

Article

Digital Twin for Experimental Data Fusion Applied to a Semi-Industrial Furnace Fed with H₂-Rich Fuel Mixtures

Alberto Procacci^{1,2,*} , Marianna Cafiero^{1,2,3}, Saurabh Sharma^{1,2}, Muhammad Mustafa Kamal^{1,2}, Axel Coussement^{1,2}  and Alessandro Parente^{1,2}

¹ Aero-Thermo-Mechanics Laboratory, École Polytechnique de Bruxelles, Université Libre de Bruxelles, 1050 Brussels, Belgium

² Brussels Institute for Thermal-Fluid Systems and Clean Energy (BRITE), Université Libre de Bruxelles and Vrije Universiteit Brussel, 1050 Brussels, Belgium

³ Institute of Mechanics, Materials, and Civil Engineering, Université Catholique de Louvain, Place du Levant 2, 1348 Louvain-la-Neuve, Belgium

* Correspondence: alberto.procacci@ulb.be

Abstract: The objective of this work is to build a Digital Twin of a semi-industrial furnace using Gaussian Process Regression coupled with dimensionality reduction via Proper Orthogonal Decomposition. The Digital Twin is capable of integrating different sources of information, such as temperature, chemiluminescence intensity and species concentration at the outlet. The parameters selected to build the design space are the equivalence ratio and the benzene concentration in the fuel stream. The fuel consists of a H₂/CH₄/CO blend, doped with a progressive addition of C₆H₆. It is an H₂-rich fuel mixture, representing a surrogate of a more complex Coke Oven Gas industrial mixture. The experimental measurements include the flame temperature distribution, measured on a 6 × 8 grid using an air-cooled suction pyrometer, spatially resolved chemiluminescence measurements of OH* and CH*, and the species concentration (i.e., NO, NO₂, CO, H₂O, CO₂, O₂) measured in the exhaust gases. The GPR-based Digital Twin approach has already been successfully applied on numerical datasets coming from CFD simulations. In this work, we demonstrate that the same approach can be applied on heterogeneous datasets, obtained from experimental measurements.

Keywords: digital twin; data fusion; dimensionality reduction



Citation: Procacci, A.; Cafiero, M.; Sharma, S.; Kamal, M.M.; Coussement, A.; Parente, A. Digital Twin for Experimental Data Fusion Applied to a Semi-Industrial Furnace Fed with H₂-Rich Fuel Mixtures. *Energies* **2023**, *16*, 662. <https://doi.org/10.3390/en16020662>

Academic Editor: Giancarlo Sorrentino

Received: 30 November 2022

Revised: 21 December 2022

Accepted: 29 December 2022

Published: 5 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The need for the rapid decarbonization of the global economy requires the development of new combustion systems that are both efficient and flexible, to allow the use of new zero-carbon fuels such as hydrogen and ammonia [1].

These requirements, coupled with the necessity of limiting the production of harmful pollutants, impose strict operating conditions on the combustion systems. This means that the design and control of these systems is crucial, and the margin of error is limited.

For this reason, the model of trial-and-error employed for the design of a traditional combustion system is both too time-consuming and error prone to be applied on the development of new combustion designs.

Luckily, the recent developments in machine-learning techniques and the increasing availability of data offer various tools that can be exploited in the design and operation of combustors.

In particular, the development of Digital Twins (DTs) has been increasingly regarded as a way to substantially improve both the knowledge of industrial systems and their control [2–4]. The DT is defined as a digital representation of a physical object that can closely simulate its behavior in the real environment [2,4]. The DT can help in the design, production and service phases of a product. In the design phase, the DT is useful in the iterative optimization and the virtual evaluation, while in the production phase it can be

employed for real-time monitoring, control and prediction. Finally, in the service phase, the DT can be used to forecast the product's maintenance [2,3].

Crucially, the DT has to provide a real-time prediction of the state of its physical counterpart, so that it can supply constant reference values.

This requirement generally excludes Computational Fluid Dynamics (CFD) simulations as techniques to rely on building DTs because the computing time is not negligible, even for relatively simple geometries.

Instead, data-driven regression models, such as linear regression, Gaussian Process Regression (GPR) and Neural Networks (NNs), are particularly suited for this kind of application because the training phase is distinct from the prediction phase. This means that the training phase, however long, can be performed offline while the trained model can predict the system's state almost instantaneously.

Among these regression methods, the GPR offers a number of advantages with respect to the other regression models. In particular, it is a nonlinear technique based on Bayesian inference with a limited amount of hyperparameters to tune. This probabilistic framework means that the predictions are complemented with the model's uncertainty [5].

Moreover, the GPR performs well with sparse datasets. The availability of sparse datasets is often the case when dealing with reactive systems because the amount of operating conditions explored is limited by the computational cost (for CFD simulations) or the operating cost (for experimental campaigns).

Other important tools to build DTs are dimensionality reduction techniques, in particular the Proper Orthogonal Decomposition (POD) [6]. The POD is a linear technique capable of finding a low-dimensional representation of the system without losing important information. This allows for greatly reducing the complexity of the model while maintaining high accuracy in the prediction.

Recently, the GPR and POD approach has been employed to develop a numerical DT of a semi-industrial furnace [7], while the POD coupled with sparse sensing was used to develop a hybrid numerical-experimental DT of the same furnace [8].

However, in some situations, it is not possible to rely on the availability of numerical data to build a DT. For example, numerical simulations of highly turbulent reacting flows still do not provide accurate predictions of pollutants and stability limits [9]. Moreover, the use of state-of-the-art Large Eddy Simulation combustion models is often too expensive to be applied to industrial facilities [10].

In these cases, an experimental campaign is often carried out with different measuring instruments to characterize the combustion system. This usually produces various datasets with vastly different spatial scales (the pressure field obtained using pressure probes has a much lower spatial resolution than a velocity field measured using particle image velocimetry, for example).

The objective of this work is to develop a DT of the ULB semi-industrial furnace which relies solely on experimental data. The DT is capable of integrating different data streams and it provides in real time the prediction of temperature, chemiluminescence and species concentration, for different degrees of equivalence ratio and benzene doping in the fuel blend. To achieve this, the experimental data are first projected into the low-dimensional space using POD. Then, the regression is performed using GPR in the low-dimensional manifold. Finally, the prediction can be obtained by projecting the solution found by the GPR model in the original space.

The experimental setup is outlined in Section 2, while the methodology for the development of the DT is explained in more detail in Section 3. The results are shown in Section 4 and discussed in Section 5.

2. Experimental Setup

The experimental data have been collected in a semi-industrial furnace, fed with $H_2/CH_4/CO$ fuel mixtures doped with C_6H_6 . Detailed information about the experimental

setup and measurement techniques are provided in previous publications [11,12], while the dataset used in this work has been published in [13].

However, some key features of the experimental set-up are also summarized below, along with the specific operating conditions used for this dataset.

The combustion furnace used in this work consists of an insulated combustion chamber. Figure 1 shows the schematic diagram of the vertical cross-section of the furnace-burner assembly, while Figure 2 shows a schematic representation of the test bench.

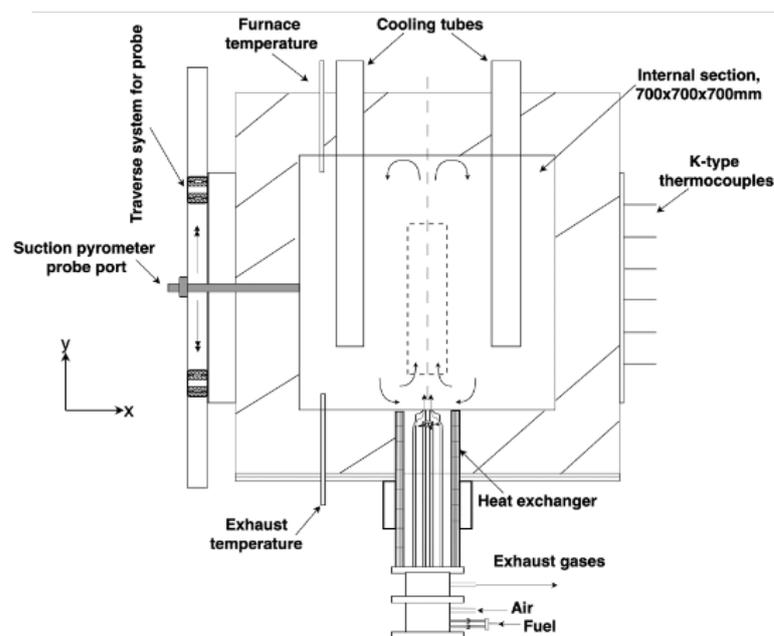


Figure 1. Vertical cross-section of the ULB semi-industrial furnace. The dashed lines represent a projection of the quartz window on the vertical cross-section plane.

The furnace has a cubic shape ($1100 \times 1100 \times 1100$ mm), and it is internally insulated with a 200 mm thick high-temperature (1300 °C) ceramic foam layer, which limits the heat lost through the walls. Therefore, the resulting inner dimensions of the combustion chamber are: $700 \times 700 \times 700$ mm.

The furnace is equipped with an air cooling system, which consists of four cooling tubes of 80 mm of outer diameter and a length of 630 mm. By varying the cooling air flow rate, different stable conditions are established inside the combustion chamber, and the effect of variable industrial loads can be tested.

On each vertical wall of the combustion chamber, there is a slot (150×600 mm) available for measurements. The details of the measurements through different openings can be found in [11,12].

At the bottom of the combustion chamber, a commercial WS® REKUMAT M150 recuperative Flame-FLOX burner with a nominal power of 20 kW is mounted. Two different air/fuel injection configurations are possible when the burner is operated in Flame or in FLOX® combustion mode [11].

The dimension of the air injection nozzle ID can be varied (ID: 16, 20, 25 mm). However, the experimental data reported in [13] have been measured while operating the burner only in Flame mode and with using a fixed air injection nozzle (ID:25 mm).

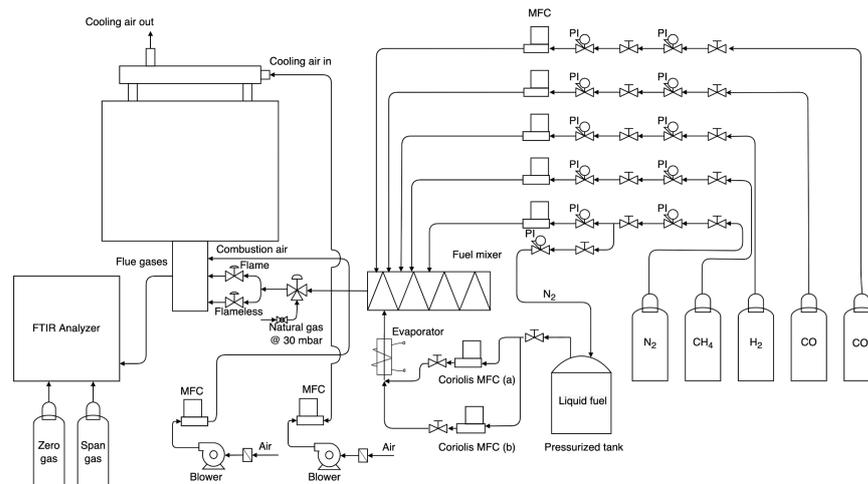


Figure 2. Schematic P & I of the ULB experimental test bench.

The semi-industrial facility is equipped with a fuel-feeding system, integrated with an evaporation system and a mixing unit that allow for creating synthetic blends. The evaporation system was used to vaporize benzene in order to homogeneously mix it with the other gases in the static mixing unit.

Brooks SLA58XX® Series mass flow controllers are used to control and monitor the flow of gases, and Brooks Quantim® Series Coriolis mass flow controllers of ranges, 320 g/h and 1500 g/h are used for the benzene.

An air-cooled suction pyrometer probe equipped with a B-type (Pt-Rh 6% - Pt-Rh 30%) thermocouple is used to perform in situ flame temperature measurements.

OH* and CH* chemiluminescence imaging was carried out by means of an Intensified Relay Optics (IRO) and a Charge-Coupled Device (CCD) camera 1.4 M (La Vision 1392 × 1040 pixels) coupled with a UV 78mm f/3.8 lens and two interferential filters to collect the chemiluminescence emitted by OH* (310 ± 10 nm) and CH* (438 ± 24 nm). The CCD camera has a maximum frame rate at a full resolution equal to 17 fps.

The flue gas composition was detected by using a Fourier Transform Infrared Spectroscopy (FTIR) analyzer from HORIBA® (HORIBA MEXA-ONE), equipped with a paramagnetic analyzer for oxygen measurement.

Figure 3 shows the position of the experimental probes inside the furnace. The position of the temperature and the chemiluminescence sensors was chosen to accurately map the reactive region and by considering the design constraints of the set-up. The species concentrations are measured at the outlet of the furnace to check the emissions of pollutants and main combustion products.

The experimental uncertainties for the temperature and the emissions are reported in Table 1.

Table 1. Experimental uncertainties (95% confidence interval).

T (K)	$X_{NO < 200}$ (Wet ppmv)	$X_{NO > 200}$ (Wet ppmv)	X_{NO_2} (Wet ppmv)	X_{CO} (Wet vol. Fraction)	X_{H_2O} (Wet vol. Fraction)	X_{CO_2} (Wet vol. Fraction)	X_{O_2} (Dry vol. Fraction)
±20	±5.56	±19.65	±3.93	±0.20%	±0.40%	±0.39%	±0.6%

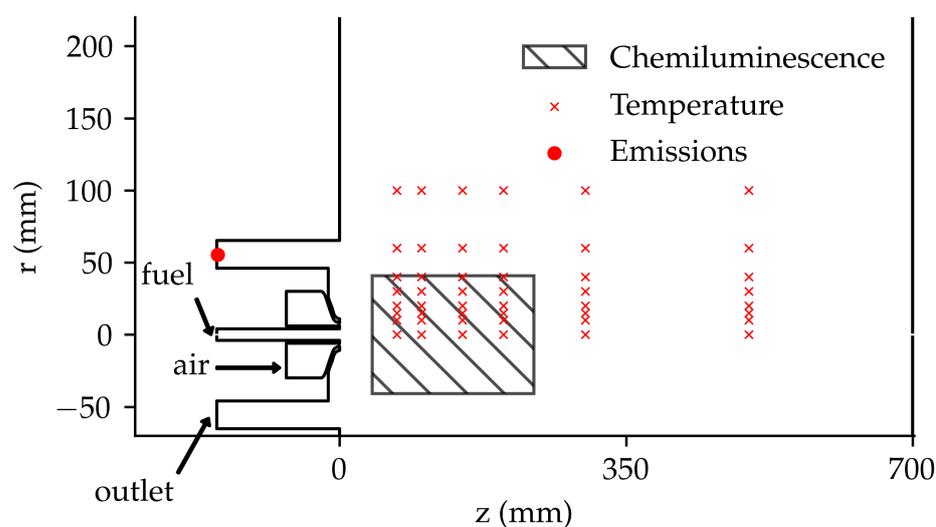


Figure 3. Position of the sensors relative to the outline of the ULB furnace.

Operating Conditions

A mixture of $H_2/CH_4/CO$ has been used as fuel. It is a H_2 -rich fuel mixture, considered as a surrogate of an industrial Coke Oven Gas (COG) mixture. As in [14], the $H_2/CH_4/CO$ fuel mixture composition was fixed by considering only the major components of the more complex industrial COG composition. Moreover, the relative ratios of H_2 , CH_4 and CO correspond to the ones in the fuel source mixture, and they are in accordance with those related to other COG compositions available in literature [15–17].

The fuel has been doped with C_6H_6 up to 5% (vol.), in order to investigate the effects of aromatic additives in H_2 -rich fuel mixtures, mainly on NO emissions and flame radiation, in a semi-industrial furnace.

The compositions of the final six fuel blends tested are provided in Table 2, along with the equivalence ratio and the benzene doping.

This was carried out to investigate the effects of aromatics doping of the under-firing H_2 -fuel mixtures, at both constant and varying air excess.

The thermal input Q_{th} was kept constant at 20 kW for all the investigated cases. Similarly, the flow rate of the cooling air was kept constant. Moreover, the temperature of the heated fuel pipeline, between the mixing unit and the inlet of the furnace, was controlled to keep a constant inlet fuel temperature of 130 °C, in order to avoid any benzene condensation before the entrance of the furnace.

Table 2. Operating conditions of the experimental investigation of $H_2/CH_4/CO$ fuel mixtures, with different levels of C_6H_6 (%v/v) doping and equivalence ratios.

Case	C_6H_6 (%v/v)	ϕ	H_2 (%v/v)	CH_4 (%v/v)	CO (%v/v)	LHV (MJ/Nm ³)
1	0	0.80	68.68	25.34	5.98	17.24
2	0.5	0.80	68.34	25.21	5.95	17.86
3	1.12	0.80	67.91	25.06	5.91	18.63
4	2	0.80	67.31	24.83	5.86	19.73
5	3	0.80	66.62	24.58	5.80	20.97
6	5	0.80	65.25	24.07	5.68	23.45
7	0	0.91	68.68	25.34	5.98	17.24
8	0.5	0.91	68.34	25.21	5.95	17.86
9	1.12	0.91	67.91	25.06	5.91	18.63

Table 2. Cont.

Case	C ₆ H ₆ (%v/v)	ϕ	H ₂ (%v/v)	CH ₄ (%v/v)	CO (%v/v)	LHV (MJ/Nm ³)
10	2	0.91	67.31	24.83	5.86	19.73
11	3	0.91	66.62	24.58	5.80	20.97
12	5	0.91	65.25	24.07	5.68	23.45
13	0	1.00	68.68	25.34	5.98	17.24
14	0.5	1.00	68.34	25.21	5.95	17.86
15	1.12	1.00	67.91	25.06	5.91	18.63
16	2	1.00	67.31	24.83	5.86	19.73
17	3	1.00	66.62	24.58	5.80	20.97
18	5	1.00	65.25	24.07	5.68	23.45
19	0	1.05	68.68	25.34	5.98	17.24
20	0.5	1.05	68.34	25.21	5.95	17.86
21	1.12	1.05	67.91	25.06	5.91	18.63
22	2	1.05	67.31	24.83	5.86	19.73
23	3	1.05	66.62	24.58	5.80	20.97
24	5	1.05	65.25	24.07	5.68	23.45
25	0	1.10	68.68	25.34	5.98	17.24
26	0.5	1.10	68.34	25.21	5.95	17.86
27	1.12	1.10	67.91	25.06	5.91	18.63
28	2	1.10	67.31	24.83	5.86	19.73
29	3	1.10	66.62	24.58	5.80	20.97
30	5	1.10	65.25	24.07	5.68	23.45
31	0	1.20	68.68	25.34	5.98	17.24
32	0.5	1.20	68.34	25.21	5.95	17.86
33	1.12	1.20	67.91	25.06	5.91	18.63
34	2	1.20	67.31	24.83	5.86	19.73
35	3	1.20	66.62	24.58	5.80	20.97
36	5	1.20	65.25	24.07	5.68	23.45

3. Digital Twin Methodology

As mentioned in Section 1, the goal of this work was to develop a DT capable of predicting simultaneously different types of measurements, such as emissions, temperatures and the chemiluminescence signal.

As a first step, 30 samples of temperature, chemiluminescence and emissions were randomly selected to build the training data. The remaining six samples were used to test the model.

After that, the temperature, chemiluminescence and emissions datasets were arranged into the data matrices $\mathbf{X}_T \in \mathbb{R}^{n_T \times p}$, $\mathbf{X}_{CL} \in \mathbb{R}^{n_{CL} \times p}$ and $\mathbf{X}_E \in \mathbb{R}^{n_E \times p}$. Each column vector contains the data for a certain operating condition, and the number of rows depends on the type of measurement. In particular, $n_T = 48$, $n_{CL} = 898,560$ and $n_E = 6$.

Then, the POD was applied to the data matrices, which were centered and scaled to unitary variance. The POD allows for decomposing the dataset into two parts, of which one is only a function of the spatial coordinates and one is only a function of the parameters:

$$\mathbf{X}_T(\mathbf{r}, \boldsymbol{\theta}) = \mathbf{U}_T(\mathbf{r})\boldsymbol{\Sigma}_T\mathbf{V}_T(\boldsymbol{\theta})^T, \quad (1)$$

$$\mathbf{X}_{CL}(\mathbf{r}, \boldsymbol{\theta}) = \mathbf{U}_{CL}(\mathbf{r})\boldsymbol{\Sigma}_{CL}\mathbf{V}_{CL}(\boldsymbol{\theta})^T, \quad (2)$$

where $\mathbf{U}_T \in \mathbb{R}^{n_T \times p}$ and $\mathbf{U}_{CL} \in \mathbb{R}^{n_{CL} \times p}$ represent the spatial POD modes while $\mathbf{V}_T, \mathbf{V}_{CL} \in \mathbb{R}^{p \times p}$ contain the parametric coefficients. The diagonal matrices $\boldsymbol{\Sigma}_T, \boldsymbol{\Sigma}_{CL} \in \mathbb{R}^{p \times p}$ contain the singular values of the POD modes. The singular values are the square root of the eigenvalues of the covariance matrix $\mathbf{S} = \frac{1}{n-1}\mathbf{X}^T\mathbf{X}$, and they reflect the amount of variance retained by each POD mode.

The threshold of 99.9% of explained variance was selected for truncation, resulting in the matrices $\mathbf{U}_{T,r} \in \mathbb{R}^{n_T, r_T}$, $\mathbf{U}_{CL,r} \in \mathbb{R}^{n_{CL}, r_{CL}}$, $\mathbf{V}_{T,r} \in \mathbb{R}^{p, r_T}$ and $\mathbf{V}_{CL,r} \in \mathbb{R}^{p, r_{CL}}$ where $r_T = 5$ and $r_{CL} = 8$.

The GPR model was applied to the global data matrix $\mathbf{X}_G \in \mathbb{R}^{n, p}$, created by concatenating the matrices \mathbf{X}_E , \mathbf{V}_T and \mathbf{V}_{CL} such that $n = n_E + r_T + r_{CL}$.

The GPR assumes that, in the classical regression framework,

$$\mathbf{y} = f(x) + \epsilon, \quad (3)$$

the function $f(x)$ is a sample from a Gaussian process

$$f(x) \sim GP(m(x), k(x, x')), \quad (4)$$

where $m(x)$ and $k(x, x')$ are the mean and covariance function of the Gaussian process, which can be thought of as a Gaussian distribution over functions [5]. In this framing, the observed and predicted data become a sample from a multivariate Gaussian distribution:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I} & \mathbf{K}(\mathbf{X}, \mathbf{X}_*) \\ \mathbf{K}(\mathbf{X}_*, \mathbf{X}) & \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix}\right), \quad (5)$$

where \mathbf{y} represents the observed target values, and $\mathbf{f}_* = f(\mathbf{X}_*)$ is the prediction in the unexplored part of the design space \mathbf{X}_* . The covariance matrix of the multivariate Gaussian distribution is built using the kernel function \mathbf{K} . The prior distribution $p(\mathbf{f})$ and the likelihood $p(\mathbf{y}|\mathbf{f})$ are both assumed to be Gaussian. The prior represents the prior knowledge before considering the data, while the likelihood is the probability of seeing the data given the choice of the model.

By employing Bayes' theorem, it is possible to calculate the posterior distribution $p(\mathbf{f}|\mathbf{y})$ given the prior and the likelihood:

$$p(\mathbf{f}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{f}) p(\mathbf{f})}{p(\mathbf{y})}. \quad (6)$$

The predicting distribution $p(\mathbf{f}_*|\mathbf{y})$ is then calculated by marginalizing the likelihood with the posterior distribution:

$$p(\mathbf{f}_*|\mathbf{y}) = \int_{-\infty}^{+\infty} p(\mathbf{f}_*|\mathbf{f}) p(\mathbf{f}|\mathbf{y}) d\mathbf{f}. \quad (7)$$

The predicting distribution is again a Gaussian distribution $p(\mathbf{f}_*|\mathbf{y}) = \mathcal{N}(\bar{\mathbf{f}}_*, \text{cov}(\bar{\mathbf{f}}_*))$ with mean and covariance:

$$\bar{\mathbf{f}}_* = \mathbf{K}_*^T [\mathbf{K} + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{y} \quad (8)$$

$$\text{cov}(\bar{\mathbf{f}}_*) = \mathbf{K}_{**} - \mathbf{K}_*^T [\mathbf{K} + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{K}_*, \quad (9)$$

where, for compactness, $\mathbf{K} = \mathbf{K}(\mathbf{X}, \mathbf{X})$, $\mathbf{K}_* = \mathbf{K}(\mathbf{X}, \mathbf{X}_*)$ and $\mathbf{K}_{**} = \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*)$. The model's hyperparameters, such as the kernel's length scale and the observations' noise, are selected by minimizing the negative log marginal likelihood:

$$-\log p(\mathbf{y}) = \frac{1}{2} \mathbf{y}^T (\mathbf{K}(\phi) + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y} + \frac{1}{2} \log |\mathbf{K}(\phi) + \sigma_n^2 \mathbf{I}| + \frac{n}{2} \log 2\pi. \quad (10)$$

Once trained, the GPR model is able to predict the target quantity \mathbf{X}_G^* in the unexplored part of the design space. In this work, the GPR has been implemented using GPyTorch [18].

The reconstruction of the temperature and chemiluminescence data are carried out by multiplying the respective rows of \mathbf{X}_G^* times the POD modes \mathbf{U}_T and \mathbf{U}_{CL} and the singular values $\Sigma_{T,r}$ and $\Sigma_{CL,r}$:

$$\mathbf{X}_E^* = \mathbf{I} \mathbf{X}_G^*[1 : n_E] \quad (11)$$

$$\mathbf{X}_T^* = \mathbf{U}_{T,r} \Sigma_{T,r} \mathbf{X}_G^{*T}[n_E : n_E + r_T] \quad (12)$$

$$\mathbf{X}_{CL}^* = \mathbf{U}_{CL,r} \Sigma_{CL,r} \mathbf{X}_G^{*T}[n_E + r_T : n], \quad (13)$$

where the notation $\mathbf{A}[i : j]$ represents the matrix built by considering only the rows of \mathbf{A} between the indices i and j .

The methodology described in this section is summarized in Figure 4.

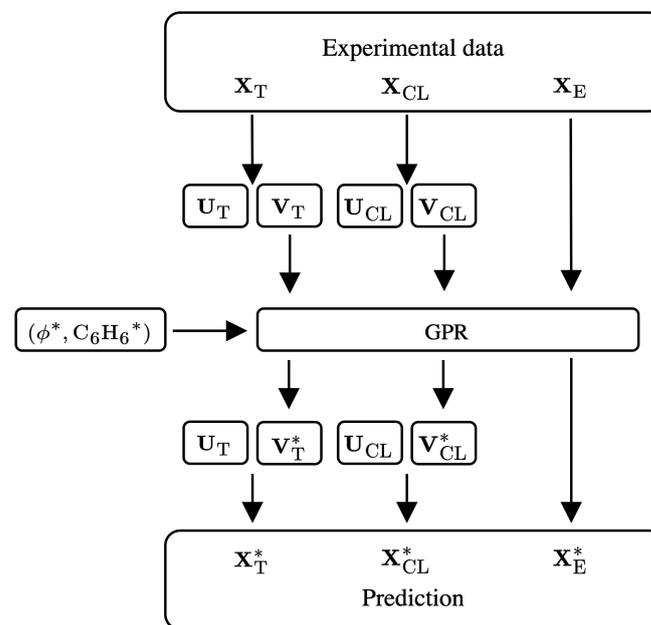


Figure 4. Diagram depicting the methodology used to build the Digital Twin.

4. Results

The testing cases randomly selected from the complete dataset are cases 2, 8, 11, 15, 16 and 29. The remaining cases are used to train the model.

The model's training takes around 10 s, while the prediction is obtained in around 1 s.

Figure A1 shows the surface response of the scalar quantities of interest as a function of the benzene doping and the equivalence ratio. The scalar quantities selected are the maximum temperature and the normalized sum of the chemiluminescence counts, along with the species concentrations.

The maximum temperature increases slightly with the benzene doping at $\phi < 1.0$, and it reaches the maximum at $\phi = 0.80$ and $C_6H_6 = 5$. However, by increasing the equivalence ratio above the stoichiometric condition, the maximum temperature drops with the addition of benzene.

The chemiluminescence signal is strongly dependent on both the benzene doping and the equivalence ratio. Similarly to the maximum flame temperature, the intensity of both OH^* and CH^* chemiluminescence emissions increases with the presence of aromatics in the $H_2/CH_4/CO$ flames, under lean conditions, showing a distinct maximum at $\phi = 0.91$ and $C_6H_6 = 5.0\%$, whereas under stoichiometric and rich conditions the chemiluminescence intensity of both OH^* and CH^* radicals decreases with benzene doping levels higher than 3%.

Figure A1 shows also that the concentration of species such as NO, CO and O₂ depends mainly on the equivalence ratio, while the concentration of NO₂, H₂O and CO₂ depends on both the equivalence ratio and the level of benzene doping. The NO concentration drops when the fuel/air mixture goes from lean to rich, while the opposite happens for the CO concentration.

Both the H₂O and the CO₂ concentration peaks at $\phi = 1.0$, and it increases with the benzene doping.

Finally, the O₂ concentration depends only on the equivalence ratio. In the lean region, the oxygen concentration drops until it reaches the minimum value at stoichiometry. Then, it increases with the equivalence ratio in the rich region.

In general, the surface response is nonlinear for all scalar quantities, which justifies the use of a nonlinear regression technique such as the GPR.

The constant mean function was selected to train the GPR model. The kernel function was chosen as the sum of the linear kernel and the radial basis function kernel:

$$K(\mathbf{x}, \mathbf{x}') = v\mathbf{x}^T \mathbf{x}' + \exp\left(-\frac{1}{2} \frac{(\mathbf{x} - \mathbf{x}')^T (\mathbf{x} - \mathbf{x}')}{\Theta^2}\right), \quad (14)$$

where v and Θ are the variance and length scale parameters.

The observed and predicted spatial distribution of the temperature and chemiluminescence signal for case 2 is reported in Figure 5, while the observed and predicted emissions are reported in Table 3.

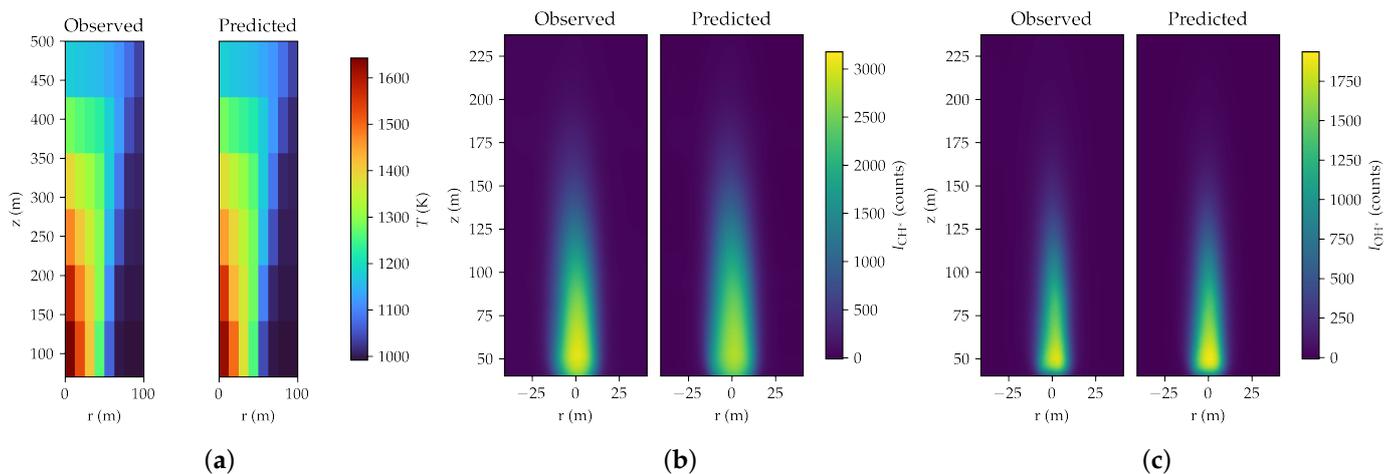


Figure 5. Comparison between the predicted and observed temperature field (a); CH* field (b); and OH* field (c) for case 2.

Table 3. Observed and predicted species emissions for case 2.

	X_{NO} (ppmv)	X_{NO_2} (ppmv)	X_{CO} (Wet vol. Fraction)	$X_{\text{H}_2\text{O}}$ (Wet vol. Fraction)	X_{CO_2} (Wet vol. Fraction)	X_{O_2} (Dry vol. Fraction)
Observed	308	3.25	0	0.222	0.0828	0.0224
Predicted	304	3.01	0	0.223	0.0814	0.0231

The predictions are generally good, especially for the temperature field. However, the model tends to overpredict the OH* signal, while it underpredicts the CH* signal. This behavior is confirmed for the other testing conditions by looking at Figure 6, which shows the parity plot for all six testing conditions. However, the R^2 is well above 0.9 for both OH* and CH*. The R^2 has been calculated by considering the points with a count greater than 100, to avoid including the non-emitting region.

Figure 7 shows the parity plots relative to the prediction of the species emissions. The predictions are good both for minor species, such as NO, and major species, such as

H₂O and O₂. The prediction is slightly worse for NO₂, which is also associated with the highest relative experimental uncertainty on such a species. This is due to the low NO₂ concentrations detected in the exhaust gases with respect to the total instrument error for this species detection (2 ppmv).

If we compare the experimental uncertainty reported in Table 1 with the model's uncertainty, we can observe that the model's uncertainty is higher than the experimental uncertainty for measurements such as temperature, NO, CO, H₂O, while it is lower for NO₂, CO₂ and O₂. This suggests that, for those measurements in which the model's uncertainty is higher than the experimental uncertainty, the number of samples collected was not enough to converge to the experimental uncertainty.

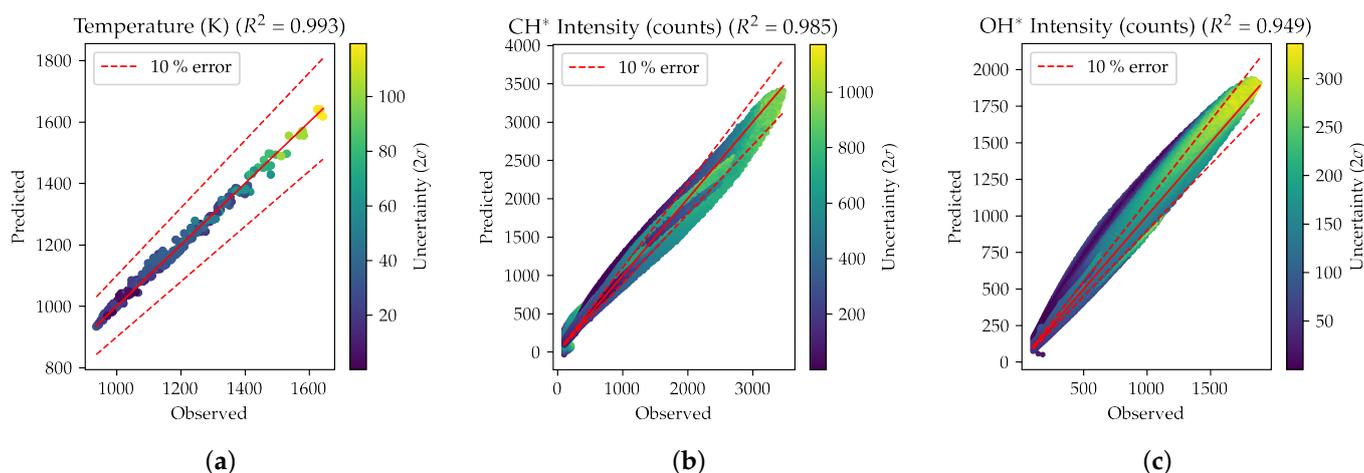


Figure 6. Parity plot for the temperature (a); CH* signal (b); and OH* signal (c), for the six testing cases. The uncertainty represents the model's uncertainty in the prediction. The R^2 for the OH* and CH* signals has been calculated by excluding the points with a count lower than 100.

To check the quality of the model's training, a 6-fold cross-validation was performed on the complete dataset. The 36 samples were randomly divided into six groups (folds), each containing six samples. A GPR model is built for each fold. This model is tested on the data contained in the corresponding fold, while the training data are comprised of all the samples, minus the testing data. The average value of the R^2 for each fold is reported in Table 4, along with the standard deviation.

The chemiluminescence signal, OH* in particular, displays the biggest variance in the R^2 . This is because the chemiluminescence signal has a highly nonlinear behavior as a function of the equivalence ratio and benzene doping, as shown in Figure A1. In addition, the nonlinear behavior is more pronounced on the edge of the operating conditions' envelope, for C₆H₆ = 5%. This means that the predictions are accurate when the model is interpolating, i.e., the testing conditions are inside the operating conditions' envelope. However, when the testing conditions are on the edge of the envelope, the model performs worse because it lacks important information on the behaviour of the chemiluminescence signal as a function of the benzene doping and equivalence ratio.

Table 4. Average and standard deviation of the R^2 for the ensemble of folds in the cross-validation algorithm.

	T	CH*	OH*	NO	NO ₂	CO	H ₂ O	CO ₂	O ₂
Average	0.988	0.870	0.770	0.966	0.974	0.889	0.945	0.955	0.981
Standard deviation	0.006	0.086	0.168	0.028	0.029	0.094	0.063	0.056	0.012

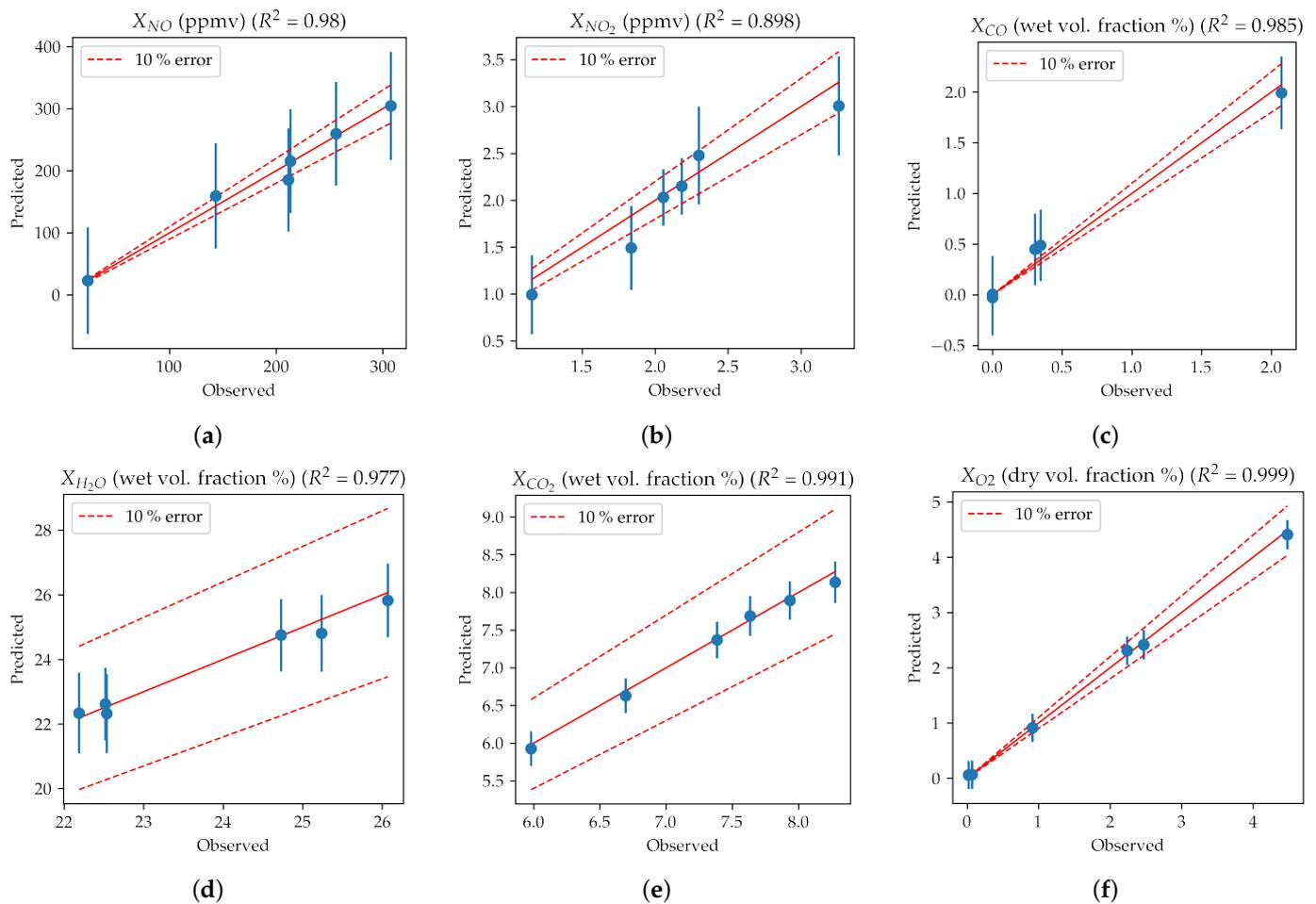


Figure 7. Parity plot for the emissions of NO (a); NO₂ (b); CO (c); H₂O (d); CO₂ (e); and O₂ (f) for the six testing cases. The uncertainty represents the model's uncertainty in the prediction.

5. Discussion

The objective of this work was to develop a GPR-based DT capable of predicting the temperature, chemiluminescence and species concentration of a semi-industrial furnace fed with an H₂-rich fuel mixture.

To integrate datasets with a vastly different spatial resolution, the POD was applied to the temperature and chemiluminescence signals. This was carried out to decouple the spatial information from the parameter-dependent information.

The GPR was selected as a regression model for its properties, mainly its nonlinearity and the ability to handle sparse datasets.

The predictions show a good level of accuracy, with a $R^2 > 0.9$ for all the signals except for the NO₂ concentration. However, even in this case, the maximum relative error is around 11%.

A 6-fold cross-validation was performed to judge the generalizability of the model. The results show that the prediction of temperature and species concentration is accurate for all combinations of testing conditions. For the prediction of the chemiluminescence signal, the model performs very well in interpolation, while it produces less accurate results in extrapolation. This means that care should be taken in the selection of samples for the model's training.

In conclusion, the results show that the DT can operate as a real-time surrogate of the corresponding physical object. This opens the possibility of employing the DT for redundant control, state and failure detection or data assimilation if coupled with low-fidelity measurements.

Author Contributions: Conceptualization, A.P. (Alberto Procacci), M.C., A.C. and A.P. (Alessandro Parente); Data curation, M.C.; Formal analysis, A.P. (Alberto Procacci); Investigation, M.C., M.M.K. and S.S.; Methodology, A.P. (Alberto Procacci), M.C., A.C. and A.P. (Alessandro Parente); Supervision, A.C. and A.P. (Alessandro Parente); Visualization, A.P. (Alberto Procacci); Writing—original draft, A.P. (Alberto Procacci) and M.C.; Writing—review and editing, S.S., M.M.K., A.C. and A.P. (Alessandro Parente). All authors have read and agreed to the published version of the manuscript.

Funding: This publication benefits from the support of the Walloon Region as part of the FRIA grant funding No. 40004712. M. Mustafa Kamal acknowledges research funding from “Le Fonds de la Recherche Scientifique—FNRS” awarded to the application file No. 40005579 (Researcher [CR]). This project has also received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program under Grant No. 714605, and the Energy Transition Fund of Belgium.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in Zenodo at <https://doi.org/10.5281/zenodo.7334331> (accessed on 18 November 2022).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DT	Digital Twin
CFD	Computational Fluid Dynamics
GPR	Gaussian Process Regression
NN	Neural Network
POD	Proper Orthogonal Decomposition
IRO	Intensified Relay Optics
CCD	Charge-Coupled Device
FTIR	Fourier Transform Infrared Spectroscopy
COG	Coke Oven Gas

Appendix A

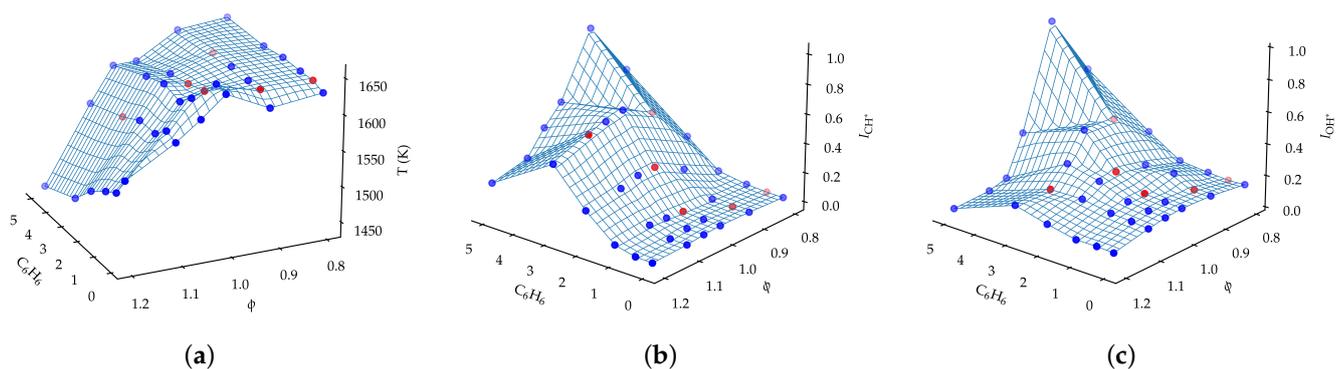


Figure A1. Cont.

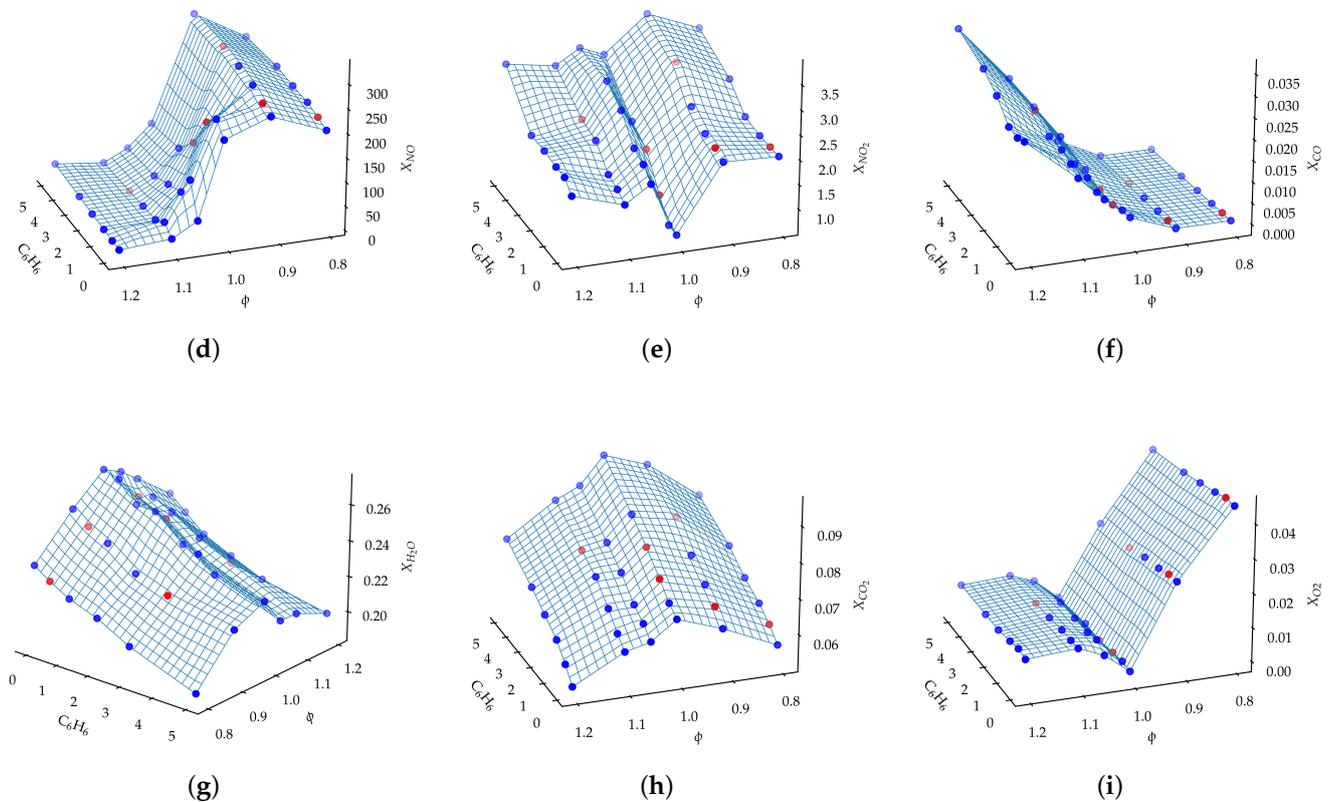


Figure A1. Maximum temperature (a); OH* normalized sum (b); CH* normalized sum (c); emissions of NO (d); NO₂ (e); CO (f); H₂O (g); CO₂ (h) and O₂ (i), as a function of the input parameters. The OH* and CH* values have been obtained by summing all the counts for each condition, and they have been normalized by the maximum value. The blue dots indicate the samples used for training while the red dots represent the samples used for testing.

References

- Chai, W.S.; Bao, Y.; Jin, P.; Tang, G.; Zhou, L. A review on ammonia, ammonia-hydrogen and ammonia-methane fuels. *Renew. Sustain. Energy Rev.* **2021**, *147*, 111254. [[CrossRef](#)]
- Liu, M.; Fang, S.; Dong, H.; Xu, C. Review of digital twin about concepts, technologies, and industrial applications. *J. Manuf. Syst.* **2021**, *58*, 346–361. [[CrossRef](#)]
- Semeraro, C.; Lezoche, M.; Panetto, H.; Dassisti, M. Digital twin paradigm: A systematic literature review. *Comput. Ind.* **2021**, *130*, 103469. [[CrossRef](#)]
- Fang, X.; Wang, H.; Liu, G.; Tian, X.; Ding, G.; Zhang, H. Industry application of digital twin: From concept to implementation. *Int. J. Adv. Manuf. Technol.* **2022**, *121*, 4289–4312. [[CrossRef](#)]
- Rasmussen, C.E., Gaussian Processes in Machine Learning. In *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, 20014 February 2003, Tübingen, Germany, 4–16 August 2003, Revised Lectures*; Bousquet, O., von Luxburg, U., Rätsch, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 63–71. [[CrossRef](#)]
- Berkooz, G.; Holmes, P.; Lumley, J.L. The proper orthogonal decomposition in the analysis of turbulent flows. *Annu. Rev. Fluid Mech.* **1993**, *25*, 539–575. [[CrossRef](#)]
- Aversano, G.; Ferrarotti, M.; Parente, A. Digital twin of a combustion furnace operating in flameless conditions: Reduced-order model development from CFD simulations. *Proc. Combust. Inst.* **2021**, *38*, 5373–5381. [[CrossRef](#)]
- Procacci, A.; Amaduzzi, R.; Coussement, A.; Parente, A. Adaptive digital twins of combustion systems using sparse sensing strategies. *Proc. Combust. Inst.* **2022**. [[CrossRef](#)]
- Giusti, A.; Mastorakos, E. Turbulent Combustion Modelling and Experiments: Recent Trends and Developments. *Flow Turbul. Combust.* **2019**, *103*, 847–869. [[CrossRef](#)]
- Raman, V.; Hassanaly, M. Emerging trends in numerical simulations of combustion systems. *Proc. Combust. Inst.* **2019**, *37*, 2073–2089. [[CrossRef](#)]
- Sharma, S.; Savarese, M.; Coussement, A.; Parente, A. Decarbonisation potential of dimethyl ether/hydrogen mixtures in a flameless furnace: Reactive structures and pollutant emissions. *Int. J. Hydrog. Energy* **2022**, *46*, 2401–2427. [[CrossRef](#)]

12. Ferrarotti, M.; De Paepe, W.; Parente, A. Reactive structures and NO_x emissions of methane/hydrogen mixtures in flameless combustion. *Int. J. Hydrog. Energy* **2021**, *46*, 34018–34045. [[CrossRef](#)]
13. Cafiero, M.; Sharma, S.; Kamal, M.M.; Coussement, A.; Parente, A. Experimental Investigation on Hydrogen-Rich Fuel Mixtures (H₂/CH₄/CO) Doped with C₆H₆ in a 20 kW Semi-Industrial Scale Furnace. 2022. Available online: <https://zenodo.org/record/7334332#.Y7bFFBVByUk> (accessed on 18 November 2022).
14. Cafiero, M.; Dias, V.; Stagni, A.; Nguyen, P.D.; Nowakowska, M.; Coussement, A.; Jeanmart, H.; Parente, A. The effect of benzene on the structure of low-pressure premixed H₂/CH₄/CO-air flames and related NO formation at different equivalence ratios. *Combust. Flame* **2021**, *232*, 111510. [[CrossRef](#)]
15. Razzaq, R.; Li, C.; Zhang, S. Coke oven gas: Availability, properties, purification, and utilization in China. *Fuel* **2013**, *113*, 287–299. [[CrossRef](#)]
16. Uribe-Soto, W.; Portha, J.F.; Commenge, J.M.; Falk, L. A review of thermochemical processes and technologies to use steelworks off-gases. *Renew. Sustain. Energy Rev.* **2017**, *74*, 809–823. [[CrossRef](#)]
17. Bermúdez, J.M.; Arenillas, A.; Luque, R.; Menéndez, J.A. An overview of novel technologies to valorise coke oven gas surplus. *Fuel Process. Technol.* **2013**, *110*, 150–159. [[CrossRef](#)]
18. Gardner, J.R.; Pleiss, G.; Bindel, D.; Weinberger, K.Q.; Wilson, A.G. GPpyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 7576–7586.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.