scientific reports

OPEN

Adaptive physics-informed neural operator for coarse-grained non-equilibrium flows

Ivan Zanardi, Simone Venturi & Marco Panesi✉

This work proposes a new machine learning (ML)-based paradigm aiming to enhance the computational efficiency of non-equilibrium reacting flow simulations while ensuring compliance with the underlying physics. The framework combines dimensionality reduction and neural operators through a hierarchical and adaptive deep learning strategy to learn the solution of multi-scale coarse-grained governing equations for chemical kinetics. The proposed surrogate's architecture is structured as a tree, with leaf nodes representing separate neural operator blocks where physics is embedded in the form of multiple soft and hard constraints. The hierarchical attribute has two advantages: (i) It allows the simplification of the training phase via transfer learning, starting from the slowest temporal scales; (ii) It accelerates the prediction step by enabling adaptivity as the surrogate's evaluation is limited to the necessary leaf nodes based on the local degree of non-equilibrium of the gas. The model is applied to the study of chemical kinetics relevant for application to hypersonic flight, and it is tested here on pure oxygen gas mixtures. In 0-D scenarios, the proposed ML framework can adaptively predict the dynamics of almost thirty species with a maximum relative error of 4.5% for a wide range of initial conditions. Furthermore, when employed in 1-D shock simulations, the approach shows accuracy ranging from 1% to 4.5% and a speedup of one order of magnitude compared to conventional implicit schemes employed in an operator-splitting integration framework. Given the results presented in the paper, this work lays the foundation for constructing an efficient ML-based surrogate coupled with reactive Navier-Stokes solvers for accurately characterizing non-equilibrium phenomena in multi-dimensional computational fluid dynamics simulations.

Accurate modeling of non-equilibrium reacting flows is critical in many engineering and science disciplines, e.g., designing hypersonic vehicles for space exploration[1,2] or material processing and manufacturing with low-temperature plasmas[3,4]. The need for describing and understanding these flows has led to the development of increasingly large and sophisticated mathematical models[5-8], describing multiple physical phenomena characterized by a broad spectrum of spatio-temporal scales.

The most physically consistent approach to model non-equilibrium flows relies on the direct numerical solution of the master equation[5,6,9-13], whereby all the relevant spatial and temporal scales resulting from chemical and radiative processes are accounted for. Indeed, the availability of quantum state-to-state (StS) chemistry models based on *ab initio* theories[14-18] enables unprecedented levels of physical accuracy[5-8], crucial for modeling flows typified by a significant degree of non-equilibrium. However, the exponentially large number of degrees of freedom (i.e., molecules' and atoms' energy levels) and the numerical restrictions (stiffness) associated with the derived system of equations make these models impracticable in large-scale multi-dimensional computational fluid dynamics (CFD) problems. To overcome these difficulties, crude "engineering" non-equilibrium models[19,20], referred to as multi-temperature (MT) models, have been developed over the years, often assembled without any rigorous derivation from fundamental kinetic equations nor consideration for physical principles and constraints. Given their interpolative nature, these cannot be used to perform predictions outside their development range.

This work targets the numerical challenges in solving such computationally intense systems of equations by surrogating the thermochemical processes characterizing non-equilibrium phenomena that conventional techniques cannot address. Surrogate and reduced-order models[21-26] can be designed and constructed by employing various techniques, such as projection-based methods[27-33], data-fit interpolation and regression[34], and machine learning (ML)-based models[35,36]. A recent application of surrogates for hypersonics has been published by Ozbenli et al.[37], who trained a feed-forward neural network (FNN) to learn a given set of the master equations'

Center for Hypersonics and Entry Systems Studies, Department of Aerospace Engineering, University of Illinois Urbana-Champaign, Urbana 61801, IL, USA. ✉email: mpanesi@illinois.edu

solution functions for a specific non-equilibrium model[38]. Their ML framework showed a great computational speed-up compared to numerical integrators, with generalization performances left unclear. Similarly, Campoli et al.[39] explored different ML algorithms to regress the source terms of the ODEs system modeling the thermochemical relaxation processes. A coupling between a conventional integrator and the ML regressor was attempted, and speed-up performances were analyzed. They also tried to infer the solution of Euler's equations for a single one-dimensional reacting shock flow scenario by leveraging a deep neural network (DNN). Scherding and coworkers[40] developed a lower-dimensional surrogate to compute the thermochemical properties of the gas mixture to be used in place of any high-dimensional look-up non-equilibrium thermodynamic library. However, despite the considerable speed-up performances and encouraging perspective, they did it only for steady-state solutions, targeting specific flow conditions and considering only chemical and not thermal non-equilibrium. The above-mentioned frameworks lack generalization performances and do not impose physical constraints during the surrogate construction, making them less suitable for CFD simulations. Instead, the present study aims to provide a prototyping tool that can replace the master equations with a surrogate that preserves the original's essential properties and physical constraints while being orders of magnitude faster and able to cover an extensive range of physical conditions. The present work augments the framework introduced by Zanardi et al.[41], and it introduces a new machine learning-based method for solving non-equilibrium flows by combining:

i. Coarse-graining, i.e., a reduced order modeling (ROM) technique that extracts meaningful physics from the master equations[10,42–45], in general, by leveraging unsupervised learning adaptation to seek the optimal grouping configuration[46]. The so-derived reduced system of equations models the dynamics of groups of states, addressing the high-dimensionality problem characterizing the StS models.
ii. Neural operators, i.e., a ML-based surrogate that approximates the integral solution operator of a family of partial differential equations (PDEs) to bypass conventional numerical integration[47].

Coarse-graining. Constructing a surrogate for high-fidelity quantum-state-specific chemistry models to describe non-equilibrium phenomena is not a simple task as they rely on the solution of an overwhelmingly large number of differential equations (order of 10^5)[5]. More importantly, the mathematical closure of these equations requires the determination of a sizeable kinetic database that often cannot be computed owing to many processes (order of 10^{16}) to be considered. Therefore, performing first a physics-preserving dimensionality reduction is of paramount importance. To this end, nonlinear manifold learning techniques such as autoencoders[48], diffusion maps[49], or kernel PCA[50] could be employed. Recently, Oommen et al.[51] proposed learning high-dimensional complex dynamics by combining neural operators and autoencoders. Their application first reduced the problem's dimensionality by training a convolutional autoencoder and then learned the low-dimensional dynamics lying in the latent space using a deep neural operator. However, although powerful in applications requiring dimensionality reduction, autoencoders lack physical interpretability and introduce spurious correlations, not necessarily guaranteeing a discrete separation of temporal scales. To overcome these limitations, our approach relies on a class of physics-based reduced-order coarse-grained (CG) models[52–54]. In chemical kinetics, coarse-grained modeling has extensively been used to describe non-equilibrium phenomena of atomic and molecular species[45,46,55–58]. The central idea in the proposed CG model is to combine the solution of the coarse-grained dynamics with the partial equilibration of the underlying microscopic structure. The concept of partial equilibrium suggests applying the *maximum entropy principles* (MEP) to reconstruct the unresolved scales or physics. This choice is of paramount importance, as it ensures the physical consistency of the model by enforcing the *principle of detailed balance* and ensuring the positivity and boundness of the distribution function.

Neural operators. The second basis of the proposed methodology aims to address the stiffness associated with thermochemical processes, characterized by a broad spectrum of temporal scales, ranging from the flow time scales to time scales that are orders of magnitude smaller. This work uses DNNs to infer the generalized solution of the governing equations to bypass the conventional numerical integration. In literature, a series of new ML-based paradigms for speeding up the numerical simulation of partial differential equations[59–65] have been proposed over the past few years. In particular, this work leverages the family of neural operators[47,66–70], DNN-based surrogates designed to learn or discover solution operators defined by the mapping between inputs of a dynamical system, such as initial or boundary conditions (ICs/BCs), and its state. We employ a parametric-based approach to operator learning, introduced first by Chen et al.[71] and then recently extended by Lu et al.[72] In their work, Lu and coworkers introduced DeepONet, a novel network architecture that effectively approximates the solution operator of linear and nonlinear parametric PDEs. DeepONets have found applications in various fields of physics[73,74], including hypersonics with the work of Mao et al.[75], who approximated the fluid flow evolution and concentration profiles downstream of a normal shock with a DeepONet-based surrogate. Although Mao et al.'s work is significant for the scientific community, it relies on a simple physical model that cannot correctly represent the non-equilibrium distribution of internal energy states, which is crucial for the current study. Additionally, their approach lacks physics constraints during the design and training phase of the model, such as physics-informed (PI) machine learning methodologies employed in this work, commonly known as PINNs[76–81]. These techniques impose constraints by penalizing deviations from governing equations, enhancing the model's generalization performance. This new class of machine learning models, called physics-informed deep neural operator (PI-DeepONet)[82–85], which combines physics-informed techniques with the DeepONet architecture, was initially introduced by Wang et al.[82] and successfully applied to construct surrogate solution operators for various partial differential equations (PDEs), demonstrating excellent results.
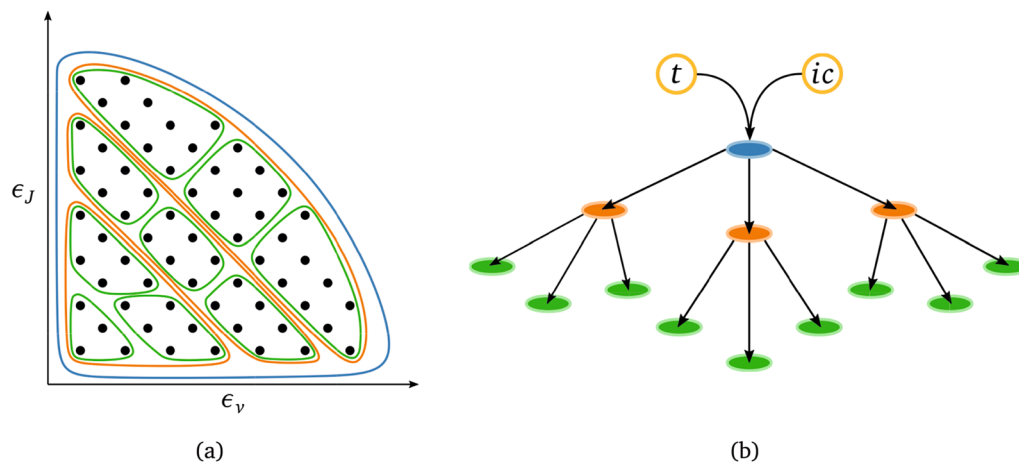
Figure 1. Schematics of the proposed approach. Combining coarse-graining (**a**) and hierarchical DeepONets (**b**). (**a**) Reduced order modeling technique based on clustering the species' quantum energy states (schematized as black dots and as functions of vibrational, ϵ_v, and rotational energy, ϵ_J) into macroscopic bins. In the figure, three different levels of hierarchical clustering are shown. (**b**) Tree visualization of the hierarchical deep learning framework, where the leaf nodes correspond to separate DeepONets (one for each macroscopic bin), which take as inputs the initial conditions, *ic*, and time, *t*.

Proposed approach. The combined use of coarse-graining and neural operators is of primary importance. On the one hand, the mere application of neural operators does not resolve the high-dimensionality problem, as it is not straightforward to design and train an efficient surrogate for thousands of coupled differential equations. On the other hand, dimensionality reduction does not solve the issues with integration, as small steps are still needed to stably integrate the reduced system of equations. For these reasons, the proposed framework (Fig. 1) is characterized by a novel physics-inspired architecture based on a hierarchy of DeepONets used to learn the solution operator for multiple coarse-grained configurations to resolve different scales of the phenomena considered. The CG surrogate herein proposed, referred to as CG-DeepONet throughout the rest of the paper, is constructed by training each scale sequentially and employing transfer learning between them. In this sense, our framework is in line with recent operator learning techniques for multi-scale systems[86–91]. Among the latest ones, Liu et al.[86] proposed a promising hierarchical time-stepper approach for solving the system dynamics. In their approach, they trained multiple neural networks to capture different timescales of the physical phenomenon by varying the integration step. We also recall the work of Migus et al.[87], who designed a multi-scale architecture based on multi-pole graph neural operators (MGNO) by embedding multi-resolution iterative methods[92]. Liu and coworkers[88] drew inspiration from hierarchical matrix methods to develop their multi-scale hierarchical transformer. Furthermore, Liu and Cai[89] integrated multi-scale deep neural networks (MscaleDNNs)[93] within the DeepONet architecture. These innovative approaches open up new possibilities for more accurate and efficient modeling of multi-scale complex systems, and the paradigm proposed in this work builds upon these advancements. Indeed, our framework allows the development of a parsimonious and autonomous tool that can quickly deliver the optimal thermochemical representation of the gas given initial conditions and time instant by adaptively choosing the most efficient and physically accurate grouping resolution. The need for adaptation is a direct consequence of different physical scenarios arising in multidimensional numerical simulations, ranging from equilibrium or near-equilibrium to strong non-equilibrium conditions. A controller-acting surrogate, identified as Neq-DeepONet in the remainder of this paper, is responsible for the model adaption to the local flow conditions. In this sense, our framework can be viewed as a multi-fidelity composition of DeepONets and shares analogies with some recent works on the topic[94–96]. However, the novelty of our approach stems from the definition of such a composition based on the maximum-entropy coarse-grained modeling, which is consistent with the underlying physics.

Physics-informed attributes of the surrogate. In this paragraphs, we highlight the physics-informed features of the proposed approach, which take the form of either soft or hard constraints imposed on the surrogate:

  i. *Dimensionality reduction in the state space.* In addition to the dimensionality reduction in the space of the initial conditions automatically carried out by the DeepONet based on the scenarios provided during training[97], a physics-based reduction is performed in the state space (i.e., in the space of the discrete energy states) by grouping states that are likely to be found in local equilibrium[46,57]. Only briefly introduced above, such a coarse-graining approach will be detailed in Section "Physical modeling".
  ii. *Physics-consistent architecture components.* A Boltzmann transformation layer is built into the surrogate to enforce the equilibrium distributions between states in the same group, as explained in Section "Neural operators".

3

iii. *Interpretable prior distributions for the network parameters.* As discussed in Section "Neural operators", the addition of Boltzmann layers allows the imposition of prior distributions for the network parameters that, when propagated to the state populations (e.g., mass fractions), produce equilibrium distributions between distinct groups of states. Therefore, such priors can provide physically consistent solutions even for un-trained surrogates.

iv. *Physics-informed loss function.* The framework employs a physics-informed loss as a soft constraint, which biases the surrogate predictions towards physically consistent solutions. In particular, the employed hybrid strategy, described in Section "Neural operators", combines data from high-fidelity simulations (or experiments) to anchor the solution to frequent or reproducible real-world scenarios and the residual of the governing laws to ensure generalizability to different unseen physical conditions.

v. *Hierarchical architecture and transfer learning.* The training strategy involves sequential fine-tuning transfer learning between different temporal scales, explained in Section "Neural operators". On the one hand, this approach allows for partially preserving the learned physics. On the other hand, it enables surrogate adaptation and knowledge transfer from one temporal scale to another, speeding up the training process of the entire network.

vi. *Physics-driven online pruning at the prediction phase.* As detailed in Section "Training strategy", an additional (controller-acting) surrogate learns the dynamics of a physically-relevant non-equilibrium control variable, determining the minimum resolution level required to accurately describe the system dynamics while avoiding explicitly computing unnecessary fine scales. During the prediction phase, this additional surrogate is responsible for selecting which component of the overall architecture needs to be queried.

The paper is structured as follows. First, in Section "Physical modeling", the basic framework and derivation of the thermochemical non-equilibrium model are provided, along with the details of the one-dimensional numerical experiment conducted in this work. Next, in Sections "Neural operators" and "Training strategy", the proposed ML framework and the developed adaptive technique are described, respectively. In the "Results" section, the accuracy and performance of the surrogate with and without adaptive inference are illustrated and discussed in detail for both 0-D and 1-D test case scenarios. Finally, in the "Conclusions" section, final remarks are presented along with possibilities for future work. Additional information can be found in the Supplementary Information for interested readers.

Methods

Physical modeling.
Modeling of chemically reacting flows relies on the solution of Navier-Stokes equations complemented by additional conservation equations accounting for changes in the chemical composition and non-equilibrium relaxation of the energy modes. This extra set of equations often represents a computational burden that makes reacting non-equilibrium flows hard to solve. An extensive discussion on non-equilibrium modeling can be found in reference[45].

The most general way to express the extra set of governing equations is

$$\frac{D}{Dt}\left(\rho_i e_i^m\right) + \nabla \cdot \mathscr{I}_i^{\ m} = \Omega_i^m \ , \tag{1}$$

where ρ_i and e_i indicate the mass density and the internal energy of the i-th pseudo-species (i.e., a particular species' internal degree of freedom treated as a state variable). Additionally, m denotes the moment order (0, 1, 2, *etc.*), Ω_i^m the reactive source terms, D/Dt the Lagrangian derivative, and $\mathscr{I}_i^{\ m}$ the dissipative/diffusion terms. Depending on the assumptions made in the definition of the chemical species indicated by i, three different models can be identified:

i. If i refers to a particular energy state, ϵ_i (i.e., rovibronic $i = (el, v, J)$), the approach is called state-to-state (StS) master equations[5,6]. In this case, m is set to 0.

ii. If ρ_i indicates the density of a group of states, the approach is named coarse-grained (CG) modeling or coarse-grained master equations (CGME)[42,45,46,57,58,98–100]. In this case, the conservation equations for mass, momentum, and energy are complemented by additional equations (i.e., $m = 0$ and/or $m = 1$) to model chemical composition and internal energy modes.

iii. In the case of binning one group per internal energy mode, which is a particular case of (ii), we have the multi-temperature (MT) models[101].

Figure 2 compares the levels of physical accuracy and resolution among the three models mentioned above for O_2+O kinetics, the only system considered in this work. A substantial loss of physical information can be noticed moving from the internal energy states distribution obtained with the StS model to the one defined by Park's two-temperature model[101], which is a particular case of the MT models, where all the states are collapsed along a straight line. Differently, the CGME approach better captures the StS distribution by modeling the dynamics of multiple clusters of states (27 in Fig. 2, namely the CGME27 model). In this work, only the coarse-grained master equations approach will be employed to construct our surrogate, which is tested in both 0-D and 1-D scenarios.

Coarse-grained modeling. The numerical solution of the master equations, whereby the dynamics of each internal energy state is captured via the direct solution of the corresponding mass conservation equations, is often impractically expensive. Moreover, it is usually not required since the internal energy distribution is generally a composition of partial equilibria rather than a complete non-equilibrium state[46]. The concept of local or partial
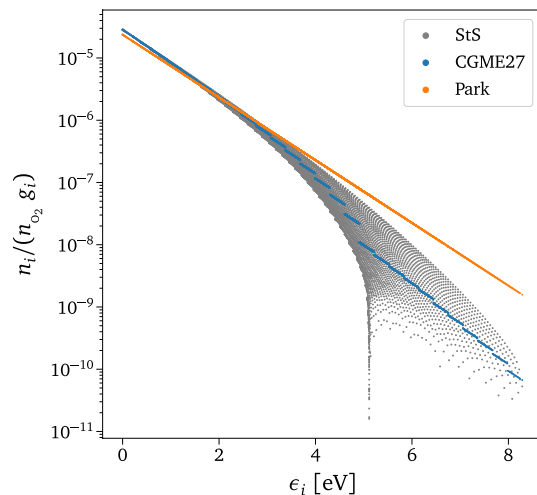
Figure 2. Normalized quasi-steady state (QSS) rovibrational states distribution for different models. The level of physical accuracy can vary significantly depending on the choice of the thermochemical model. The orange dots are determined by low-fidelity Park's two-temperature model[102], e.g., a particular MT model, the blue ones by the 27-groups CG grouping strategy, and the grey ones by the high-fidelity StS modeling. Initial conditions used for the 0-D simulation: $P_0 = 3\,000$ Pa, $X_{O_0} = 0.2$, $T_{int_0} = 1000$ K, $T = 10\,000$ K.

equilibrium suggests the application of the *principle of maximum entropy* to reconstruct the unresolved scales of physics[10,44,45]. The construction of a coarse-grained model is accomplished by adopting a two-step procedure which goes as follows[103]:

i. Group energy states into $\mathcal{N}_{\mathcal{G}}$ macroscopic bins according to a specific strategy;
ii. Prescribe a bin-wise distribution function to represent the population within each group together with a series of moment constraints.

This work employs a log-linear form for the bin-wise distribution function, which results in a thermalized local Boltzmann distribution within individual bins, defined as follows

$$\mathscr{F}_P^i(\epsilon_i): \quad \log\left(\frac{g_i}{n_i}\right) = \alpha_P + \beta_P \epsilon_i \,, \tag{2}$$

where the bin-specific coefficients α_P and β_P are expressed in function of the macroscopic group constraints (i.e., number density, energy, etc.). The total population and energies of the different bins are the set of unknowns for the reduced-order system. The governing equations for these macroscopic constraints can be derived by taking successive moments of the StS master equations, using $(\epsilon_i)^m$ for $m = 0, 1, \ldots$ as weights (see Supplementary Sect. S.1.2 for more details).

While more accurate strategies have been developed during the past few years[46,57], the model-reduction approach employed in this work is the rovibrational energy-based grouping technique[99,100], which lumps together energy states with similar internal energy regardless of their rotational and vibrational quantum numbers.

Zero-dimensional chemical reactor. We wish to investigate the behavior of oxygen molecules in their electronic ground state undergoing dissociation when subjected to sudden heating in an ideal chemical reactor. We make the following assumptions:

i. The 0-D reactor is plunged into a thermal bath maintained at constant temperature T.
ii. The translational energy mode of the atoms and molecules is assumed to follow a Maxwell-Boltzmann distribution at the temperature T of the thermal bath.
iii. At the beginning of the numerical experiment, the population of the rovibrational energy levels is assumed to follow a Boltzmann distribution at the internal temperature T_{int_0}.
iv. The volume of the chemical reactor is kept constant during the experiment, and the thermodynamic system is closed (no mass exchange with the surrounding environment).
v. Only α_P in Eq. (2) is modeled for each bin P, while $\beta_P = 1/(k_B T_P)$ is kept constant during the 0-D simulation, with k_B being the Boltzmann's constant and $T_P = T$.

Therefore, Eq. (1) reduces to

$$\frac{d\rho_i}{dt} = \Omega_i^0(\rho_i, T) \; ,$$
$$\rho_i(t = 0) = f_i \tag{3}$$

where f_i refers to the corresponding Maxwell-Boltzmann equilibrium value of specie i at temperature T_{int_0}.

Since the goal is to learn the integral solution operator of the rovibrational CG master equations to be able to deliver accurate predictions in multidimensional CFD simulations characterized by a wide range of physical scenarios, we aim to generalize over the space of initial conditions (ICs) and time domain. The ICs are generated by defining the initial pressure P_0, the initial molar fraction of atomic oxygen X_{O_0}, the translational temperature T, and the initial internal temperature T_{int_0} for which a Boltzmann distribution is prescribed for the O_2 bins. In this work, the domain in which the initial conditions have been sampled is defined in table 1 as minimum-maximum pair values. For all the possible sampling scenarios, T is greater than T_{int_0}, which implies that thermal excitation and dissociation processes are the dominant phenomena occurring in the reactor.

Regarding the time domain, we train the model over an interval of $[0, 10^{-2}]$ s, covering most excitation and dissociation processes for the non-equilibrium problem under investigation.

One-dimensional numerical experiment. Following the approach used by Zanardi et al.[104], a one-dimensional shock case scenario is employed to test the ML-based framework proposed in this work. The governing equations for the dynamics of inviscid, one-dimensional gas flows are given by the Euler equations:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = \mathbf{S} \; , \tag{4}$$

where t represents time and x represents space. It is worth noting that Eq. (1) is the Lagrangian version of Eq. (4), including an additional diffusive term. The vectors $\mathbf{U}$, $\mathbf{F}$, and $\mathbf{S}$ represent the conservative variables, inviscid fluxes, and source terms, respectively. They are defined as follows:

$$\mathbf{U} = [\rho_i \quad \rho u \quad \rho E]^{\mathrm{T}} \; , \tag{5}$$

$$\mathbf{F} = [\rho_i u \quad p + \rho u^2 \quad \rho u H]^{\mathrm{T}} \; , \tag{6}$$

$$\mathbf{S} = [\Omega_i^0 \quad 0 \quad 0]^{\mathrm{T}} \; , \tag{7}$$

where the total energy and enthalpy per unit-mass are $E = e + u^2/2$ and $H = E + p/\rho$, respectively. The thermodynamics of the system is explained in detail in the Supplementary Sect. S.1.1, and the variables e, u, p, and ρ have their usual meanings in the context of gas dynamics. The source term Ω_i^0 represents the mass production term, which is the same one as defined in Eq. (3) and described in detail in the Supplementary Sect. S.1.2.

The flow governing equations (4) are discretized in space using the finite volume method, with inviscid fluxes evaluated using van Leer's flux vector splitting in conjunction with the second-order upwind-biased MUSCL reconstruction procedure[105,106]. The time integration method is based on the operator-splitting technique proposed by Strang[107]. This method integrates the transport operator, $\mathcal{T}(\mathbf{U}) = \partial \mathbf{F}/\partial x$, and the reaction operator, $\mathcal{R}(\mathbf{U}) = \mathbf{S}$, sequentially in a symmetric fashion:

$$\partial_t \mathbf{U}^{(1)} = \mathcal{T}\left(\mathbf{U}^{(1)}\right) \; , \qquad \mathbf{U}^{(1)}(t_n) = \mathbf{U}_n \tag{8}$$

$$\partial_t \mathbf{U}^{(2)} = \mathcal{R}\left(\mathbf{U}^{(2)}\right) \; , \qquad \mathbf{U}^{(2)}(t_n) = \mathbf{U}^{(1)}(t_n + \Delta t/2) \tag{9}$$

$$\partial_t \mathbf{U}^{(3)} = \mathcal{T}\left(\mathbf{U}^{(3)}\right) \; , \qquad \mathbf{U}^{(3)}(t_n + \Delta t/2) = \mathbf{U}^{(2)}(t_n + \Delta t) \tag{10}$$

$$\mathbf{U}_{n+1} = \mathbf{U}^{(3)}(t_n + \Delta t) \; , \tag{11}$$

where Δt is the time step. The splitting formulation is second-order accurate, strongly stable, and symplectic for non-linear equations. Its convergence and stability properties have been extensively studied for reacting flow simulations[108–111]. The use of an operator-splitting approach facilitates the straightforward insertion of the constructed neural operator into the framework described by Eqs. (8) to (11). Instead of using an implicit scheme to integrate the stiff reaction step described by Eq. (9), a simple evaluation of the trained surrogate is performed

	P_0 [Pa]	X_{O_0}	T_{int_0} [K]	T [K]
Min	1 000	0	1 000	8 000
Max	10 000	0.95	8 000	15 000

Table 1. Space of initial conditions. Minimum-maximum pair values for each IC variable.

to evolve the solution in time. The surrogate takes the solution from the first flux integration step as input and provides the evolved gas state resulting from the reaction operator to the last step of the splitting scheme.

The main configuration details of the one-dimensional shock case scenario used herein are given below.

– *Initial and boundary conditions.* Table 2 presents the piece-wise initial conditions. On the left side, freestream conditions corresponding to a hot gas at $T = 3000$ K and $u = 3000$ m/s are imposed. This choice is made because, at this temperature, the equilibrium state of the gas results in a reasonable amount of dissociated oxygen. It is important to note that this condition is not a requirement of the method itself but rather a consequence of only modeling the O_2+O kinetics without considering the O_2+O_2 system, where molecular oxygen alone is sufficient to activate the thermochemical processes. On the right side, the initial solution is set equal to the post-shock equilibrium state. A supersonic inflow boundary condition (BC) is imposed on the left side, where all characteristics are incoming, by prescribing all flow variables. A subsonic outflow BC is imposed on the right side with a specified pressure value.
– *Time and space grid.* The one-dimensional domain length is set to $L = 0.1$ m, and the spatial discretization uses a space step of $\Delta x = 4 \times 10^{-4}$ m, resulting in a total of 250 cells. The integration is performed until the shock profile is fully developed, using a total of 500 iterations with a constant time step of $\Delta t = 1.33 \times 10^{-7}$ s determined by the freestream velocity while maintaining a maximum CFL number of 1 to ensure numerical stability.

To ensure that the left and right equilibrium conditions are fully guaranteed and avoid any error accumulation due to even minor discrepancies in the surrogate's predictions, the inference is performed only for those cells experiencing non-local thermodynamic equilibrium (NLTE) effects, meaning for gas thermochemical states different from the ones shown in table 2. However, to fairly compare the numerical integrator's and the surrogate's performance, the inference is performed for the whole 1-D domain, and the predictions for those cells in the same conditions as in table 2 are simply disregarded.

To ensure physical consistency, the surrogate must learn the integral solution of the zero-dimensional formulation of Eq. (4), specifically Eq. (9), which describes an adiabatic thermodynamic system without energy or mass exchange. Consequently, the isothermal assumption made in the 0-D analysis does not apply to this particular test. To accurately represent the adiabatic case, an additional DeepONet is required on top of the surrogate described in the next section. This additional DeepONet is employed to model the translational temperature T, enabling a more comprehensive and accurate representation of the complex thermochemical dynamics in the 1-D domain. Therefore, a distinct surrogate is constructed specifically for this simulation, with detailed information on data generation and network construction provided in the Supplementary Sect. S.3.1.

Neural operators. *DeepONet.* Building upon the original formulation of the DeepONet by Lu et al.[72], whereby the solution operator G maps an input function $\boldsymbol{u}$ and the continuous coordinates $\boldsymbol{y}$ of $G(\boldsymbol{u})$ to a real scalar value, this work extends the DeepONet framework to accommodate the high-dimensional nature of the master equations, thus obtaining an output vector $G(\boldsymbol{u})(\boldsymbol{y}) \in \mathbb{R}^D$, where D is the number of the output variables[41,67]. As illustrated in Fig. S1 in the Supplementary Information, the DeepONet architecture is characterized by two different deep neural networks: the "branch net" and the "trunk net". The modified version is characterized by multiple branches, one for each output variable, which takes $\boldsymbol{u}$ as input and returns a feature embedding $\boldsymbol{\alpha} \in \mathbb{R}^p$ as output. Instead, the trunk net takes the continuous coordinates $\boldsymbol{y}$ as inputs and outputs another feature embedding $\boldsymbol{\phi} \in \mathbb{R}^p$. This block is shared between different branches[67,97], gaining computational efficiency. In the framework of operator learning for ODEs, $\boldsymbol{u}$ represents the space of initial conditions, whereas $\boldsymbol{y}$ is the time variable. To obtain a continuous and differentiable representation of the output functions of the DeepONet, the outputs of each branch and the trunk networks are merged via dot product as follows:

$$\widehat{G}^{(i)}(\boldsymbol{u})(\boldsymbol{y}) = \sum_{k=1}^{p} \alpha_k^{(i)}(\boldsymbol{u})\phi_k(\boldsymbol{y}) \qquad \text{for } i = 1, \ldots, D . \tag{12}$$

One can notice that Eq. (12) reminds the proper orthogonal decomposition (POD) formulation[112], as highlighted by Lu et al.[67], and more generally Eq. (12) can be related to the singular value decomposition (SVD) factorization, as explained by Venturi and Casey[97]. From this perspective, the trunk net learns the p most important modes of the dynamical system, $\boldsymbol{\phi}$, while the branch net learns the coefficients $\boldsymbol{\alpha}$ of the expansion. Under this perspective, the shared-trunk version of the DeepONet works reasonably well only when the dynamics of the modeled variables are similar to each other such that they can share the same basis $\boldsymbol{\phi}$[97].

|	x [m]	u [m/s]	p [Pa]	X_O	T [K]
Freestream	≤ 0.04	3 000	2 940	0.12	3 000
Post-shock (LTE)	> 0.04	492	29 598	0.39	4 245

Table 2. Initial conditions for one-dimensional shock case scenario.

Multi-scale hierarchical coarse-grained model. Similar to what is done in adaptive mesh refinement (AMR) techniques used in CFD, the accuracy of the CG model can be improved by increasing the number of groups but at a higher computational cost. The improvement in accuracy is explained by the larger range of scales (or kinetic processes) that can be resolved. Indeed, taking as an example the rovibrational energy-based grouping strategy employed in this work, if we recursively split the energy space of the internal states by following a cascade in the groups, all the micro-groups inside the corresponding macro-group quickly reach the same equilibrium value, showing a fast dynamical behavior. Consistently, we leveraged the multi-scale nature of the physical problem to construct a physics-inspired ML-based surrogate (see Supplementary Sect. S.2.2 for all the details) by sequentially learning the different timescales of the thermochemical phenomena occurring inside a 0-D reactor.

– *Timescale 1.* Chemical dissociation of O_2 molecules (irrespective of their internal excitation) and creation of O atoms are the slowest processes that can be learned. As shown in Fig. 3a, the outputs of the DeepONet employed for this first timescale, denominated as CG-DeepONet$^{(1,1)}$ (i.e., the surrogate's component in charge of predicting the group number one in the scale number one), are simply the mass fractions of O and O_2. So, we are assuming that all the internal states can be clustered in one unique group, but we do not solve for the rovibrational-translation energy transfer phenomena. As concerns the physical input of the model, $\boldsymbol{u}$ represents the initial conditions of the reactor, which is characterized by translational temperature, T, reactor density, ρ, and initial mass fraction of O_2, while the independent variable, $\boldsymbol{y}$, of the operator $G(\boldsymbol{u})$ is the time, t:

$$
\begin{aligned}
\boldsymbol{u} &= \begin{bmatrix} T, \ \rho, \ Y_{O_{20}} \end{bmatrix} & \in \mathbb{R}^3 \\
\boldsymbol{y} &= t & \in \mathbb{R}^1 \\
\widehat{G}(\boldsymbol{u})(\boldsymbol{y}) &= \begin{bmatrix} \widehat{Y}_O|_{\boldsymbol{u}}(t), \ \widehat{Y}_{O_2}^{(1,1,1)}|_{\boldsymbol{u}}(t) \end{bmatrix} & \in \mathbb{R}^2
\end{aligned}
\tag{13}
$$

In (13) and Fig. 3a, a series of two or three superscripts have been used, where the first one corresponds to the *timescale* investigated, the second the DeepONet *index*, and the last one the O_2 *group*. They will help to
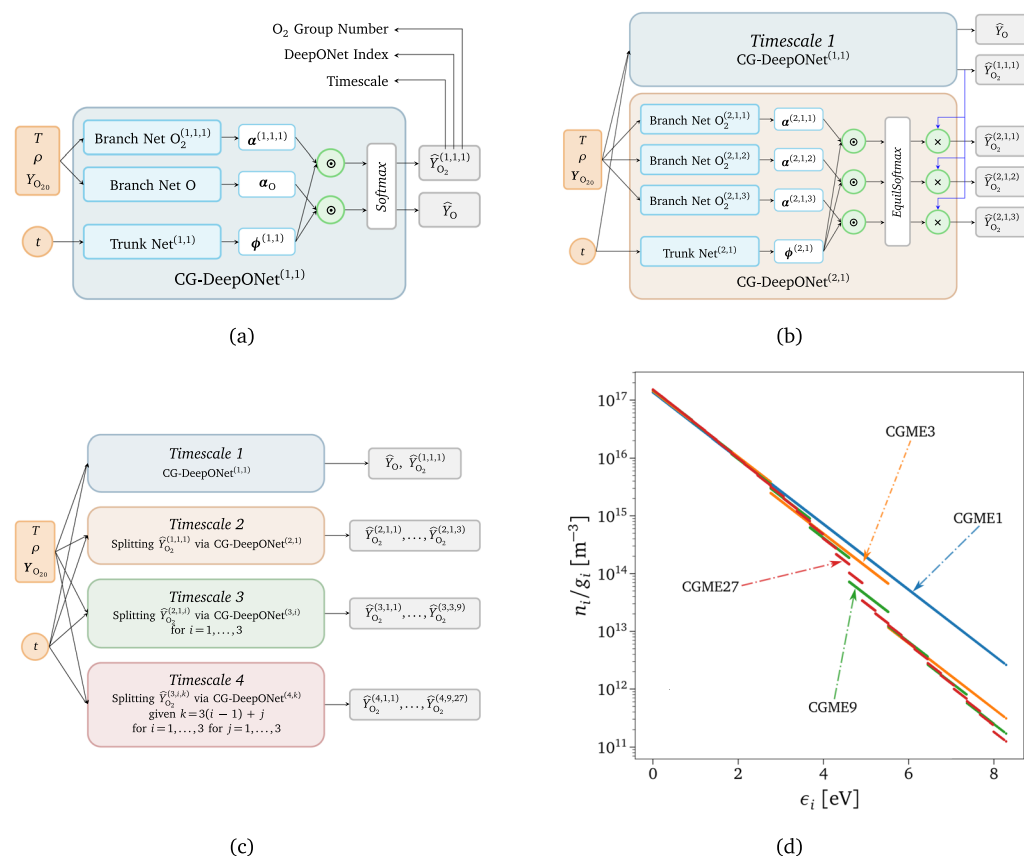
Figure 3. Hierarchical surrogate for multi-scale coarse-grained dynamics. (**a**) Schematics of the network architecture for 1-group CG master equations (CGME1). (**b**) Schematics of the network architecture for 3-groups CG master equations with the additional *EquilSoftmax* layer. (**c**) Schematics of the complete hierarchical network architecture. (**d**) O_2 rovibrational distribution for 1-group (CGME1), 3-groups (CGME3), 9-groups (CGME9), and 27-groups (CGME27) coarse-grained grouping.

identify the different variables and DeepONets used for each timescale. The *Softmax* function in Fig. 3a is applied to the dot product outputs after these being linearly transformed. It guarantees the mass fractions to be positive values and the mass to be conserved.

– *Timescale 1-2*. In the following timescale, we start modeling the energy exchange processes for O_2. To do so, the internal states are clustered into three groups (CGME3) which is equivalent to uniformly splitting the energy space covered by the unique group from the previous timescale (CGME1) into three parts, as shown in Fig. 3d. To learn the dynamics of this new system, the information learned from the previous timescale is leveraged by adopting transfer learning for the calibrated weights of CG-DeepONet$^{(1,1)}$. The new DeepONet is designed to learn the 3-group normalized distribution. The mass fractions of the three bins are then obtained by multiplying the modeled distribution by the total mass fractions of O_2 predicted by CG-DeepONet$^{(1,1)}$, as shown in Fig. 3b, ensuring the conservation of mass across the two scales. In terms of architecture, two are the difference between *Timescale 1* and *Timescale 2*. The first is related to the inputs, $\boldsymbol{u}$, of the branch net, which considers the initial mass fractions of all the three groups, $\boldsymbol{Y}_{O_{2_0}}$. Since *Timescale 1* takes as an input the total mass fraction of O_2 as described in (13), the three values are summed to get the correct input for CG-DeepONet$^{(1,1)}$. The second aspect concerns the replacement of the *Softmax* layer with the *EquilSoftmax* one. The latter can be considered as an extension of the former, and it has the following formulation:

$$\frac{\widehat{Y}_{O_2}^{(2,1,i)}}{\widehat{Y}_{O_2}^{(1,1,1)}} = EquilSoftmax(\boldsymbol{x})_i = \frac{\exp\left(x^{(2,1,i)}\right)Q_i(T)}{\sum_i \exp\left(x^{(2,1,i)}\right)Q_i(T)} \qquad \text{for } i = 1, 2, 3 \,, \tag{14}$$

where $Q_i(T)$ is the internal partition function of group i. Therefore, if $x^{(2,1,i)} = 0 \,\forall\, i$, all the groups are in equilibrium at the translational temperature T. In the case of isothermal reactors, T is provided as one of the inputs $\boldsymbol{u}$. Conversely, for adiabatic systems like the 1-D test case scenario considered in this work, T is predicted by a separate DeepONet. This additional transformation layer, referred to as the Boltzmann layer in the introductory section, enforces local equilibrium distributions between states in the same group by construction. Moreover, it positively impacts the regularization of the network by providing a physically consistent prior distribution to anchor the network parameters, specifically a zero-valued distribution, which can be effectively regulated using L^2 regularization. This ensures that the surrogate predictions remain closely aligned with the known reference equilibrium state, preventing excessive divergence and enhancing the robustness of the surrogate. It is worth highlighting that during the joint training process, all the parameters of CG-DeepONet$^{(1,1)}$ are re-trained together with the ones of CG-DeepONet$^{(2,1)}$, rather than being kept frozen. This is performed by employing fine-tuning transfer learning with L^1-*SP* and L^2-*SP* regularization as described in reference[113].

– *Faster Timescales*. It is possible to increase the accuracy of the CG model by further splitting the energy space into a higher number of clusters. Therefore, by sequentially repeating the same procedure that has been done for augmenting the model from *Timescale 1* to *Timescale 2*, we can construct a surrogate that can predict the dynamics of high-resolution CG models. In our case, we further split each bin into three more bins, obtaining first a 3-group CG modeling for *Timescale 2*, then a 9-group CG modeling for *Timescale 3*, and finally a 27-group CG modeling for *Timescale 4*. We treat each group's triplet with a single DeepONet, and we apply the *EquilSoftmax* layer at the output of each entire timescale block. As explained in the previous paragraph, the predicted mass fraction of each macro-group multiplies the distribution of the corresponding three micro-groups, obtaining a hierarchical surrogate for multi-scale coarse-grained dynamics, as shown in Fig. 3c.

Training strategy.   Physics-informed neural networks (PINNs)[76] can integrate data and physical governing laws by adding PDE residuals to the loss function of neural networks by relying on automatic differentiation. This capability can also be incorporated into the DeepONet framework (physics-informed DeepONet or PI-Deep-ONet)[82,83]. Specifically, the following composite loss function is minimized to train the network parameters, $\boldsymbol{\theta}$:

$$\mathscr{L}(\boldsymbol{\theta}) = \lambda_d \mathscr{L}_d(\boldsymbol{\theta}) + \lambda_r \mathscr{L}_r(\boldsymbol{\theta}) + \lambda_{ic} \mathscr{L}_{ic}(\boldsymbol{\theta}) + \Lambda(\boldsymbol{\theta}) \,, \tag{15}$$

where $\mathscr{L}_d(\boldsymbol{\theta})$ is computed based on the discrepancy between predicted and given data points, $\mathscr{L}_r(\boldsymbol{\theta})$ is the residual loss, $\mathscr{L}_{ic}(\boldsymbol{\theta})$ is the loss over the initial conditions of the 0-D reactor, and $\Lambda(\boldsymbol{\theta})$ contains the L^1 and L^2 regularization loss. These terms can be expressed as follows:

$$\mathscr{L}_d(\boldsymbol{\theta}) = \frac{1}{N_d} \sum_{i=1}^{N_d} \ell\left(\boldsymbol{Y}\left(\boldsymbol{u}_d^i, t_d^i\right), \widehat{\boldsymbol{Y}}\left(\boldsymbol{u}_d^i, t_d^i\right)\right), \tag{16}$$

$$\mathscr{L}_r(\boldsymbol{\theta}) = \frac{1}{N_r} \sum_{i=1}^{N_r} r\left(t_r^i, \rho_r^i, \widehat{\boldsymbol{Y}}\left(\boldsymbol{u}_r^i, t_r^i\right), T_r^i\right), \tag{17}$$

$$\mathscr{L}_{ic}(\boldsymbol{\theta}) = \frac{1}{N_{ic}} \sum_{i=1}^{N_{ic}} \ell\left(\boldsymbol{Y}\left(\boldsymbol{u}_{ic}^i, 0\right), \widehat{\boldsymbol{Y}}\left(\boldsymbol{u}_{ic}^i, 0\right)\right), \tag{18}$$

where N_d, N_r, and N_{ic} denote the batch sizes of the training data. $\boldsymbol{Y}$ are the exact mass fraction values from direct numerical simulation of the CG master equations (CGME), whereas $\widehat{\boldsymbol{Y}}$ are the predicted ones from the surrogate.

The parameters λ_d, λ_r, and λ_{ic} correspond to weight coefficients in the loss function that can effectively assign a different learning rate to each loss term. In this study, the error function ℓ is expressed as follows:

$$\ell\left(Y, \widehat{Y}\right) = \frac{1}{|\widehat{Y}|} \left\| \frac{\ln\left(Y\right) - \ln\left(\widehat{Y}\right)}{\ln\left(Y\right)} \right\|_1 , \tag{19}$$

while the residual $r \in \mathbb{R}$ is

$$r\left(t, \rho, \widehat{Y}, T\right) = \frac{1}{|\widehat{Y}|} \left\| \rho \frac{d\widehat{Y}}{dt} - \mathbf{\Omega^0}\left(\rho, \widehat{Y}, T\right) \right\|_1 , \tag{20}$$

with $\mathbf{\Omega^0}$ being the right hand side of Eq. (3).

Given the hierarchical structure of the proposed surrogate, the parameters of the entire network are trained by adopting a multi-step procedure:

i. *Fully data-driven optimizations*. In this first step, the surrogate is trained sequentially from the slowest to the fastest timescale with only anchor and ICs points ($\lambda_d = 1$, $\lambda_r = 0$, $\lambda_{ic} = 1$) obtained from the numerical solution of the coarse-grained master equations:

   a) Training only *Timescale 1* with data generated by solving CGME1;
   b) Training jointly *Timescales 1-2* with data generated by solving CGME3;
   c) Training jointly *Timescales 1-2-3* with data generated by solving CGME9;
   d) Training jointly *Timescales 1-2-3-4* with data generated by solving CGME27.

   At each training step, the knowledge acquired from the previous iterations is preserved and used as a prior by employing fine-tuning transfer learning with L^1-*SP* and L^2-*SP* regularization as described in reference[113]. For instance, in step (b), the calibrated weights for *Timescale 1* from step (a) are kept and finely retrained with the newly initialized parameters of *Timescale 2*.

ii. *Hybrid physics-informed and data-driven optimization*. The governing equations describing the CGME27 model are now enforced in the trained surrogate from step (i.d) using the hybrid loss formulation shown in Eq. (15). The weight coefficients λ_i are automatically tuned using the learning rate annealing technique described in reference[114]. The tuning procedure involves balancing the gradients of different loss terms during back-propagation using λ_i as a re-scaling factor of the learning rate corresponding to each loss term. This technique ensures that the model's parameters are updated in a balanced manner, giving equal importance to all the loss terms. The complete training history of the parameter values λ_i can be found in the Supplementary Sect. S.2.2.3.

The decision to incorporate the residual loss only in the final step is intended to accelerate the training of the entire surrogate. Data from numerical simulations serves as anchor points for frequent or commonly seen scenarios, while the residual of the governing laws ensures the model's ability to generalize to different, unseen physical conditions.

Adaptive pruning technique.   Flow simulations are often characterized by regions of strong and weak non-equilibrium conditions of the gas. When the extent of non-equilibrium is large, the highest resolution is needed to resolve all the physical processes accurately. However, there are conditions for which the fine scales (or micro-groups) corresponding to the highest resolution CG model are in equilibrium with other neighboring groups or states. For these cases, adding resolution penalizes the computational efficiency rather than improving the model's accuracy. In fact, under these conditions, the population distribution can be approximated with a Boltzmann distribution, and the low-fidelity CG model can accurately resolve their dynamics. Figure 4 illustrates the concept described above, where all the reconstructed low-lying energy states from different coarse-grained (CG) models are considered to be in equilibrium. As a result, it is sufficient to predict the values of the first group of the CGME3 model, without needing to resolve all the timescales.

These observations indicate the need to introduce a controller in the algorithm that accurately determines the resolution level needed to describe the dynamics of the system, without explicitly computing unnecessary fine scales. In the following, the design procedure for the additional controller-acting surrogate is firstly outlined, including the definition of the control variable and the network architecture. Subsequently, the adaptive inference technique is described, which involves the dynamic pruning of unnecessary nodes in the CG-DeepONets hierarchical architecture. This online pruning process enhances computational efficiency by selectively skipping the evaluation of specific nodes based on the local thermochemical state of the gas.

– *Physically-relevant non-equilibrium control variable*. First, defining a metric that can quantify the physical information lost due to the coarse-graining procedure is crucial. This work employs the Euclidean distance between the Boltzmann reconstructed states of the highest resolution CG model available (i.e., *Timescale 4*) and the remaining low-fidelity ones. Since only the zeroth-order moment of the master equations is considered, the bin-specific coefficient α in Eq. (2) is selected to construct our metric, which can be expressed as follows:

(a)

(b)
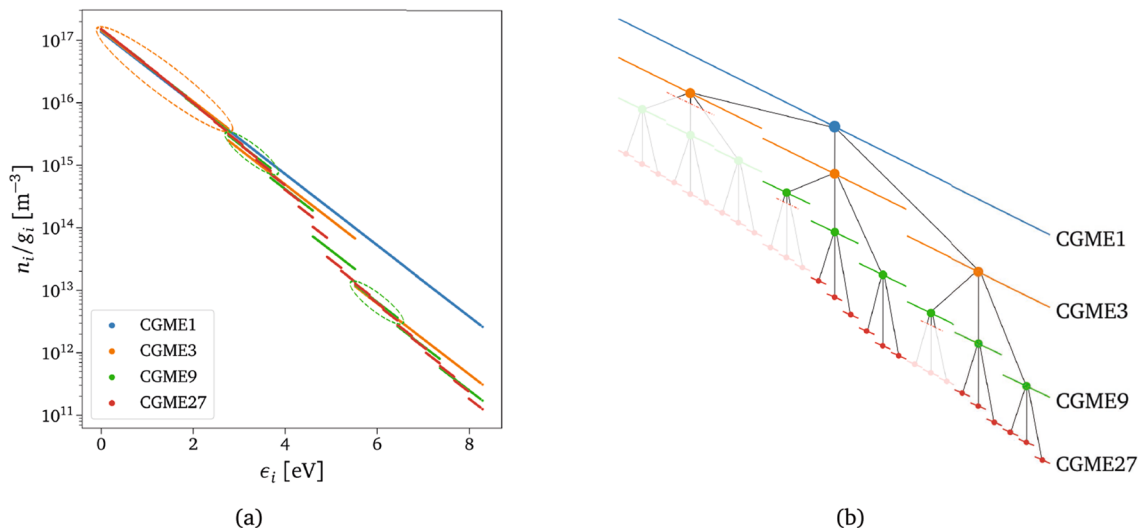
Figure 4. Example of the adaptive strategy. (**a**) O_2 QSS rovibrational distribution for 1-group (CGME1), 3-groups (CGME3), 9-groups (CGME9), and 27-groups (CGME27) coarse-graining. The dashed ovals identify those CG high-resolution groups that can be accurately reconstructed from the low-resolution ones. (**b**) Exploded view of the groups' graph. The opaque dots represent the Boltzmann-reconstructed groups that correspond to the ovals in (**a**) and do not require evaluations of CG-DeepONets' high-resolution components.

$$\delta^{(ts,\cdot,P)} = \frac{1}{\mathcal{N}_p} \sum_{p,\mathscr{I}^{(4,\cdot,p)} \subset \mathscr{I}^{(ts,\cdot,P)}} \left(\alpha^{(ts,\cdot,P)} - \alpha^{(4,\cdot,p)} \right)^2 , \tag{21}$$

where *ts* and *P* (or *p*) refer to the timescale and its specific group, respectively. Equation (21) involves the computation of the difference between the offsets of the log-linear Boltzmann distribution functions described in Eq. (2). The sum in Eq. (21) is performed over all the $\mathcal{N}_p$ micro-groups of *Timescale 4* that belong to the macro-group *P* of timescale *ts*. Figure 5a provides a visual intuition of Eq. (21) for the first CGME3-group, which consists of the sum of the drawn dashed black lines. We briefly mention that other options for constructing the metric could have relied on the Kullback-Leibler divergence computed between population or energy distributions at the different temporal scales.

– *Controller-acting surrogate architecture.* Given the defined metric, the design of the non-equilibrium controller-acting surrogate requires a specific architecture. To maintain consistency with the coarse-grained operator network described in Section "Neural operators", we again leverage the multi-scale connotation of

(a)

(b)
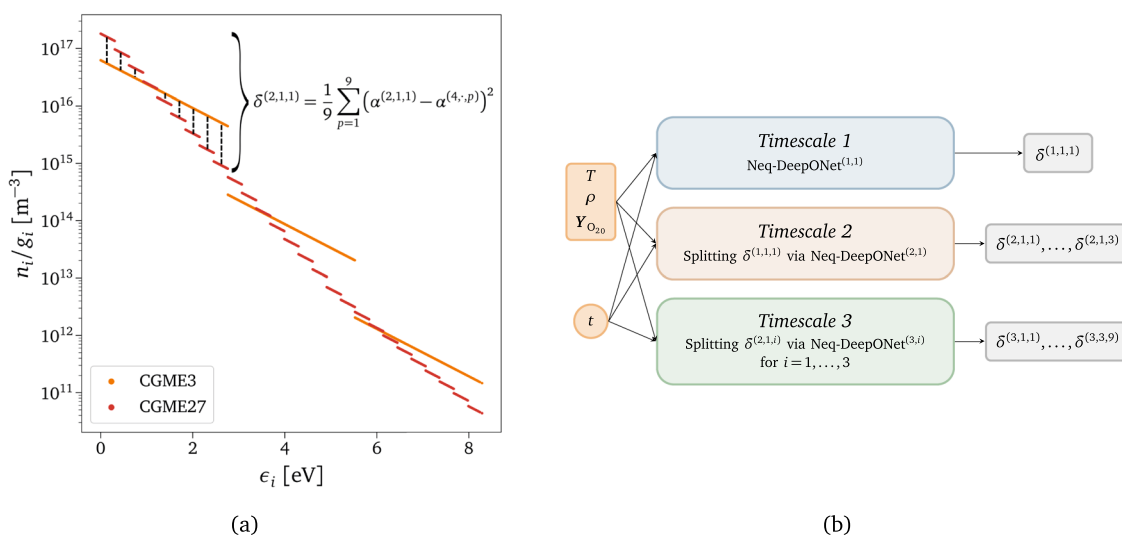
Figure 5. Adaptive inference design. (**a**) Euclidean distance metric, δ, used to quantify the physical information lost due to the equilibrium assumption imposed in a too-large subspace in the energy phase. α represents the zeroth-order term, i.e., the offset of the log-linear Boltzmann distribution function defined in Eq. (2). (**b**) Schematics of the multi-scale network architecture of the controller-acting surrogate responsible for adapting the required coarse-grained model resolution based on the local flow conditions.

the physical problem by separately modeling the underpredicted non-equilibrium values for each CG low-fidelity model, as illustrated in Fig. 5b. An exponential transformation is applied to the surrogate outputs, and a single DeepONet is used for each triplet of values, following a similar approach as used for the CG-DeepONets. More details can be found in the Supplementary Sect. S.2.3.

– *Physics-driven online pruning*. The composition of coarse-grained deep operator networks (CG-DeepONets) and non-equilibrium controller-acting DeepONets (Neq-DeepONets) allows the development of a technique that, given IC and time instant, adaptively predicts the groups' distribution with the highest accuracy and lowest computational cost possible. This technique can be summarized as a two-step procedure which goes as follows:

    i.   The first step involves querying the Neq-DeepONets to obtain the non-equilibrium control variable δ for each CG resolution level. This variable reflects the inaccuracy of the low-fidelity CG models in describing the non-equilibrium state of the gas at the upcoming time instant.

    ii.   The predicted δ is then compared with a user-chosen tolerance level, δ_{tol}. If the predicted value is lower than the tolerance, the resolution level of the specific low-fidelity CG model is deemed sufficient to accurately represent the reactor dynamics. In such a case, the leaf nodes of the corresponding dependent tree in the CG-DeepONets model are temporarily pruned and not evaluated, as exemplified in Fig. 4b.

At this point, we highlight the twofold advantage of CG-DeepONets' hierarchical structure. In fact, other than simplifying the training stage, the presence of the controller boosts the inference phase, as the surrogate relies only on the CG-DeepONets' components that are truly required to characterize the non-equilibrium distributions. The details of the adaptive algorithm are presented in the Supplementary Sect. S.2.3.

Results

The framework discussed in the previous sections is used to construct a surrogate for an ideal chemical reactor. The first part of this section provides the details of the training and testing of the surrogate in isothermal 0-D scenarios, demonstrating its ability to learn the differential operator governing the physics of the reactor. The surrogate's predictions are then compared against the solutions obtained from the numerical integration of the governing equations. Observables such as time-resolved distributions and its moments, including densities and energies, are employed for evaluation. Furthermore, details regarding the adaptive technique and a preliminary analysis of computational savings are provided. At the end of the section, the results of the one-dimensional numerical experiment are analyzed in terms of surrogate accuracy and performance.

Inference.   As explained in Section "Physical modeling", different initial conditions have been uniformly sampled from Table 1 to train and test the proposed ML framework. Figure 6 shows the broad ranges of the space of ICs for pressure, P_0, molar fraction of atomic oxygen, X_{O_0}, and internal temperature, T_{int_0}. A fourth dimension should be considered since the translational temperature of the reactor, T, also varies. In Fig. 6, the red dots represent unseen test scenarios, whereas the black crosses represent the training points.

Figure 7a compares the exact solution computed by the numerical integrator and the surrogate's predictions for one unseen scenario taken from the test data set in Fig. 6. The isolated blue line represents the evolution of the atomic oxygen taken from *Timescale 1*. In contrast, the others describe the dynamics of the 27 rovibrational energy-based groups predicted by *Timescale 4*. The inference has been performed by querying the CG-DeepONet
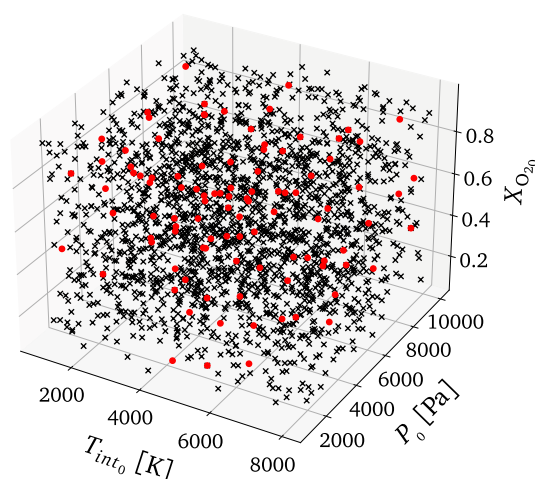
Figure 6. Space of initial conditions. The black crosses represent the set of training points, while the red dots identify the testing data set. Note: the figure is missing the last fourth dimension in the space of initial conditions, i.e., the translational temperature of the reactor, T.
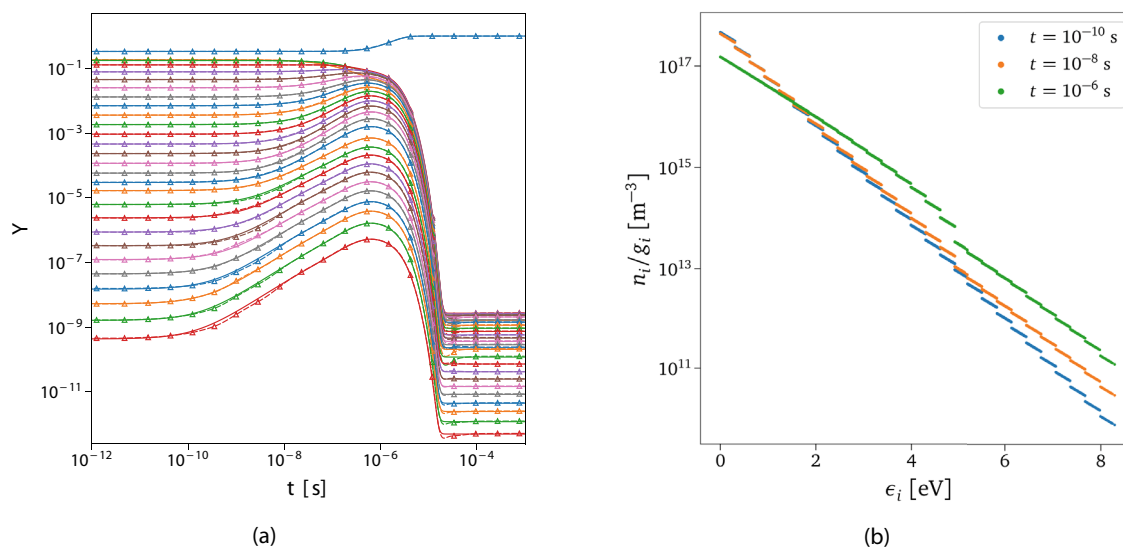
Figure 7. Inferred solution. (**a**) Reference solution (solid line) versus prediction from the trained model (dashed line with markers). The isolated blue line represents the evolution of the atomic oxygen, while the remaining are the 27 groups of O_2 in ascending order of energy content per particle (top-down). Initial conditions used: $P_0 = 3\,617.24$ Pa, $X_{O_0} = 0.4996$, $T_{int_0} = 4\,333.16$ K, $T = 9\,742.13$ K. (**b**) Predicted O_2 rovibrational states distribution at three different time instants taken from the dynamics shown in (**a**).

based on the vector of time instants generated from the numerical integrator and the given initial conditions, defined by $\left\{ \left[T,\ \rho,\ \boldsymbol{Y}_{O_{2_0}} \right],\ t_k \right\}_{k=1}^{M}$, with M the number of evaluation points.

From Fig. 7a, it can be observed that the predicted and exact solutions show excellent agreement. This indicates that the trained model is capable of accurate predictions for different and unseen initial conditions (additional test cases are presented in the Supplementary Sect. S.2.2.2). Negligible discrepancies can be noticed in various regions of the dynamics of the heat bath, which can be improved by further refining the trained model. To the author's best knowledge, this work provides the first application of PI-DeepONets to a dynamical system containing many such degrees of freedom. The main reason for such good surrogation of the dynamics is that the hierarchical structure of the proposed deep learning framework embodies the multi-scale connotations of the problem, showing higher accuracy and robustness compared to a vanilla DeepONet architecture (details provided in the Supplementary Sect. S.2.1.1). The micro-groups inside each macro-groups equilibrate faster between each other than with other ones outside it. For this reason, they show very similar behavior in their dynamics, which can be captured by the few modes discovered by the shared trunk. This aspect facilitates reaching high levels of accuracy with a relatively small number of network parameters. Indeed, the surrogate correctly predicts the dynamics of almost thirty species spanning a wide range of orders of magnitude (around 12) in mass fractions values. Additionally, to expand the initial conditions' space even further by keeping such a high accuracy level and relatively small network architecture, one could consider constructing multiple surrogates. Each of these surrogates can be built with the same architecture but specialized for a local sub-domain in the space of the initial conditions.

Accuracy. The relative L^2-norm has been used as the error metric to evaluate the accuracy of the surrogate, consistently with reference[82]. In particular, the employed test error corresponds to the mean relative error of the surrogate's predictions for *Timescale 4* over all the examples in the test data set:

$$\varepsilon^{(j)} = \frac{1}{N} \sum_{i=1}^{N} \frac{\left\| \widehat{Y}_i^{(4,\cdot,j)}(t) - Y_i^{(4,\cdot,j)}(t) \right\|_2}{\left\| Y_i^{(4,\cdot,j)}(t) \right\|_2} \qquad \text{for } j \in \{1, \ldots, \mathcal{N}_{\mathcal{G}} \}, \tag{22}$$

where $\mathcal{N}_{\mathcal{G}} = 27$ represents the number of groups, $N = 100$ denotes the number of testing cases, and t represents a set of log-uniformly spaced points in the time domain. For this analysis, $1\,000$ points in time have been sampled from each testing scenario. The four highest errors of the inferred solution are presented in table 3. Once again, the reported values confirm the excellent agreement between the numerically integrated master equations and the predicted solutions, with a maximum relative L^2-norm error of approximately 4.5%.
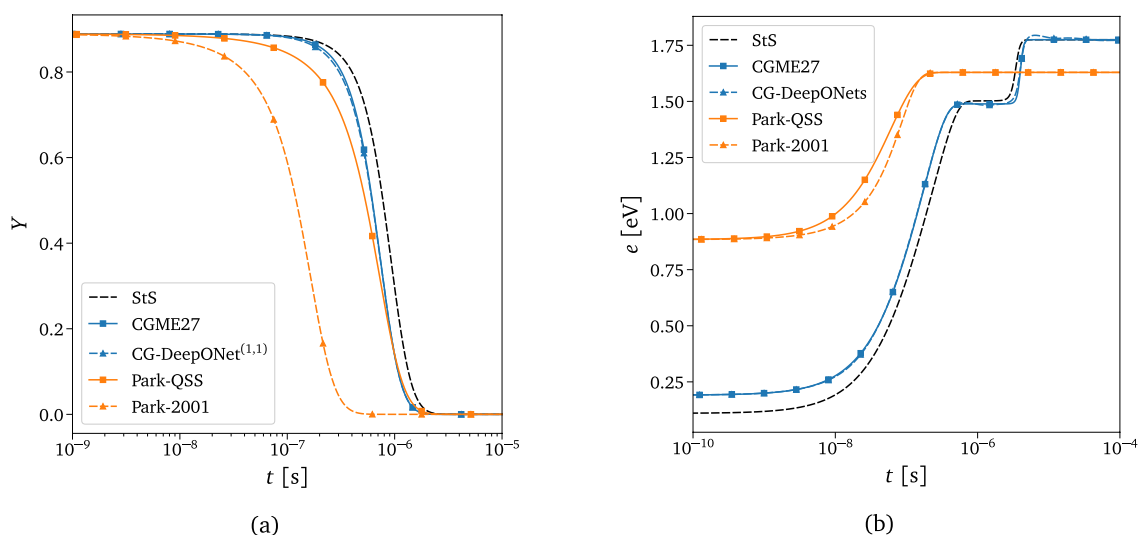
Surrogate predictions *vs.* numerically-integrated thermochemical models. To demonstrate the level of physical accuracy of the coarse-grained surrogate discussed in this study, a comparison is made against the reference CG solution, the high-fidelity state-to-state solution, and the computationally cheaper two-temperature model of Park, which is a specific case of the multi-temperature models described in Section "Physical modeling". The exact CG, StS, and Park's solutions have been computed with traditional numerical integrators.

Group	Rel. error [%]
$\widehat{Y}^{(4,2,6)}$	4.52 ± 2.44
$\widehat{Y}^{(4,3,9)}$	4.12 ± 2.64
$\widehat{Y}^{(4,9,27)}$	3.82 ± 1.59
$\widehat{Y}^{(4,6,21)}$	3.35 ± 1.54

Table 3. *Test error.* The four highest mean relative L^2-norm testing errors (with standard deviations) of the trained model for *Timescale 4*.

In Fig. 8, two different approaches are considered for Park's model, one employing the less accurate but still widely used kinetics from reference[102], derived from empirical methods or experimental data, and the other using the more recent QSS approach[5], whose kinetic database is directly computed from state-to-state calculations. Figure 8 shows the evolution of the total mass fraction and internal energy content per particle of O_2 for the different models considered. It is evident from the figure that the coarse-grained grouping strategy employed in this work provides the closest solution to state-to-state modeling. Only *Timescale 1* (or CG-DeepONet$^{(1,1)}$) of the proposed surrogate has been queried to produce the evolution of the total mass fraction of O_2 shown in Fig. 8a, which is in excellent agreement with the numerically-integrated CG solution. This is because CG-DeepONet$^{(1,1)}$ implicitly contains all the information about the energy transfer processes between the 27 groups, as it has been trained with data from the integration of CGME27. However, while using only *Timescale 1* is sufficient for accurately predicting the dynamics of the total mass fraction of the reactor species, the same approach may not be accurate for predicting the total internal energy content of the molecule. This is because CG-Deep-ONet$^{(1,1)}$ is specifically designed to model only the zeroth-order moment of the master equations and may not capture higher-order moments, such as the total internal energy content, with sufficient accuracy. Therefore, this quantity generally requires the evaluation of the overall surrogate, which includes the low-scale components CG-DeepONet$^{(2:4,:)}$. The discrepancy between the CG surrogate's predictions and the StS numerical solution in Fig. 8 is almost exclusively determined by the physical simplifications made by the CG model. In particular, the energy difference that can be noticed at the initial time instants is caused by the fact that the reconstructed states within each bin follow a Boltzmann distribution at the translational temperature T (for the assumptions made in Section "Physical modeling"). In contrast, the quantum energy levels for the StS solution follow a distribution at temperature T_{int_0}.

The proposed hierarchical architecture could be upgraded to model higher-order moments of the master equations. This improvement could involve replicating the same architecture as the CG-DeepONets to model the internal energy content of every single bin. Consequently, CG-DeepONet$^{(1,1)}$ could correctly predict both zeroth-, i.e., total mass, and first-order moment, i.e., internal energy, of O_2. In such a case, the low-scale components CG-DeepONet$^{(2:4,:)}$ would not be required to predict the solution shown in Fig. 8b, but they might still be


Figure 8. Surrogate predictions compared to numerically-integrated thermochemical models. ◦ Black dashed line: state-to-state exact solution ◦ Blue line with markers: coarse-grained exact solution ◦ Blue dashed line with markers: coarse-grained surrogate predictions ◦ Orange line with markers: Park's two-temperature model with QSS approach[5] ◦ Orange dashed line with markers: Park's two-temperature model with kinetics from reference[102]. Initial conditions used: $P_0 = 3\,000$ Pa, $X_{O_0} = 0.2$, $T_{int_0} = 1000$ K, $T = 10\,000$ K. (a) Evolution of total mass fraction of O_2. (b) Weighted rovibrational energy evolution per particle of O_2.

necessary for providing the correct distribution function of the quantum energy states when considering other physical phenomena, such as radiation.

Adaptive inference.   The advantage of the hierarchical architecture proposed in this work is the ability to tailor the model complexity to the specific localized flow conditions to obtain a computationally efficient yet accurate physical model. Figure S4 in the Supplementary Information shows an example of the dynamics of underpredicted non-equilibrium Euclidean metric computed via Eq. (21) for *Timescale 1* and *Timescale 3* for the same test case shown in Fig. 9. The values plotted in Fig. S4 can be considered a good reference for the space the proposed metric can span, as the analyzed test case exhibits considerable initial thermal and chemical non-equilibrium. It should be noted that the values of $\delta^{(1,1,1)}$ reported in Fig. S4a are almost an order of magnitude larger than Fig. S4b due to the more accurate modeling adopted in the latter. Overall, the trend is decreasing by approaching the equilibrium, except for the evident QSS region starting around 10^{-6} s, where all the quantities remain constant. $\delta^{(3,2,6)}$ shows an interesting behavior in Fig. S4b, which corresponds to the sixth group of the 9-groups rovibrational energy-based coarse-grained grouping strategy for *Timescale 3*, the one close to the dissociation energy (5.115 eV). By observing the highly non-equilibrium StS dynamics at QSS of the states in this group (e.g., Fig. 2), it is clear that the highest resolution possible is necessary for that region of the energy space to model the dynamics of those states accurately[5].

The solution obtained with the adaptive technique is compared with the exact one in Fig. 9 for two different values of the underpredicted non-equilibrium metric tolerance, δ_{tol}. This value acts as a discriminant for assuming equilibrium inside each macro-group for all the timescales modeled. For $\delta_{\text{tol}} = 0.1$, the adaptation starts playing effect just before the QSS region, as can also be deduced from Fig. S4a, whereas for $\delta_{\text{tol}} = 0.5$, it already acts at the beginning of the dynamics. We can assert that for a value of $\delta_{\text{tol}} = 0.1$, the solution looks very similar to the exact one, supporting the effectiveness of the adaptive technique in terms of physical accuracy. The adaptive solutions shown in Fig. 9 have been obtained by solving the number of groups dictated by the respective δ reported in Fig. 10a as functions of time. From Fig. 10a, it is evident that the number of the solved groups decreases considerably by increasing the tolerance value, confirming the validity of the proposed adaptive technique. As already demonstrated in the previous section, the prediction of the total mass fraction of O_2 is independent of the tolerance used since our model has been trained such that even the lowest-fidelity coarse-grained model can correctly predict the actual mass of the reactor species. However, in the case of energy, the choice of the proper tolerance can play an essential role in predicting its correct value, as shown in Fig. 10b.

Figure 10c presents a preliminary performance analysis of the adaptive technique for the different tolerance values based on a comparison with the standalone CG-DeepONet model. The reported timings are obtained as the mean of 1000 different inference evaluations of the model per each physical time instant, conducted with a single central processing unit (CPU) core. The computations shown in Fig. 10c have been performed in the TensorFlow[115] environment, which means that a large part of the network evaluation time involves Python call overhead. The bar plot illustrates that the adaptive technique outperforms the standalone surrogate at later stages of the system's dynamical evolution, particularly when the composition approaches the asymptotic equilibrium value. The opaque bar chunks in Fig. 10c represent the contribution to the inference cost due to the Neq-Deep-ONets surrogate. A great advantage of this methodology is also its flexibility, as computational costs and physical accuracy can be easily balanced by tuning the tolerance value, δ_{tol}. Moreover, inference with physics-informed

(a)   (b)
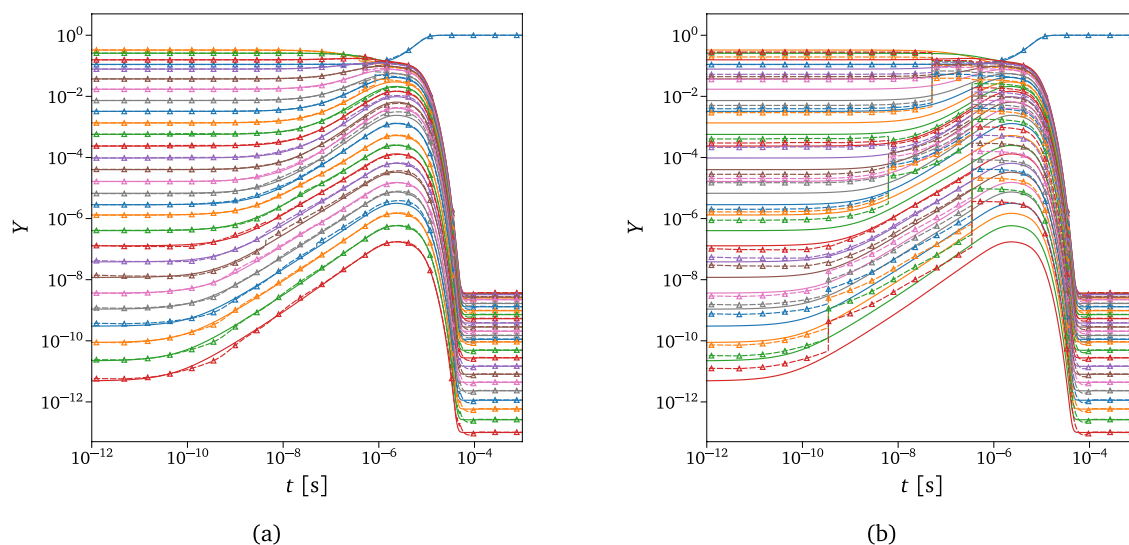
Figure 9. Exact vs. adaptive solution for different δ_{tol}. Exact solution (solid line) versus prediction from the trained surrogate (dashed line with markers) using the adaptive technique. The isolated blue line represents the evolution of the atomic oxygen, while the remaining are the 27 groups of O_2 in ascending order of energy content per particle (top-down). Initial conditions used: $P_0 = 3\,000$ Pa, $X_{O_0} = 0.2$, $T_{int_0} = 3\,500$ K, $T = 8\,000$ K. (a) Underpredicted non-equilibrium tolerance used: $\delta_{\text{tol}} = 0.1$. (b) Underpredicted non-equilibrium tolerance used: $\delta_{\text{tol}} = 0.5$.

(a) (b) (c)
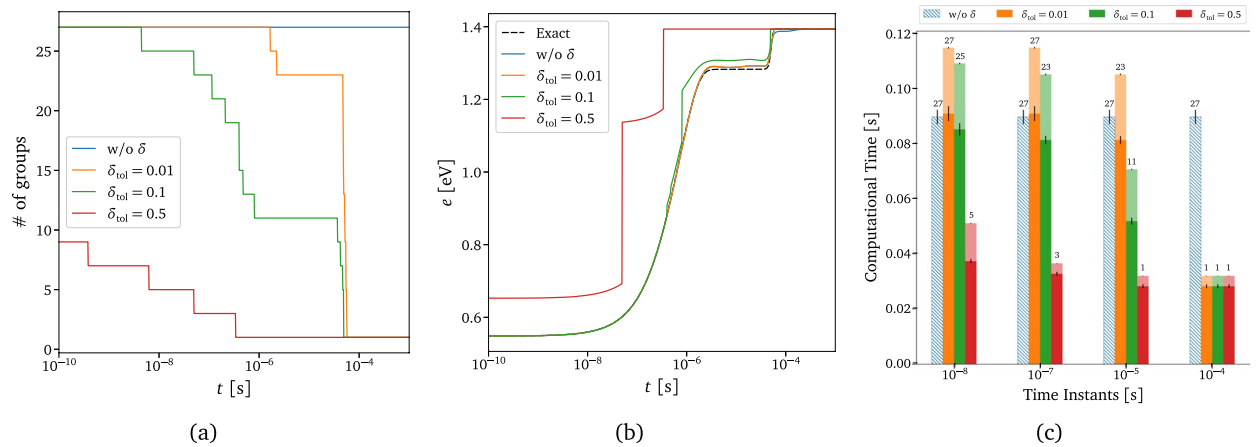
Figure 10. Adaptive solution for different δ_{tol}. The evolution of the total mass fraction of O_2 is not reported here, as the model accurately predicts the solution regardless of the tolerance value used. (**a**) Total number of actually modeled O_2 groups. (**b**) Weighted sum of the grouped-specific O_2 rovibrational energy, with weights given by the groups' distribution. (**c**) Computational cost comparison between the adaptive technique and the standalone CG-DeepONet model for single-time instant evaluation. The numbers on top of the bars are the corresponding number of groups, also shown in (**a**).

DeepONets is trivially parallelizable with graphics processing units (GPUs), which can remarkably boost the inference timings shown in Fig. 10c. Wang et al.[82,83] have already demonstrated that PI-DeepONets can outperform and replace conventional numerical solvers even for long-time integration.

One-dimensional shock case scenario. In this section, preliminary results of a one-dimensional numerical experiment are presented, where the constructed surrogate is tested both with and without the adaptive technique.

Figure 11a,b present the final temperature and mass fraction profiles in the shock reference frame for the test case scenario described in Sect. 2.1. In both figures, the exact solution obtained using a thermochemical library is represented with black dashed lines, while the solution obtained using the surrogate without adaptation and employing adaptive inference with tolerance values of $\delta_{tol} = 0.01$ and $\delta_{tol} = 0.05$ are represented by blue, orange, and green lines, respectively. The integration using the surrogate produces physically correct solutions, with the largest differences noticed at the tail of the temperature profile, in particular when the tolerance value is high. As already explained in the previous section and demonstrated in Fig. 10b, the reason for these small discrepancies is due to the incorrect predictions of internal energy, which can result in incorrect temperature profiles while the conservation equation for total energy is integrated in time. The reconstructed microscopic distribution is also presented in Fig. 11c, showing a good agreement of the surrogate predictions with and without adaptation compared to the numerically integrated solution.


(a) (b) (c)
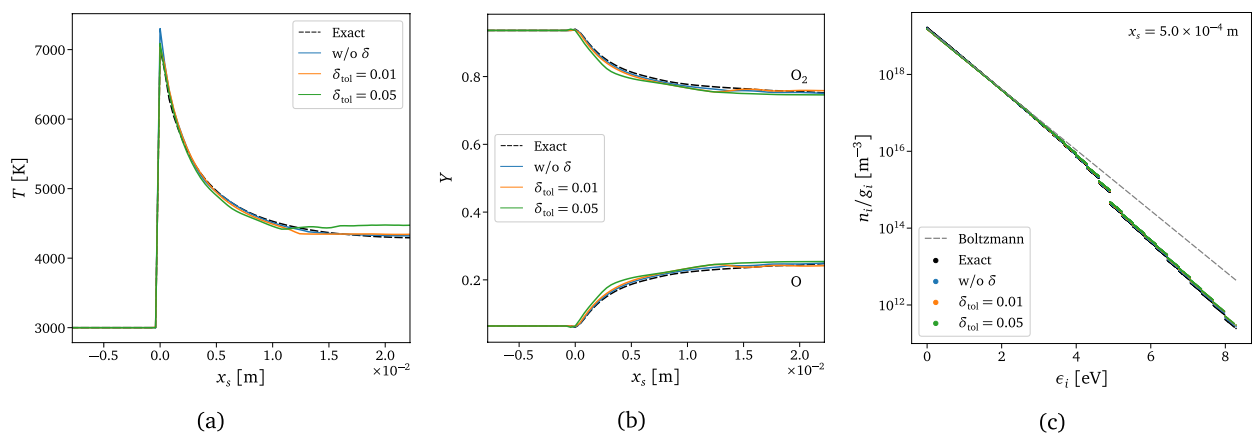
Figure 11. One-dimensional shock solution. Comparison between exact and predicted final solutions in the shock reference frame, x_s. ○ Black dashed line: exact solution ○ Blue line: surrogate predictions without adaptation ○ Orange line: adaptive surrogate predictions with $\delta_{tol} = 0.01$ ○ Green line: adaptive surrogate predictions with $\delta_{tol} = 0.05$. (**a**) Translational temperature. (**b**) Total mass fractions of O_2 and O. (**c**) Reconstructed O_2 rovibrational states distribution at $x_s = 5 \times 10^{-4}$ m and corresponding Boltzmann equilibrium distribution function.

Figure 12a,b provide a preliminary performance analysis of surrogate inference with and without adaptation. The timings are computed by evaluating only the integration time for the reactive step in Eq. (9) using a single CPU core within Fortran 2008 environment. The corresponding statistics, i.e., mean and standard deviation, are calculated over 500 iterations and averaged over the number of cells in the 1-D domain. The speedup statistics are then obtained using the formula proposed by Díaz and Rubio[116], which approximates the ratio of two independent normal random variables with a normal distribution. In Fig. 12a, the speedup of the standalone surrogate is presented as a function of time step, Δt, which has been varied by changing only the number of cells and keeping everything else fixed. The surrogate inference is at least eight times faster than the serial integration performed with a conventional implicit scheme, in this case, the second-order backward differentiation formula (BDF2). Furthermore, the maximum speedup is reached when the integration time is much longer, which is expected since the integrator may need more steps to reach the final time, unlike the surrogate inference, which is independent of the total integration time. The computed speedup depends on various factors, such as the dimension of the network, the stiffness associated with the system of equations, the scheme and tolerances used for the ODEs integration, and the length of the integrated physical time. All these details for this particular test can be found in the Supplementary Sect. S.3. In Fig. 12b, a comparison is shown between the varying speedup with δ_{tol} obtained with the adaptive inference technique (light blue) and the constant one obtained with the standalone surrogate (light orange) for $\Delta t = 1.33 \times 10^{-7}$ s. As expected, increasing the tolerance values leads to higher speedups, which is consistent with the reported timings in Fig. 10c. However, this comes at a cost of reduced accuracy, as shown in Fig. 12c, which presents the increasing mean relative error for temperature and total mass fraction of O_2 with increasing δ_{tol}. The reference error values for the surrogate without adaptation are $\varepsilon_{Y_{O_2}} = 0.93\%$ and $\varepsilon_T = 0.58\%$. It is noteworthy that the computation of the error does not include points in the domain where the gas experiences the left or right equilibrium thermochemical states, as the surrogate predictions are not considered in those regions. The increasing error is again related to the inaccurate prediction of internal energy, as observed in the previous analysis of the temperature profile in Fig. 11a, and it may be exacerbated by the error accumulation issue, also shown by Zanardi et al.[104] This highlights the importance of upgrading the surrogate to also model the internal energy content of each individual bin, as it can lead to improved accuracy in terms of the macroscopic quantities of interest. Nevertheless, this approach holds promise when scaled to multi-dimensional CFD simulations with millions of unknowns. For example, in hypersonic simulations, most domain points may lie in the equilibrium or near-equilibrium regions, while only a few points may be in strong non-equilibrium regions (such as shock proximity) where the evaluation of the entire surrogate is needed. In light of these considerations and the performance analysis performed, the adaptive technique has the potential to outperform the standalone model in a multi-dimensional simulation framework.

Conclusions

We proposed a new machine learning-based paradigm inspired and constrained by physical laws for solving multiscale non-equilibrium flows. The designed model (CG-DeepONet) sequentially learned the integral solution operator for multi-fidelity coarse-grained master equations by employing a physics-inspired hierarchical architecture, where physics-informed DeepONet (PI-DeepONet) represents the core element. Furthermore, we developed a controller-acting surrogate (Neq-DeepONet) to learn the dynamics of the underpredicted degree of non-equilibrium to tailor the model's accuracy to the local non-equilibrium conditions. Finally, by combining the two, we designed a novel adaptive pruning inference technique for non-equilibrium thermochemical processes, which showed flexibility in balancing accuracy and computational cost.
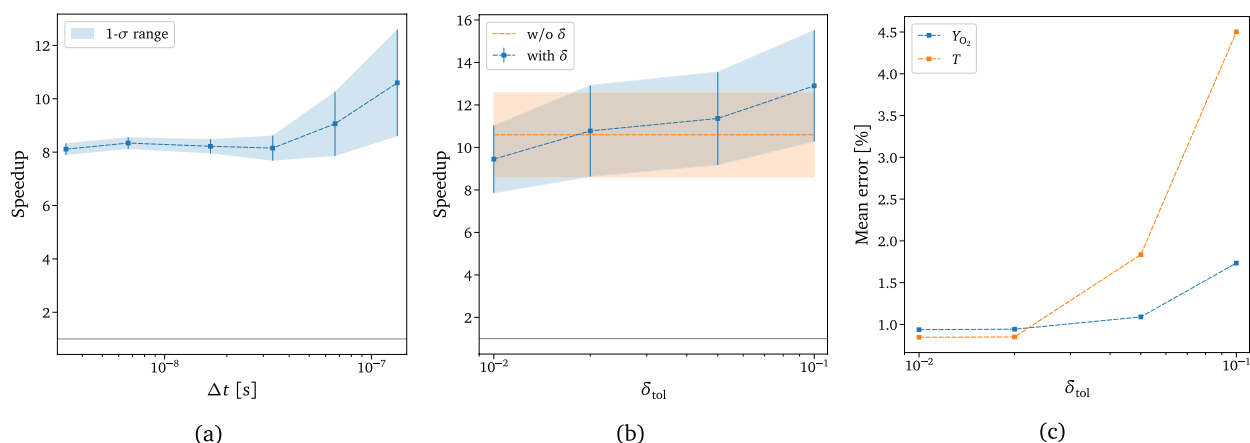
Figure 12. Surrogate performances and accuracy in one-dimensional simulation. (**a**) Speedup achieved using the surrogate without adaptation with increasing Δt, which is equivalent to reducing the number of cells while maintaining a constant CFL. (**b**) Speedup achieved using adaptive surrogate inference with increasing δ_{tol}. (**c**) Mean relative percentage error for Y_{O_2} and T as a function of δ_{tol}. The constant reference values for the surrogate without adaptation are $\varepsilon_{Y_{O_2}} = 0.93\%$ and $\varepsilon_T = 0.58\%$.

Overall, the proposed framework incorporates different key elements that enforce the underlying physics into the surrogate: (i) the physics-based dimensionality reduction in the state space; (ii) the additional layers enforcing the Boltzmann distribution functions, which in turn allow the imposition of prior distributions for the network parameters. When propagated to the state populations (e.g., mass fractions), such priors provide physically consistent solutions even when the surrogate is not trained (i.e., equilibrium distributions); (iii) the physics-informed loss; (iv) the hierarchical architecture and the related sequential fine-tuning transfer learning between different time scales, with mass conservation enforced; (v) the online pruning of the surrogate at the prediction phase through a parsimony-based approach that relies on an additional controller-acting surrogate informed by a non-equilibrium variable.

The methodology was applied to the study of chemical kinetics relevant for application to hypersonic flight and was tested on oxygen mixtures. However, the framework is not constrained to the chosen thermochemical configuration, but it can be extended to Air-5 mixtures (i.e., simultaneously with N_2, O_2, NO, N, and O species) or even other fields of physics spanning a wide range of temporal scales, such as electromagnetism, magnetohydrodynamics, and more generally, plasma physics. The proposed framework was tested in 0-D and 1-D configurations, and the following results were obtained:

- In 0-D scenarios, the CG-DeepONet surrogate alone showed excellent physical accuracy compared to the numerical integration of the master equation, with a maximum relative error of 4.5%. It also exhibited good computational efficiency when the adaptive method was used, gaining more than 3X speedup in the regions of weak non-equilibrium.
- The 1-D numerical experiment demonstrated the flexibility of the proposed method in capturing complex dynamics and confirmed the good performances and accuracy of both standalone and adaptive versions of the constructed surrogate. The relative error was in the range of 1–4.5% with a corresponding 8X-13X speedup compared to conventional implicit schemes employed in an operator-splitting integration framework. As expected, the choice of high tolerances for the adaptive schemes and the consequent lack of degrees of freedom in characterizing the rovibrational distribution generated error accumulations in the predictions of the overall O_2 internal energy. In future work, we will treat the group temperatures as state variables together with the species mass fractions. This addition will have two benefits. Firstly, it will allow us to achieve comparable accuracy with fewer groups. Secondly, it will enable the accurate prediction of the O_2 internal energy by relying only on the first scale (i.e., CG-DeepONet$^{(1,1)}$), similar to what was achieved for the mass fractions (e.g., Fig. 8).

Future work will extend and test the framework to 2-D and 3-D simulations, leveraging its ability to be designed and constructed independently of geometric features of the problem. Additionally, alternative neural operator approaches other than DeepONets will be explored to mitigate the issue of error accumulation. Beyond the application and the numerical outcomes, this work serves as an example on how physics and machine learning can enhance each other, aiming for more interpretable and robust ML-based tools for the scientific community.

Data availability
The dataset used in the current study is available from the corresponding author upon reasonable request.

Code availability
The code used in the current study is available from the corresponding author upon reasonable request.

References
1. Gnoffo, P. A. Planetary-entry gas dynamics. *Annu. Rev. Fluid Mech.* **31**, 459–494. https://doi.org/10.1146/annurev.fluid.31.1.459 (1999).
2. Johnston, C. O. & Panesi, M. Impact of state-specific flowfield modeling on atomic nitrogen radiation. *Phys. Rev. Fluids* **3**, 013402. https://doi.org/10.1103/PhysRevFluids.3.013402 (2018).
3. Harpale, A., Panesi, M. & Chew, H. B. Communication: Surface-to-bulk diffusion of isolated versus interacting C atoms in Ni(111) and Cu(111) substrates: A first principle investigation. *J. Chem. Phys.* **142**, 061101. https://doi.org/10.1063/1.4907716 (2015).
4. Harpale, A., Panesi, M. & Chew, H. B. Plasma-graphene interaction and its effects on nanoscale patterning. *Phys. Rev. B* **93**, 035416. https://doi.org/10.1103/PhysRevB.93.035416 (2016).
5. Panesi, M., Jaffe, R. L., Schwenke, D. W. & Magin, T. E. Rovibrational internal energy transfer and dissociation of $N_2(^1\Sigma_g^+)$-$N(^4S_u)$ system in hypersonic flows. *J. Chem. Phys.* **138**, 044312. https://doi.org/10.1063/1.4774412 (2013).
6. Panesi, M., Munafò, A., Magin, T. E. & Jaffe, R. L. Nonequilibrium shock-heated nitrogen flows using a rovibrational state-to-state method. *Phys. Rev. E* **90**, 013009. https://doi.org/10.1103/PhysRevE.90.013009 (2014).
7. Munafò, A., Lani, A., Bultel, A. & Panesi, M. Modeling of non-equilibrium phenomena in expanding flows by means of a collisional-radiative model. *Phys. Plasmas* **20**, 073501. https://doi.org/10.1063/1.4810787 (2013).
8. Kustova, E. & Mekhonoshina, M. Models for bulk viscosity in carbon dioxide. *AIP Conf. Proc.* **2132**, 150006. https://doi.org/10.1063/1.5119646 (2019).
9. Nagnibeda, E. A. & Kustova, E. *Non-equilibrium Reacting Gas Flows. Heat and Mass Transfer* (Springer, 2009).
10. Panesi, M., Magin, T. E., Bourdon, A., Bultel, A. & Chazot, O. Electronic excitation of atoms and molecules for the FIRE II flight experiment. *J. Thermophys. Heat Transfer* **25**, 361–374. https://doi.org/10.2514/1.50033 (2011).
11. Macdonald, R. L., Munafò, A., Johnston, C. O. & Panesi, M. Nonequilibrium radiation and dissociation of CO molecules in shock-heated flows. *Phys. Rev. Fluids* **1**, 043401. https://doi.org/10.1103/PhysRevFluids.1.043401 (2016).
12. Capitelli, M. *et al. Fundamental Aspects of Plasma Chemical Physics* Vol. 85 (2016).

13. Macdonald, R. L., Torres, E., Schwartzentruber, T. E. & Panesi, M. State-to-State master equation and direct molecular simulation study of energy transfer and dissociation for the N_2-N system. *J. Phys. Chem. A* **124**, 6986–7000. https://doi.org/10.1021/acs.jpca.0c04029 (2020).

14. Wang, D. *et al.* Quantal study of the exchange reaction for $N+N_2$ using an ab initio potential energy surface. *J. Chem. Phys.* **118**, 2186–2189. https://doi.org/10.1063/1.1534092 (2003).

15. Esposito, F., Armenise, I. & Capitelli, M. N-N_2 state to state vibrational-relaxation and dissociation rates based on quasiclassical calculations. *Chem. Phys.* **331**, 1–8. https://doi.org/10.1016/j.chemphys.2006.09.035 (2006).

16. Galvão, B. R. L. & Varandas, A. J. C. Accurate double many-body expansion potential energy surface for $N_3(^4A")$ from correlation scaled ab initio energies with extrapolation to the complete basis set limit. *J. Phys. Chem. A* **113**, 14424–14430. https://doi.org/10.1021/jp903719h (2009).

17. Jaffe, R. L., Schwenke, D. W. & Chaban, G. Theoretical analysis of N_2 collisional dissociation and rotation-vibration energy transfer. In *47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition* (American Institute of Aeronautics and Astronautics, 2009). https://doi.org/10.2514/6.2009-1569.

18. Venturi, S., Jaffe, R. L. & Panesi, M. Bayesian machine learning approach to the quantification of uncertainties on Ab initio potential energy surfaces. *J. Phys. Chem. A* **124**, 5129–5146. https://doi.org/10.1021/acs.jpca.0c02395 (2020).

19. Hammerling, P., Teare, J. D. & Kivel, B. Theory of radiation from luminous shock waves in nitrogen. *Phys. Fluids* **2**, 422. https://doi.org/10.1063/1.1724413 (1959).

20. Knab, O., Fruehauf, H.-H. & Messerschmid, E. W. Theory and validation of the physically consistent coupled vibration-chemistry-vibration model. *J. Thermophys. Heat Transfer* **9**, 219–226. https://doi.org/10.2514/3.649 (1995).

21. Zhu, Y., Zabaras, N., Koutsourelakis, P.-S. & Perdikaris, P. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. *J. Comput. Phys.* **394**, 56–81. https://doi.org/10.1016/j.jcp.2019.05.024 (2019).

22. Haghighat, E., Raissi, M., Moure, A., Gomez, H. & Juanes, R. A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics. *Comput. Methods Appl. Mech. Eng.* **379**, 113741. https://doi.org/10.1016/j.cma.2021.113741 (2021).

23. Sun, L., Gao, H., Pan, S. & Wang, J.-X. Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data. *Comput. Methods Appl. Mech. Eng.* **361**, 112732. https://doi.org/10.1016/j.cma.2019.112732 (2020).

24. Choi, Y., Brown, P., Arrighi, W., Anderson, R. & Huynh, K. Space-time reduced order model for large-scale linear dynamical systems with application to Boltzmann transport problems. *J. Comput. Phys.* **424**, 109845. https://doi.org/10.1016/j.jcp.2020.109845 (2021).

25. You, H., Yu, Y., Trask, N., Gulian, M. & D'Elia, M. Data-driven learning of nonlocal physics from high-fidelity synthetic data. *Comput. Methods Appl. Mech. Eng.* **374**, 113553. https://doi.org/10.1016/j.cma.2020.113553 (2021).

26. Mai, C. V., Spiridonakos, M. D., Chatzi, E. N. & Sudret, B. Surrogate modeling for stochastic dynamical systems by combining nonlinear autoregressive with exogenous input models and polynomial chaos expansions. *Int. J. Uncertain. Quantif.* **6**, 313–339. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2016016603 (2016).

27. Rozza, G., Huynh, D. B. P. & Patera, A. T. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Arch. Comput. Methods Eng.* **15**, 229–275. https://doi.org/10.1007/s11831-008-9019-9 (2008).

28. Benner, P., Gugercin, S. & Willcox, K. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Rev.* **57**, 483–531. https://doi.org/10.1137/130932715 (2015).

29. Amsallem, D. & Farhat, C. Stabilization of projection-based reduced-order models. *Int. J. Numer. Methods Eng.* **91**, 358–377. https://doi.org/10.1002/nme.4274 (2012).

30. Huang, C., Wentland, C. R., Duraisamy, K. & Merkle, C. Model reduction for multi-scale transport problems using model-form preserving least-squares projections with variable transformation. *J. Comput. Phys.* **448**, 110742. https://doi.org/10.1016/j.jcp.2021.110742 (2022).

31. Swischuk, R., Mainini, L., Peherstorfer, B. & Willcox, K. Projection-based model reduction: Formulations for physics-based machine learning. *Comput. Fluids* **179**, 704–717. https://doi.org/10.1016/j.compfluid.2018.07.021 (2019).

32. Choi, Y. & Carlberg, K. Space-time least-squares Petrov–Galerkin projection for nonlinear model reduction. *SIAM J. Sci. Comput.* **41**, A26–A58. https://doi.org/10.1137/17M1120531 (2019).

33. Carlberg, K., Bou-Mosleh, C. & Farhat, C. Efficient non-linear model reduction via a least-squares Petrov–Galerkin projection and compressive tensor approximations. *Int. J. Numer. Methods Eng.* **86**, 155–181. https://doi.org/10.1002/nme.3050 (2011).

34. Forrester, A. I. J., Sóbester, A. & Keane, A. J. *Engineering Design via Surrogate Modelling* (Wiley, 2008).

35. Xu, J. & Duraisamy, K. Multi-level convolutional autoencoder networks for parametric prediction of spatio-temporal dynamics. *Comput. Methods Appl. Mech. Eng.* **372**, 113379. https://doi.org/10.1016/j.cma.2020.113379 (2020).

36. Kim, Y., Choi, Y., Widemann, D. & Zohdi, T. A fast and accurate physics-informed neural network reduced order model with shallow masked autoencoder. *J. Comput. Phys.* **451**, 110841. https://doi.org/10.1016/j.jcp.2021.110841 (2022).

37. Ozbenli, E., Vedula, P., Vogiatzis, K. & Josyula, E. *Numerical Solution of Hypersonic Flows via Artificial Neural Networks* (American Institute of Aeronautics and Astronautics, 2020).

38. Colonna, G., Armenise, I., Bruno, D. & Capitelli, M. Reduction of state-to-state kinetics to macroscopic models in hypersonic flows. *J. Thermophys. Heat Transfer* **20**, 477–486. https://doi.org/10.2514/1.18377 (2006).

39. Campoli, L., Kustova, E. & Maltseva, P. Assessment of machine learning methods for state-to-state approach in nonequilibrium flow simulations. *Mathematics* **10**, 928. https://doi.org/10.3390/math10060928 (2022).

40. Scherding, C., Rigas, G., Sipp, D., Schmid, P. J. & Sayadi, T. Data-driven framework for input/output lookup tables reduction: With application to hypersonic flows in chemical non-equilibrium. https://doi.org/10.48550/ARXIV.2210.04269 (2022).

41. Zanardi, I., Venturi, S. & Panesi, M. Towards efficient simulations of non-equilibrium chemistry in hypersonic flows: a physics-informed neural network framework. In *AIAA SCITECH 2022 Forum* (American Institute of Aeronautics and Astronautics, 2022). https://doi.org/10.2514/6.2022-1639.

42. Panesi, M. & Lani, A. Collisional radiative coarse-grain model for ionization in air. *Phys. Fluids* **25**, 057101. https://doi.org/10.1063/1.4804388 (2013).

43. Munafò, A., Panesi, M. & Magin, T. E. Boltzmann rovibrational collisional coarse-grained model for internal energy excitation and dissociation in hypersonic flows. *Phys. Rev. E* **89**, 023001. https://doi.org/10.1103/PhysRevE.89.023001 (2014).

44. Munafò, A., Liu, Y. & Panesi, M. Modeling of dissociation and energy transfer in shock-heated nitrogen flows. *Phys. Fluids* **27**, 127101. https://doi.org/10.1063/1.4935929 (2015).

45. Liu, Y., Panesi, M., Sahai, A. & Vinokur, M. General multi-group macroscopic modeling for thermo-chemical non-equilibrium gas mixtures. *J. Chem. Phys.* **142**, 134109. https://doi.org/10.1063/1.4915926 (2015).

46. Sahai, A., Lopez, B., Johnston, C. O. & Panesi, M. Adaptive coarse graining method for energy transfer and dissociation kinetics of polyatomic species. *J. Chem. Phys.* **147**, 054107. https://doi.org/10.1063/1.4996654 (2017).

47. Kovachki, N. *et al. Neural Operator: Learning Maps Between Function Spaces.* https://doi.org/10.48550/ARXIV.2108.08481 (2020).

48. Kingma, D. P. & Welling, M. *Auto-encoding Variational Bayes.* https://doi.org/10.48550/ARXIV.1312.6114 (2013).

49. Coifman, R. R. *et al.* Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc. Natl. Acad. Sci. USA* **102**, 7426–7431. https://doi.org/10.1073/pnas.0500334102 (2005).

50. Schölkopf, B., Smola, A. & Müller, K.-R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **10**, 1299–1319. https://doi.org/10.1162/089976698300017467 (1998).

51. Oommen, V., Shukla, K., Goswami, S., Dingreville, R. & Karniadakis, G. E. Learning two-phase microstructure evolution using neural operators and autoencoder architectures. https://doi.org/10.48550/ARXIV.2204.07230 (2022).

52. Merchant, B. A. & Madura, J. D. A review of coarse-grained molecular dynamics techniques to access extended spatial and temporal scales in biomolecular simulations. In *Annual Reports in Computational Chemistry* (ed. Wheeler, R. A.) 67–87 (Elsevier, Amsterdam, 2011).

53. Kmiecik, S. *et al.* Coarse-grained protein models and their applications. *Chem. Rev.* **116**, 7898–7936. https://doi.org/10.1021/acs.chemrev.6b00163 (2016).

54. Boniecki, M. J. *et al.* SimRNA: A coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* **44**, e63–e63. https://doi.org/10.1093/nar/gkv1479 (2016).

55. Heritier, K. L., Jaffe, R. L., Laporta, V. & Panesi, M. Energy transfer models in nitrogen plasmas: Analysis of $N_2(^1\Sigma_g^+)$-$N(^4S_u)$-e interaction. *J. Chem. Phys.* **141**, 184302. https://doi.org/10.1063/1.4900508 (2014).

56. Esposito, F., Capitelli, M. & Gorse, C. Quasi-classical dynamics and vibrational kinetics of $N+N_2(v)$ system. *Chem. Phys.* **257**, 193–202. https://doi.org/10.1016/S0301-0104(00)00155-5 (2000).

57. Venturi, S., Sharma Priyadarshini, M., Lopez, B. & Panesi, M. Data-inspired and physics-driven model reduction for dissociation: Application to the O_2+O system. *J. Phys. Chem. A* **124**, 8359–8372. https://doi.org/10.1021/acs.jpca.0c04516 (2020).

58. Sharma Priyadarshini, M., Liu, Y. & Panesi, M. Coarse-grained modeling of thermochemical nonequilibrium using the multi-group maximum entropy quadratic formulation. *Phys. Rev. E* **101**, 013307. https://doi.org/10.1103/PhysRevE.101.013307 (2020).

59. Jagtap, A. D. & Karniadakis, G. E. Extended physics-informed neural networks (XPINNs): A generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations. *Commun. Comput. Phys.* **28**, 2002–2041. https://doi.org/10.4208/cicp.OA-2020-0164 (2020).

60. Bar, L. & Sochen, N. Unsupervised Deep Learning Algorithm for PDE-based Forward and Inverse Problems. https://doi.org/10.48550/ARXIV.1904.05417 (2019).

61. Bhatnagar, S., Afshar, Y., Pan, S., Duraisamy, K. & Kaushik, S. Prediction of aerodynamic flow fields using convolutional neural networks. *Comput. Mech.* **64**, 525–545. https://doi.org/10.1007/s00466-019-01740-0 (2019).

62. Zhu, Y. & Zabaras, N. Bayesian deep convolutional encoder-decoder networks for surrogate modeling and uncertainty quantification. *J. Comput. Phys.* **366**, 415–447. https://doi.org/10.1016/j.jcp.2018.04.018 (2018).

63. Sirignano, J. & Spiliopoulos, K. DGM: A deep learning algorithm for solving partial differential equations. *J. Comput. Phys.* **375**, 1339–1364. https://doi.org/10.1016/j.jcp.2018.08.029 (2018).

64. Duvall, J., Duraisamy, K. & Pan, S. Discretization-independent surrogate modeling over complex geometries using hypernetworks and implicit representations. https://doi.org/10.48550/ARXIV.2109.07018 (2021).

65. Gao, H., Sun, L. & Wang, J.-X. PhyGeoNet: Physics-informed geometry-adaptive convolutional neural networks for solving parameterized steady-state PDEs on irregular domain. *J. Comput. Phys.* **428**, 110079. https://doi.org/10.1016/j.jcp.2020.110079 (2021).

66. Kissas, G. *et al.* Learning Operators with Coupled Attention. https://doi.org/10.48550/ARXIV.2201.01032 (2022).

67. Lu, L. *et al.* A comprehensive and fair comparison of two neural operators (with practical extensions) based on FAIR data. *Comput. Methods Appl. Mech. Eng.* **393**, 114778. https://doi.org/10.1016/j.cma.2022.114778 (2022).

68. Li, Z. *et al.* Neural Operator: Graph Kernel Network for Partial Differential Equations. https://doi.org/10.48550/ARXIV.2003.03485 (2020).

69. Li, Z. *et al.* Fourier Neural Operator for Parametric Partial Differential Equations, https://doi.org/10.48550/ARXIV.2010.08895 (2020).

70. You, H., Yu, Y., D'Elia, M., Gao, T. & Silling, S. Nonlocal kernel network (NKN): A stable and resolution-independent deep neural network. *J. Comput. Phys.* **469**, 111536. https://doi.org/10.1016/j.jcp.2022.111536 (2022).

71. Chen, T. & Chen, H. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Trans. Neural Netw.* **6**, 911–917. https://doi.org/10.1109/72.392253 (1995).

72. Lu, L., Jin, P., Pang, G., Zhang, Z. & Karniadakis, G. E. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat. Mach. Intell.* **3**, 218–229. https://doi.org/10.1038/s42256-021-00302-5 (2021).

73. Ranade, R., Gitushi, K. & Echekki, T. Generalized Joint Probability Density Function Formulation in Turbulent Combustion using DeepONet. https://doi.org/10.48550/ARXIV.2104.01996 (2021).

74. Sharma Priyadarshini, M., Venturi, S., Zanardi, I. & Panesi, M. Efficient Quasi-Classical Trajectory Calculations by means of Neural Operator Architectures. https://doi.org/10.26434/chemrxiv-2022-fs3rv (2022).

75. Mao, Z., Lu, L., Marxen, O., Zaki, T. A. & Karniadakis, G. E. DeepM &Mnet for hypersonics: Predicting the coupled flow and finite-rate chemistry behind a normal shock using neural-network approximation of operators. *J. Comput. Phys.* **447**, 110698. https://doi.org/10.1016/j.jcp.2021.110698 (2021).

76. Raissi, M., Perdikaris, P. & Karniadakis, G. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045 (2019).

77. Karniadakis, G. E. *et al.* Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440. https://doi.org/10.1038/s42254-021-00314-5 (2021).

78. Shukla, K., Jagtap, A. D. & Karniadakis, G. E. Parallel physics-informed neural networks via domain decomposition. *J. Comput. Phys.* **447**, 110683. https://doi.org/10.1016/j.jcp.2021.110683 (2021).

79. Wang, J.-X., Wu, J.-L. & Xiao, H. Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data. *Phys. Rev. Fluids* **2**, 034603. https://doi.org/10.1103/PhysRevFluids.2.034603 (2017).

80. Mao, Z., Jagtap, A. D. & Karniadakis, G. E. Physics-informed neural networks for high-speed flows. *Comput. Methods Appl. Mech. Eng.* **360**, 112789. https://doi.org/10.1016/j.cma.2019.112789 (2020).

81. Jagtap, A. D., Kharazmi, E. & Karniadakis, G. E. Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems. *Comput. Methods Appl. Mech. Eng.* **365**, 113028. https://doi.org/10.1016/j.cma.2020.113028 (2020).

82. Wang, S., Wang, H. & Perdikaris, P. Learning the solution operator of parametric partial differential equations with physics-informed DeepONets. *Sci. Adv.* **7**, 8605. https://doi.org/10.1126/sciadv.abi8605 (2021).

83. Wang, S. & Perdikaris, P. Long-time integration of parametric evolution equations with physics-informed DeepONets. https://doi.org/10.48550/ARXIV.2106.05384 (2021).

84. Wang, S., Wang, H. & Perdikaris, P. Improved architectures and training algorithms for deep operator networks. *J. Sci. Comput.* **92**, 35. https://doi.org/10.1007/s10915-022-01881-0 (2022).

85. Goswami, S., Bora, A., Yu, Y. & Karniadakis, G. E. Physics-Informed Deep Neural Operator Networks. https://doi.org/10.48550/ARXIV.2207.05748 (2022).

86. Liu, Y., Kutz, J. N. & Brunton, S. L. Hierarchical deep learning of multiscale differential equation time-steppers. *Philos. Trans. R. Soc. A* **380**, 20210200. https://doi.org/10.1098/rsta.2021.0200 (2022).

87. Migus, L., Yin, Y., Mazari, J. A. & Gallinari, P. Multi-scale Physical Representations for Approximating PDE Solutions with Graph Neural Operators. https://doi.org/10.48550/ARXIV.2206.14687 (2022).

88. Liu, X., Xu, B. & Zhang, L. HT-Net: Hierarchical Transformer based Operator Learning Model for Multiscale PDEs. https://doi.org/10.48550/ARXIV.2210.10890 (2022).

89. Liu, L. & Cai, W. Multiscale DeepONet for Nonlinear Operators in Oscillatory Function Spaces for Building Seismic Wave Responses. https://doi.org/10.48550/ARXIV.2111.04860 (2021).

90. Lin, C. *et al.* Operator learning for predicting multiscale bubble growth dynamics. *J. Chem. Phys.* **154**, 104118. https://doi.org/10.1063/5.0041203 (2021).

91. Lütjens, B., Crawford, C. H., Watson, C. D., Hill, C. & Newman, D. Multiscale Neural Operator: Learning Fast and Grid-independent PDE Solvers. https://doi.org/10.48550/ARXIV.2207.11417 (2022).

92. Jaysaval, P., Shantsev, D. V., de la Kethulle de Ryhove, S. & Bratteland, T. Fully anisotropic 3-D EM modelling on a Lebedev grid with a multigrid pre-conditioner. *Geophys. J. Int.* **207**, 1554–1572. https://doi.org/10.1093/gji/ggw352 (2016).

93. Liu, Z., Cai, W. & Xu, Z.-Q.J. Multi-scale deep neural network (MscaleDNN) for solving Poisson-Boltzmann equation in complex domains. *Commun. Comput. Phys.* **28**, 1970–2001. https://doi.org/10.4208/cicp.OA-2020-0179 (2020).

94. Thakur, A., Tripura, T. & Chakraborty, S. Multi-fidelity wavelet neural operator with application to uncertainty quantification. https://doi.org/10.48550/ARXIV.2208.05606 (2022).

95. Howard, A. A., Perego, M., Karniadakis, G. E. & Stinis, P. Multifidelity Deep Operator Networks. https://doi.org/10.48550/ARXIV.2204.09157 (2022).

96. Lu, L., Pestourie, R., Johnson, S. G. & Romano, G. Multifidelity deep neural operators for efficient learning of partial differential equations with application to fast inverse design of nanoscale heat transport. *Phys. Rev. Res.* **4**, 023210. https://doi.org/10.1103/PhysRevResearch.4.023210 (2022).

97. Venturi, S. & Casey, T. SVD perspectives for augmenting DeepONet flexibility and interpretability. *Comput. Methods Appl. Mech. Eng.* **403**, 115718. https://doi.org/10.1016/j.cma.2022.115718 (2023).

98. Munafò, A. *et al.* QCT-based vibrational collisional models applied to nonequilibrium nozzle flows. *Eur. Phys. J. D* **66**, 188. https://doi.org/10.1140/epjd/e2012-30079-3 (2012).

99. Macdonald, R. L., Jaffe, R. L., Schwenke, D. W. & Panesi, M. Construction of a coarse-grain quasi-classical trajectory method. I. Theory and application to N_2-N_2 system. *J. Chem. Phys.* **148**, 054309. https://doi.org/10.1063/1.5011331 (2018).

100. Macdonald, R. L., Grover, M. S., Schwartzentruber, T. E. & Panesi, M. Construction of a coarse-grain quasi-classical trajectory method. II. Comparison against the direct molecular simulation method. *J. Chem. Phys.* **148**, 054310. https://doi.org/10.1063/1.5011332 (2018).

101. Park, C. *Nonequilibrium Hypersonic Aerothermodynamics* (Wiley, 1990).

102. Park, C., Jaffe, R. L. & Partridge, H. Chemical-kinetic parameters of hyperbolic earth entry. *J. Thermophys. Heat Transfer* **15**, 76–90. https://doi.org/10.2514/2.6582 (2001).

103. Munafò, A., Venturi, S., Sharma Priyadarshini, M. & Panesi, M. Reduced-Order Modeling for Non-equilibrium Air Flows. In *AIAA Scitech 2020 Forum* (American Institute of Aeronautics and Astronautics, 2020). https://doi.org/10.2514/6.2020-1226.

104. Zanardi, I., Venturi, S. & Panesi, M. Towards Efficient Simulations of Non-Equilibrium Chemistry in Hypersonic Flows: Application of Physics-Informed DeepONet to Shock-Heated Flow Scenarios. In *AIAA SCITECH 2023 Forum* (American Institute of Aeronautics and Astronautics, 2023). https://doi.org/10.2514/6.2023-1202.

105. van Leer, B. Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method. *J. Comput. Phys.* **32**, 101–136. https://doi.org/10.1016/0021-9991(79)90145-1 (1979).

106. Hirsch, C. *Numerical Computation of Internal and External Flows* 1st edn. (Elsevier, 2007).

107. Strang, G. On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.* **5**, 506–517. https://doi.org/10.1137/0705041 (1968).

108. Knio, O. M., Najm, H. N. & Wyckoff, P. S. A semi-implicit numerical scheme for reacting flow. *J. Comput. Phys.* **154**, 428–467. https://doi.org/10.1006/jcph.1999.6322 (1999).

109. Singer, M. A., Pope, S. B. & Najm, H. N. Operator-splitting with ISAT to model reacting flow with detailed chemistry. *Combust. Theory Model.* **10**, 199–217. https://doi.org/10.1080/13647830500307501 (2006).

110. Ren, Z., Xu, C., Lu, T. & Singer, M. A. Dynamic adaptive chemistry with operator splitting schemes for reactive flow simulations. *J. Comput. Phys.* **263**, 19–36. https://doi.org/10.1016/j.jcp.2014.01.016 (2014).

111. Wu, H., Ma, P. C. & Ihme, M. Efficient time-stepping techniques for simulating turbulent reactive flows with stiff chemistry. *Comput. Phys. Commun.* **243**, 81–96. https://doi.org/10.1016/j.cpc.2019.04.016 (2019).

112. Berkooz, G., Holmes, P. & Lumley, J. L. The proper orthogonal decomposition in the analysis of turbulent flows. *Annu. Rev. Fluid Mech.* **25**, 539–575. https://doi.org/10.1146/annurev.fl.25.010193.002543 (1993).

113. Li, X., Grandvalet, Y. & Davoine, F. A baseline regularization scheme for transfer learning with convolutional neural networks. *Pattern Recogn.* **98**, 107049. https://doi.org/10.1016/j.patcog.2019.107049 (2020).

114. Wang, S., Teng, Y. & Perdikaris, P. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. arXiv:2001.04536 (2020).

115. Abadi, M. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. https://doi.org/10.48550/ARXIV.1603.04467 (2016).

116. Díaz-Francés, E. & Rubio, F. J. On the existence of a normal approximation to the distribution of the ratio of two independent normal random variables. *Stat. Pap.* **54**, 309–323. https://doi.org/10.1007/s00362-012-0429-2 (2013).

Acknowledgements

Author contributions

I.Z. and S.V. conceptualization and methodology; I.Z. software; M.P. supervision and funding acquisition. All authors participated in the data analysis, paper writing, and manuscript revision.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-41039-y.

Correspondence and requests for materials should be addressed to M.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.