



HAL
open science

Accelerating hypersonic reentry simulations using deep learning-based hybridization (with guarantees)

Paul Novello, Gaël Poëtte, David Lugato, Simon Peluchon, Pietro Marco Congedo

► **To cite this version:**

Paul Novello, Gaël Poëtte, David Lugato, Simon Peluchon, Pietro Marco Congedo. Accelerating hypersonic reentry simulations using deep learning-based hybridization (with guarantees). *Journal of Computational Physics*, 2024, 498, pp.112700. 10.1016/j.jcp.2023.112700 . hal-04323551v2

HAL Id: hal-04323551

<https://inria.hal.science/hal-04323551v2>

Submitted on 5 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Accelerating hypersonic reentry simulations using deep learning-based hybridization (with guarantees)

Paul Novello^{a,b,c,*}, Gaël Poëtte^a, David Lugato^a, Simon Peluchon^a, Pietro Marco Congedo^{b,c}

^a*CESTA, CEA, Le Barp, France*

^b*INRIA Saclay, Palaiseau, France*

^c*CMAP, Ecole Polytechnique, Palaiseau, France*

Abstract

In this paper, we are interested in the acceleration of numerical simulations. We focus on a hypersonic planetary reentry problem whose simulation involves coupling fluid dynamics and chemical reactions. Simulating chemical reactions takes most of the computational time but, on the other hand, cannot be avoided to obtain accurate predictions. We face a trade-off between cost-efficiency and accuracy: the numerical scheme has to be sufficiently efficient to be used in an operational context but accurate enough to predict the phenomenon faithfully. To tackle this trade-off, we design a hybrid numerical scheme coupling a traditional fluid dynamic solver with a neural network approximating the chemical reactions. We rely on their power in terms of accuracy and dimension reduction when applied in a big data context and on their efficiency stemming from their matrix-vector structure to achieve important acceleration factors ($\times 10$ to $\times 18.6$). This paper aims to explain how we design such cost-effective hybrid numerical schemes in practice. Above all, we describe methodologies to ensure accuracy guarantees, allowing us to go beyond traditional surrogate modeling and to use these schemes as references.

*Corresponding author. Now at IRT Saint Exupery and Artificial and Natural Intelligence Toulouse Institute (ANITI), DEEL team

Email addresses: paul.novello@outlook.fr (Paul Novello), gael.poette@cea.fr (Gaël Poëtte), david.lugato@cea.fr (David Lugato), simon.peluchon@cea.fr (Simon Peluchon), pietro.congedo@inria.fr (Pietro Marco Congedo)

Keywords: Reentry, Chemical reactions, Machine Learning, Deep neural networks

1 Introduction

In this paper, we are interested in the acceleration of numerical simulations. More specifically, we focus on a hypersonic planetary reentry problem [1, 2, 3, 4, 5, 6]: during a high-speed planetary atmosphere reentry, a shock wave forms ahead of the entering object leading to an increase of temperature and pressure of the fluid across the shock (see figure 1 for a general sketch).

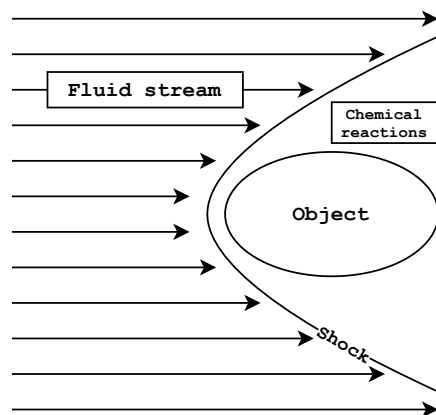
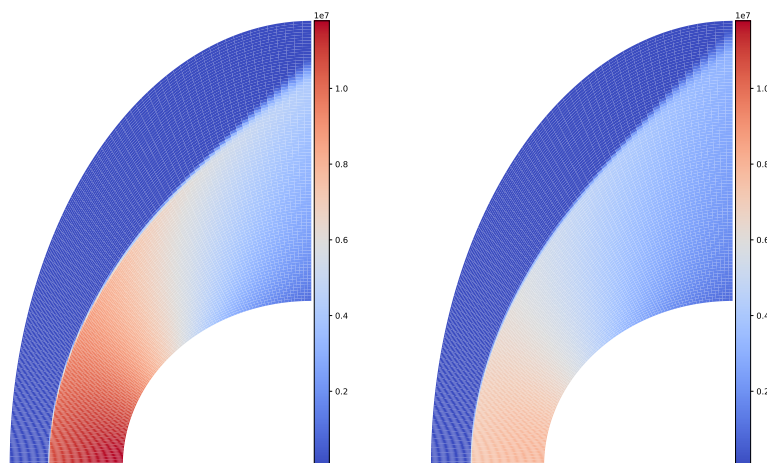


Figure 1: General sketch of a reentry problem: the object entering the atmosphere is subject to a high-speed fluid stream creating a shock ahead of the object. The temperature and pressure rise between the shock and the object leading to chemical dissociation reactions. An accurate prediction of the flow field is mandatory in order to design efficient protections and ensure the integrity of the object.

This increase generates chemical dissociation reactions within the shocked fluid and changes its composition. The composition strongly affects the thermodynamic quantities between the shock and the object [4, 5, 7]: in other words, we here face a strong coupling problem between gas dynamics and reactions. Several other physical phenomenon are certainly also important, such as turbulence [8, 9], ablation [4, 5], pyrolysis [4, 6] etc. Although it may be necessary to

simulate the phenomenon accurately, the coupling of these different physics can make the computations prohibitively intensive. In the following, we focus on the coupling between compressible gas dynamics and the physics of reacting fluids. Despite being simpler than simulating the full physics, this test case is sufficiently challenging to emphasize the computational problems arising when coupling different physics. Let us now give an idea of the difficulty of such simulation by considering a simple motivating example: figure 2 presents two pressure fields (the same scale is used on both pictures) of air around a sphere entering (a simplified) earth's atmosphere at a normal velocity of $4930.83 \text{ m.s.}^{-1}$ (Mach 16). The complete details of the simulations are given later on in section 5, let us here focus on the results. The two simulations of figures 2 (a) and (b) only differ



(a) Without chemical reactions

run-time: 81 s.

(b) With chemical reactions

run-time: 4090 s.

Figure 2: Pressure fields (in Pa) for a sphere entering earth's atmosphere with and without taking into account chemical reactions (see details in section 5).

from the fact that on figure 2 (a), chemical reactions are neglected whereas on figure 2 (b), reaction $N + O \rightleftharpoons NO$ is taken into account. The pressure on figure

2 (a) is higher than on figure 2 (b) and the position of the shock is different. By neglecting the chemical reactions, the pressure on the boundary of the object is overestimated in this case, which can lead to sub-optimal object conception. As an example, designing protections for such an object based on the simulations that do not take chemical reactions into account would lead to heavier designs, with propellers consuming more fuel. However, simulating chemical reactions (even in this simplified case where we only consider $N + O \rightleftharpoons NO$) comes with a cost: the simulation of figure 2 (b) is about $\times 50$ more costly than the one of figure 2 (a). Simulating chemical equilibrium takes most of the computational time of the scheme, but on the other hand, Figure 2 shows that we cannot avoid simulating it to obtain accurate predictions. There are consequently high stakes in accelerating the computations related to the chemistry of the problem (let us keep in mind that to design a new reentry object, more than one computation is needed, for example, to propagate uncertainties [10, 11, 12, 13] or in order to optimize designs under uncertainties [14, 15]). Some uncertain situations are considered in section 5.

The previous example helps to understand our objective: we want to be able to perform accurate predictions for reentry problems by taking into account chemical reactions (at equilibrium in this paper), together with comprehensive parametric studies. Hence, we need to accelerate our simulations. To achieve this goal, in this paper, we study the design of a neural network-based hybrid scheme.

Neural networks have already been intensively used in computational physics for surrogate modeling, such as, for instance, in molecular simulations [16, 17, 18], biological simulations [19, 20], fluid dynamic simulations [21, 22, 23, 8, 24]. Even though supervised approaches are often followed, some emerging techniques are becoming increasingly used, such as physics-informed deep learning [25, 26, 27, 28, 29], deep learning aided simulations [30, 31, 32] or hybridization [33, 34, 35, 36, 37, 38].

In this work, we study the last approach and leverage the coupling structure of the numerical scheme to replace the chemical reaction solver with a neural

network. The advantages of this approach are twofold. First, the definition of neural networks allows us to easily vectorize their calls on array-like data structures such as meshes. It is a solid computational advantage compared to the original numerical scheme, which has to call the chemical reaction solver in each mesh cell. Second, the chemical reaction solver is not costly when executed as a standalone application. Consequently, it is possible to build a large training database and substantially improve the neural network’s accuracy without additional cost during inference. Besides, the neural network only approximates the chemical reactions of the problem, thereby being applicable, once trained, to any simulation which involves the same chemical reactions.

Still, these gains are not necessarily easily earned as the neural network built to approximate the solution of the system of chemical reactions must be both accurate and cost-effective in order to accelerate reentry computations. The *aim of this paper is to explain how we build such cost-effective neural networks in practice and how we progressively test them before embedding them within the numerical scheme. Above all, how we use them so that guarantees of accuracy are ensured with the hybrid numerical scheme.* As an important point, the methodologies introduced in this paper are not specific to the studied reentry test case: they can be applied to any numerical scheme involving a strong coupling between different physics.

The paper is organized as follows. Section 2 describes the physical model we consider in this paper. The model is relatively simple but representative of the difficulties encountered in real-life applications. Furthermore, its simplicity allows for obtaining reference solutions in practical times. Section 3 focuses on Machine Learning (ML) and explains why we choose to consider (deep) neural networks in order to accelerate the numerical schemes rather than more classical ML models such as polynomial regression or Kriging. We also provide a methodology to anticipate the potential gains in terms of acceleration before plugging any model within the numerical scheme. Then, we emphasize the trade-off between accuracy and describe how to construct neural networks that are both accurate and cost-effective based on a comprehensive goal-oriented sensitivity analysis

of their hyperparameters. Section 5 is devoted to numerical results and gives a practical description of the methodology we apply in this paper. We assess the hybrid scheme and find that the obtained predictions are both quantitatively and qualitatively very promising while being 18.7 times faster. Above all, we explain how we *obtain guarantees with the hybrid scheme*. Finally, section 6 is a concluding section: in particular, it summarizes the conditions under which the described methodology can be applied to any other physics/simulation schemes.

2 Coupling compressible gas dynamics with chemical reactions at equilibrium

In this section, we describe the set of partial differential equations (PDEs) solved in order to produce the results of figure 2 together with its resolution strategy. It corresponds to the coupling of compressible gas dynamics with chemical reactions at equilibrium. The fluid dynamic is modeled thanks to the Euler equations in 2D spatial dimension. It describes the behavior of non-viscous compressible gas. In reentry problems, Navier-Stokes' equations are generally considered a finer model, but Euler's system is enough for the purpose of this paper, and the following material can easily be applied to any other fluid model without more difficulties. The Euler system in 2D spatial dimension solved in the spatial domain \mathcal{D} is given by

$$\begin{cases} \partial_t U(x, y, t) + \nabla \cdot F(U(x, y, t)) = 0, & \forall (x, y) \in \mathcal{D}, \\ U(x, y, t) = b(x, y, t), & \forall (x, y) \in \partial\mathcal{D}. \end{cases} \quad (1)$$

In the above equation, b corresponds to the boundary conditions: typically, it corresponds to an incoming flux boundary condition everywhere on $\partial\mathcal{D}$ except on the boundary of the object where it corresponds to a no-slip one (this is what has been used in the results of figure 2).

The details of this system and the involved physical variables are left in **Appendix C**. Here, it is enough to emphasize that the system is not closed and needs an additional equation to close the system. A first modeling choice can

be made by using the perfect gas equation. However, this hypothesis is coarse (see figure 2 and the related comments). In a more general case, the pressure p can be accurately computed by simulating the chemical equilibrium between the species produced by the N_r chemical reactions that occur during the dynamic. It can be achieved through the minimization of the Gibbs free energy, also detailed in **Appendix C**.

In our numerical scheme, the Euler counterpart (8) of the splitting is solved using the numerical scheme presented in [4, 39, 40]: it is a Lagrange+remap scheme. The main idea of the splitting is to separate the acoustic and dissipative phenomenon from the transport one. In Low Mach computations (as we aim at treating liquid ablation in further work [4, 5]), an implicit treatment of the Lagrangian step is done since the fast acoustic waves would induce very small time steps otherwise. The remapping step is explicit and performed with a finite volume scheme. The overall scheme resulting from this splitting operator strategy is very robust, conservative, and preserves contact discontinuities.

The Gibbs free energy minimization counterpart (9) is solved thanks to the library Mutation++ [7]. This library provides accurate and efficient computation of physicochemical properties associated with partially ionized gases in various degrees of thermal nonequilibrium. The users can compute thermodynamic and transport properties, multiphase linearly-constrained equilibria, chemical production rates, energy transfer rates, and gas-surface interactions (i.e. Mutation++ is also a promising tool for other test cases than atmospheric reentry). The framework is based on an object-oriented design in C++, allowing users to plug and play various models, algorithms, and data as necessary. Mutation++ is available open-source under the GNU Lesser General Public License v3.0.

Mutation++ allows performing accurate reference solutions for our reentry problem but remains costly for our needs. Remember the example of figure 2: we only consider the reaction $N + O \rightleftharpoons NO$. In that case, there are only three species, N , O , and NO , and the computations are already computationally intensive. Those three species are far from being sufficient to characterize Earth's atmosphere. A relevant set of species for Earth needs at least $n_s = 18$ species

and is even more computationally intensive (see section 3.3). Besides, it does not even take into account the species ejected from the ablating surface of the object [4, 5, 6].

Let us present a sketch of the numerical scheme: algorithm 1 presents the main steps of the resolution. We insist on the fact that algorithm 1 certainly stands for a coarse description of the numerical scheme. But it is enough in order to present the methodology applied in this paper. It also testifies to the simplicity of application of the material of this paper.

```

1 # general initialization (mesh, quantities on mesh etc.)
2 initialise_guess_vector_of_unknowns_on_mesh( $U^0, \mathbf{x}^0$ )
3 while convergence_criterion_not_satisfied do
4    $U^{n+\frac{1}{2}} = \text{solve\_Euler\_equations}(U^n, \mathbf{x}^n)$ 
5   for  $i \in \{1, \dots, N_{\mathcal{D}}\}$  do
6      $\mathbf{x}_i^{n+1}, U^{n+1} = \text{minimize\_Gibbs\_free\_energy\_with\_mutation}++(U_i^{n+\frac{1}{2}}, \mathbf{x}_i^n)$ 
7   end
8    $U^n \leftarrow U^{n+1}$ 
9    $\mathbf{x}^n \leftarrow \mathbf{x}^{n+1}$ 
10 end

```

Algorithm 1: Core of the reentry scheme.

First, in algorithm 1, a mesh has to be built and the different quantities must be initialized on this mesh. Of course, the closer to the stationary solution the initialization, the faster the resolution in terms of iterations. In practice, we rely on uniformly initialized quantities which are certainly far from the solution to the problem. All the information is condensed in function `initialise_guess_vector_of_unknowns_on_mesh` in line 2 of algorithm 1. Then comes the *while* loop: a convergence criterion must be chosen but it is not central in this paper so we choose not to describe it. While convergence is not fulfilled,

the scheme solves the Euler equations (function `solve_Euler_equations`) before feeding the updated field $U^{n+\frac{1}{2}}$ into the minimization of the Gibbs free energy. Note that Navier Stokes' system could be solved instead of the Euler one in this paper, and this would not change the methodology described in the next lines. Note also that the minimization must occur within each cell $i \in \{1, \dots, N_{\mathcal{D}}\}$ where $N_{\mathcal{D}}$ corresponds to the total number of cells. The minimization is made in the function `minimize_Gibbs_free_energy_with_mutation++`, which is nothing more than a call of `Mutation++`. It takes as inputs ρ , the density, ϵ , the mixture energy, and the mole fractions of the elements initially found in the fluid (i.e. the information contained in U). It outputs $\{\mathbf{x}_1, \dots, \mathbf{x}_{n_s}\}$ the mass fractions of the mixture of n_s chemical species but also additional quantities such as c , the speed of sound, C_p the heat at constant pressure, C_v , the heat at constant volume, p the pressure, and T the temperature after the equilibrium of the reactions is fulfilled. In a sense, the call to `Mutation++`, denoted by \mathbf{M}^{++} , can be summarized as a function of \mathbb{R}^{n_e+2} in \mathbb{R}^{n_s+5}

$$\mathbf{M}^{++} : \begin{pmatrix} \mathbf{x}_1 \\ \dots \\ \mathbf{x}_{n_e} \\ \rho \\ \epsilon \end{pmatrix} \in \mathbb{R}^{n_e+2} \longrightarrow \begin{pmatrix} \mathbf{x}_1 \\ \dots \\ \mathbf{x}_{n_s} \\ P \\ T \\ C_p \\ C_v \\ c \end{pmatrix}, \in \mathbb{R}^{n_s+5}, \quad (2)$$

with the outputs such that $\Delta G = 0$. As highlighted by the example of section 1, the minimization of the Gibbs free energy is necessary for model accuracy but very costly. We would like to build a surrogate model to replace the call to `Mutation++` by approximating \mathbf{M}^{++} , just as in algorithm 2, and hopefully accelerate the reentry scheme without impacting its accuracy.

```

1 # general initialization (mesh, quantities on mesh etc.)
2 initialise_guess_vector_of_unknowns_on_mesh( $U^0, \mathbf{x}^0$ )
3 while convergence_criterion_not_satisfied do
4    $U^{n+\frac{1}{2}} = \text{solve\_Euler\_equations}(U^n, \mathbf{x}^n)$ 
5    $\mathbf{x}^{n+1}, U^{n+1} = \text{call\_surrogate\_model}(U^{n+\frac{1}{2}}, \mathbf{x}^n)$ 
6    $U^n \leftarrow U^{n+1}$ 
7    $\mathbf{x}^n \leftarrow \mathbf{x}^{n+1}$ 
8 end

```

Algorithm 2: Core of the scheme with a call to a surrogate model of Mutation++.

Obviously, for efficiency, the surrogate model has to be well-chosen. The classical reflex at this stage would be to build and use some abacuses, as *offline* calls to Mutation++ can be made. This is classical for tabulated equations of state for example. Those abacuses can then be loaded into memory and interpolated during the simulation [41, 42]. However, this method becomes intractable when the input/output dimensions increases (here we have $d_{\text{in}} = n_e + 2 \gg 1$ or $d_{\text{out}} = n_s + 5 \gg 1$, see section 3.3) because the number of points N needed to obtain a fine interpolation increases exponentially fast with it, together with a complexity for the search in the database which strongly depends on N too. For this reason, in the following section, we study the possibility of building different surrogate models of Mutation++ from gathered data. In particular, we explain why we are interested in neural networks. As can be seen with algorithm 2, our methodology is *intrusive*, we need to modify a few lines of the numerical scheme as we are going to plug a neural network in it, hence the *hybrid* denomination.

3 Neural networks as approximators for hybridization

In the previous section, we formalized the problem of replacing Mutation++ with a surrogate model as an approximation problem. In section 3.1, we investigate the different types of potential surrogate models allowing us to reach our needs and explain why we consider neural networks. In section 3.3, we study

the capabilities of acceleration of neural networks on several benchmarks/atmospheres and verify their behaviors in terms of complexity with respect to the input and output dimensions (d_{in} and d_{out}). Neural networks are promising (see section 3.3), but their design has a strong impact on the final performances of the hybrid scheme: section 4 is devoted to explaining how we look for *accurate and cost-effective* neural networks.

3.1 Approximation problem

With the last paragraph of section 2, we formalised our problem as approximating a function of $X \in \mathbb{R}^{d_{\text{in}}} \rightarrow u(X) \in \mathbb{R}^{d_{\text{out}}}$ from N available data $(X_i, u(X_i))_{i \in \{1, \dots, N\}}$. Of course, for our application, u is nothing more than the call to \mathbf{M}^{++} . In other words, we face an *approximation theory* problem. In approximation theory (and in ML, which largely intersects with this field), it is classical to look for a parametric function

$$(X, \theta) \in \mathbb{R}^{d_{\text{in}}} \times \mathbb{R}^{d_{\theta}} \rightarrow u(X, \theta) \in \mathbb{R}^{d_{\text{out}}},$$

which has to be the closest possible to $u(X)$ in a certain metric L . This goal is achieved with the optimization of θ driven by the minimization of

$$J(\theta) = \int L(u(X), u(X, \theta)) d\mathcal{P}_X,$$

where $d\mathcal{P}_X$ is the measure of the input space. In practice, we do not have access to $J(\theta)$ because the measure $d\mathcal{P}_X$ is unknown. Hence, we approximate $J(\theta)$ with an experimental design $(X_i, w_i)_{i \in \{1, \dots, N\}}$, which is a discretisation¹ of $(X, d\mathcal{P}_X)$, as:

$$J(\theta) = \int L(u(X), u(X, \theta)) d\mathcal{P}_X \approx J_N(\theta) = \sum_{i=1}^N w_i L(u(X_i), u(X_i, \theta)), \quad (3)$$

where X is a random vector (of potentially correlated components) and $d\mathcal{P}_X$ is its probability measure.

¹In the sense that $\forall f \in L_2, \sum_{i=1}^N w_i f(X_i) \xrightarrow[N \rightarrow \infty]{L_2} \int f(X) d\mathcal{P}_X$.

3.2 The advantages of neural networks

There exist many different types of surrogate models. Amongst the most classical ones in numerical and uncertainty analysis, we can count² **Lagrange interpolation or collocation, Gaussian Process regression or Kriging, Polynomial regression and generalised Polynomial Chaos. A detailed comparison of these types of surrogate models can be found in Appendix D.**

In this work, we focus on neural networks. They are nonlinear approximators leading to a non-convex loss function J [43, 44]. They are based on the choices:

- $L(x, y)$ can be general (L_2 -norm, L_1 -norm, cross-entropy etc., see [43]).
- For a shallow neural network³, which is nonlinear with respect to $\theta = (\theta^1, \theta^2)$. The function $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ is called the *activation function*. In practice, it only has to be unbounded and non-constant [45, 46].
- Deep neural networks with L layers correspond to L compositions of the above expression: an L -layer feed forward neural network is defined recursively as $d_\theta = 2(P + 1) \times L$ parameters⁴ and rely on the recursive formula:

$$\begin{aligned} u^0(X, \theta) &= X, \\ u^l(X, \theta) &= \sum_{k=0}^P \theta_k^l \sigma(\theta^{l-1} u^{l-1}(X, \theta)), \forall l \in \{1, \dots, L-1\}, \\ u(X, \theta) &= \theta^{L-1} u^{L-1}(X, \theta), \end{aligned}$$

where σ is applied element-wise. Note that in this formula, the number of neurons per layers $P = \text{n_units}$ is considered constant but may change with $l \in \{1, \dots, L = \text{n_layers}\}$.

- The number of parameters d_θ is not constrained by N .

²The lists of references in the following points are not exhaustive and have been chosen because of the proximity of their application domain.

³Bias are taken into account with this notation, $u(X, \theta) = \sum_{k=0}^P \theta_k^1 \sigma(\theta^2 \cdot X)$, see [44].

⁴Or $d_\theta = 3(P + 1) \times L$ parameters if bias is considered [44].

The convergence of the approximation is guaranteed under the hypothesis of Hornik’s theorem [45] (or Barron’s one [46]) as the number of neurons grows; or under the hypothesis of [47] for deep neural networks as the number of layers L grows. In each of the previous theoretical results, the existence of a set of parameters ensuring convergence is guaranteed. Yet, a difference remains with the other approaches in that the loss function J (and J_N) may have a lot of local minima [48], and we consequently have to find the set of parameters ensuring convergence.

Neural networks share many similarities with classical surrogate models. However, they differ on several points.

- For *classical surrogate modeling*, some pre-processing is needed in order to take into account correlated input variables [49, 50, 51]. It can be a problem since training data may come from previous simulations and uncontrolled (possibly correlated) distributions.
- In *classical surrogate modeling*, when the space of output is of size d_{out} , d_{out} surrogate models must be built. This is problematic when the training and inference time of the model increase. For *neural networks*, only the last linear layer depends on d_{out} .
- *Classical surrogate models* are very sensitive to the curse of dimensionality as the number of parameters d_θ may grow exponentially fast with both P and d_{in} . As a consequence, these models are not well suited to high-dimensional problems. *Neural networks* complexity scales linearly with d_{in} , and is famous for its recent breakthrough on high dimensional test cases like image or text processing. In addition, the complexity of one prediction is independent of N , so we can leverage very large databases.
- The implementation of the inference of *neural networks* boils down to a succession of matrix-vector products, which can be easily vectorized. As a result, they can process array-like data structures very efficiently, which makes them perfectly suited to computations on meshes. *Classical*

surrogate models do not offer such implementation properties.

In the next section, we introduce a profiling experiment that gives an idea about the potential computational gains of neural network-based hybridization.

3.3 Assessing the capabilities of neural networks in terms of accelerations

To test the potential computational gains of hybridization, we construct a simple benchmarking code that compares the run-times of neural networks and Mutation++ for a given number of input points. In our case, this number, $N_{\mathcal{D}}$, is the size of the simulation mesh. This code (C++) only corresponds to the extraction of the *for* loop of algorithm 1 for Mutation++ and of the `call_surrogate_model` of algorithm 2. The neural networks are implemented within this code *via* the Tensorflow C API and a wrapper, CppFlow⁵.

In the next studies, the neural networks have `n_layers` = 5 hidden layers, and the number of neurons in each layer `n_units` is constant per layer and chosen as a parameter of the code. In other words, we only study the influence of hyperparameter `n_units`, the number of neurons per layer, even if many others exist (dropout rate, the different architectures, and their parameters, the different optimizers and their parameters, etc. see [43]) together with the influence of operational conditions $N_{\mathcal{D}} \in \{10, 10^2, 10^3, 10^4, 10^5, 10^6\}$, the numbers of cells of the grid and $n_s \in \{3, 18, 38, 64\}$, the number of species. In order to study the influence of n_s , we consider 4 different atmospheres (arbitrarily constructed out of the elements found in each atmosphere to obtain increasingly complex test cases):

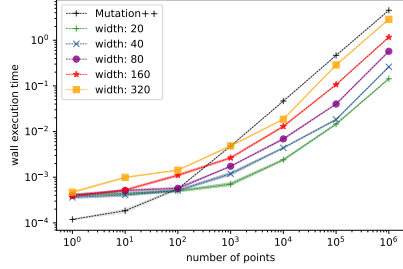
- Toy problem: the toy problem corresponds to the conditions mentioned in section 1 with 2 elements and 3 species N, O, NO .
- Earth: it corresponds to the case where the fluid is air with 2 elements but where 18 species are considered: $N, NO, O, N_2, O_2, e^-, N^+, O^+, N_2^+, O_2^+, NO^+, NO_3^-, NO_3, NO_2, O_3, NO_2^-, O^-, O_2^-$, i.e. $n_s = 18$ species.

⁵<https://github.com/serizba/cppflow>

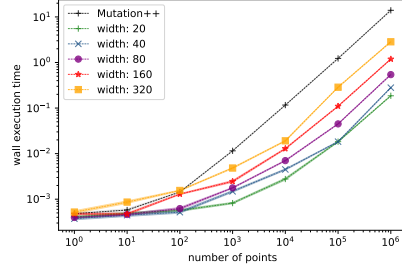
- Jupiter: Jupiter’s atmosphere corresponds to considering 4 different elements: O , C , He and H , and 36 corresponding species O , O_2 , C , e^- , C^+ , O^+ , O_2^+ , CO^+ , C_2 , CO , CO_2 , H , CH , OH , H^+ , CH^+ , OH^+ , H_2O^+ , H_2O , H_2 , H_2^+ , CH_4 , He , He^+ , CH_2 , H^- , HCO^+ , CH_3 , C_2H , HCO , C^- , O^- , OH^- , C_2^- , O_2^- , C_2O .
- Cloudy Jupiter: Jupiter’s clouds are made of water but also of ammonium hydrosulfide and ammonia. Hence, N and S are added as input elements, and 64 species are considered: O , O_2 , C , e^- , C^+ , O^+ , O_2^+ , CO^+ , C_2 , CO , CO_2 , H , CH , OH , H^+ , CH^+ , OH^+ , H_2O^+ , H_2O , H_2 , H_2^+ , CH_4 , He , He^+ , CH_2 , H^- , HCO^+ , CH_3 , C_2H , HCO , C^- , O^- , OH^- , C_2^- , O_2^- , C_2O , N , NO , O , N_2 , O_2 , e^- , N^+ , O^+ , N_2^+ , O_2^+ , NO^+ , NH_3 , NO_3^- , NO_3 , NH_4^+ , NO_2 , N_2H_2 , H^- , HNO , O_3 , HNO_2 , HNO_3 , NO_2^- , NH_2 , S , S^+ , S^- , CS , CS_2 , COS , $CNCOCN$, CN , CN^+ , CN^- .

The cloudy Jupiter scenario may appear far fetched but considering it allows progressively increasing the number of species n_s from 3 (toy), 18 (Earth), 38 (cloudy Earth) to 64 (cloudy Jupiter) and study the impact of the number of outputs of the neural networks. Moreover, with 64 species, the *cloudy Jupiter* scenario is close to some operational conditions in which the species of the atmosphere are mixed with some from the ablating surface of the entering object.

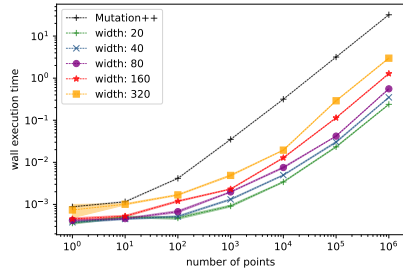
On Figure 3, we plot the run-times of Mutation++ and neural networks of different widths $n_units \in \{20, 40, 80, 160, 320\}$ for each atmospheres with respect to the number of grid points $N_{\mathcal{D}}$. Note that 10 repetitions are carried out for each curve to check for their stability (standard deviations are plotted, but they are barely visible due to the low value of the variance). First, for $N_{\mathcal{D}} = 10^0 = 1$, we can compare the *sequential* run-times of Mutation++ and of the neural networks for the different atmospheres: the neural networks are not always faster than the calls to Mutation++. However, within the scheme, Mutation++ is called sequentially, while the neural networks are executed on the whole array of cells in a batch fashion. Figure 3 illustrates how neural networks



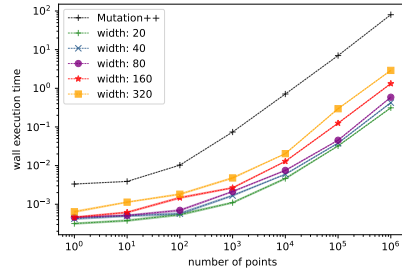
(a) Toy test case



(b) Earth test case



(c) Jupiter test case



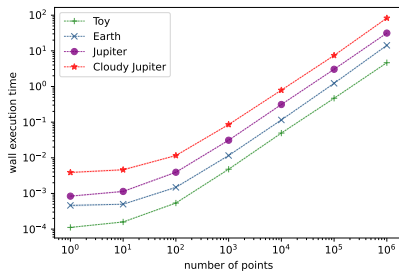
(d) Cloudy Jupiter test case

Figure 3: Execution time of Mutation++ vs neural networks of different widths for each test case with respect to the number of input points (with log axes).

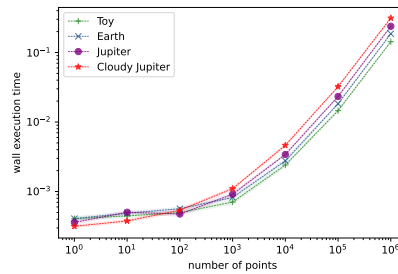
exploit vectorial acceleration: as the number of cells $N_{\mathcal{D}}$ increases, their run-times become competitive with respect to Mutation++. In (a) it happens at $N_{\mathcal{D}} > 10^3$. Of course, for some test cases, neural networks are competitive even sequentially (in (c) and (d)). In a general manner, for $N_{\mathcal{D}}$ between 10^4 and 10^5 - which turns out to be the orders of cells numbers per mesh block (4000 and 12000 in our experiment of section 5), neural networks performs much better than Mutation++ in terms of computational time.

In Figure 4 we visualize the run-times of neural networks and Mutation++ run-times with respect to n_s . We can see that the higher the number of species, the higher the run-time of Mutation++. This effect is less marked for the neural networks. For conciseness, we only present the plots for `n_units` of 20 and 80,

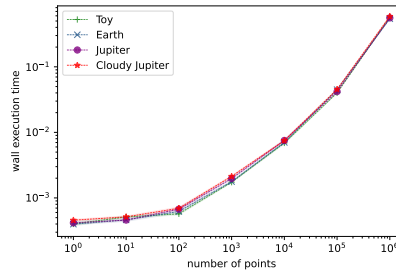
which show that even for low width, the effect of n_s on the run-time is limited, and above a width of 80, it can be hardly distinguished. We recover experimentally the fact that the computational complexity of the operations between the hidden layers is insensitive to the dimensions of the problem. The input and output dimensions can be increased with a limited impact on the run-time, which illustrates how neural networks mitigate the curse of dimensionality.



(a) Test cases with Mutation++



(b) Test cases with a network of width 20



(c) Test cases with a network of width 80

Figure 4: Execution time of Mutation++ and a neural network of a given width for the different test cases with respect to the number of input points (with log axes).

Figure 5 summarizes the gain factor that we can hope for each test case, with respect to $N_{\mathcal{D}}$ and n_s . It also emphasizes that the higher the number of species n_s , the more important the gain with factors going up to $\times 275$ for `n.units = 20` for *cloudy Jupiter's* atmosphere. Of course, this factor of gain will be relevant only if a good accuracy can be reached for a moderate width. It motivates the next

section, which deals with the optimization of neural networks’ hyperparameters that have a strong influence on both their accuracy and cost-efficiency.

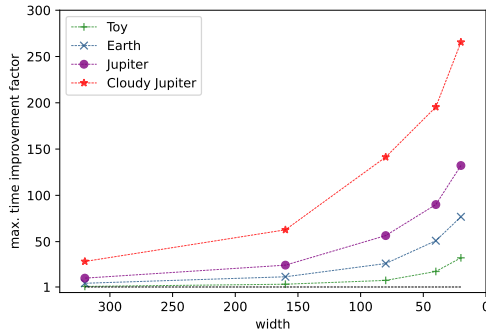


Figure 5: Gains in terms of run-times of neural networks compared to the ones of Mutation++, with respect to the width `n_units`.

The following section describes how we design neural networks to maximize this computational gain factor while maintaining a good approximation accuracy.

4 Design of accurate and cost-effective neural networks with goal-oriented sensitivity analysis of hyperparameters

In this section, we emphasize the impact of neural networks’ hyperparameters on their cost-efficiency and accuracy. To that end, we consider the approximation of Mutation++ in the conditions of the atmosphere of the toy, [Earth](#), and [Jupiter reentry problems as defined in section 3.3](#).

The hyperparameters involved in the training are given in [Appendix B](#). We also use the work of [52], which studies the link between the neural network error and the variance of the output to learn. They define a sampling scheme and its weighting counterparts to account for this variance and improve the error. In the previous section, we explained we want to learn as much as possible from operational conditions of (ρ, ϵ) . But for this problem, relying on the previous independent probability measures gives accurate enough models (as will be seen later on) and considerably eases the reproducibility of the results

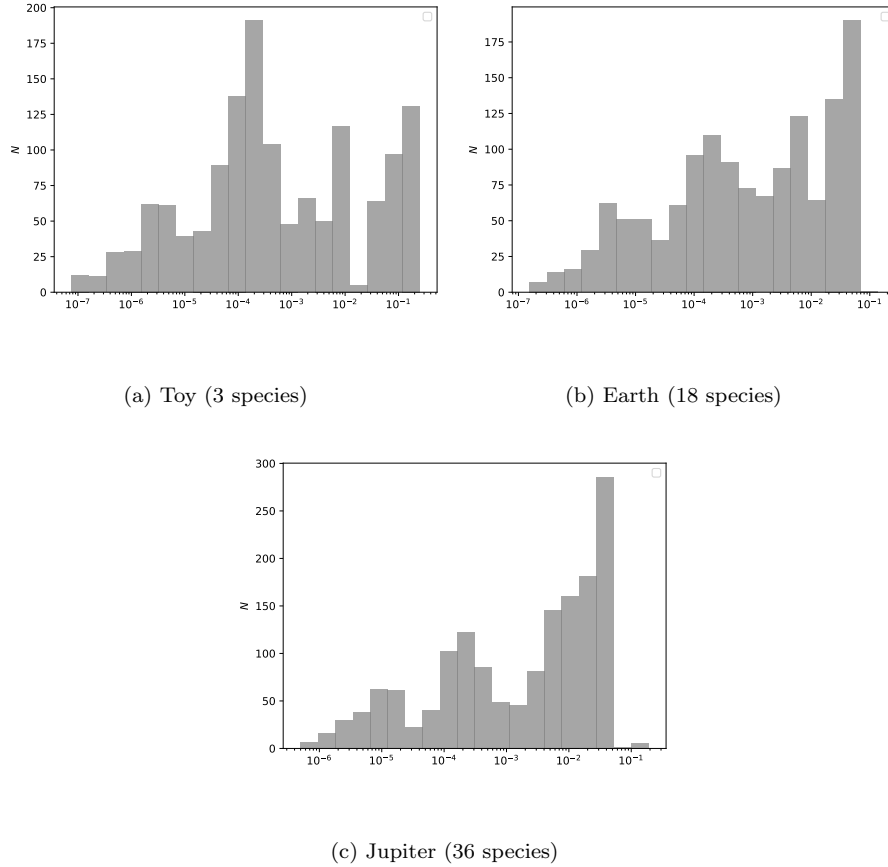


Figure 6: Histograms of normalized L_2 validation error for approximating \mathbf{M}^{++} for $N_{MC} = 1400$ random hyperparameter configurations in the three test cases.

of this paper. The neural networks are trained using Tensorflow in python on a training dataset of 170000 points, and the hyperparameters are selected using a test set of 20000 points. The training and test sets are constructed by sampling ρ and ϵ , respectively, uniformly and log-uniformly within the intervals $[0.1, 3.8] \times [2.07503 \times 10^7, 3 \times 10^8]$, defined based on the execution of the numerical scheme without chemical reactions.

Figure 6 displays the histogram of the L_2 errors obtained after a random search [53], i.e. a uniform Monte Carlo sampling of $N_{MC} = 1400$ samples on the

hyperparameter space described. It shows several important properties:

- first, there is a non-zero probability of having poor performance in terms of accuracies: the L_2 errors go up to $\approx 10^{-1}$. For this reason, it is important performing an optimization of the hyperparameters and not only rely on one test.
- The errors within the range $[10^{-5}, 10^{-1}]$ are more probable than others, but they are certainly not enough for our reentry application.
- Finally, there is a non-zero probability of having very good accuracies, with errors going down to 10^{-7} .

In a nutshell, neural networks allow a wide range of errors (**6 orders**) depending on the hyperparameters use depending on the hyperparameters used. **For these problems, the best errors are summarized in Table 1.**

	Toy	Earth	Jupiter
L_2 error	7.59×10^{-8}	1.55×10^{-7}	4.99×10^{-7}

Table 1: **Best L_2 validation errors obtained for each test case**

The question now is: is such an accurate neural network cost-effective and competitive with respect to a call of Mutation++? **Before answering this question, note that from now on, we focus on the Toy test case because**

- **It is easier to manipulate and to embed in the simulation code**
- **The experiments show that the best achievable error is close for each test case, so we can reasonably expect that the behavior in the Toy problem is representative of the other two.**

The neural network yielding the previous performances has depth `n_layers = 9` and width `n_units = 191` units, which is close to the upper boundary of the search space for these hyperparameters. Using such depth and width could significantly affect the expected cost efficiency improvement. Indeed, in figure

5, with `n_layers = 5` and `n_units = 191`, only a gain of approximately 7 is achievable.

Now, in histogram 6, there are other values of errors that are probably acceptable (for example, errors in the range $[10^{-8}, 10^{-6}]$). The question is: are there cost-effective neural networks allowing us to reach such errors?

In order to answer this question, we rely on the work of [54], based on papers [55, 15]: In [54], the authors study the use of goal-oriented sensitivity analysis, based on the Hilbert-Schmidt Independence Criterion (HSIC), for hyperparameter analysis and optimization. They design a robust analysis index that is able to quantify hyperparameters' relative impact on an NN's final error. This tool allows a better understanding of the hyperparameters' effects on both the error and the run-time. It is able to identify which hyperparameter(s) is (are) responsible for explaining the lowest errors. Once this/these are identified, it allows focusing the optimization on the hyperparameters having a significant impact on the error while considering the constraints of having viable run-times. The application of the methodology only consists of quick post-processing of the random search results. It can give fast and accurate insights from the previously performed random search. We do not give more details on the matter of paper [54]; we only present the results and the interpretations.

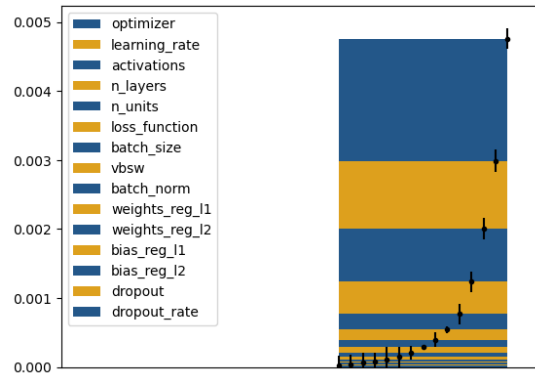


Figure 7: Goal-oriented sensitivity indices of the hyperparameters as in [54] obtained by post-processing the random search for Mutation++ for the toy atmosphere.

Figure 7 presents the aforementioned goal-oriented sensitivity indices for each hyperparameter. The indices are stacked and sorted by decreasing importance from top to bottom. For instance, the choice of the `optimizer` is the most important hyperparameter to reach the best 10% error percentile, and `dropout_rate` is the least important. Besides, estimation error bars for these indices are provided, attesting to relatively converged results with respect to the number of points in the random search $N_{MC} = 3 \times 10^3$. Note that the error bars of the last hyperparameters intersect, so we will not allow ourselves to interpret them. Nonetheless, according to the error bars, we can focus on the primary hyperparameters, at least the 5 most important. In other words, hyperparameters `optimizer`, `learning_rate` and `activation` are by far the three most influential hyperparameters in order to reach the 10% best errors. This information is of great value: these hyperparameters do not impact the run-time once the parameters θ are tuned. This means that within the 10% best results, there are probably cheap (i.e. shallow and tight) neural networks. The number of layers `n_layers` and of units `n_units` only come at the 4th and 5th position of relative importance. By using the methodologies of [54], it is possible to:

- Select cost effective values for `n_layers` and `n_units`, as well as other low-impactful hyperparameters, with a limited impact on the error,
- Focus subsequent hyperparameter optimization on the three most impactful hyperparameters, `optimizer`, `learning_rate` and `activations`.

We perform a Gaussian Process-based bayesian optimization in low dimension - see the TS-GPBO methodology of [54]. The obtained neural network reaches an L_2 error of 8.48×10^{-8} , which is even lower than with the previous random search, with only `n_layers` = 5 layers and `n_units` = 20 neurons. This network has a competitive error with far fewer parameters and much shorter run times. We suggest now plugging this neural network into the reentry scheme and revisiting the problem of section 1 with a hybrid numerical scheme.

5 Deep Learning-based hybridization with guarantees

In this section, we revisit the motivating example of section 1 with a hybrid reentry numerical scheme. The sketch of the scheme is recalled below in the algorithm 3.

```

1 # general initialization (mesh, quantities on mesh etc.)
2 initialise_guess_vector_of_unknowns_on_mesh( $U^0, \mathbf{x}^0$ )
3 while convergence_criterion_not_satisfied do
4      $U^{n+\frac{1}{2}} = \text{solve\_Euler\_equations}(U^n, \mathbf{x}^n)$ 
5      $\mathbf{x}^{n+1}, U^{n+1} = \text{call\_neural\_network}(U^{n+\frac{1}{2}}, \mathbf{x}^n)$ 
6      $U^n \leftarrow U^{n+1}$ 
7      $\mathbf{x}^n \leftarrow \mathbf{x}^{n+1}$ 
8 end

```

Algorithm 3: Core of the scheme with a call to a neural network surrogate model of Mutation++.

In function call_neural_network, we plug the neural network approximating Mutation++ described at the end of section 3 using the Tensorflow C API and a wrapper, CppFlow⁶. Note that Algorithm 3 is slightly different from Algorithm 1 because, in the former, the neural network is called in a batch fashion to take advantage of the vectorial optimizations, while in the latter, Mutation++ has to be called sequentially. We qualitatively verify that the results obtained thanks to the neural network are in agreement with the ones of Mutation++ by displaying the maps $(\rho, \epsilon) \rightarrow \alpha(\rho, \epsilon)$ for $\alpha \in \{\mathbf{x}_O, \mathbf{x}_N, \mathbf{x}_{NO}, P, T, C_p, C_v, c\}$ in Appendix **A**.

For each output physical observable of figure 8, the predictions are compared to the reference values computed by \mathbf{M}^{++} . The results are in very good agree-

⁶<https://github.com/serizba/cppflow>

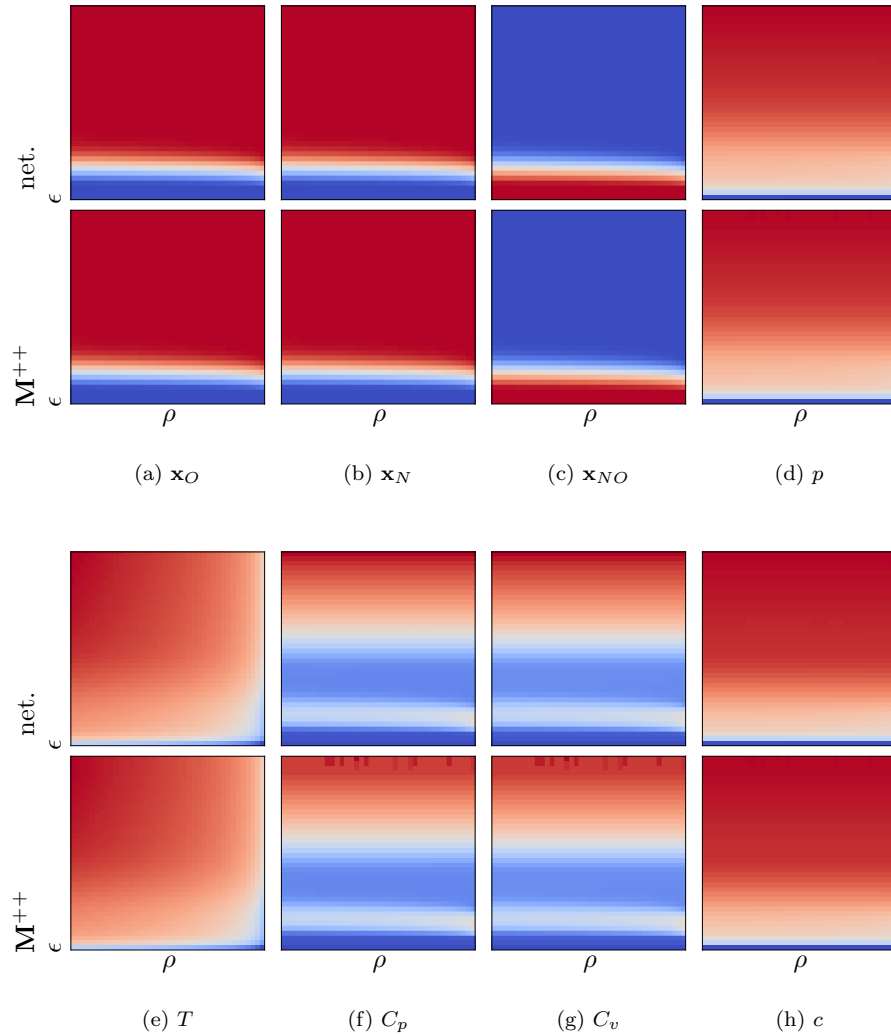


Figure 8: Predictions of the neural network (top) and predictions of \mathbf{M}^{++} (bottom) on the input domain. Axis values are omitted for clarity but we recall that ρ and ϵ are defined on $[0.1, 3.8]$ and $[2.07503 \times 10^7, 3 \times 10^8]$ respectively.

ment (remember we have a L_2 error close to 10^{-8} with this neural network). Still, in figure 9 (f) and (g), for C_p and C_v , there are numerical artifacts in the predictions of \mathbf{M}^{++} at the top of the domain (for high values of ϵ) whereas these are not observable with the neural network. These numerical instabilities

\mathbf{M}^{++} may disturb the reentry computation in practice. The neural network does not seem to be subject to such instabilities. We will see that this point is of importance in the following numerical results.

From now on, this section is articulated as follows. In section 5.1, we present the results and accelerations obtained with the hybrid reentry scheme. However, in critical decision-making, the use of machine learning is often controversial because it lacks natural accuracy guarantees. Sections 5.2 and 5.3 are dedicated to alleviating this problem. Notably, in section 5.2, we explain how to make sure that the hybrid scheme recovers *exactly* the same results as the native one (together with ensuring a $\times 10$ factor of acceleration). Finally, in section 5.3, we suggest a way to validate the reliability of the hybrid numerical scheme as such based on proper analysis and study of the different sources of errors and uncertainties of the original numerical scheme.

5.1 Acceleration of the reentry scheme

In this section, we come back to the configuration of section 1. The reference simulations are provided by the native reentry numerical scheme described in section 2 (and algorithm 1).

In the configuration of interest, see table 2, we consider a sphere of radius $r_{\text{sphere}} = 10^{-2}$ entering a simplified Earth’s atmosphere (species N, O, NO). The

Input	value
Elements, $n_e = 2$ (elem:fraction)	O:0.2, N:0.8
Upstream pressure	35737.40Pa
Upstream temperature	216.57K
Upstream velocity (Mach 16)	4930.83 $m.s^{-1}$
Chemical species ($n_s = 3$)	N, O, NO

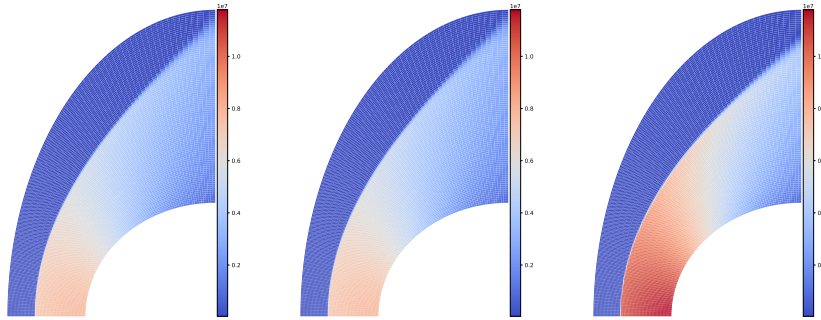
Table 2: Simulation parameters and boundary conditions for the toy example.

boundary conditions are:

- no-slip boundary conditions on the sphere incoming in Earth’s atmosphere,
- incoming flux upstream of the sphere, see table 2 for the considered nominal values of the upstream velocity, pressure, and temperature.

Let us first perform some comparisons on the pressure fields in the same conditions (same mesh) as in section 1 with the three different reentry schemes:

- PG (for perfect gas) denotes the results obtained with the reentry scheme without simulating any chemical reactions. For air, which is a mixture of gases, it is commonly accepted that the adiabatic index γ can be assumed to be 1.4 with only a small error over a wide temperature range [56]. This value is used in all the PG simulations.
- MPP (for Mutation++) denotes the results obtained with the reentry scheme with the simulation of chemical reactions using Mutation++.
- NN (for neural network) denotes the results obtained with the hybrid reentry scheme with the simulation of chemical reactions using the neural network approximating Mutation++ obtained with the methodology described in section 3.



(a) MPP, run-time 4090 s. (b) NN, run-time 220 s. (c) PG, run-time 81 s.

Figure 9: Pressure field for MPP, NN and PG, with the same scales.

Figure 9 echoes figure 2 and presents the results obtained with the three above reentry schemes. As in section 1, we can observe, by comparing figure 9 (a) and (c) that the perfect gas closure (PG) is a coarse model for our reentry problem as its results in terms of pressure field considerably differ from the one obtained with chemical reactions (MPP), i.e. which takes into account finer physics. On the other hand, the pressure field obtained with the hybrid reentry scheme (NN) is not visually distinguishable from the results obtained with Mutation++. Now, in terms of run-time, with about the same accuracy, the hybrid reentry scheme ensures a gain of a factor $\times \frac{4090}{220} \approx 18.6$.

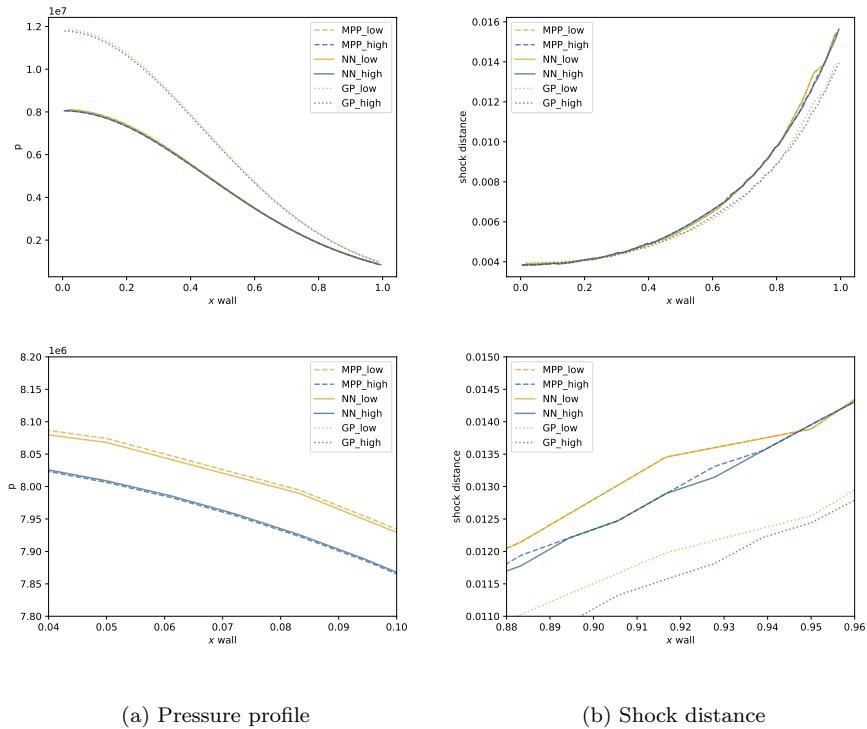


Figure 10: Top: pressure profile and shock distance projected on the surface of the object. Bottom: same quantities with a zoom on the highest error area for the hybrid scheme.

The results of figure 9 are rather qualitative. Let us progressively switch

to more quantitative ones. For this, on figure 10, we extract the pressure on the surface of the object (figure 10 left) and the distance of the shock (figure 10 right) for the three different schemes. Note that the suffixes `_high` and `_low` corresponds to two different meshes:

`_low` is for low-resolution mesh (30×100),

`_high` is for high-resolution mesh (90×100) and is 3 times more refined in the direction of the shock.

First, with figure 10 (right column), we can see that the perfect gas closure is also coarse for the shock distance: the model error (i.e. between MPP and PG independently of the observable of interest) is way more important than the discretization error (as the error between the two meshes is small in comparison to the differences between PG and MPP). This justifies taking the chemistry of the phenomenon into account for the simulation. Now, for those two observables at the surface of the object, we can see that the NN error is even smaller than the discretization error as the dotted lines are the closest to the full lines. Finally, without zooming, no visual differences can be detected between MPP and NN. To assess the prediction error of NN, we have to look at more quantitative results: table 3 displays the L_2 and L_∞ normalized errors for these curves with respect to MPP`_low` and MPP`_high` together with the run-times of the different schemes: Several comments can be made regarding these results:

- The errors between PG (`_low` and `_high`) and MPP`_high` are at least three decades higher than the errors between the results of NN and MPP, regardless of the resolution, which illustrates the need for simulating chemical reactions.
- The errors of NN`_high` and NN`_low` are comparable to that of MPP`_low` when MPP`_high` is taken as the reference. The error between NN`_low` and MPP`_low` is even lower.
- However, the run-times of NN are closer to those of PG with 220s. and 529s. for NN and 81s. and 211s. for PG, against 4090s. and 9478s. for

MPP. Note also that NN_high is still faster than the MPP_low while being more accurate.

	MPP_low (ref)	NN_low	PG_low
Time (s)	4090	220	81
Impr. (\times)	-	18.7	58.5
Pressure			
L^2	-	6.06×10^{-7}	7.74×10^{-2}
L^∞	-	1.13×10^{-3}	4.64×10^{-1}
Shock dist.			
L^2	-	7.85×10^{-7}	1.70×10^{-3}
L^∞	-	4.85×10^{-3}	9.89×10^{-2}

	MPP_high (ref)	MPP_low	NN_high	NN_low	PG_high	PG_low
Time (s)	9478	4090	529	220	211	81
Impr. (\times)	-	2.3	17.9	43.1	44.9	117
Pressure						
L^2	-	3.19×10^{-5}	1.13×10^{-6}	4.00×10^{-5}	7.65×10^{-2}	8.15×10^{-2}
L^∞	-	9.07×10^{-3}	2.22×10^{-3}	9.82×10^{-3}	4.62×10^{-1}	4.75×10^{-1}
Shock dist.						
L^2	-	9.14×10^{-5}	6.55×10^{-6}	9.19×10^{-5}	1.65×10^{-3}	1.29×10^{-3}
L^∞	-	3.60×10^{-2}	1.08×10^{-2}	3.60×10^{-2}	1.08×10^{-1}	9.00×10^{-2}

Table 3: Execution times and normalized errors for the different schemes.

To sum up, on the one hand, MPP and NN are comparable in terms of error, and on the other hand, NN is almost one decade faster than MPP. These results are comforting and motivate us to study the method further and discuss the possibility of having guarantees with the hybrid scheme.

The error of NN was comparable to that of MPP_low on this prediction, but it was on one single prediction. This evaluation process is insufficient to state whether the error is acceptable. In order to ensure prediction guarantees, which are mandatory for using schemes in production, we have to go deeper into the

analysis.

In the next sections, we introduce two ways to obtain guarantees on the predictions of the hybrid scheme. The first ensures to have exactly the same prediction accuracy as the fine reentry scheme (i.e. MPP) but brings additional computations. The second compares the error made with the hybridization of the reentry scheme with other sources of errors that are ubiquitous in numerical simulation to assess the acceptability of the hybrid scheme. If the hybridization error is lower than other errors, it is then possible to use the hybrid scheme (NN) at full speed.

5.2 Zero-error guarantees of the hybrid scheme

As we mentioned earlier, the reentry numerical scheme is an iterative solver, see algorithm 1. It is initialized with a guess solution, which is uninformative - usually, the same value over the entire mesh - and iterations are made until a certain convergence criterion is reached. In this section, we suggest first executing the hybrid scheme (NN), using its prediction as initialization for the classical scheme (MPP). Then, the classical solver may hopefully converge in fewer iterations since the initialization is supposed to be close to the convergence point. The new structure of the scheme is presented in algorithm 4.

```

1 # general initialization (mesh, quantities on mesh etc.)
2 initialise_guess_vector_of_unknowns_on_mesh( $U^0, \mathbf{x}^0$ )
3 # 1st step with the neural network surrogate model
4 while convergence_criterion_not_satisfied do
5    $U^{n+\frac{1}{2}}$  =solve_Euler_equations( $U^n, \mathbf{x}^n$ )
6    $\mathbf{x}^{n+1}, U^{n+1}$  =call_NN_surrogate_model( $U^{n+\frac{1}{2}}, \mathbf{x}^n$ )
7    $U^n \leftarrow U^{n+1}$ 
8    $\mathbf{x}^n \leftarrow \mathbf{x}^{n+1}$ 
9 end
10 # 2nd step with Mutation++ using the NN results as the initial guess
11 while convergence_criterion_not_satisfied do
12    $U^{n+\frac{1}{2}}$  =solve_Euler_equations( $U^n, \mathbf{x}^n$ )
13   for  $i \in \{1, \dots, N_{\mathcal{D}}\}$  do
14      $\mathbf{x}_i^{n+1}, U^{n+1}$  =minimize_Gibbs_free_energy_with_mutation++( $U_i^{n+\frac{1}{2}}, \mathbf{x}_i^n$ )
15   end
16    $U^n \leftarrow U^{n+1}$ 
17    $\mathbf{x}^n \leftarrow \mathbf{x}^{n+1}$ 
18 end

```

Algorithm 4: Core of the scheme with a call to a neural network surrogate model of Mutation++ in order to initialize the guess of a classical computation.

The strategy described in algorithm 4 gives the following results. Once the prediction of NN_low is given as a guess of the original scheme, MPP_low, the latter reaches convergence in 163s.. If we sum the run times of both schemes, in that case, the simulation takes 383s. which is $\times 10.6$ faster than MPP_low alone, for a prediction whose accuracy is *exactly* the same as the ones of the original scheme. This strategy is denoted by NN+MPP in the following paragraphs. In that case, there is no need to compare the pressure profiles or the shock distances between the different options of the scheme since **the prediction are indistinguishable from one another**. In other words, we obtained an

acceleration of a factor 10.6 for the exact same accuracy, i.e. *with the same guarantees as the original scheme.*

Remark 1. *This approach is not specific to our reentry problem: it can be applied to any stationary simulations or in a general manner to any computation scheme involving an iterative solver whose iterations are not of interest as simulation outputs. As an example, [57] applies this idea for stochastic dynamical systems simulating chaotic flows.*

5.3 Guarantees of acceptable error for the hybrid scheme

As we have seen in the previous section, for our reentry simulations, it is possible to have the same guarantees with the hybrid scheme as with the original one at the price of a smaller acceleration ($\times 10$ with guarantees instead of 18.6). Now, one could be interested in this 18.6 factor of acceleration instead of $\times 10.6$ (or one could not have a stationary problem). In order to make sure the error coming from neural network approximation remains acceptable, it is mandatory to be able to quantify it and compare it to the other sources of errors that are classically found in numerical simulations. In this section, we compare the error due to the use of a neural network within the hybrid scheme with these other errors, namely the model error, the discretization error, and the fluctuations relative to the sources of uncertainty.

Let us formalize this: suppose we are interested in observable y (it can be the pressure on the surface of the object, the distance to the shock etc.). Let us denote by y_r the ground-truth value of y that we want to predict with a model. We denote the original scheme MPP by \mathcal{M} and the hybrid scheme NN by $\widehat{\mathcal{M}}$, with \mathcal{M}_{high} denoting the model of MPP_high, and so on for the other MPP and NN schemes. The models take parameters \mathbf{x} as input and output a prediction $\mathcal{M}(\mathbf{x})$ and $\widehat{\mathcal{M}}(\mathbf{x})$. In our case, the input vector \mathbf{x} contains the upstream pressure, temperature, and speed. The predictions can be expressed as

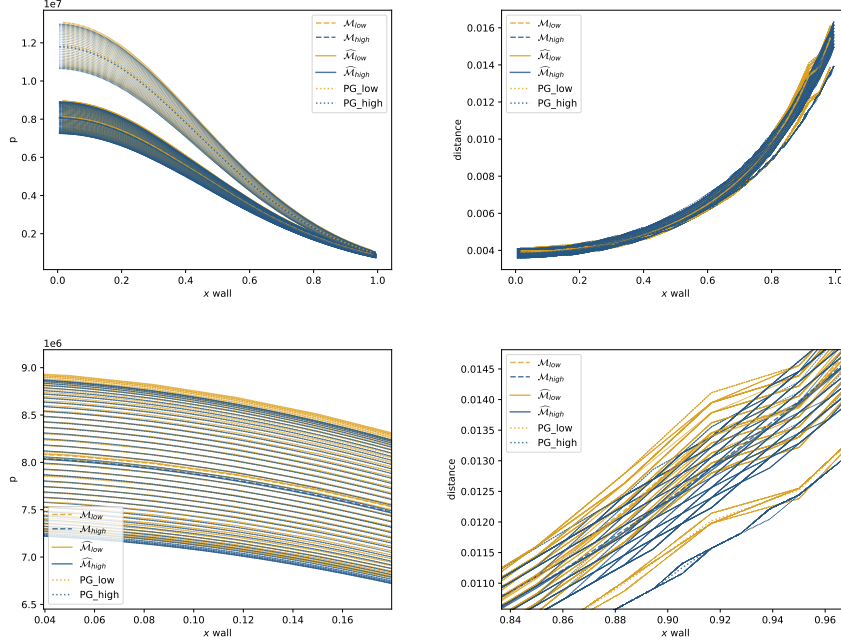
$$\begin{cases} \mathcal{M}(\mathbf{x}) = y_r + e_{\mathcal{M}}(\mathbf{x}) = y_r + \delta_{\Delta, \mathcal{M}}(\mathbf{x}) + \delta_{\mathbf{x}, \mathcal{M}}(\mathbf{x}), \\ \widehat{\mathcal{M}}(\mathbf{x}) = y_r + e_{\widehat{\mathcal{M}}}(\mathbf{x}) = y_r + \delta_{\Delta, \widehat{\mathcal{M}}}(\mathbf{x}) + \delta_{\mathbf{x}, \widehat{\mathcal{M}}}(\mathbf{x}) + \delta_{\theta, \widehat{\mathcal{M}}}(\mathbf{x}). \end{cases} \quad (4)$$

Equation (4) emphasizes three different types of errors :

- The discretization errors $\delta_{\Delta, \mathcal{M}}$ and $\delta_{\Delta, \widehat{\mathcal{M}}}$. The choice of the mesh used to run the simulation has an impact on the prediction error. A low mesh resolution may degrade the error, as we saw with MPP_low, but the geometry of the mesh also has its impact.
- The parameters' errors $\delta_{\mathbf{x}, \mathcal{M}}$ and $\delta_{\mathbf{x}, \widehat{\mathcal{M}}}$. In practice, we conduct numerical simulations because we are interested in the output of a phenomenon under specific conditions of interest. Nonetheless, we may have imperfect knowledge of these conditions of interest. This translates into uncertainties on the input vector \mathbf{x} , which contains the values of the parameters that define the conditions of the simulation. These uncertainties have an impact on the model prediction and, therefore, on their prediction error.
- The neural network approximation error $\delta_{\theta, \widehat{\mathcal{M}}}$. In the hybrid scheme, NN, the neural network approximates \mathbf{M}^{++} with a certain error. This error propagates through the hybrid scheme, thereby affecting its prediction error.

In order to ensure that NN yields reliable predictions, the neural network approximation error must be compared with parameters and discretization errors. If the former is at most similar to the two others, NN could be used safely.

Remark 2. *Another type of error is often specified when decomposing the error of numerical schemes. This error is called modeling error and refers to the error that stems from modeling choices. In our case, the model error would be the error between PG and MPP, coming from the choice not to simulate the chemistry in PG. Another model error could come from the choice not to simulate the chemistry, to simulate it with more or fewer species, or to use Navier-Stokes equations rather than Euler equations. In addition, round-off errors could also be taken into account. In this work, we suppose they are negligible with respect to the other sources of errors. The reader interested in ways to quantify them can refer to [58] for example.*



(a) Pressure profile

(b) Shock distance

Figure 11: Pressure profile and shock distance projected on the wall of the object, for 40 different values of the upstream speed. In the bottom line, zoom of the curves in the highest error area.

5.3.1 Uncertainty propagation for reliable error comparison

To compare the different errors, we introduce a perturbation in \mathbf{x} , modeling the uncertainty on the upstream speed. Let $\mathbf{x} = (x, y, z)$, with $x \sim \mathcal{U}(0.95x, 1.05x)$, $x = 4930.83$ (the nominal value of the initial test case). The random variable x traduces the uncertainty on the speed of the upstream field, and the values of y and z are the upstream pressure and temperature. We simulate the random variable x on $N = 40$ Gauss quadrature points $\{x_1, \dots, x_N\}$, with $x_i \in [0.95x, 1.05x]$, that are used to evaluate $\mathbb{E}[\mathcal{M}(\mathbf{x})]$ with $\mathcal{M} \in \{\mathcal{M}_{low}, \mathcal{M}_{high}, \widehat{\mathcal{M}}_{low}, \widehat{\mathcal{M}}_{high}\}$. Such an analysis allows us to study the effect of parameter uncertainty on the error, as well as statistically compare the

different sources of error.

The mean $\mathbb{E}[\mathcal{M}(\boldsymbol{x})]$ and each of the 40 curves are plotted in Figure 11. These graphs are quite loaded, but they highlight that the variability induced by parameter uncertainty is much higher than that coming from approximation and even discretization errors.

Comparing discretization and neural network approximation errors is more subtle because it is not clear-cut in Figure 11. To do so, we directly compare the discretization error of \mathcal{M}_{low} with the approximation error of $\widehat{\mathcal{M}}_{low}$ and $\widehat{\mathcal{M}}_{high}$ under parameters uncertainty. First, we plot $\|\mathcal{M}_{low} - \mathcal{M}_{high}\|$ and $\|\widehat{\mathcal{M}}_{low} - \mathcal{M}_{high}\|$ for the 40 values of x_i in Figure 12. Here, $\|\cdot\|$ is the normalized absolute difference evaluated point-wise in the pressure profile and the shock distance. At first sight, the approximation error seems to be lower (for the pressure profile) or equivalent (for the shock distance) to the discretization error.

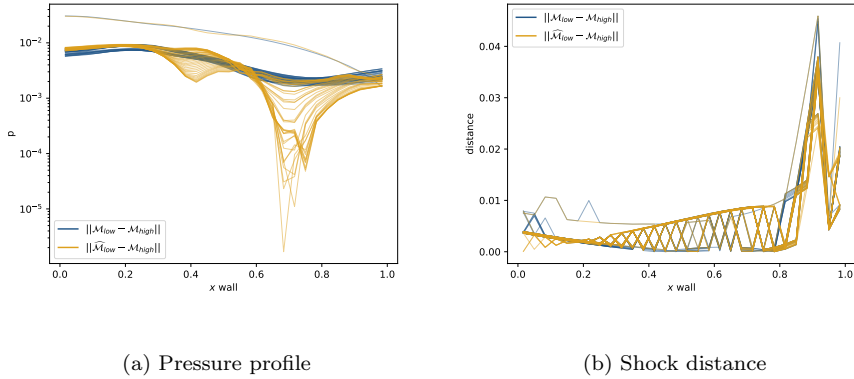


Figure 12: Discretization errors $\|\mathcal{M}_{low} - \mathcal{M}_{high}\|$ and $\|\widehat{\mathcal{M}}_{low} - \mathcal{M}_{high}\|$ for the pressure profile and the shock distance for each of the 40 different values of the upstream speed.

We confirm this observation by plotting $\|\widehat{\mathcal{M}}_{low} - \mathcal{M}_{low}\|$ and $\|\widehat{\mathcal{M}}_{high} - \mathcal{M}_{high}\|$ in Figure 13. We plot the 40 curves corresponding to each x_i and the mean estimated using the Gauss quadrature. It clarifies the comparison and strengthens the conclusion that approximation error is lower than both discretization and parameters error.

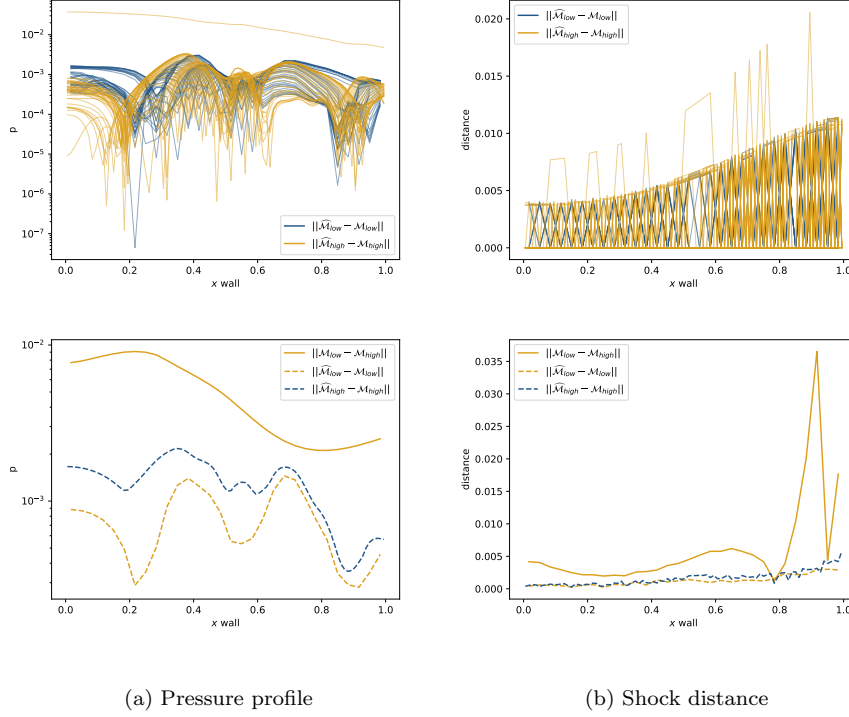


Figure 13: Top: approximation errors $\|\widehat{\mathcal{M}}_{low} - \mathcal{M}_{low}\|$ and $\|\widehat{\mathcal{M}}_{high} - \mathcal{M}_{high}\|$ for the pressure profile and the shock distance for each of the 40 different values of the upstream speed. Bottom: comparison of the mean of the discretization and approximation errors.

This error study shows that the neural network’s approximation error can be small compared to other types of errors. In this case, the hybrid scheme is reliable, which is a strong argument in favor of the use of hybrid numerical schemes.

5.3.2 Benefits of NN+MPP to obtain reference predictions

To conduct this experiment, we never executed \mathcal{M}_{low} and \mathcal{M}_{high} entirely, but always used an initialization from the prediction of $\widehat{\mathcal{M}}_{low}$ and $\widehat{\mathcal{M}}_{high}$ (NN+MPP, as described in Section 5.2). The advantages were twofold. First, the study was much faster (approximately by a factor of 10). Second, for some $\{x_1, \dots, x_N\}$, MPP did not converge, perhaps because of numerical instabilities. Initializing

MPP using the hybrid scheme solved this problem. This echoes remark 1 on the artefacts in predictions of C_p and C_v by \mathbf{M}^{++} .

6 Discussion and Perspectives

In this work, we studied the acceleration of a numerical scheme involving the coupling between hypersonic fluid dynamic and chemical equilibrium. The numerical scheme is computationally expensive because of the chemical equilibrium simulator (Mutation++) that has to be called at each cell of the simulation mesh. This motivated using neural networks as surrogate models approximating chemical reactions because (1) their execution can be easily vectorized, so they can be called efficiently on a whole mesh, and (2) they can be trained on an extensive database constructed out of Mutation++.

By taking care of constructing neural networks not only accurate but also cost-effective thanks to the methodology of [54], we achieve an acceleration factor of 18.7 for the hybrid numerical scheme (NN). The obtained prediction is qualitatively indistinguishable and quantitatively very close to that of the original numerical scheme (MPP).

Nonetheless, though promising, these results are insufficient to ensure NN guarantees for reliable use in production. We describe two simple methodologies for obtaining guarantees to tackle that problem. The first methodology ensures the same guarantees as MPP by initializing it with the prediction of NN. However, the acceleration factor decreases to 10. The second methodology relies on the consideration that many sources of errors affect the prediction of numerical scheme, even without hybridization. We conduct a statistical study of the effect of some of these errors, namely the parameter’s uncertainty and discretization errors, and compare them to the neural network’s approximation error. It turns out that in this test case, the neural network’s error is negligible with respect to the others, so we tend to recommend using the hybrid scheme with its full acceleration factor of 18.7 safely. We would like to conclude the paper with a discussion on the perspectives of this work.

6.1 *Towards a general approximation of Mutation++*

In this section, the methodology for approximating Mutation++ consists of constructing a training database and fitting a neural network. This is a strong advantage of this method since the neural network can be used in any numerical scheme involving the same chemical equilibrium setting (as we saw in section 5.3).

However, the neural network is trained for a fixed output dimension corresponding to the number of species. It cannot be used for test cases that involve different species because it requires constructing a new training database for each different chemical setting.

It would be interesting to investigate the use of transfer learning to make the approach easily applicable to other chemical reactions. One could pre-train a neural network once for a high number of different species and with a large database constructed out of Mutation++. Then, one could find a simple way to adapt this neural network for each different test case, for instance, with a least-squares linear regression on the feature space of the pre-trained network, using a smaller data set.

6.2 *A general pattern for hybridization*

Our approach is not specific to hypersonic reentry coupled with chemical equilibrium. The idea of constructing a hybrid scheme based on both numerical simulation and machine learning has already been explored in previous works. In molecular simulations, [59] use a neural network approximation of potential energy surface, and [60] use Gaussian processes to sample Gibbs free energy surface, opening the avenue for applications based on such methodologies [34, 35, 36, 37]. [33] use neural networks to approximate physical components of multi-physics problems for electro-thermal simulation when conducting electrosurgery. In [23, 38], the authors approximate non-local thermodynamic equilibrium in the simulation of inertial confinement fusion.

More formally, a multi-physics numerical scheme often solves a coupled system of several components that model different physics. To simplify the framework,

we only consider a scheme with two coupled systems of equations. The system can be written :

$$\begin{cases} F_1(U, \mathbf{x}, \boldsymbol{\alpha}) = 0, & (5a) \\ F_2(U, \mathbf{x}, \boldsymbol{\alpha}) = 0, & (5b) \end{cases}$$

where

- U and \mathbf{x} are vectors of unknowns,
- $\boldsymbol{\alpha}$ is a vector of physical parameters, that are not computer during the simulation (e.g. physical or chemical constants),
- F_1 and F_2 are mathematical (possibly differential) operators. In our test case, equation (5a) is Euler equations and equation (5b) is the Gibbs free energy minimization equation, which is behind Mutation++.

In such cases, most of the time, the solver needs to solve equation (5b) repeatedly in order to solve equation (5a). As a result, equation (5a) is costly to solve because it regularly calls for the resolution of equation (5b). The approach of approximating F_2 with a neural network and leveraging its implementation to process an entire mesh in a batch input fashion is, therefore, more general than the present hypersonic reentry test case. Numerical simulations, in general, could benefit from such an approach, strengthened by the described methodology for guaranteeing the results of the obtained hybrid scheme.

6.3 Hybrid numerical schemes as an additional lever for acceleration

We would like to point out that even if the obtained hybrid scheme no longer uses the scheme part that is approximated by a neural network, this part is still crucial for constructing the hybrid scheme. Indeed, a good training set is mandatory to ensure neural network accuracy, and the original scheme part is key to achieving that accuracy. That is why we do not claim deep learning to replace numerical schemes. Instead, we argue that it should be seen as an additional step in constructing numerical schemes, allowing for significant accelerations.

Bibliography

- [1] A. Martin, I. Boyd, Strongly coupled computation of material response and nonequilibrium flow for hypersonic ablation, *Journal of Spacecraft and Rockets* 52. doi:10.2514/1.A32847.
- [2] F. S. Milos, Y.-K. Chen, Ablation, thermal response, and chemistry program for analysis of thermal protection systems, *Journal of Spacecraft and Rockets* 50 (2013) 137–149.
- [3] D. Bianchi, F. Nasuti, E. Martelli, Navier–stokes simulations of hypersonic flows with coupled graphite ablation, *Journal of Spacecraft and Rockets - J SPACECRAFT ROCKET* 47 (2010) 554–562. doi:10.2514/1.47995.
- [4] S. Peluchon, Approximation numérique et modélisation de l’ablation liquide, Theses, Université de Bordeaux (Nov. 2017).
URL <https://tel.archives-ouvertes.fr/tel-01684676>
- [5] M. Latige, Simulation numérique de l’ablation liquide, Theses, Université Sciences et Technologies - Bordeaux I (Sep. 2013).
URL <https://tel.archives-ouvertes.fr/tel-00858600>
- [6] J. Yin, H. Zhang, X. Xiong, B. Huang, J. Zuo, Ablation properties of carbon/carbon composites with tungsten carbide, *Applied Surface Science* 255 (9) (2009) 5036–5040. doi:10.1016/j.apsusc.2008.12.063.
- [7] J. B. Scoggins, V. Leroy, G. Bellas-Chatzigeorgis, B. Dias, T. E. Magin, Mutation++: Multicomponent thermodynamic and transport properties for ionized gases in c++, *SoftwareX* 12. doi:10.1016/j.softx.2020.100575.
- [8] F. Danvin, M. Olazabal, F. Pinna, Laminar to turbulent transition prediction in hypersonic flows with neural networks committee, 2021. arXiv:<https://arc.aiaa.org/doi/pdf/10.2514/6.2019-2837>, doi:10.2514/6.2019-2837.
URL <https://arc.aiaa.org/doi/abs/10.2514/6.2019-2837>

- [9] M. Olazabal-Loumé, F. Chedevergne, F. Danvin, J. Mathiaud, Roughness corrections applied to the simulation of turbulent hypersonic flows, in: EUCASS 2019, MADRID, Spain, 2019.
URL <https://hal.archives-ouvertes.fr/hal-02344348>
- [10] O. P. Le Maître, O. M. Knio, Uncertainty Propagation using Wiener-Haar Expansions, *J. Comp. Phys.* 197 (2004) 28–57.
- [11] S. Hosder, R. W. Walters, R. Perez, A Non Intrusive Polynomial Chaos Method for Uncertainty Propagation in CFD Simulations, 44th AIAA Aerospace Sciences Meeting and Exhibit AIAA 2006-891.
- [12] G. Poëtte, B. Després, D. Lucor, Uncertainty Quantification for Systems of Conservation Laws, *J. Comp. Phys.* 228 (7) (2009) 2443–2467. doi:<http://dx.doi.org/10.1016/j.jcp.2008.12.018>.
- [13] D. Lucor, J. Meyers, P. Sagaut, Sensitivity Analysis of LES to Subgrid-Scale-Model Parametric Uncertainty using Polynomial Chaos, *J. Fluid Mech.* 585 (2007) 255–279.
- [14] P. Congedo, C. Corre, J.-M. Martinez, Shape optimization of an airfoil in a bzt flow with multiple-source uncertainties, *Computer Methods in Applied Mechanics and Engineering* 200 (1) (2011) 216–232. doi:<https://doi.org/10.1016/j.cma.2010.08.006>.
URL <https://www.sciencedirect.com/science/article/pii/S0045782510002392>
- [15] S. Da Veiga, F. Wahl, F. Gamboa, Local polynomial estimation for sensitivity analysis on models with correlated inputs, *Technometrics* 51 (2009) 452 – 463.
- [16] N. Plattner, S. Doerr, G. De Fabritiis, F. Noé, Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling, *Nature Chemistry* 9 (10) (2017) 1005–1011.

doi:10.1038/nchem.2785.

URL <http://www.nature.com/articles/nchem.2785>

- [17] M. I. Zimmerman, G. R. Bowman, Fast conformational searches by balancing exploration/exploitation trade-offs, *Journal of Chemical Theory and Computation* 11 (12) (2015) 5747–5757. doi:10.1021/acs.jctc.5b00737. URL <https://pubs.acs.org/doi/10.1021/acs.jctc.5b00737>
- [18] S. Doerr, G. De Fabritiis, On-the-fly learning and sampling of ligand binding by high-throughput molecular simulations, *Journal of Chemical Theory and Computation* 10 (5) (2014) 2064–2069. doi:10.1021/ct400919u. URL <https://pubs.acs.org/doi/10.1021/ct400919u>
- [19] W. A. Lahoz, B. Khattatov, R. Ménard (Eds.), *Data assimilation: making sense of observations*, Springer, Heidelberg, New York, 2010, oCLC: ocn499067426.
- [20] R. E. Baker, J.-M. Peña, J. Jayamohan, A. Jérusalem, Mechanistic models versus machine learning, a fight worth fighting for the biological community?, *Biology Letters* 14 (5) (2018) 20170660. arXiv:<https://royalsocietypublishing.org/doi/pdf/10.1098/rsbl.2017.0660>, doi:10.1098/rsbl.2017.0660. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rsbl.2017.0660>
- [21] S. Koziel, L. Leifsson, Knowledge-based airfoil shape optimization using space mapping, in: *30th AIAA Applied Aerodynamics Conference*, 2012. arXiv:<https://arc.aiaa.org/doi/pdf/10.2514/6.2012-3016>, doi:10.2514/6.2012-3016. URL <https://arc.aiaa.org/doi/abs/10.2514/6.2012-3016>
- [22] X. Guo, W. Li, F. Iorio, Convolutional neural networks for steady flow approximation, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, Association

for Computing Machinery, New York, NY, USA, 2016, p. 481–490. doi:
10.1145/2939672.2939738.

URL <https://doi.org/10.1145/2939672.2939738>

- [23] G. Kluth, K. D. Humbird, B. K. Spears, J. L. Peterson, H. A. Scott, M. V. Patel, J. Koning, M. Marinak, L. Divol, C. V. Young, Deep learning for nlte spectral opacities, *Physics of Plasmas* 27 (5) (2020) 052707. arXiv:
<https://doi.org/10.1063/5.0006784>, doi:10.1063/5.0006784.
URL <https://doi.org/10.1063/5.0006784>

- [24] P. J. Milan, J.-P. Hickey, X. Wang, V. Yang, Deep-learning accelerated calculation of real-fluid properties in numerical simulation of complex flowfields, *Journal of Computational Physics* 444 (2021) 110567. doi:
<https://doi.org/10.1016/j.jcp.2021.110567>.
URL <https://www.sciencedirect.com/science/article/pii/S0021999121004629>

- [25] M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations, arXiv preprint arXiv:1711.10561.

- [26] M. Raissi, P. Perdikaris, G. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational Physics* 378 (2019) 686–707. doi:
<https://doi.org/10.1016/j.jcp.2018.10.045>.
URL <https://www.sciencedirect.com/science/article/pii/S0021999118307125>

- [27] S. Cai, Z. Mao, Z. Wang, M. Yin, G. E. Karniadakis, Physics-informed neural networks (pinns) for fluid mechanics: A review (2021). arXiv:2105.09506.

- [28] L. Lu, P. Jin, G. Pang, Z. Zhang, G. E. Karniadakis, Learning nonlinear operators via deeponet based on the universal approximation theorem

of operators, *Nature Machine Intelligence* 3 (3) (2021) 218–229. doi:
10.1038/s42256-021-00302-5.

URL <https://doi.org/10.1038/s42256-021-00302-5>

- [29] W. S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, *The bulletin of mathematical biophysics* 5 (4) (1943) 115–133. doi:10.1007/BF02478259.

URL <https://doi.org/10.1007/BF02478259>

- [30] Y. Li, Z. Zhou, S. Ying, Delisa: Deep learning based iteration scheme approximation for solving pdes, *Journal of Computational Physics* 451 (2022) 110884. doi:<https://doi.org/10.1016/j.jcp.2021.110884>.

URL <https://www.sciencedirect.com/science/article/pii/S0021999121007798>

- [31] J. Huang, H. Wang, H. Yang, Int-deep: A deep learning initialized iterative method for nonlinear problems, *Journal of Computational Physics* 419 (2020) 109675. doi:<https://doi.org/10.1016/j.jcp.2020.109675>.

URL <https://www.sciencedirect.com/science/article/pii/S0021999120304496>

- [32] Z. Wang, Z. Zhang, A mesh-free method for interface problems using the deep learning approach, *Journal of Computational Physics* 400 (2020) 108963. doi:<https://doi.org/10.1016/j.jcp.2019.108963>.

URL <https://www.sciencedirect.com/science/article/pii/S0021999119306680>

- [33] Z. Han, Rahul, S. De, A deep learning-based hybrid approach for the solution of multiphysics problems in electrosurgery, *Computer Methods in Applied Mechanics and Engineering* 357 (2019) 112603. doi:10.1016/j.cma.2019.112603.

URL <https://linkinghub.elsevier.com/retrieve/pii/S0045782519304797>

- [34] E. Schneider, L. Dai, R. Q. Topper, C. Drechsel-Grau, M. E. Tuckerman, Stochastic Neural Network Approach for Learning High-Dimensional Free Energy Surfaces, *Physical Review Letters* 119 (15) (2017) 150601. doi:10.1103/PhysRevLett.119.150601.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.119.150601>
- [35] L. Mones, N. Bernstein, G. Csányi, Exploration, Sampling, And Reconstruction of Free Energy Surfaces with Gaussian Process Regression, *Journal of Chemical Theory and Computation* 12 (10) (2016) 5100–5110. doi:10.1021/acs.jctc.6b00553.
URL <https://pubs.acs.org/doi/10.1021/acs.jctc.6b00553>
- [36] T. Bereau, R. A. DiStasio, A. Tkatchenko, O. A. von Lilienfeld, Non-covalent interactions across organic and biological subsets of chemical space: Physics-based potentials parametrized from machine learning, *The Journal of Chemical Physics* 148 (24) (2018) 241706. doi:10.1063/1.5009502.
URL <http://aip.scitation.org/doi/10.1063/1.5009502>
- [37] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, K.-R. Müller, Bypassing the Kohn-Sham equations with machine learning, *Nature Communications* 8 (1) (2017) 872. doi:10.1038/s41467-017-00839-3.
URL <http://www.nature.com/articles/s41467-017-00839-3>
- [38] G. Kluth, K. Humbird, B. Spears, H. Scott, M. Patel, L. Peterson, J. Koning, M. Marinak, L. Divol, C. Young, Deep Learning for Non-Local Thermodynamic Equilibrium in hydrocodes for ICF, in: *APS Division of Plasma Physics Meeting Abstracts*, Vol. 2019 of *APS Meeting Abstracts*, 2019, p. BO5.009.
- [39] S. Peluchon, G. Gallice, L. Mieussens, A robust implicit–explicit acoustic-transport splitting scheme for two-phase flows, *Journal of Computational Physics* 339 (2017) 328–355. doi:<https://doi.org/10.1016/j.jcp.2017.03.019>.

URL <https://www.sciencedirect.com/science/article/pii/S0021999117302061>

- [40] S. Peluchon, G. Gallice, P.-H. Maire, Some acoustic-transport splitting schemes for two-phase compressible flows, 2016, pp. 1151–1164. doi:10.7712/100016.1876.7453.
- [41] J.-F. Sigrist, Numerical Simulation, An Art of Prediction 1, 2019. doi:10.1002/9781119686798.
- [42] J.-F. Sigrist, Numerical Simulation, An Art of Prediction 2: Examples, 2020. doi:10.1002/9781119694731.
- [43] I. J. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, Cambridge, MA, USA, 2016, <http://www.deeplearningbook.org>.
- [44] J. Berner, P. Grohs, G. Kutyniok, P. Petersen, The modern mathematics of deep learning (2021). arXiv:2105.04026.
- [45] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (5) (1989) 359 – 366.
- [46] A. R. Barron, Approximation and estimation bounds for artificial neural networks, Machine Learning 14 (1) (1994) 115–133. doi:10.1007/BF00993164. URL <https://doi.org/10.1007/BF00993164>
- [47] Z. Lu, H. Pu, F. Wang, Z. Hu, L. Wang, The expressive power of neural networks: A view from the width, CoRR abs/1709.02540. arXiv:1709.02540. URL <http://arxiv.org/abs/1709.02540>
- [48] P. Auer, M. Herbster, M. K. Warmuth, Exponentially many local minima for single neurons, Departement of computer science.
- [49] R. Lebrun, A. Dutfoy, A Generalization of the Nataf Transformation to Distributions with Elliptical Copula, Prob. Eng. Mech. 24,2 (2009) 172–178.

- [50] R. Lebrun, A. Dutfoy, An Innovating Analysis of the Nataf Transformation from the Copula viewpoint, *Prob. Eng. Mech.* 24,3 (2009) 312–320.
- [51] A. Dutfoy, R. Lebrun, Practical approach to dependence modelling using copulas, *Journal of Risk and Reliability* 223 (4) (2009) 347–361.
URL <http://EconPapers.repec.org/RePEc:sae:risrel:v:223:y:2009:i:4:p:347-361>
- [52] P. Novello, G. Poëtte, D. Lugato, P. Congedo, Variance based sample weighting for supervised learning, in: Submitted to International Conference on Learning Representations, 2021, under review.
URL <https://openreview.net/forum?id=3F0Qm7TzNDM>
- [53] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, *Journal of Machine Learning Research* 13 (10) (2012) 281–305.
URL <http://jmlr.org/papers/v13/bergstra12a.html>
- [54] P. Novello, G. Poëtte, D. Lugato, P. M. Congedo, Goal-Oriented Sensitivity Analysis of Hyperparameters in Deep Learning, working paper or preprint (Dec. 2021).
URL <https://hal.archives-ouvertes.fr/hal-03128298>
- [55] A. Gretton, O. Bousquet, A. Smola, B. Schölkopf, Measuring statistical dependence with hilbert-schmidt norms, in: Proceedings of the 16th International Conference on Algorithmic Learning Theory, ALT’05, Springer-Verlag, Berlin, Heidelberg, 2005, p. 63–77. doi:10.1007/11564089_7.
URL https://doi.org/10.1007/11564089_7
- [56] H. W. Liepmann, A. Roshko, *Elements of Gasdynamics*, Dover, Mineola, New York, 2001.
- [57] Z. Wang, J. Xin, Z. Zhang, Deepparticle: Learning invariant measure by a deep neural network minimizing wasserstein distance on data generated from an interacting particle method, *Journal of Computational Physics* 464 (2022) 111309. doi:<https://doi.org/10.1016/j.jcp.2022.111309>.

URL <https://www.sciencedirect.com/science/article/pii/S0021999122003710>

- [58] N. Demeure, Gestion du compromis entre la performance et la précision de code de calcul, Theses, Université Paris-Saclay (Jan. 2021).
URL <https://tel.archives-ouvertes.fr/tel-03116750>
- [59] J. Behler, M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, *Phys. Rev. Lett.* 98 (2007) 146401. doi:10.1103/PhysRevLett.98.146401.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.98.146401>
- [60] T. Stecher, N. Bernstein, G. Csányi, Free Energy Surface Reconstruction from Umbrella Samples Using Gaussian Process Regression, *Journal of Chemical Theory and Computation* 10 (9) (2014) 4079–4097. doi:10.1021/ct500438v.
URL <https://pubs.acs.org/doi/10.1021/ct500438v>
- [61] J. Anderson, Hypersonic and High-temperature Gas Dynamics, AIAA education series, American Institute of Aeronautics and Astronautics, 2006.
URL <https://books.google.fr/books?id=UgWmQgAACAAJ>
- [62] J. Scoggins, Development of numerical methods and study of coupled flow, radiation, and ablation phenomena for atmospheric entry, Ph.D. thesis, PhD thesis, Université Paris-Saclay and VKI (2017).
- [63] A. Loeven, J. Witteveen, H. Bijl, Efficient Uncertainty Quantification using a Two-Step Approach with Chaos Collocation, ECCOMAS CFD.
- [64] F. Nobile, R. Tempone, C. Webster, A Sparse Grid Stochastic Collocation Method for Partial Differential Equations with Random Input Data, *SIAM J. Numer. Anal.* 46 (5) (2008) 2309–2345.
- [65] D. Xiu, J. Hesthaven, High-Order Collocation Methods for Differential Equations with Random Inputs, *J. Sci. Comput.* 27(3) (2005) 1118–1139.

- [66] G. Loeven, H. Bijl, Airfoil Analysis with Uncertain Geometry Using the Probabilistic Collocation Method, 49th AIAA Aerospace Sciences Meeting and Exhibit AIAA 2008-2070.
- [67] G. Loeven, J. A. S. Witteveen, H. Bijl, Probabilistic Collocation: an Efficient Non Intrusive Approach for Arbitrarily Distributed Parametric Uncertainties, 45th AIAA Aerospace Sciences Meeting and Exhibit AIAA 2007-317.
- [68] B. Ganapathysubramanian, N. Zabararas, Sparse Grid Collocation Schemes for Stochastic Natural Convection Problems, *J. Comp. Phys.* 225 (2007) 652–685.
- [69] J. F. Epperson, On the runge example, *The American Mathematical Monthly* 94 (4) (1987) 329–341.
URL <http://www.jstor.org/stable/2323093>
- [70] E. Cheney, W. Light, *A Course in Approximation Theory*, Graduate studies in mathematics, American Mathematical Soc.
URL <https://books.google.fr/books?id=mukLPHQsQLUC>
- [71] N. Wiener, The Homogeneous Chaos, *Amer. J. Math.* 60 (1938) 897–936.
- [72] D. Xiu, G. E. Karniadakis, The wiener–askey polynomial chaos for stochastic differential equations, *SIAM J. Sci. Comput.* 24 (2) (2002) 619–644. doi: 10.1137/S1064827501387826.
URL <https://doi.org/10.1137/S1064827501387826>
- [73] G. Blatman, B. Sudret, Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach, *Comptes Rendus Mécanique* 336 (6) (2008) 518–523. doi:10.1016/j.crme.2008.02.013.
URL <https://linkinghub.elsevier.com/retrieve/pii/S1631072108000582>
- [74] L. D., X. D., S. C.H., K. G.E., Predictability and uncertainty in CFD, *International Journal for Numerical Methods in Fluids* 43 (2003) 483–505.

- [75] D. Lucor, C. Enaux, H. Jourden, P. Sagaut, Multi-Physics Stochastic Design Optimization: Application to Reacting Flows and Detonation, *Comp. Meth. Appl. Mech. Eng.* 196 (2007) 5047–5062.
- [76] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, *Reliability Engineering & System Safety* 93 (7) (2008) 964–979. doi:10.1016/j.ress.2007.04.002.
URL <https://linkinghub.elsevier.com/retrieve/pii/S0951832007001329>
- [77] X. Wan, G. Karniadakis, Beyond Wiener-Askey Expansions: Handling Arbitrary PDFs, *SIAM J. Sci. Comp.* 27(1-3).
- [78] G. Poëtte, D. Lucor, Non Intrusive Iterative Stochastic Spectral Representation with Application to Compressible Gas Dynamics, *J. of Comput. Phys.* DOI information: 10.1016/j.jcp.2011.12.038.
- [79] A. Birolleau, G. Poëtte, D. Lucor, Adaptive bayesian inference for discontinuous inverse problems, application to hyperbolic conservation laws, *Communications in Computational Physics* 16 (1) (2014) 1–34. doi:10.4208/cicp.240113.071113a.
- [80] Ernst, Oliver G., Mugler, Antje, Starkloff, Hans-Jörg, Ullmann, Elisabeth, On the convergence of generalized polynomial chaos expansions, *ESAIM: M2AN* 46 (2) (2012) 317–339. doi:10.1051/m2an/2011045.
URL <https://doi.org/10.1051/m2an/2011045>
- [81] G. Poëtte, A comparative study of generalized Polynomial Chaos based Approximations: integration vs. regression vs. collocation vs. kriging, *International Journal for Uncertainty Quantification*.
URL <https://hal.archives-ouvertes.fr/hal-01831191>
- [82] C. E. Rasmussen, C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*, The MIT Press, 2005.

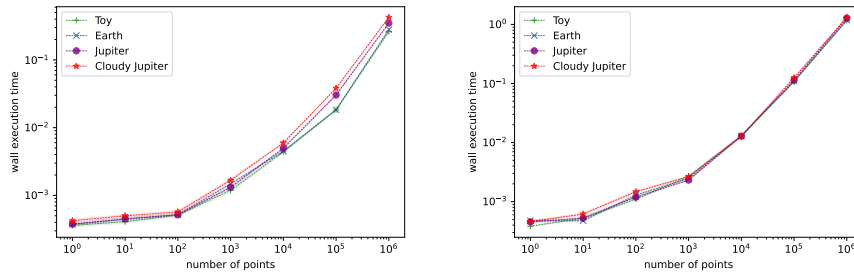
- [83] J. E. Oakley, A. O’Hagan, Probabilistic sensitivity analysis of complex models: a bayesian approach, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66 (3) (2004) 751–769. [arXiv:https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2004.05304.x](https://arxiv.org/abs/https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2004.05304.x), doi:<https://doi.org/10.1111/j.1467-9868.2004.05304.x>.
URL <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2004.05304.x>
- [84] A. Marrel, N. Marie, M. De Lozzo, Advanced surrogate model and sensitivity analysis methods for sodium fast reactor accident assessment, *Reliability Engineering & System Safety* 138 (2015) 232–241. doi:<https://doi.org/10.1016/j.ress.2015.01.019>.
URL <https://www.sciencedirect.com/science/article/pii/S0951832015000290>
- [85] M. C. Kennedy, A. O’Hagan, Bayesian calibration of computer models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (3) (2001) 425–464. [arXiv:https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9868.00294](https://arxiv.org/abs/https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9868.00294), doi:<https://doi.org/10.1111/1467-9868.00294>.
URL <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00294>
- [86] R. B. Gramacy, *Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences*, Chapman Hall/CRC, Boca Raton, Florida, 2020, <http://bobby.gramacy.com/surrogates/>.
- [87] F. Bachoc, Estimation paramétrique de la fonction de covariance dans le modèle de krigeage par processus gaussiens : application à la quantification des incertitudes en simulation numérique, Ph.D. thesis, thèse de doctorat dirigée par Garnier, Josselin *Mathématiques appliquées Paris 7* 2013 (2013).
URL <http://www.theses.fr/2013PA077111>
- [88] F. Bachoc, Asymptotic analysis of the role of spatial sampling for covariance

parameter estimation of Gaussian processes, ArXiv e-printsarXiv:1301.4321.

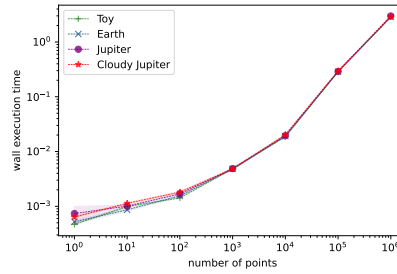
- [89] G. Poëtte, A comparative study of generalized Polynomial Chaos based Approximations: integration vs. regression vs. collocation vs. kriging, working paper or preprint (Jul. 2018).
URL <https://hal.archives-ouvertes.fr/hal-01831191>
- [90] D. R. Jones, M. Schonlau, W. J. Welch, Efficient global optimization of expensive black-box functions, *J. of Global Optimization* 13 (4) (1998) 455–492. doi:10.1023/A:1008306431147.
URL <https://doi.org/10.1023/A:1008306431147>
- [91] T. J. Santner, W. B., N. W., *The Design and Analysis of Computer Experiments*, Second Edition, Springer-Verlag, 2018.
- [92] R. Shöbi, B. Sudret, J. Wiart, Polynomial-chaos-based kriging.
- [93] P. Kersaudy, B. Sudret, N. VARSIER, O. Picon, J. Wiart, A new surrogate modeling technique combining Kriging and polynomial chaos expansions – Application to uncertainty analysis in computational dosimetry, *Journal of Computational Physics* 286 (2015) 130–117. doi:10.1016/j.jcp.2015.01.034.
URL <https://hal.archives-ouvertes.fr/hal-01143146>
- [94] R. Schöbi, B. Sudret, S. Marelli, Rare Event Estimation Using Polynomial-Chaos Kriging, *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 3 (2) (2017) D4016002. doi:10.1061/AJRUA6.0000870.
- [95] F. Bachoc, Asymptotic analysis of covariance parameter estimation for Gaussian processes in the misspecified case, ArXiv e-printsarXiv:1412.1926.

Appendix A: Additional plots

In this appendix, we gather additional plots of the effect of the neural network input dimension on its execution time for different widths. These plots are complementary with those of Section 3.3.



(a) Test cases with a network of width 40 (b) Test cases with a network of width 160



(c) Test cases with a network of width 320

Figure 14: Execution time of a neural network of a given width for the different test cases with respect to the number of input points (with log axes).

Appendix B: Hyperparameter search space

This table shows the hyperparameters considered in the hyperparameter search of Section 4, as well as their possible values.

Appendix C: Detailed description of the Physical phenomenon

In equation (1), the different quantities are defined by (we drop the spatial and time dependences for the sake of conciseness)

$$U = \begin{pmatrix} \rho_1 \\ \dots \\ \rho_{n_e} \\ \rho v \\ \rho w \\ \rho E \end{pmatrix}, F_x(U) = \begin{pmatrix} \rho_1 v \\ \dots \\ \rho_{n_e} v \\ \rho v^2 + p \\ \rho v w \\ \rho u(E + \frac{p}{\rho}) \end{pmatrix}, F_y(U) = \begin{pmatrix} \rho_1 w \\ \dots \\ \rho_{n_e} w \\ \rho v w \\ \rho w^2 + p \\ \rho w(E + \frac{p}{\rho}) \end{pmatrix}.$$

The first n_e equations stand for the conservation of the mass of the different elements of the fluid (N and O typically for the two elements of reaction $N + O \rightleftharpoons NO$). The partial density of element k can be expressed as $\rho_k = \sum_{i=1}^{n_s} a_i^k \frac{m_k^e}{m_i} \rho_i$, where m_k^e and m_i are molar masses of element k and species i while a_i^k is the number of the k^{th} element in species i . Partial mass of the n_s species are denoted by $(\rho_i)_{i \in \{1, \dots, n_s\}}$. The density of the fluid can be deduced from the partial densities of the elements as $\rho = \sum_{k=1}^{n_e} \rho_k$. Besides, v and w are respectively the horizontal and vertical velocities of the fluid so that the equations on ρv and ρw in equation (1) ensure the conservation of momentum. Finally, $E = \epsilon + \frac{v^2 + w^2}{2}$ is the total energy of the fluid and ϵ its internal energy. The last equation of (1), on ρE , ensures the conservation of the total energy of the fluid. The system remains to be closed: we need one last equation to relate the pressure p to the other quantities.

The first solution for the latter purpose is to make the hypothesis that chemical reactions have a negligible effect. It is the case, for example, when the fluid is considered a perfect gas, in which case p is related to ϵ and directly to ρ with

$$p = (\gamma - 1)\rho\epsilon,$$

where γ is the Laplace constant. In the case of a diatomic gas, for example, $\gamma = 1.4$. When chemical reactions are neglected, the closure is simple and

computationally fast, and the system can be closed without updating the partial densities of the elements $(\rho_i)_{i \in \{1, \dots, n_s\}}$. However, this hypothesis is coarse (see figure 2 and the related comments). In a more general case, the pressure p can be accurately computed by simulating the chemical equilibrium between the species produced by the N_r chemical reactions that occur during the dynamic. Every elementary chemical reaction $r \in \{1, \dots, N_r\}$ can be described through the general formula,

$$\sum_{i=1}^{n_s} \nu_{ir} A_i \rightleftharpoons 0,$$

where n_s is the number of species (i.e. $n_s = 3$ in the case of reaction $N + O \rightleftharpoons NO$ for N, O and NO) and where $(\nu_{ir})_{i \in \{1, \dots, n_s\}}$ are the forward minus reverse stoichiometric coefficients for species $(A_i)_{i \in \{1, \dots, n_s\}}$ in reaction r (i.e. $\nu_{ir} > 0$ if A_i disappears in reaction r , $\nu_{ir} < 0$ if A_i appears in reaction r and $\nu_{ir} = 0$ if A_i is not involved in reaction r). In this case, the pressure is given by (see [61])

$$\begin{cases} p(\rho, \epsilon, \mathbf{x}) = \frac{\rho R T(\epsilon, \mathbf{x})}{\sum_{i=1}^{n_s} x_i m_i}, \\ T(\epsilon, \mathbf{x}) = \frac{\epsilon m - \sum_{i=1}^{n_s} x_i m_i h_i^0}{\sum_{i=1}^{n_s} x_i m_i C_{vi}}, \\ \Delta G(\mathbf{x}, U) = 0. \end{cases} \quad (6)$$

In the above equations, R is the universal gas constant, $(h_i^0)_{i \in \{1, \dots, n_s\}}$ and $(C_{vi})_{i \in \{1, \dots, n_s\}}$ are respectively the mass enthalpies of formation and the mass heat capacities at constant volume of the n_s species. These expressions also involve their molar fractions and molar masses $\{x_i\}_{i \in \{1, \dots, n_s\}}$ and $\{m_i\}_{i \in \{1, \dots, n_s\}}$. Finally, the vector of mass fractions \mathbf{x} is obtained by minimizing the Gibbs free energy G , which is implicitly recalled by the last equation of the system (6), i.e. *via* the fact that \mathbf{x} cancels ΔG , the differential of G . Note that the minimization of the Gibbs free energy depends on the vector of unknowns of the Euler system U and on the molar fractions \mathbf{x} . In a nutshell, at equilibrium, we must have

$$\Delta G(\mathbf{x}, U) = \sum_{r=1}^{N_r} \sum_{i=1}^{n_s} \nu_{ir} G_i(x_i, p, T) = \sum_{i=1}^{n_s} \nu_i G_i(x_i, p, T) = 0,$$

where $(G_i)_{i \in \{1, \dots, n_s\}}$ are the Gibbs free energy of species $i \in \{1, \dots, n_s\}$ per mole of i , see [61, 7, 62] for more details.

Finally, the resolution of the whole strongly coupled system of equations can be summed up by $\forall t \in \mathbb{R}^+$

$$\begin{cases} \partial_t U(x, y, t) + \nabla \cdot F(U(x, y, t), \mathbf{x}(x, y, t)) = 0, & \forall x, y \in \mathcal{D}, \\ \Delta G(\mathbf{x}(x, y, t), U(x, y, t)) = 0, & \forall x, y \in \mathcal{D}, \\ U(x, y, t) = b(x, y, t), & \forall x, y \in \partial\mathcal{D}, \end{cases} \quad (7)$$

in which the dependence on \mathbf{x} is made explicit in the expression of the flux F .

Now, for our reentry problem, system (7) must be solved for long times (stationary problems). In practice, a second order in time splitting is operated so that one iteration of the resolution, which is closely related to a time step $[t^n, t^{n+1} = t^n + \Delta t]$, consists in the resolution of

$$\begin{cases} \partial_t U(x, y, t) + \nabla \cdot F(U(x, y, t), \mathbf{x}(x, y, t^n)) = 0, & \forall x, y \in \mathcal{D}, \\ U(x, y, t) = b(x, y, t), & \forall x, y \in \partial\mathcal{D}, \end{cases} \quad (8)$$

during time step $\forall t \in [t^n, t^{n+\frac{1}{2}}]$, followed by the resolution of

$$\Delta G(\mathbf{x}(x, y, t^{n+1}), U(x, y, t^{n+\frac{1}{2}})) = 0, \quad \forall x, y \in \mathcal{D}, \quad (9)$$

during time step $\forall t \in [t^{n+\frac{1}{2}}, t^{n+1}]$.

Appendix D: A taxonomy of surrogate models

Lagrange interpolation or collocation. (and higher order ones such as Hermite interpolation) [63, 64, 65, 66, 67, 68, 69]: they are based on the choices

- $L(x, y) = (x - y)^2$ in equation (3),
- $u(X, \theta) = \sum_{k=0}^P \theta_k X^k$, i.e. a polynomial approximation and a linear application $\theta \rightarrow u(X, \theta)$ with respect to variable θ .
- together with $N = P + 1 = d_\theta$.

This surrogate model ensures $u(X_i) = u(X_i, \theta) \forall i \in \{1, \dots, N\}$. Besides, spectral convergence can be achieved [65], hence very good accuracies, but the convergence behavior strongly depends on the choice of the experimental design $(X_i, w_i)_{i \in \{1, \dots, N\}}$ discretising $(X, d\mathcal{P}_X)$ (see the divergence of the approximation of Runge's function with uniform points [70, 69]). Finally, if N is huge (i.e. in a big data context), $P \sim N$ is huge and the run-time of the ML model strongly depends on the size of the database. It may become prohibitive for the desired accuracy.

Polynomial regression and generalised Polynomial Chaos. [71, 72, 73, 10, 11, 74, 12, 75, 14, 76, 13, 77] are very popular, especially in problems of Uncertainty Quantification (UQ). Since the seminal work of [10], it is extensively used in (non-intrusive) uncertainty propagation [11, 74, 12, 75, 14, 78, 79, 77]. These surrogate models are based on the choices:

- $L(x, y) = (x - y)^2$ in equation (3),
- $u(X, \theta) = \sum_{k=0}^P \theta_k \phi_k(X)$, is linear with respect to θ and $(\phi_k(X))_{k \in \{0, \dots, P\}}$ are orthonormal polynomials with respect to the scalar product defined by $d\mathcal{P}_X$.
- The number of parameters $d_\theta = P + 1$ is not constrained by N .

Spectral convergence with respect to P is ensured [80]. The orthonormality of the basis helps with round-off errors and conditioning [81] while having a model for which N and P are not correlated anymore (i.e. we can take $N \gg P$, big datasets, with a model having a run-time depending on $P \ll N$).

Gaussian Process regression or Kriging. Popularized for ML by [82] and intensively used in Uncertainty Quantification (UQ) [83, 15, 84, 85, 86, 87, 88, 89, 90, 91], this technique has established to a leading position in surrogate modeling. It can be summed up as taking

- $L(x, y) = (x - y)^2$ in equation (3),
- $u(X, \theta) = \sum_{k=0}^P \theta_k \phi_k(X) + Z(\theta_{P+1}, \dots, \theta_{d_\theta})$,
 - where $(\phi_k(X))_{k \in \{0, \dots, P\}}$ can be orthonormal polynomials with respect to the scalar product defined by $d\mathcal{P}_X$ as in [92, 93, 94, 81] or classical polynomials [87],
 - and where Z is a gaussian process conditioned to satisfy $u(X_i) = u(X_i, \theta) \forall i \in \{1, \dots, N\}$. Some particular shapes of Z are *a priori* determined by choosing particular covariance functions for the process [87, 88, 95].
 - The ML model is linear with respect to $(\theta_0, \dots, \theta_P)$ and nonlinear with respect to $(\theta_{P+1}, \dots, \theta_{d_\theta})$.
- The number of parameters d_θ is not constrained by N .

This type of ML model can be understood as a way to make the best of the two previous approaches as we have $u(X_i) = u(X_i, \theta) \forall i \in \{1, \dots, N\}$ together with convergence properties for the mean [92, 93, 94, 81] and with having $N > d_\theta$. But the run-time of such an ML model still strongly depends on N . For huge databases, these models can be very accurate but far from being cost-effective.

hyperparameter	type	values
<code>n_layers</code>	integer	$\in \{1, \dots, 10\}$
<code>n_units</code>	integer	$\in \{7, \dots, 512\}$
<code>activation</code>	categorical	<code>elu</code> , <code>relu</code> , <code>tanh</code> or <code>sigmoid</code>
<code>dropout</code>	boolean	<code>true</code> or <code>false</code>
<code>dropout_rate</code>	continuous	$\in [0, 1]$
<code>batch_norm</code>	boolean	<code>true</code> or <code>false</code>
<code>learning_rate</code>	continuous	$\in [1 \times 10^{-6}, 1 \times 10^{-2}]$
<code>weights_reg_l1</code>	continuous	$\in [1 \times 10^{-6}, 0.1]$
<code>weights_reg_l2</code>	continuous	$\in [1 \times 10^{-6}, 0.1]$
<code>bias_reg_l1</code>	continuous	$\in [1 \times 10^{-6}, 0.1]$
<code>bias_reg_l2</code>	continuous	$\in [1 \times 10^{-6}, 0.1]$
<code>batch_size</code>	integer	$\in \{1, \dots, 500\}$
<code>loss_function</code>	categorical	L_2 error or L_1 error
<code>optimizer</code>	categorical	<code>adam</code> , <code>sgd</code> , <code>rmsprop</code> , <code>adagrad</code> or <code>nadam</code>
<code>amsgrad</code>	boolean	<code>true</code> or <code>false</code>
<code>1st_moment_decay</code>	continuous	$\in [0.8, 1]$
<code>2nd_moment_decay</code>	continuous	$\in [0.8, 1]$
<code>centered</code>	boolean	<code>true</code> or <code>false</code>
<code>nesterov</code>	boolean	<code>true</code> or <code>false</code>
<code>momentum</code>	continuous	$\in [0.5, 0.99]$
<code>n_seeds</code>	integer	$\in \{1, \dots, 10\}$
<code>sampling</code> [52]	boolean	<code>true</code> or <code>false</code>
<code>weighting</code> [52]	boolean	<code>true</code> or <code>false</code>

Table 4: Hyperparameters values for Mutation++ approximation. We also include the weighting and the sampling scheme of [52] as hyperparameters.