Self-learning Digital Twin of a combustion furnace through the Kalman Filter method

L. Donato[1,2,3], A. Procacci[1,2], C. Galletti[3], A. Coussement[1,2] and A. Parente[1,2]
laura.donato@ulb.be
[1]Aero-Thermo-Mechanics Lab., Université Libre de Bruxelles, Avenue F. D. Roosevelt, 1050, Brussels, Belgium
[2]BRITE - Brussels Institute for Thermal-fluid systems and clean Energy, Brussels, Belgium
[3]Department of Civil and Industrial Engineering, University of Pisa, Largo L. Lazzarino 2, 56122, Pisa, Italy

Abstract
This study aims to develop a self-learning Digital Twin (DT) model using a data assimilation approach, i.e., the Kalman Filter, which can adjust a numerical model with the sparse sensing (SpS) model. The numerical model, i.e., the DT, is obtained by coupling a dimensionality-reduction method, the Proper Orthogonal Decomposition (POD), and the Gaussian Process Regression (GPR) model for several Reynolds-averaged Navier-Stokes (RANS) simulations of a semi-industrial combustion furnace. The SpS technique leverages dimensionality reduction to predict the state of the system using experimental measurements as input. Finally, the DA framework has been used to provide a solution with lower uncertainty bounds. The results show that this framework can be used to upgrade the DT by considering a SpS model built from experimental values.

Introduction
Meeting society's energy needs is challenging due to intermittent renewable sources and the requirement for long-term storage and high-density energy for transportation and manufacturing. One solution is storing excess renewable energy as synthetic fuels, integrated with combustion systems. Novel combustion technologies such as Moderate and Intense Low-oxygen Dilution (MILD) combustion offer high efficiency and fuel flexibility with low emissions [1]. While CFD tools have advanced significantly, the use of physics-based reduced-order models (ROMs) is becoming increasingly attractive for real-time simulations. These models can simplify the relationship between inputs and outputs, allowing for faster simulations of complex combustion regimes like MILD combustion. Additionally, the development of digital twins (DTs), for industrial systems presents numerous opportunities, including using data to anticipate system response and using simulations to develop new technologies through virtual prototyping [2]. Accurately assessing the reliability of numerical models is a crucial and challenging task, given a lot of uncertainties that are not yet fully understood or quantified. To improve design models, a blend of experimental data and numerical models is necessary. Data

assimilation (DA) offers this type of solution by combining both experimental and numerical models to enhance accuracy in numerical forecasts and minimize errors related to both sources [3]. This work introduces a novel framework for DA that efficiently integrates experimental data with a digital twin (DT) developed for a combustion furnace. The proposed DA scheme combines results collected from experiments with DT predictions to enhance the DT's accuracy. This goal is achieved by developing the DT as a ROM using Gaussian Process Regression (GPR) and Proper Orthogonal Decomposition (POD) and comparing it to the results obtained using a Sparse Sensing (SpS) method. This GPR-based ROM methodology, demonstrated by Aversano et al. [4] to develop ROMs of turbulent reacting flow applications, was applied to build the DT of a MILD combustion furnace. The SpS method [5] is used to construct a hybrid experimental-numerical DT for MILD combustion by coupling POD modes, recovered from numerical simulations, with experimental measurements [6].

Numerical and Experimental set-up

The test case used for this study is the ULB semi-industrial combustion furnace, with a nominal power of 20 kW. The furnace is composed of a cubic combustion chamber insulated with a thick high-temperature ceramic foam layer. The furnace incorporates a burner with a heat exchanger to recover heat from the exhaust gases and preheat the combustion air. Fuel injection occurs through a central nozzle, surrounded by a coaxial air jet. The unit is equipped with four air cooling tubes located inside the furnace. An opening is available for measurements on each vertical wall of the chamber. Fig. 1 reports the schematic representation of the furnace. In-flame temperature profiles are sampled at different axial and radial locations of the furnace, with a suction pyrometer equipped with N-type thermocouples. Electrochemical sensors with nominal accuracy for different species are used to measure the exhaust gas composition. All the experiments are performed at steady-state operating conditions. More details can be found in [7].

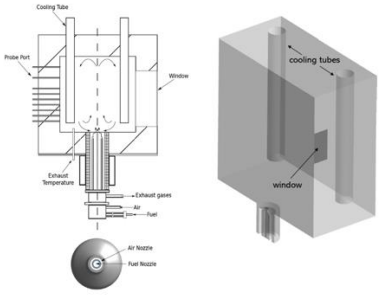

Figure 1. Vertical cross-section (top left), burner nozzle (bottom left), and 3D half geometric representation (right) of the ULB MILD combustion furnace.

41 RANS simulations were generated to obtain the necessary samples for developing

the ROM of the furnace. Three input parameters were considered to produce the simulation samples: the content of hydrogen, 0-100%, the equivalence ratio ($\emptyset = 0.7 - 1$) and air nozzle diameter (ID = 16, 20 and 25 mm). Regarding the setup of the numerical simulations, we consider the standard $k - \varepsilon$ turbulence model with the Partially Stirred Reactor (PaSR) model for the turbulence-chemistry interactions with a C_{mix} of 0.5 and the KEE (17 species and 58 reactions) for the kinetic mechanism. The computational domain is a 45° angular sector of the 3D geometry of the chamber, given the symmetry of its domain. The selected grid contains about 200k cells. More detailed information on the simulation can be found in [4, 7].

Methodology

The goal of the ROM is to predict the complete thermo-chemical state of the furnace at unexplored operating conditions. This outcome can be accomplished by combining dimensionality reduction through POD with regression conducted via GPR. The POD (also known as Principal Component Analysis [8]) is a data compression method where the data matrix $\boldsymbol{X} \subset \mathbb{R}^{n,p}$ is decomposed into:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{A}^T$$

with $\boldsymbol{U} \subset \mathbb{R}^{n,p}$ and $\boldsymbol{A} \subset \mathbb{R}^{p,p}$, which are orthogonal matrices that constitute a basis for the columns and rows of $\boldsymbol{X}$, respectively. POD is used to reduce dimensionality since the truncation of the new basis to the q-order minimizes the reconstruction error in the l_2 norm. This implies that the matrix $\boldsymbol{X}$ can be approximately estimated as:

$$\boldsymbol{X} \approx \boldsymbol{U}_q \boldsymbol{A}_q^T$$

where $\boldsymbol{U}_q \subset \mathbb{R}^{n,p}$ and $\boldsymbol{A}_q \subset \mathbb{R}^{p,p}$ are the q-order truncation of $\boldsymbol{U}$ and $\boldsymbol{A}$. The matrix $\boldsymbol{X}$ is constructed by gathering the outcomes of the numerical simulations, by varying the three parameters explained in the previous section: each column of $\boldsymbol{X}$ contains the results of a simulation, and the number of rows is the number of computational cells times the number of variables. The variables considered are the temperature inside the furnace and the species mass fraction. To predict the solution in the unexplored region of the design space, a GPR model is employed. This regression problem is expressed as $y = f(x) + \varepsilon$, where y is the observed value, f(x) is the underlying function we are modeling (a sample from a Gaussian Process (GP)), x is the independent variable and is the noise ε formulated as an independent, identically distributed (i.i.d.) random variable. More details about GPR method can be found in [9]. After training the model, the ROM can map each simulation made of a specific set of parameters to the low-dimensional representation a_{GPR}. The solution is produced by projecting a_{GPR} in the original higher dimensional space:

$$x_{GPR} = \boldsymbol{U}_q \, a_{q,GPR}$$

Regarding the experimental point of view, the SpS model is a powerful framework that leverages the concept of sparsity to efficiently capture and reconstruct signals or data. By exploiting the inherent sparsity in a signal, which refers to the fact that only

a small number of elements contribute significantly to its representation, sparse sensing techniques enable the reconstruction of high-quality signals using a reduced number of measurements. The mathematical formulation of SpS is described in detail in [6]. To summarize, the goal of SpS is to solve this linear system $\mathrm{y} = \mathbf{C}\mathrm{x}$, with $y \subset R^s$ contains the measured values, $\boldsymbol{C} \subset R^{s,n}$ represents the measurement matrix and $x \subset R^n$ is the state of the system. Assuming a sparse representation of the system in a different basis, we can approximate the variable $x \approx \boldsymbol{\Phi}_q a_q$, where $\boldsymbol{\Phi}_q \subset R^{n,q}$ represents the q-order truncation of the transforming basis, and a_q is the projection of x onto this truncated basis. The transforming basis is determined by applying the Proper Orthogonal Decomposition (POD) to the dataset. Consequently, we can rephrase the original linear system as:

$$y = \boldsymbol{\Theta} a_{q,SpS}$$

with $\boldsymbol{\Theta} = \boldsymbol{C}\boldsymbol{U}_q$, where $\boldsymbol{U}_q$ is computed through Singular Value Decomposition (SVD) applied to the data matrix.

Data Assimilation

Data assimilation aims to merge observations with model predictions in a way that optimally integrates the available information by considering the uncertainties in both the model and the observations. By doing so, DA methods can provide improved estimates of the current state of a system and even make better predictions about its future behaviour. The assimilation between the GPR model and the SpS model is done by applying the Kalman filter [10]. The Kalman filter can produce the optimal estimate of the system's true state given the model's prediction and the experimental observations.

$$a_{q,adj} = a_{q,GPR} + \boldsymbol{K}\left(a_{q,SpS} - a_{q,GPR}\right)$$
$$\boldsymbol{C}_{adj} = \boldsymbol{C}_{GPR} - \boldsymbol{K}\boldsymbol{C}_{GPR}$$
$$\boldsymbol{K} = \frac{\boldsymbol{C}_{GPR}}{\boldsymbol{C}_{GPR} + \boldsymbol{C}_{SpS}}$$

where $a_{q,adj}$ is a linear combination through the so-called Kalman gain, **K**, between $a_{q,GPR}$ and $a_{q,SpS}$, which represent low-dimensional representation of the GPR and the SpS model respectively. In addition, $\boldsymbol{C}_{GPR}$ and $\boldsymbol{C}_{SpS}$ are diagonal matrices containing the GPR and SpS model uncertainty, respectively.

Results

As mentioned before, the DT model has been applied to the matrix containing 41 simulations obtained with different combinations of the three parameters: $\emptyset, ID_{air}, \%H_2$. The feature included in the data matrix is the temperature field inside the furnace. Fig. 2a shows the temperature given by CFD simulation and the one predicted by the DT for the case: $\emptyset = 0.93, ID_{air} = 16mm\,, \%H_2 = 60$.

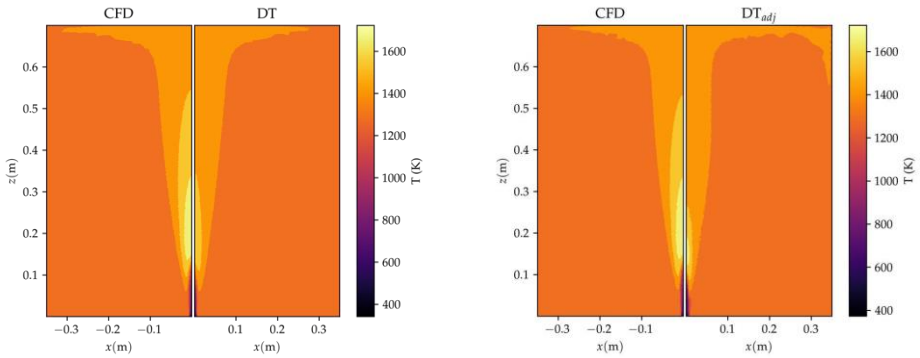

Figure 2. a) Comparison between the original temperature field (on the left) and the temperature field predicted by the GPR ROM (on the right); b) Comparison between the original temperature field (on the left) and the temperature field predicted by the GPR ROM (on the right), after data assimilation.

By assimilating the GPR-POD solution with the one derived from SpS, an adjusted solution can be obtained through the KF algorithm. The adjusted solution is therefore visualized in Fig. 2b. Through assimilation, the model's uncertainty is reduced by comparing the uncertainties associated with the GPR and SpS models. This expression can be illustrated in Fig. 3 for several radial profiles of temperature at different axial positions. The GPR model's prediction is characterized by a significantly higher uncertainty level than the SpS model. Consequently, the adjusted solution experiences a minor reduction in uncertainty.

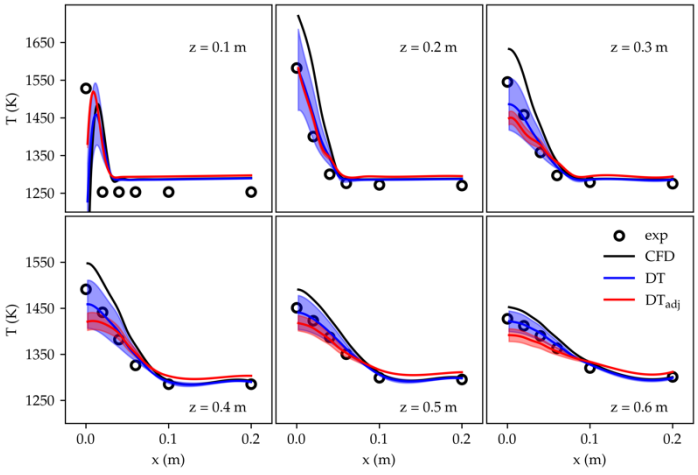

Figure 3. Several radial profiles of measured temperature along with the adjusted value obtained by assimilating the GPR and SpS models.

Conclusions

The aim of this study is to create a DA framework that combines a DT, which is a ROM obtained by coupling GPR and POD, with SpS to obtain an improved DT, with lower uncertainty. To achieve this, a dataset consisting of 41 RANS simulations of a semi-industrial combustion furnace has been used as the training data, each simulation performed with different combinations of a set of three parameters $(\emptyset, \boldsymbol{ID_{air}}, \%\boldsymbol{H_2})$. First, POD has been applied to the dataset, and the GPR method has been trained on the set of POD coefficients, enabling the construction of a model capable of predicting the POD coefficients for unexplored combinations of the parameters. Then, to include experimental information, a SpS model has been constructed. This approach is aimed at finding the set of POD coefficients that minimizes the l_2 distance with the experimental measurements. Finally, DA is employed to blend the DT model with the SpS one to have an adjusted DT with lower uncertainty than both the initial sources.

References

[1] Cavaliere A., De Joannon M. Mild combustion. Progress in Energy and Combustion science. (2004).
[2] E. Glaessgen, D. Stargel, The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles.
[3] M. Asch, M. Bocquet, M. Nodet, Data Assimilation: Methods, Algorithms, and Applications. (2016)
[4] Aversano G, Ferrarotti M, Parente A. Digital twin of a combustion furnace operating in flameless conditions: reduced-order model development from cfd simulations. Proceedings of the Combustion Institute (2021)
[5] Manohar K, Brunton BW, Kutz JN, et al. Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. IEEE Control Systems Magazine (2018)
[6] Procacci A, Amaduzzi R, Coussement A, et al. Adaptive digital twins of combustion systems using sparse sensing strategies. Proceedings of the Combustion Institute (2022)
[7] M. Ferrarotti, M. F¨urst, E. Cresci, W. de Paepe, A. Parente, Key modeling aspects in the simulation of a quasi-industrial 20 kw moderate or intense low-oxygen dilution combustion chamber, Energy & Fuels. (2018)
[8] Parente A, Sutherland JC. Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity. Combustion and Flame 160(2):340–350. (2013)
[9] Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press. (2005)
[10] Kalman RE. A New Approach to Linear Filtering and Prediction Problems. Journal of Basic Engineering. (1960).