2.1, where high K–L divergence between their igniting and nonigniting marginal probability distributions reveals how ignition likelihood varies. As the number of samples dramatically increases from that afforded by the CFD IR data stream alone, the estimated probability fields converge to stable contours. Crucially, decision boundaries corresponding to ignition success probabilities of $Pr = 0.5, 0.75$, and 0.9 can be identified directly within these landscapes, as shown in Figure 13. We find in the converged

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

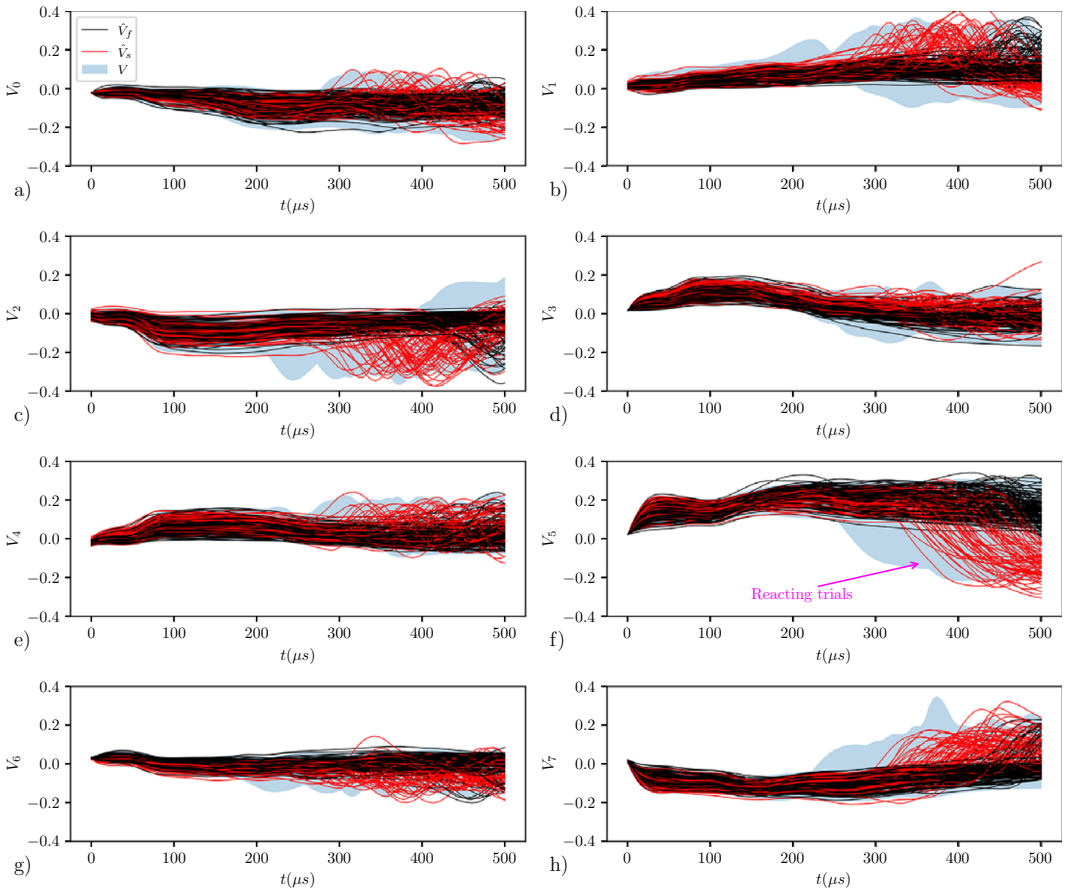


Figure 12. *Generative latent-space trajectory predictions for unseen cases output by the DnAE, classified as either igniting or nonigniting; a reduced subset is shown. Each panel corresponds to one component of the latent vector $\widehat{\mathbf{V}}(t)$. For reference, the range of trajectories on the low-dimensional manifold extracted from compressed CFD IR data is shaded in the background, where $\widehat{\mathbf{V}}_f(t)$ denotes predicted latent trajectories associated with ignition failure and $\widehat{\mathbf{V}}_s(t)$ denotes those associated with ignition success.*

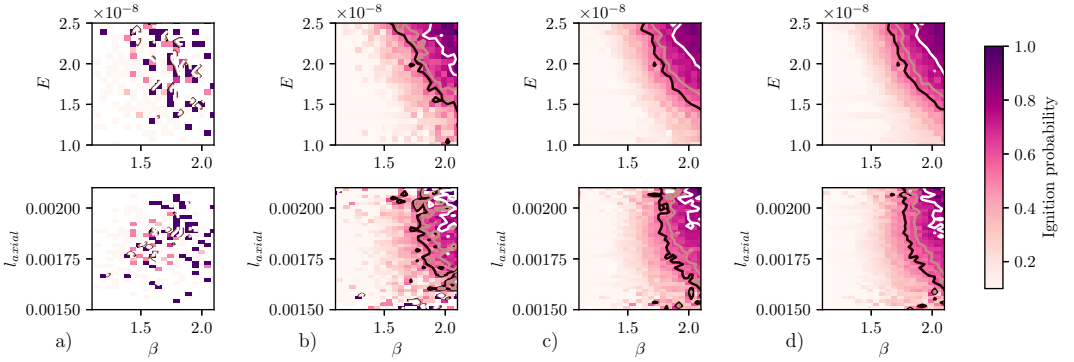


Figure 13. *Joint ignition probability distributions, estimated via Monte Carlo sampling of the NODE latent space trajectories, as a function of the input parameter pairs (β,E) (top row) and (β,l_{axial}) (bottom row), obtained using (a) 300 samples, (b) 200,000 samples, (c) 500,000 samples, and (d) 1,000,000 samples. Additionally marked are decision boundaries for $Pr(ign) = 0.5,0.75,0.9$.*

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

Figure 13(d) that a combination of high laser energy (ξ_5), deposited over an increased axial length (ξ_4), and with deposition lobe asymmetry (ξ_2) leads to high likelihood of ignition success. These provide laser design thresholds and recommended operating conditions. This capability transforms the surrogate model from a mere predictor of individual trajectories into a quantitative tool for uncertainty-informed decision making, where reliability can be assessed in terms of probability contours rather than binary outcomes.

5. Conclusions

In this paper, we present dynamical autoencoders (DnAE) that extend the generative cAE paradigm previously employed for steady-state aerodynamics. This surrogate model first compresses CFD-based IR field representations into a low-dimensional latent manifold $\mathbf{V}$ through an AE, yielding a learnable mapping between the input parameter vector $\boldsymbol{\xi}$ and the latent manifold $\mathbf{V}$. The dynamic extension is achieved through parameterized neural ODEs in the latent space to forecast complex ignition trajectories in the context of laser-ignited rocket combustors. The DnAE approach dramatically reduced computational cost compared to a forward pass of a scale-resolving simulation and remained capable of emulating the spatiotemporal evolution of the path-integrated IR fields that lead to critical ignition events.

A methodology devised for learning the complex trajectories that individual trials follow in the latent manifold was essential for capturing the bistable ignition fate of the system. We demonstrated that a curriculum learning strategy, which progressively extends the neural ODE forecast to longer trajectory horizons, was necessary to stabilize the learning process because standard training methods failed. Inspection of the latent space suggested a clear region of the latent space manifold where igniting cases (specifically in V_5) had been attracted, clearly separating successful ignition and failure branches. The decoder component $d(\cdot)$ further allowed for the reconstruction of physically realistic thermal imaging sequences $\hat{\mathbf{x}}(t)$, enabling one-to-one comparison against integrated quantities derived from reference large eddy simulation data and verifying that the model captures the correct mechanisms of kernel growth.

The utility of the developed surrogate model is a shift from sparse samples in the input parameter space to converged joint probability landscapes of ignition success in response to the operating laser input parameters. The DnAE framework facilitated the generation of $N_{\mathrm{MC}} = 10^6$ ignition trials, giving statistical power for these findings. The outcome is the demarcation of complex decision boundaries within the laser operating parameter space $\boldsymbol{\xi}$. By quantifying the probability of ignition success $Pr(\mathrm{ign})$ with high statistical confidence, this work provides a robust tool for uncertainty-informed design.

Acknowledgments. The authors also extend their gratitude to Dr. Gianluca Geraci for helpful discussions pertaining to this work.

Author contribution. Conceptualization: T.Z., G.I., M.C., and D.R. Methodology: T.Z., E.S., M.C., D.B., and D.R. Software: T.Z., M.C., and D. B. Data curation: T.Z., E.S., M.C., D.B., and D.R. Investigation: T.Z. Validation: T.Z. and E.S. Visualization: T.Z. and D.R. Writing – original draft: T.Z. Writing – review & editing: all authors. Supervision: D.R. and G.I. All authors approved the final submitted draft.

Competing interests. The authors declare none.

Data availability statement. The datasets used to generate the findings of this study are available from Zenodo (Zahtila, 2026) at https://doi.org/10.5281/zenodo.18091817. The code that illustrates training and prediction for latent-space forecasting is available in the L-NeuralODE GitHub repository (Zahtila, 2026) at https://github.com/tzahtila1/L-NeuralODE.

Funding statement. The authors acknowledge financial support from the US Department of Energy's National Nuclear Security Administration via the Stanford PSAAP-III Center for the prediction of laser ignition of a rocket combustor (DE-NA0003968).

Ethical standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

References

Alfonsi G and Primavera L (2007) The structure of turbulent boundary layers in the wall region of plane channel flow. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 463*(2078), 593–612. https://doi.org/10.1098/rspa.2006.1785.

Barwey S, **Prakash S**, **Hassanaly M and Raman V** (2021) *Data-driven classification and modeling of combustion regimes in detonation* Waves. *Flow Turbulence and Combustion 106*(4), 1065–1089. https://doi.org/10.1007/s10494-020-00176-4.

Brouzet DJ, **Zahtila T**, **Rossinelli D**, **Voci A**, **Iaccarino G**, **Passiatore D**, **Strelau R**, **Gejji R and Slabaugh CD** (2025) Large-eddy simulations of a laser-ignited subscale rocket combustor: Modeling strategies and experimental comparison. *AIAA SCITECH 2025 Forum*, 2088. https://doi.org/10.2514/6.2025-2088.

Brouzet D, **Rossinelli D**, **Zahtila T**, **Voci A**, **Strelau R**, **Warner A**, **Gejji RM**, **Slabaugh CD and Iaccarino G** (2026) Thermal expansion-driven laser ignition in a gas subscale rocket combustor. *Combustion and Flame 284*, 114692. https://doi.org/10.1016/j.combustflame.2025.114692.

Brunton SL, **Noack BR and Koumoutsakos P** (2020) Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics 52*(1), 477–508. https://doi.org/10.1146/annurev-fluid-010719-060214.

Bucci MA, **Semeraro O**, **Allauzen A**, **Chibbaro S and Mathelin L** (2023) Curriculum learning for data-driven modeling of dynamical systems. *The European Physical Journal E 46*(3), 12. https://doi.org/10.1140/epje/s10189-023-00269-8.

Bui-Thanh T, **Damodaran M, and Willcox K** (2003) *Proper Orthogonal Decomposition Extensions for Parametric Applications in Compressible Aerodynamics.* 21st AIAA applied aerodynamics conference, p. 4213.

Chakraborty D, **Chung SW**, **Arcomano T and Maulik R** (2024) Divide and conquer: Learning chaotic dynamical systems with multistep penalty neural ordinary differential equations. *Computer Methods in Applied Mechanics and Engineering 432*, 117442. https://doi.org/10.1016/j.cma.2024.117442.

Chen RTQ, **Rubanova Y**, **Bettencourt J and Duvenaud DK** (2018) Neural ordinary differential equations. *Advances in Neural Information Processing Systems 31*, 6572–6583.

Chung WT, **Akoush B**, **Sharma P**, **Tamkin A**, **Jung KS**, **Chen J**, **Guo J**, **Brouzet D**, **Talei M**, **Savard B**, **Poludnenko AY**, **Akoush B**, **Sharma P and Tamkin A** (2023) Turbulence in focus: Benchmarking scaling behavior of 3d volumetric super-resolution with blastnet 2.0 data. *Advances in Neural Information Processing Systems 36*, 77430–77484.

Chung WT, **Laurent C**, **Passiatore D and Ihme M** (2024) Ensemble predictions of laser ignition with a hybrid stochastic physics-embedded deep-learning framework. *Proceedings of the Combustion Institute 40*(1–4), 105304. https://doi.org/10.1016/j.proci.2024.105304.

Cutforth M and Mirjalili S (2025) *Convolutional Autoencoders for the Reconstruction of Three-Dimensional Interfacial Multiphase Flows*, arXiv preprint arXiv:2508.04084

Cutforth M, **Fan T**, **Zahtila T**, **Doostan A and Darve E** (2026) Bi-fidelity interpolative decomposition for multimodal data. *Computer Methods in Applied Mechanics and Engineering 451*, 118706. https://doi.org/10.1016/j.cma.2025.118706.

Di Renzo M, **Fu L and Urzay J** (2020) HTR solver: An open-source exascale-oriented task-based multi-GPU high-order code for hypersonic aerothermodynamics. *Computer Physics Communications 255*, 107262. https://doi.org/10.1016/j.cpc.2020.107262.

Duan Y, **Cooling C**, **Ahn JS**, **Jackson C**, **Flint A**, **Eaton MD and Bluck MJ** (2019) Using a gaussian process regression inspired method to measure agreement between the experiment and CFD simulations. *International Journal of Heat and Fluid Flow 80*, 108497. https://doi.org/10.1016/j.ijheatfluidflow.2019.108497.

Fan T, **Cutforth M**, **D'Elia M**, **Cortiella A**, **Doostan A, and Darve E** (2025) *Physically Interpretable Representation and Controlled Generation for Turbulence Data*, arXiv preprint arXiv:2502.02605.

Ferziger JH and Perić M (2002) *Computational Methods for Fluid Dynamics*, *586*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-56026-2.

Forrester A, **Sobester A and Keane A** (2008) *Engineering Design Via Surrogate Modelling: A Practical Guide*. New York: Wiley. https://doi.org/10.1002/9780470770801.

Garsdal ML, **Søgaard V, and Sørensen SM** (2022) *Generative Time Series Models Using Neural Ode in Variational Auto-encoders.* arXiv preprint arXiv:2201.04630.

He K, **Zhang X**, **Ren S and Sun J** (2016) Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. https://doi.org/10.1109/CVPR.2016.90.

Jaroslawski T, **Patil A and McKeon BJ** (2025) Predicting turbulence structure in street-canyon flows using deep generative modeling. *International Journal of Heat and Fluid Flow 115*, 109849. https://doi.org/10.1016/j.ijheatfluidflow.2025.109849.

Kim Y and Heiland J (2023) *Convolutional Autoencoders, Clustering and Pod for Low-Dimensional Parametrization of Navier-Stokes Equations.* arXiv preprint arXiv:2302.01278.

Krishnapriyan A, **Gholami A**, **Zhe S**, **Kirby R and Mahoney MW** (2021) Characterizing possible failure modes in physics-informed neural networks. *Advances in Neural Information Processing Systems 34*, 26548–26560.

Kullback S and Leibler RA (1951) On information and sufficiency. *The Annals of Mathematical Statistics 22*(1), 79–86. https://doi.org/10.1214/aoms/1177729694.

Lazzara M, **Chevalier M**, **Lapeyre C and Teste O** (2023) *NeuralODE-based latent trajectories into autoencoder architecture for surrogate modelling of parametrized high-dimensional dynamical systems*. In Iliadis L, Papaleonidas A, Angelov P and Jayne C (eds), *Artificial Neural Networks and Machine Learning – ICANN 2023*. Cham: Springer, pp. 497–508.

LeCun YA, **Bottou L**, **Orr GB, and Müller KR** (2012) *Efficient Backprop* Berlin, Heidelberg: Springer, pp. 9–48.

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

Lee K and Parish EJ (2021) Parameterized neural ordinary differential equations: Applications to computational physics problems. *Proceedings of the Royal Society A 477*(2253), 20210162. https://doi.org/10.1098/rspa.2021.0162.

Lee J, **Chan L**, **Zahtila T**, **Lu W**, **Iaccarino G and Ooi A** (2025) Surrogate models for multiregime flow problems. *Physical Review Fluids 10*(2), 024703. https://doi.org/10.1103/PhysRevFluids.10.024703.

Linot AJ and Graham MD (2022) *Data-driven reduced-order modeling of spatiotemporal chaos with neural ordinary differential* equations. *An Interdisciplinary Journal of Nonlinear Science 32*(7). https://doi.org/10.1063/5.0069536.

Lu H and Tartakovsky DM (2020) Prediction accuracy of dynamic mode decomposition. *SIAM Journal on Scientific Computing 42*(3), A1639–A1662. https://doi.org/10.1137/19M1259948.

Lu W, **Chan L and Ooi A** (2025) Spectral analysis of confined cylinder wakes. *Fluids 10*(4), 84. https://doi.org/10.3390/fluids10040084.

Lu W, **Zahtila T**, **Chan L**, **Nguyen QD**, **Lei C**, **Iaccarino G and Ooi A** (2025) Modeling of uncertainties from spanwise asymmetries in upstream conditions and measurement plane location for flow past a circular cylinder confined within a duct. *Physical Review Fluids 10*(6), 064601. https://doi.org/10.1103/PhysRevFluids.10.064601.

Lu W, **Lam WK and Zahtila T** (2026) An autoencoder-based bi-fidelity method utilising frozen latent spaces. Application to flow past a confined cylinder. *International Journal of Heat and Fluid Flow 118*, 110192. https://doi.org/10.1016/j.ijheatfluidflow.2025.110192.

McGreivy N and Hakim A (2024) Weak baselines and reporting biases lead to overoptimism in machine learning for fluid-related partial differential equations. *Nature Machine Intelligence 6*(10), 1256–1269. https://doi.org/10.1038/s42256-024-00897-5.

Milano M and Koumoutsakos P (2002) Neural network modeling for near wall turbulent flow. *Journal of Computational Physics 182*(1), 1–26. https://doi.org/10.1006/jcph.2002.7146.

Mohammadi K, **Immonen J**, **Blackburn LD**, **Tuttle JF**, **Andersson K and Powell KM** (2023) A review on the application of machine learning for combustion in power generation applications. *Reviews in Chemical Engineering 39*(6), 1027–1059. https://doi.org/10.1515/revce-2021-0107.

Molana M, **Darougheh S**, **Biglar A**, **Chamkha AJ and Zoldak P** (2024) Machine learning approaches for predicting ignition delay in combustion processes: A comprehensive review. *Industrial & Engineering Chemistry Research 63*(6), 2509–2518. https://doi.org/10.1021/acs.iecr.3c04097.

Muralidhar SD, **Podvin B**, **Mathelin L and Fraigneau Y** (2019) Spatio-temporal proper orthogonal decomposition of turbulent channel flow. *Journal of Fluid Mechanics 864*, 614–639. https://doi.org/10.1017/jfm.2019.48.

Owoyele O and Pal P (2022) ChemNODE: A neural ordinary differential equations framework for efficient chemical kinetic solvers. *Energy and AI 7*, 100118. https://doi.org/10.1016/j.egyai.2021.100118.

Özalp E, **Nóvoa A, and Magri L** (2025) *Real-Time Forecasting of Chaotic Dynamics from Sparse Data and Autoencoders.* arXiv preprint arXiv:2508.08729.

Passiatore D, **Wang JM**, **Rossinelli D**, **Di Renzo M and Iaccarino G** (2024) Computational study of laser-induced modes of ignition in a coflow combustor. *Flow Turbulence and Combustion 113*(4), 1055–1079. https://doi.org/10.1007/s10494-024-00575-x.

Portwood GD, **Mitra PP**, **Ribeiro MD**, **Nguyen TM**, **Nadiga BT**, **Saenz JA**, **Chertkov M**, **Garg A**, **Anandkumar A**, **Dengel A**, **Baraniuk R and Schmidt DP** (2019) *Turbulence Forecasting via Neural Ode.* arXiv preprint arXiv:1911.05180.

Rojas CJG, **Dengel A, and Ribeiro MD** (2021) *Reduced-Order Model for Fluid Flows via Neural Ordinary Differential Equations.* arXiv preprint arXiv:2102.02248.

Rossinelli D, **Hu J**, **Marshall Jr L**, **Kozak N**, **Zahtila T**, **Brouzet D**, **Cutforth M**, **Mungal MG and Iaccarino G** (2026a) *Digital Enhancement of Experimental Pipelines*, Center for Turbulence Research Annual Research Briefs. https://web.stanford.edu/group/ctr/ResBriefs/2025/10_Rossinelli.pdf (accessed 30 June 2026)

Rossinelli D, **Li G**, **Voci A**, **Marshall Jr L**, **Hu J**, **Brouzet D**, **Fan T**, **Williams C**, **Martin O**, **Khanwale M**, **Vignat G**, **Zahtila T**, **Cutforth M and Iaccarino G** (2026b) *High-Fidelity Modeling of Shadowgraphy and Schlieren Imaging.* Center for Turbulence Research Annual Research Briefs. Available at https://web.stanford.edu/group/ctr/ResBriefs/2025/10_Rossinelli.pdf (accessed 30 June 2026).

Saetta E, **Tognaccini R and Iaccarino G** (2022) Machine learning to predict aerodynamic stall. *International Journal of Computational Fluid Dynamics 36*(7), 641–654. https://doi.org/10.1080/10618562.2023.2171021.

Saetta E, **Tognaccini R, and Iaccarino G** (2024a) *Autoencoders with Sequential Training and Adaptive Resolution.* Available at SSRN 4898782. Available at https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=4898782 (accessed 30 June 2026).

Saetta E, **Tognaccini R and Iaccarino G** (2024b) Uncertainty quantification in autoencoders predictions: Applications in aerodynamics. *Journal of Computational Physics 506*, 112951. https://doi.org/10.1016/j.jcp.2024.112951.

Schmid PJ (2010) Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics 656*, 5–28. https://doi.org/10.1017/S0022112010001217.

Schmid PJ (2022) Dynamic mode decomposition and its variants. *Annual Review of Fluid Mechanics 54*(1), 225–254. https://doi.org/10.1146/annurev-fluid-030121-015835.

Shankar V, **Portwood G**, **Mohan A**, **Mitra P**, **Rackauckas C**, **Wilson L**, **Schmidt D, and Viswanathan V** (2020) *Learning Non-Linear Spatio-Temporal Dynamics with Convolutional Neural Odes*, Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020). Available at https://ml4physicalsciences.github.io/2020/files/NeurIPS_ML4PS_2020_125.pdf (accessed 30 June 2026).

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

Sharma AJ, **Johnson RF**, **Kessler DA, and Moses A**, (2020) *Deep Learning for Scalable Chemical Kinetics.* AIAA scitech 2020 forum, p. 0181.

Sholokhov A, **Liu Y**, **Mansour H and Nabi S** (2023) Physics-informed neural ODE (PINODE): Embedding physics into models using collocation points. *Scientific Reports 13*(1), 10166. https://doi.org/10.1038/s41598-023-36799-6.

Siegel SG, **Seidel J**, **Fagley C**, **Luchtenburg DM**, **Cohen K and McLaughlin T** (2008) Low-dimensional modelling of a transient cylinder wake using double proper orthogonal decomposition. *Journal of Fluid Mechanics 610*, 1–42. https://doi.org/10.1017/S0022112008002115.

Sochopoulos A, **Gienger M, and Vijayakumar S** (2024) *Learning Deep Dynamical Systems Using Stable Neural Odes.* IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2024, pp. 11163–11170.

Solera-Rico A, **Vila CS**, **Gómez-López M**, **Wang Y**, **Almashjary A**, **Dawson STM and Vinuesa R** (2024) β-variational autoencoders and transformers for reduced-order modelling of fluid flows. *Nature Communications 15*(1), 1361. https://doi.org/10.1038/s41467-024-45578-4.

Strelau RM, **Frederick MD**, **Winter TR**, **Senior WCB**, **Gejji RM and Slabaugh CD** (2024) Laser induced spark ignition of a gaseous methane–oxygen model rocket combustor. *Combustion and Flame 265*, 113463. https://doi.org/10.1016/j.combustflame.2024.113463.

Sudret B, **Marelli S, and Wiart J** (2017) *Surrogate Models for Uncertainty Quantification: An Overview.* 11th European conference on antennas and propagation (EUCAP), IEEE, 2017, pp. 793–797.

Turan EM and Jäschke J (2021) Multiple shooting for training neural differential equations on time series. *IEEE Control Systems Letters 6*, 1897–1902.

Vinograd MY and Di Leoni PC (2025) Reduced representations of Rayleigh–Benard flows via autoencoders. *Journal of Fluid Mechanics 1006*, A10. https://doi.org/10.1017/jfm.2025.21.

Vlachas P-R, **Pathak J**, **Hunt BR**, **Sapsis TP**, **Girvan M**, **Ott E and Koumoutsakos P** (2020) Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics. *Neural Networks 126*, 191–217. https://doi.org/10.1016/j.neunet.2020.02.016.

Vlachas PR, **Arampatzis G**, **Uhler C and Koumoutsakos P** (2022) Multiscale simulations of complex systems by learning their effective dynamics. *Nature Machine Intelligence 4*(4), 359–366. https://doi.org/10.1038/s42256-022-00464-w.

Wan K, **Barnaud C**, **Vervisch L and Domingo P** (2021) Machine learning for detailed chemistry reduction in DNS of a syngas turbulent oxy-flame with side-wall effects. *Proceedings of the Combustion Institute 38*(2), 2825–2833. https://doi.org/10.1016/j.proci.2020.06.047.

Wang JM, **Buchta DA and Freund JB** (2020) Hydrodynamic ejection caused by laser-induced optical breakdown. *Journal of Fluid Mechanics 888*, A16. https://doi.org/10.1017/jfm.2019.1066.

Wang JM, **Di Renzo M**, **Iaccarino G**, **Wang H and Urzay J** (2024) Laser-induced indirect ignition of non-premixed turbulent shear layers. *Combustion and Flame 264*, 113426. https://doi.org/10.1016/j.combustflame.2024.113426.

Wilson EB (1927) Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association 22*(158), 209–212. https://doi.org/10.1080/01621459.1927.10502953.

Wu J, **Wang X**, **Zhang G**, **Liu J**, **Li X**, **Zhang Y**, **Zhang H**, **Lyu J**, **Wang B, and Wu Y** (2025) *Physics-Informed Machine Learning for Combustion: A Review.* arXiv preprint arXiv:2509.03347.

Zahtila T (2026) *[dataset] Data for: Generative Prediction of Laser-Induced Rocket Ignition with Dynamic Latent Space Representations.* https://doi.org/10.5281/zenodo.18091817.

Zahtila T (2026) *L-NeuralODE: Latent Space Forecasting with Neural Ordinary Differential Equations.* GitHub repository. Available at https://github.com/tzahtila1/L-NeuralODE (accessed 30 June 2026).

Zahtila T, **Passiatore D, and Iaccarino G** (2023) *Progress on Multifidelity Simulations of a Laser-Ignited Rocket Combustor.* Center for Turbulence Research Annual Research Briefs 57–68. Available at https://web.stanford.edu/group/ctr/ResBriefs/2023/08_Zahtila.pdf (accessed 30 June 2026).

Zahtila T, **Saetta E**, **Brouzet D**, **Cutforth M**, **Rossinelli D and Iaccarino G** (2024) *Neural Ordinary Differential Equation based Representations of Unsteady Laser-Deposited Energy Kernels.* Center for Turbulence Research Annual Research Briefs. Available at https://web.stanford.edu/group/ctr/ResBriefs/2024/11_Zahtila.pdf (accessed 30 June 2026).

Zahtila T, **Cutforth M**, **Brouzet DJ**, **Passiatore D**, **Rossinelli D and Iaccarino G** (2025) Bi-fidelity data ensembles of a rocket ignition system with stochastic interpolative decomposition. *AIAA SCITECH 2025 Forum*, 2300.

Zeng K, **Linot AJ and Graham MD** (2022) Data-driven control of spatiotemporal chaos with reduced-order neural ode-based models and reinforcement learning. *Proceedings of the Royal Society A 478*(2267), 20220297. https://doi.org/10.1098/rspa.2022.0297.

Zhou L, **Song Y**, **Ji W and Wei H** (2022) Machine learning for combustion. *Energy and AI 7*, 100128. https://doi.org/10.1016/j.egyai.2021.100128.

A. Uncertainty parameters characterization

The input parameters for each ignition trial are samples from system variabilities in Table A1.

These correspond to laser operating variabilities ($\xi_0 - \xi_5$), aleatoric uncertainty associated with instantaneous turbulence realization (ξ_6), fuel load (ξ_7), LES model form uncertainties ($\xi_8 - \xi_{11}$), operating conditions of the rocket ($\xi_{12} - \xi_{13}$), and modeled

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press

Table A1. *Summary of the* 15 *uncertain input parameters, ξ_i, used in the ignition-trial uncertainty quantification, together with their physical descriptions, provenance, and assigned probability distributions*

Uncertainty ID	Description	Source	Distribution
$\xi_0 : \Delta x_l$	Streamwise focal imprecision	E	$N \sim (0.29\,\text{mm}, 0.04\,\text{mm}^2)$
$\xi_1 : \Delta x_r$	Radial focal location	E	$N \sim (-0.54\,\text{mm}, 0.20\,\text{mm}^2)$
$\xi_2 : \beta = R1/R2$	Lobe radii ratio	I	$U \sim [1.1, 2.3]$
$\xi_3 : \alpha = l_{axial}/2R1$	Aspect ratio	I	$U \sim [2.0, 2.5]$
$\xi_4 : l_{axial}$	Laser axial length	I	$U \sim [1.44\,\text{mm}, 2.16\,\text{mm}]$
$\xi_5 : E$	Energy deposited	E	$U \sim [20\,\text{mJ}, 54\,\text{mJ}]$
$\xi_6 : \tau_L$	Lag time	I	$U \sim [0\,\mu\text{s}, 284\,\mu\text{s}]$
$\xi_7 : m_{fuel}$	Methane system mass	I	$U \sim [5mg, 7mg]$
$\xi_8 : TF_\beta$	Thickened flame beta	M	$U \sim [0.5, 0.60]$
$\xi_9 : TF_{SL,0}$	Laminar flame speed	M	$U \sim [0.0098, 0.011]$
$\xi_{10} : \dot{\omega}$	Reaction rate	M	$U \sim [3.4 \times 10^9, 3.8 \times 10^9]$
$\xi_{11} : Cs$	Smagorinsky constant	E	$U \sim [0.15, 0.17]$
$\xi_{12} : \dot{m}_{ox}$	Mass flow rate oxygen	M	$U \sim [6.12\,\text{g/s}, 6.83\,\text{g/s}]$
$\xi_{13} : \dot{m}_{fuel}$	Mass flow rate fuel	M	$U \sim [1.97\,\text{g/s}, 2.17\,\text{g/s}]$
$\xi_{14} : S$	Squircularity	M	$U \sim [0.65, 0.97]$

Note. Here, $U \sim [a,b]$ denotes a uniform distribution on $[a,b]$, and $N \sim (\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. Source of uncertainty range: E (experimentally measured), I (inferred from simulation matching), or M (model-form assumption).

geometry (ξ_{14}). The parameters are marked in the source column to indicate whether the statistical distribution was obtained by fit to the experimental data collected in the companion gas-phase model rocket combustor campaign (Strelau et al., 2024), inferred from companion simulations (Brouzet et al., 2025), or model form assumptions.

Cite this article: Zahtila T, Saetta E, Cutforth M, Brouzet D, Rossinelli D and Iaccarino G (2026). Generative prediction of laser-induced rocket ignition with dynamic latent space representations. *Data-Centric Engineering*, 7, e28. doi:10.1017/dce.2026.10063

https://doi.org/10.1017/dce.2026.10063 Published online by Cambridge University Press