CA22151: Cyber-Physical systems and digital twins for the decarbonisation of energy-intensive industries

Deliverable D4

Report on multi-fidelity data fusion for the construction of self-updating digital twins

Authors: Alessandro Parente (ULB), Aysu Özden (ULB), Anh Khoa Doan (TU Delft), Katarzyna Bizon (PK)

Funded by the European Union

Table of Contents

1. Introduction 3
2. Methodology 3
2.1. Dimensionality Reduction 4
2.2 Manifold Alignment 4
2.3 Regression 5
3. Multi-Fidelity Applications 6
4. Conclusion 8
Acknowledgments 8
References 9

1. Introduction

The increasing complexity of engineering systems has driven the need for accurate yet inexpensive tools to explore and predict a potentially wide range of operational conditions and device geometries. In recent years, data-driven approaches have gained increasing attention within the computational science and engineering communities, particularly when traditional numerical and experimental tools prove insufficient for tasks like design exploration, optimization, or providing real-time predictions. These limitations often arise from the need for numerous model evaluations, making conventional methods inefficient or impractical.

The design of complex systems is an iterative process that requires a certain number of model evaluations and exploring "what if" scenarios. As such, it cannot rely solely on high-fidelity tools, admitting that they are available for the scale of interest. A combination of approaches is needed to bridge information from high and low-fidelity models into a reduced representation of a complex asset.

The concept of multi-fidelity modeling has gained considerable attention within the scientific community, as it allows the use of many inexpensive and simple, yet potentially inaccurate, low-fidelity observations or simulations while enhancing their accuracy by integrating a smaller set of expensive but accurate high-fidelity data. These models enable substantial computational gains and provide accurate predictions by leveraging the cross-correlations between the low- and high-fidelity data through machine learning techniques for problems that would be nearly impossible to solve if one relied solely on the expensive high-fidelity data.

In the literature, simulations are predominantly employed as sources of low- and high-fidelity data. In this context, it is possible to obtain different levels of fidelity by varying the physical model, the size of the computational grid, the computational time step, the convergence level of the simulations, and/or combining numerical data with experimental results [1]. Often, multiple types of low-fidelity models exist for a high-fidelity model, including coarse-grid approximations, projection-based reduced models, data-fitting interpolation and regression models, machine learning-based models, and simplified models. These low-fidelity models differ in terms of error and cost.

2. Methodology

In a multi-fidelity framework, although both low- and high-fidelity datasets represent the same problem, differences in fidelity present challenges for knowledge transfer between them. However, the assumption that both datasets share a common low-dimensional latent manifold due to their inherent similarities enables comparison of the embedded data. Besides allowing the comparison of low- and high-fidelity datasets, working in a lower dimensional space allows filtering noise and makes the model more general, as the focus is on the relevant features controlling its behavior and evolution. Moreover, when the problem dimensionality becomes very large, dimensionality reduction allows us to tackle the so-called curse of dimensionality by focusing on a smaller number of features.

2.1. Dimensionality Reduction

Dimensionality reduction techniques are widely used across different disciplines. The most popular dimensionality reduction technique, Proper Orthogonal Decomposition (POD), also known as Principal Component Analysis (PCA) [2], identifies linear combinations of the variables of interest in state space and captures the most significant patterns and variations in the dataset. POD and PCA are multi-linear dimensionality reduction techniques, and, as such, they can become underoptimal for data showing strong non-linearities. In this case, non-linear dimensionality reduction techniques are preferred. The latter can be broadly classified into two categories: (1) autoencoders and (2) manifold learning methods [3]. Autoencoders aim to find a compressed representation of the data while retaining as much information as possible. POD can be viewed as a linear autoencoder, while classic examples of non-linear autoencoders include Artificial Neural Networks (ANN) [4,5] and kernel PCA [6,7]. On the other hand, manifold learning methods seek to discover low-dimensional representations that preserve a certain measure of similarity within the data without focusing on the loss of information during reconstruction [8]. Prominent techniques in this category include Locally Linear Embedding (LLE) [9,10,11], ISOMAPs [12,13], and spectral submanifolds [14].

2.2 Manifold Alignment

If low- and high-fidelity data are available and latent spaces are explored for both, the latter will likely be different, and suitable approaches should be developed to assess their difference. Characterizing the discrepancy between the latent space is key to model correction terms that can be used to increase the fidelity of a low-fidelity model. For that purpose, manifold alignment [15,16] represents a very appealing technique to identify this shared latent space, facilitating effective information transfer between the datasets. Wang and Mahadevan [16] describe two types of manifold alignment techniques. The first approach involves finding the intrinsic low-dimensional latent space for each dataset using any dimensionality reduction method, followed by aligning these latent manifolds through Procrustes analysis of the embedded data. The second approach identifies a shared latent space by constructing a joint graph Laplacian, aiming to preserve both individual and shared features of the datasets simultaneously within the resulting aligned latent space [17]. Manifold alignment techniques have been utilized in a variety of transfer learning problems such as protein alignment [18], image matching [15], magnetic resonance image classification [19], hyperspectral imaging visualization [20], automatic machine translation [21], etc.

2.3 Regression

Modeling the discrepancy between low- and high-fidelity data can be handled with regression models. This allows the correction of the predictions from a low-fidelity model at conditions not present in the original training data, allowing for exploring the design space. In multi-fidelity reduced-order modeling, the regression model combines values from different fidelities at observed locations to predict one variable, meaning the regression model must incorporate a multi-fidelity formulation. Common methods used in multi-fidelity frameworks include bridge functions with either multiplicative [22] or additive corrections [23,24], ANNs [25], as well as CoKriging [26,27,28] and Hierarchical Kriging [29], which offers accuracy comparable to CoKriging but with the advantage of being relatively simpler to implement [17]. Both CoKriging and Hierarchical Kriging are extensions of Kriging, also known as the Gaussian Process Regression (GPR) [30], which is a very popular technique to predict a distribution over possible values, which includes a mean value and a measure of uncertainty for a given input. The CoKriging and Hierarchical Kriging techniques found in the literature adapt the autoregressive model of Kennedy and O'Hagan [26]. This Markov property indicates that for the same training point, if high-fidelity information is available, low-fidelity data cannot improve prediction accuracy. An example of CoKriging is shown in Figure 1 using the toy datasets provided in [31]. The autoregressive characteristic is evident from the fact that the mean prediction passes through the high-fidelity samples with zero uncertainty at those locations.

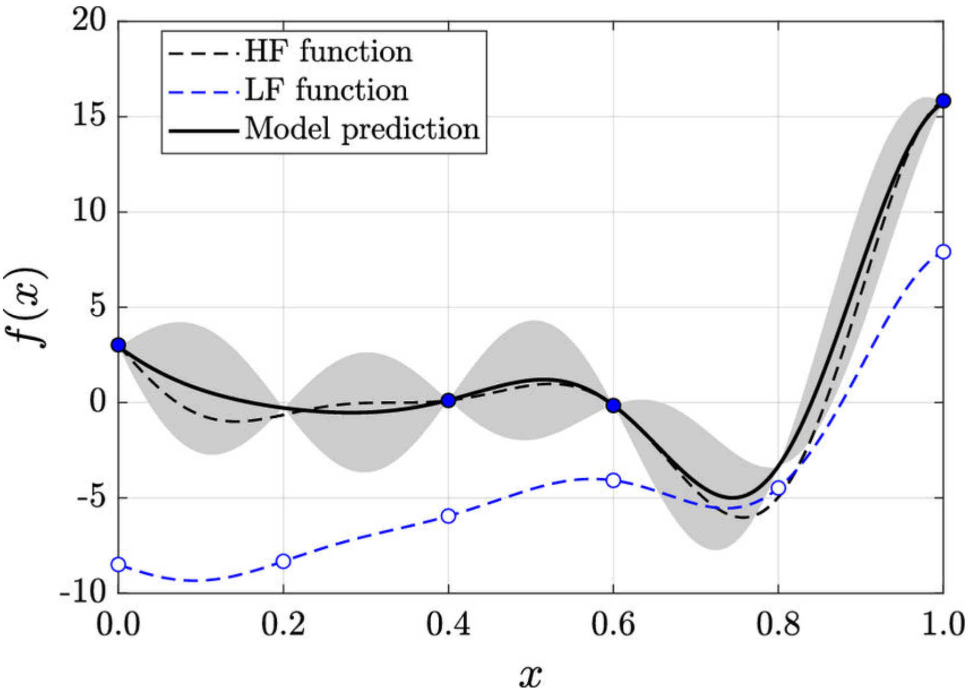

Figure 1 - Example of CoKriging regression [32]. The autoregressive characteristic is shown by mean prediction passing through the high-fidelity samples with zero uncertainty at those locations.

3. Multi-Fidelity Applications

Multi-fidelity frameworks have been applied across various scientific disciplines, as illustrated by recent literature. Yang et al. [33] proposed a multi-fidelity reduced-order model and global optimization method for the rapid and accurate simulation and design of microfluidic concentration gradient generators. Perron et al. [34] utilized a multi-fidelity framework to study high-dimensional displacement and stress fields from a structural analysis involving differences in discretization size and structural topology. Liu et al. [35] performed a hull form optimization process for resistance and wake performance of a Japan bulk carrier (JBC) by using a multi-fidelity Co-Kriging surrogate model. Nony et al. [36] evaluated atmospheric flow dispersion by integrating Large Eddy Simulations (LES) with Reynolds Averaged Navier Stokes with transport equation (RANS-TE) in a multi-fidelity framework. Demo et al. [25] successfully combined the DeepONet architecture with POD and gappy POD for sensor data, testing their approach on a parametric benchmark function and a nonlinear parametric Navier-Stokes problem. Guo et al. [37] developed a multi-fidelity Gaussian process, inspired by hierarchical Kriging, to model flame frequency response, combining data from harmonic and broadband forcing.

Tong et al. [38] studied the combustion performance of an aero-engine combustion chamber with a POD-Hierarchical Kriging multi-fidelity framework. Özden et al. [39,40] were the first to propose this method to build a multi-fidelity digital shadow of a methane-hydrogen and ammonia combustion performances in a semi-industrial furnace and a stagnation point reverse flow combustion. The proposed framework for multi-fidelity digital-twin development is summarised in Figure 2.

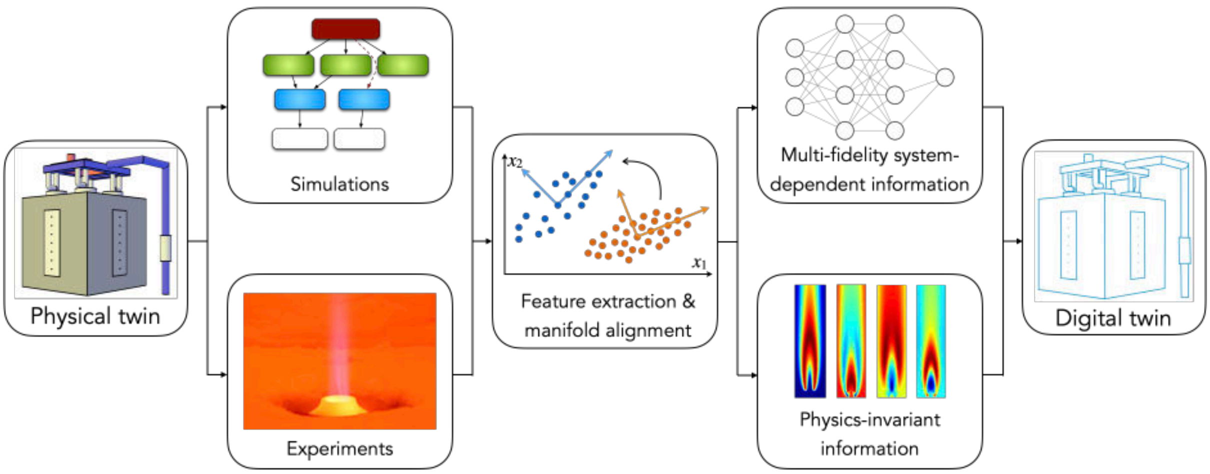

Figure 2 – Strategy for multi-fidelity digital twin development.

The impact of the high-fidelity simulations on the overall error associated with the multi-fidelity reduced-order model (ROM) is shown in Figure 3.

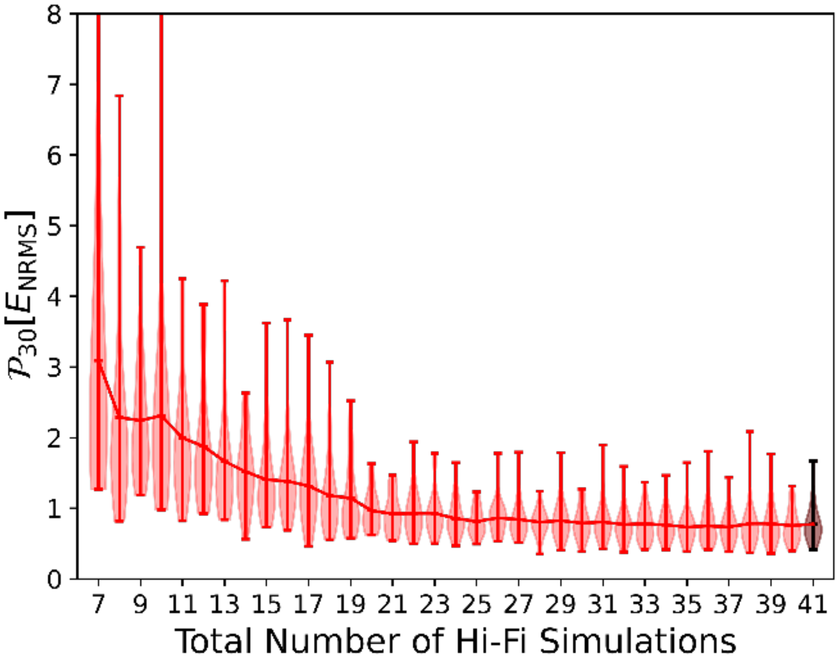

Figure 3 - 30-fold statistics of the NRMS error with increasing HiFi samples employing an uncertainty-based strategy for the selection of the HiFi simulations [39].

The selection of an optimal number of high-fidelity simulations to balance accuracy and training cost is key when building the multi-fidelity ROM. In [39], incremental sampling strategies based on the variance of the co-kriging model were developed to introduce high-fidelity simulations where the gain in terms of uncertainty reduction was higher. Figure 3 demonstrates that around half of the training cost can be saved while maintaining comparable error values with an appropriate choice of high-fidelity simulations.

Figure 4 shows the comparison of the OH radical (flame marker) obtained with the full CFD simulation (left) and the multi-fidelity ROM prediction using 41 low-fidelity simulations with (a) 9 and (b) 27 high-fidelity simulations. The OH field topology improves with additional HiFi simulations, reaching a satisfying flame location and peak OH value with 27 high-fidelity simulations (out of 41).

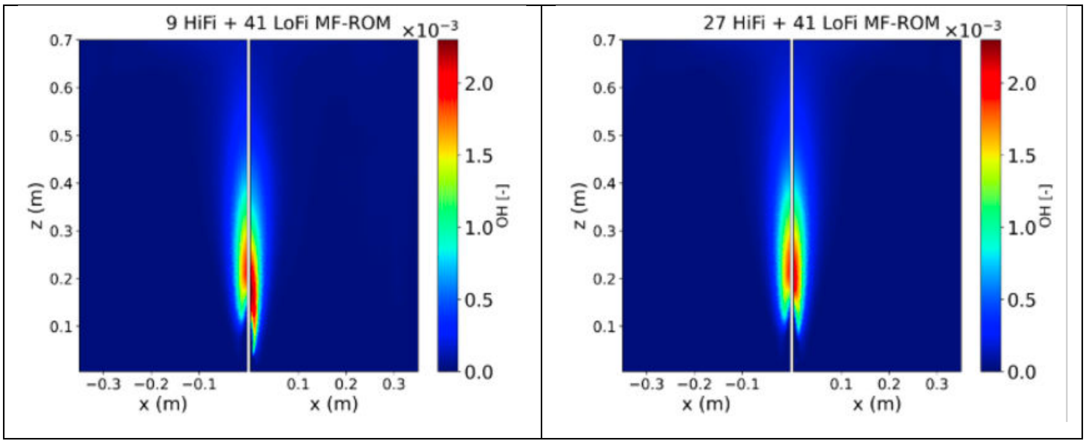

Figure 4 – Comparison of the full CFD simulation (left) and the multi-fidelity ROM prediction using 41 low-fidelity simulations with (a) 9 and (b) 27 high-fidelity simulations [39].

4. Conclusion

The proposed multi-fidelity framework relying on dimensionality reduction and non-linear regression is effective in building digital shadows of complex combustion systems. The combination of low- and high-fidelity data holds promise to reduce the computational cost associated with the generation of the numerical simulations required to generate the ROM. Future efforts from the working group members will focus on strategies to update the ROM in time, to adapt to changes in boundary and operating conditions.

Acknowledgments

This deliverable is based upon work from COST Action CA22151 – CYPHER, supported by COST (European Cooperation in Science and Technology). COST is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation. For more information, visit www.cost.eu.

References

[1] Y. Kuya, K. Takeda, X. Zhang, and A. I. Forrester, "Multifidelity surrogate modeling of experimental and computational aerodynamic data sets," *AIAA journal*, vol. 49, no. 2, pp. 289–298, 2011.

[2] A. Parente and J. C. Sutherland, "Principal component analysis of turbulent combustion data: Data preprocessing and manifold sensitivity," *Combustion and Flame*, vol. 160, no. 2, pp. 340–350, 2013.

[3] M. A. Mendez, "Linear and nonlinear dimensionality reduction from fluid mechanics to machine learning," *Measurement Science and Technology*, vol. 34, no. 4, p. 042001, 2023.

[4] G. Alain and Y. Bengio, "What regularized auto-encoders learn from the data-generating distribution," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3563–3593, 2014.

[5] I. Goodfellow, B. Yoshua, and A. Courville, "Deep learning," 2016.

[6] B. Ghojogh, F. Karray, and M. Crowley, "Eigenvalue and generalized eigenvalue problems: Tutorial," arXiv:1903.11240, 2023.

[7] B. Schölkopf, A. Smola, and K.R. Müller, "Kernel principal component analysis," in *International conference on artificial neural networks*, pp. 583–588, Springer, 1997.

[8] W.S. Zheng, J.H. Lai, and P. C. Yuen, "Weakly supervised learning on pre-image problem in kernel methods," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 2, pp. 711–715, IEEE, 2006.

[9] B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley, "Locally linear embedding and its variants: Tutorial and survey," arXiv:2011.10925, 2020.

[10] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[11] L. K. Saul and S. T. Roweis, "An introduction to locally linear embedding," unpublished. Available at: http://www.cs.toronto.edu/~roweis/lle/publications.html, 2001.

[12] B. Ghojogh, M. Crowley, F. Karray, and A. Ghodsi, "Multidimensional scaling, Sammon Mapping, and Isomap,"in *Elements of Dimensionality Reduction and Manifold Learning*. Springer, Cham. 2023.

[13] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[14] G. Haller, and S. Ponsioen, "Nonlinear normal modes and spectral submanifolds: existence, uniqueness and use in model reduction," *Nonlinear Dyn*, vol. 86, pp. 1493–1534, 2016.

[15] J. Ham, D. Lee, and L. Saul, "Semisupervised alignment of manifolds," in *International Workshop on Artificial Intelligence and Statistics*, pp. 120–127, PMLR, 2005.

[16] C. Wang and S. Mahadevan, "A general framework for manifold alignment," in *2009 AAAI Fall Symposium Series*, pp. 53–58, 2009.

[17] C. Perron, D. Rajaram, and D. Mavris, "Development of a multi-fidelity reduced-order model based on manifold alignment," in *AIAA Aviation 2020 Forum*, p. 3124, 2020.

[18] C. Wang and S. Mahadevan, "Manifold alignment using procrustes analysis," in *Proceedings of the 25th international conference on Machine learning*, pp. 1120–1127, 2008.

[19] R. Guerrero, C. Ledig, and D. Rueckert, "Manifold alignment and transfer learning for classification of alzheimer's disease," in *International Workshop on Machine Learning in Medical Imaging*, pp. 77–84, Springer,2014.

[20] D. Liao, Y. Qian, and J. Zhou, "Visualization of hyperspectral imaging data based on manifold alignment," in *22nd International Conference on Pattern Recognition*, pp. 70–75, IEEE, 2014.

[21] F. Diaz and D. Metzler, "Pseudo-aligned multilingual corpora.," in *IJCAI*, pp. 2727–2732, Citeseer, 2007.

[22] K. J. Chang, R. T. Haftka, G. L. Giles, and P.-J. Kao, "Sensitivity-based scaling for approximating structural response," *Journal of Aircraft*, vol. 30, no. 2, pp. 283–288, 1993.

[23] S. Choi, J. J. Alonso, and I. M. Kroo, "Two-level multifidelity design optimization studies for supersonic jets," *Journal of Aircraft*, vol. 46, no. 3, pp. 776–790, 2009.

[24] C. Tang, K. Gee, and S. Lawrence, "Generation of aerodynamic data using a design of experiment and data fusion approach," in *43rd AIAA Aerospace Sciences Meeting and Exhibit*, p. 1137, 2005.

[25] N. Demo, M. Tezzele, and G. Rozza, "A deeponet multi-fidelity approach for residual learning in reduced order modeling," *Advanced Modeling and Simulation in Engineering Sciences*, vol. 10, no. 1, p. 12, 2023.

[26] M. C. Kennedy and A. O'Hagan, "Predicting the output from a complex computer code when fast approximations are available," *Biometrika*, vol. 87, no. 1, pp. 1–13, 2000.

[27] J. Laurenceau and P. Sagaut, "Building efficient response surfaces of aerodynamic functions with kriging and cokriging," *AIAA journal*, vol. 46, no. 2, pp. 498–507, 2008.

[28] Z.-H. Han, Zimmermann, and S. Görtz, "Alternative cokriging method for variable-fidelity surrogate modeling," AIAA journal, vol. 50, no. 5, pp. 1205–1210, 2012.

[29] Z.-H. Han and S. Görtz, "Hierarchical kriging model for variable-fidelity surrogate modeling," *AIAA journal*, vol. 50, no. 9, pp. 1885–1896, 2012.

[30] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA, 2006.

[31] A. I. Forrester, A. Sóbester, and A. J. Keane, "Multi-fidelity optimization via surrogate modelling," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 463, no. 2088, pp. 3251–3269, 2007.

[32] L.G. Ribeiro, E. Parente, and A.M.C. de Melo, "Alternative variable-fidelity acquisition functions for efficient global optimization of black-box functions," *Structural and Multidisciplinary Optimization*, vol. 66, no. 147, 2023.

[33] H. Yang, S. H. Hong, G. Wang, and Y. Wang, "Multi-fidelity reduced-order model for gpu-enabled microfluidic concentration gradient design," *Engineering with Computers*, vol. 39, no. 4, pp. 2869–2887, 2023.

[34] C. Perron, D. Sarojini, D. Rajaram, J. Corman, and D. Mavris, "Manifold alignment-based multi-fidelity reduced-order modeling applied to structural analysis," *Structural and Multidisciplinary Optimization*, vol. 65, no. 8, p. 236, 2022.

[35] X. Liu, D. Wan, and L. Lei, "Multi-fidelity model and reduced-order method for comprehensive hydrodynamic performance optimization and prediction of jbc ship," *Ocean Engineering*, vol. 267, p. 113321, 2023.

[36] B. X. Nony, M. C. Rochoux, T. Jaravel, and D. Lucor, "Reduced-order model for microscale atmospheric dispersion combining multi-fidelity les and rans data," in *ECCOMAS Proceedia*, UNCECOMP 2023, 2023.

[37] S. Guo, C. F. Silva, and W. Polifke, "Robust identification of flame frequency response via multi-fidelity gaussian process approach," *Journal of Sound and Vibration*, vol. 502, p. 116083, 2021.

[38] S. Tong, Y. Ma, M. Guo, Y. Tian, W. Song, H. Wang, J. Le, and H. Zhang, "Optimization of aero-engine combustion chambers with the assistance of hierarchical-kriging surrogate model based on pod downscaling method," *Advances in Aerodynamics*, vol. 5, no. 1, p. 20, 2023.

[39] A. Özden, A. Procacci, R. M. Galassi, F. Contino, and A. Parente, "Incremental sampling methods for multifidelity surrogate modeling: Application on a furnace operating in mild combustion conditions," *Applied Thermal Engineering*, vol. 246, p. 122902, 2024.

[40] A. Özden, M. Savarese, L. Giuntini, A. Procacci, R. M. Galassi, A. Coussement, F. Contino, and A. Parente, "A multi-fidelity framework for developing digital twins of combustion systems from heterogeneous data: Application to ammonia combustion," Proceedings of the Combustion Institute, vol. 40, no. 1-4, p. 105608, 2024.