Development of a multi-species real fluid modelling approach using a machine learning method, application to combustion.

[1]Bruno Delhom*, [1]Chaouki Habchi, [1]Olivier Colin and [1]Julien Bohbot;

[1]IFP Energies Nouvelles, Institut Carnot Transports Energies, 1 et 4 Avenue de Bois-Préau, 92852 Rueil-Malmaison, France

* Corresponding author: bruno.delhom@ifpen.fr

Keywords: Subcritical, supercritical, transcritical flow, real fluid model, VLE, Neural network, ISAT

Abstract. Two-phase flow configurations present many challenges for numerical modelling, requiring the development of Real Fluid Models (RFM) able to simulate flows in subcritical, transcritical or supercritical regimes. Such a RFM model has been recently developed at IFPEN. This paper proposes a neural network (NN) approach to overcome the limitations of lookup tables of thermodynamic properties. We also propose a methodology for generating an optimized data set by combining a vapor-liquid equilibrium (VLE) thermodynamic solver and the in Situ Adaptive Tabulation (ISAT) method which aims to improve the neural network training process for combustion simulations where too many species are present. This NN methodology has been implemented in the CONVERGE CFD V4.1 solver and validated using an academic H_2/O_2 combustion case. The results show that this NN approach makes H_2 combustion simulation possible when coupled to the RFM framework and using a 10-species kinetic mechanism. This validation paves the way for more complex combustion cases, currently under study.

1. Introduction

In principle, the use of a real-fluid model (RFM) for the computation of the required physical properties in a CFD solver can be done through different methods such as direct coupling to vapor-liquid equilibrium (VLE) solver, lookup tabulation, or using an Artificial Neural Networks (ANN) model [1-3]. Direct VLE coupling can be very expensive in terms of CPU cost [1-2], while tabulation may require large memory allocations, making the RFM framework difficult to extend as the number of chemical species increases. ANNs have emerged as a highly effective technique for classification and response prediction, surpassing traditional linear methods like regression analysis. Their superior prediction abilities have been extensively demonstrated, leading to successful applications in various domains of engineering and science. We have already demonstrated in a previous communication [9] that ANNs are suitable and efficient for CFD simulation of cryogenic LOx/GH2 mixing layer in the RFM framework.

In this work, we propose an improved artificial neural network (ANN) method capable of handling dozens of chemical species produced during combustion. A key challenge is the generation of a thermodynamic database that enables a fine-tuned description of the neural networks. As with all neural network training processes, the dataset is the most critical input and must be carefully selected. To accomplish this, we employed an in situ adaptive tabulation (ISAT)-based sorting tree [10]. In this paper, the proposed RFM-ANN model is applied to the CFD simulation of gaseous H_2/O_2 combustion in a 0D case using a 10-species mechanism. The paper is structured as follows: Section 2 provides a brief overview of the equations governing the RFM model. Section 3 details the database generation process using ISAT. Section 4 presents the ANN design methods. Section 5 discusses the results of H_2/O_2 combustion simulations, followed by conclusions in Section 6.

2. The RFM governing equations

The diffused interface two-phase flow model adopted in the current work is a four-equation model that is fully compressible and considers multi-component in both phases under the assumptions of thermal and mechanical equilibrium, as follows,

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u_i}{\partial x_i} = 0 \tag{1}$$

$$\frac{\partial \rho u_i}{\partial t} + \frac{\partial \rho u_i u_j}{\partial x_j} = -\frac{\partial P}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j}, \quad \tau_{ij} = \mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{2}{3}\frac{\partial u_k}{\partial x_k}\delta_{ij}\right) \tag{2}$$

$$\frac{\partial \rho e}{\partial t} + \frac{\partial \rho e u_j}{\partial x_j} = -P\frac{\partial u_j}{\partial x_j} + \tau_{ij}\frac{\partial u_i}{\partial x_j} + \frac{\partial}{\partial x_j}\left(\lambda\frac{\partial T}{\partial x_j}\right) + \frac{\partial}{\partial x_j}\left(\rho D\sum\nolimits_k h_k\frac{\partial Y_k}{\partial x_j}\right) \tag{3}$$

$$\frac{\partial \rho Y_k}{\partial t} + \frac{\partial \rho Y_k u_j}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\rho D\frac{\partial Y_k}{\partial x_j}\right) \tag{4}$$

where (τ_{ij}) is the viscous stress tensor, (ρ) is the density, (u_i) is the velocity, (P) is the pressure, (e) is the specific internal energy, (T) is the temperature. Besides, (Y_k, h_k) are the mass fraction and specific enthalpy of species k, respectively. The thermal conductivity (λ) and the dynamic viscosity (μ) covers laminar and turbulent contributions. The laminar contribution of (λ, μ) is computed by Chung et al. [4] correlations. The turbulent conductivity is calculated using a given turbulent Prandtl number and the turbulent viscosity is given by the adopted turbulence model. Finally, the laminar and turbulent mass diffusion coefficients are estimated using a given Schmidt number.

Originally, the RFM model is closed by a thermodynamic equilibrium tabulation method [3-6]. The thermodynamic table is generated using the in-house IFPEN CARNOT library [5] based on VLE calculations coupled with a real-fluid equation of state (EoS), such as Peng-Robinson (PR-EoS) and Soave-Redlich-Kwong (SRK-EoS) [4-5]. The thermodynamic and transport properties of the mixture, such as the density (ρ), the specific internal energy (e), *Cp*, *Cv*, speed of sound (*Cs*), thermal conductivity (λ) and the dynamic viscosity (μ) are calculated as a function of temperature (*T*), pressure (*P*), and mass fraction of the species in the mixture (Y_k) and stored into a table.

Initially, the RFM model employed a lookup procedure using an interpolation on a regular discretization of temperature, pressure and mass fraction [4-6]. However, for cases involving more than three species, the size of the generated table became prohibitively large, leading to excessive memory consumption and computational overhead. To overcome these limitations, a neural network-based methodology is proposed as an alternative to the traditional tabulation approach. This ANN framework enables efficient and scalable CFD combustion simulations, even for cases involving a large number of species, by significantly reducing storage requirements while maintaining accuracy and computational efficiency.

3. Data base generation using an ISAT method

To generate the database for the initial binary system (O2/H2) of interest in this study, we first used CARNOT [5] to generate a regular table for the system (O2/H2) in the same way as for the binary or ternary system tabulation method used previously. The density in the plan P=105 bars of this regular database is depicted in Figure 1(a). It can be observed in this figure that many points describe the same density solution, and that in areas where density increases rapidly (near to liquid-vapor interface), the description will be fragmentary and less refined. We therefore need to adopt a new strategy for generating training databases. The proposed new strategy is based on the use of the ISAT method to identify and better refine high-gradient areas and exclude areas with low density variations, for example as shown in Figure 1(b). This method has the advantage of being able to sort the data on the fly in a binary tree construction.

The function used for the ISAT sorting tree is the CARNOT(P,T, $\boldsymbol{Y_k}$)->[$\boldsymbol{\rho}$, $\boldsymbol{e}$,Cv,Cp,Cs, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$]. At each

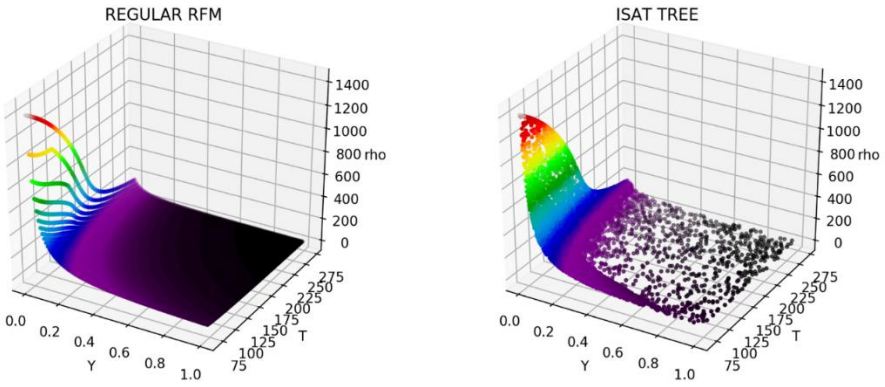

Figure 1 (a) Regular Data generation vs (b) ISAT Data Generation (isobar scatterplot for P≈10500000.0).

intermediate node, we store a midpoint coordinate and a normal vector. Each leaf node contains an input vector along with its corresponding function solution. If the solution being tested is sufficiently close to an existing one, it is discarded. Otherwise, a new leaf is created, the existing one is moved, and a new midpoint along with a new normal vector is stored. This iterative process efficiently sorts solutions while retaining only the most relevant quantities [Figure 2], X denotes the input vector (P,T, $\boldsymbol{Y_k}$) and Y the output vector ($\boldsymbol{\rho}$, $\boldsymbol{e}$,Cv,Cp,Cs, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$).

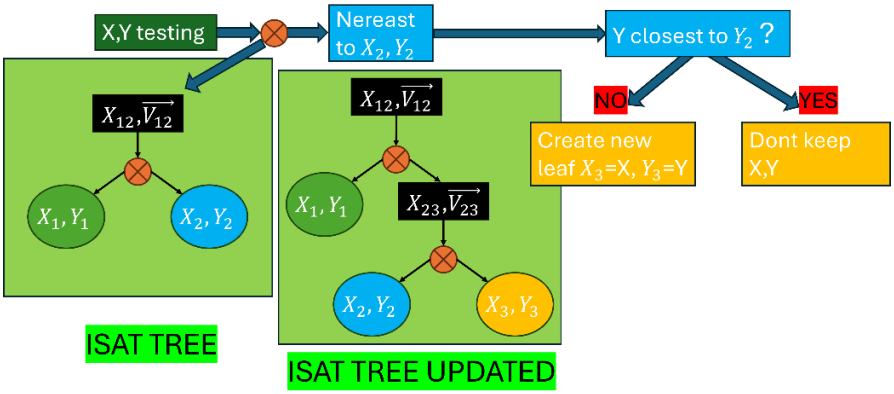

Figure 2 ISAT processus to sort the database

We immediately observe that the density (denoted ρ) is described with higher resolution in regions of large gradient values, whereas flatter regions (plateaus) require fewer data points in the sorted database. In this context, the number of points required to define the training database was reduced from 1,030,301 points (101 × 101 × 101) in the regular database to 557,277 points in the ISAT-sorted database.

4. Artificial Neural Network generation

The architecture of our artificial neural network (ANN) models was systematically optimized using the OPTUNA software (see [7] for example). This powerful hyperparameter optimization tool allowed us to explore a wide range of network configurations and identify the most suitable architecture for our specific problem. By leveraging OPTUNA's automated search capabilities, we were able to fine-tune key parameters such as the number of layers, the number of neurons per layer, the activation functions, and the learning rate. This approach ensured that our ANN models achieved optimal performance while maintaining computational efficiency. Initially, we employed a single artificial neural network (ANN) model to predict various thermodynamic parameters essential for computational fluid dynamics (CFD) simulations. While this approach proved successful for certain parameters, it was insufficient for others,

particularly density (ρ) and internal energy (*Sie*). These two parameters play a crucial role in ensuring numerical stability and convergence in CFD simulations, making their accurate prediction essential.

The limitations of the single ANN model led us to explore alternative strategies to enhance the predictive capabilities of our neural network framework. To facilitate the development of a more robust neural model, we implemented a hybrid solver capable of running both the RFM-Tabulation model and the RFM-ANN model simultaneously. This hybrid approach allowed us to systematically assess and control the error introduced by the ANN predictions in the CFD computations. By comparing ANN-generated results with those obtained using the tabulation model, we were able to identify discrepancies and refine our methodology accordingly. This validation step was critical in ensuring that the ANN models did not introduce significant numerical errors that could compromise the accuracy of the CFD simulations. Through extensive error analysis, we observed that a single ANN structure was unable to provide uniformly high accuracy across all target parameters. Some parameters exhibited higher sensitivity to neural network approximations, which led to greater deviations in their predicted values. As a result, we determined that it was more effective to decompose the ANN output space into three distinct sets, each handled by a separate neural network. This approach allowed for specialized training and optimization of each ANN, ensuring that the predictions remained as accurate as possible while maintaining computational efficiency. By adopting this multi-network approach, we significantly improved the robustness and reliability of our ANN-based solver. The division of outputs into three separate networks not only reduced overall prediction errors but also enhanced the stability of CFD simulations, leading to better numerical convergence. This structured methodology highlights the importance of tailoring ANN architectures to the specific requirements of CFD applications, ensuring both accuracy and computational efficiency in large-scale simulations. The first two artificial neural networks (ANNs) are dedicated to predicting the two most critical parameters for computational fluid dynamics (CFD) simulations: density (ρ) and internal energy (*Sie*). These parameters are essential for ensuring numerical stability and convergence, making their accurate estimation a priority. The third neural network is responsible for predicting the remaining thermodynamic parameters, collectively denoted as d, as illustrated schematically in Figure 3. This strategic subdivision of the ANN model architecture was implemented to enhance both accuracy and computational efficiency. By allocating separate networks to different parameter groups, we optimized each ANN's learning process to focus on a specific subset of variables. This specialization allowed for more precise predictions, as each network could be fine-tuned to the unique characteristics and sensitivities of its respective parameters. Moreover, this multi-network approach helped mitigate potential sources of numerical error that arise when attempting to train a single ANN to handle a wide range of parameters with varying degrees of complexity. The tailored structure improved generalization across different CFD conditions while maintaining computational efficiency, ultimately leading to more reliable and stable simulation outcomes.

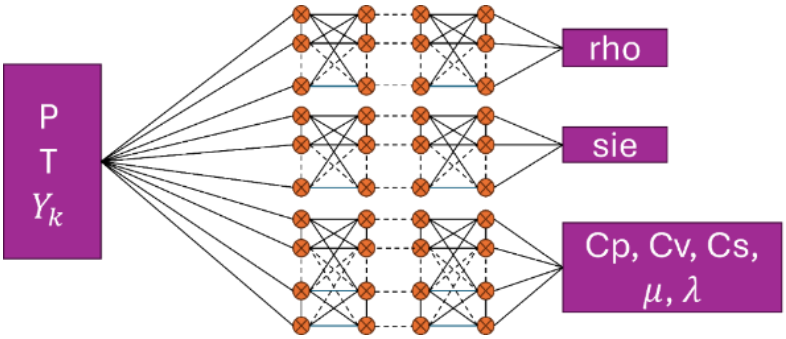

Figure 1 Optimized ANN's design using three dedicated models for the density ρ, internal energy (Sie) and the rest.

5. H2/O2 combustion test case

The SAGE detailed chemistry solver in CONVERGE CFD 4.1 [8] computes reaction rates using local conditions and the fundamental principles of chemical kinetics. Fully coupled with the flow solver, SAGE provides a highly accurate and predictive approach to modelling combustion processes. It effectively captures key combustion phenomena, including ignition dynamics and turbulent flame propagation, making it a robust tool for various combustion simulations. To evaluate the proposed methodology, we simulate a fundamental combustion case. This case corresponds to a homogeneous reactor with an initial gas mixture and thermodynamic conditions that allow auto-ignition of the mixture (initial temperature at 1200 °K). The selected test case is a 0D homogeneous H_2/O_2 mixture, using a detailed 10-species kinetic scheme. Due to the simplicity of this setup, which consists of only eight computational cells, direct coupling with the Carnot model is feasible. This allows us to perform a comparative analysis by simulating the H_2/O_2 combustion process using two different approaches: (1) a reference simulation with SAGE, a detailed chemistry solver coupled to our neural network-based approach; and (2) direct coupling with Carnot. By conducting these comparisons, we aim to assess the accuracy and computational efficiency of the neural network method relative to both a high-fidelity chemistry solver and a direct physics-based approach. Figures 4 and 5 illustrate the pressure, temperature and species mass fractions histories obtained using both methods. As observed in these figures, the RFM-ANN approach successfully models combustion while being coupled with the RFM framework. Table 1 presents the execution times, highlighting a significant reduction in computational time when using the RFM-ANN approach compared to the conventional RFM-coupled method. This demonstrates the efficiency gains achieved through the neural network-based approach. Looking ahead, future implementations will leverage GPU-based inference to further accelerate ANN computations, drastically reducing the solving time and enhancing overall performance.

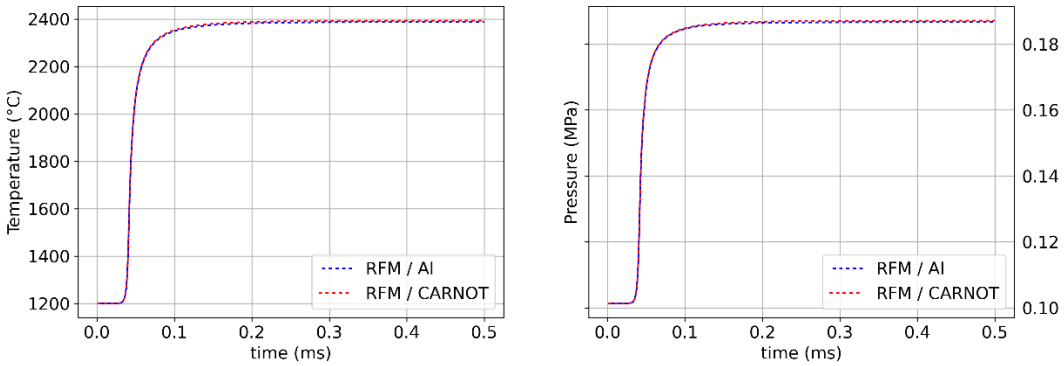

Figure 4: (Left) Temperature and (Right) Pressure evolution, without RFM, with the RFM/ANN and with the RFM/CARNOT

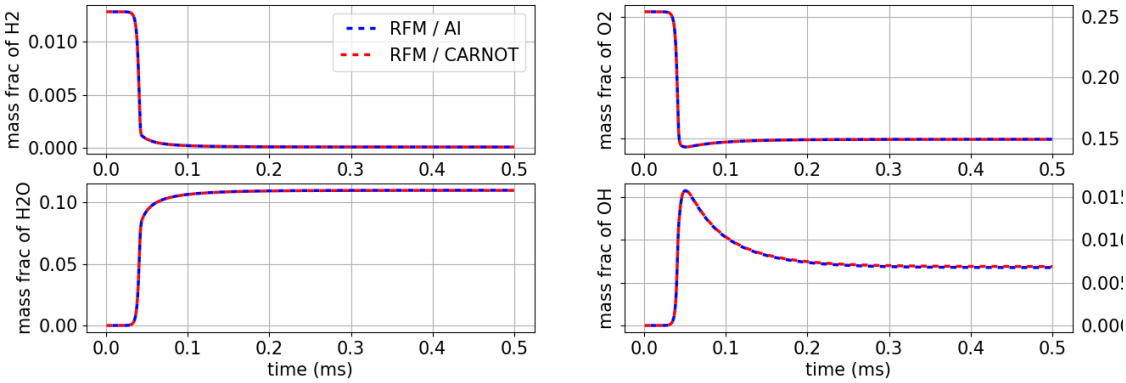

Figure 5: Mass Fraction evolution for H2/O2/H2O and OH, without RFM, with the RFM/ANN and with the RFM/CARNOT

	RFM/AI	RFM/CARNOT
Execution time	115,66 s	2300,62 s
Normalise exec time	1	19,89

Table 1 Execution Time

6. Conclusion

In this paper, we presented an innovative multi-species real fluid modelling methodology using an artificial neural network (ANN)-based approach. The integration of this methodology into the CONVERGE CFD solver has significantly improved the accuracy and efficiency of H_2/O_2 combustion simulations while drastically reducing computational costs compared to traditional thermodynamic tabulation methods. The use of In Situ Adaptive Tabulation (ISAT) for optimized database generation played a key role in enhancing the performance of the ANN models by targeting high-gradient thermodynamic regions while minimizing redundant data. Furthermore, decomposing the ANN model into three specialized networks increased the robustness and stability of the simulations, ensuring more accurate predictions of critical parameters such as density and internal energy. Results obtained from a homogeneous H_2/O_2 combustion test case demonstrate that our approach effectively captures the evolution of thermodynamic properties and species mass fractions while providing an efficient alternative to conventional tabulation methods. These advancements pave the way for more complex studies considering realistic configurations with liquid and gas. Ultimately, this methodology could be extended to other applications in combustion and fluid dynamics, enabling faster and more accurate modelling of thermodynamic phenomena in complex multi-species environments.

7. References

[1] Yang, Songzhi et al. “Real-fluid injection modeling and LES simulation of the ECN Spray A injector using a fully compressible two-phase flow approach.” International Journal of Multiphase Flow 122 (2020): 103145.
[2] Ping, Yi & Yang, Songzhi & Habchi, Chaouki. (2019). A multicomponent real-fluid fully compressible four-equation model for two-phase flow with phase change. Physics of Fluids. 31. 026102. 10.1063/1.5065781.
[3] Jafari, Sajad & Gaballa, Hesham & Habchi, Chaouki & Hemptinne, Jean-Charles & Mougin, Pascal. (2022). Exploring the Interaction between Phase Separation and Turbulent Fluid Dynamics in Multi-Species Supercritical Jets using a Tabulated Real-Fluid Model. The Journal of Supercritical Fluids. 184. 105557. 10.1016/j.supflu.2022.105557.
[4] Chung, Ting Horng et al. “Generalized multiparameter correlation for nonpolar and polar fluid transport properties.” Industrial & Engineering Chemistry Research 27 (1988): 671-679.
[5] Jean-Charles de Hemptinne, Nicolas Ferrando, Martha Hajiw-Riberaud, Véronique Lachet, Saheb Maghsoodloo, et al.. Carnot : a thermodynamic library for energy industries. Science and Technology for Energy Transition, 2023, 78, pp.30. ⟨10.2516/stet/2023023⟩. ⟨hal-04289506⟩
[6] Jafari, Sajad & Gaballa, Hesham & Habchi, Chaouki & Hemptinne, Jean-Charles. (2021). Towards Understanding the Structure of Subcritical and Transcritical Liquid-Gas Interfaces Using a Tabulated Real Fluid Modeling Approach. Energies. 14. 38. 10.3390/en14185621.
[7] Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M. (2019) Optuna: A Next-Generation Hyperparameter Optimization Framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, 4-8 August 2019, 2623-2631.
[8] Richards, K.J., Senecal, P.K., and Pomraning, E., CONVERGE 4.0.2, Convergent Science, Madison, WI (2025).
[9] Delhom, Bruno & Faney, Thibault & Mcginn, Paul & Habchi, Chaouki & Bohbot, Julien. (2023). Development of a multi-species real fluid modelling approach using a machine learning method. ILASS-EUROPE, NAPLES.
[10] Pope, S. B. (1997). Computationally efficient implementation of combustion chemistry using in situ adaptive tabulation. Combustion Theory and Modelling, 1(1), 41–63. https://doi.org/10.1080/713665229