The 9th European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS Congress 2024
3–7 June 2024, Lisboa, Portugal

DEEP LEARNING MODEL ORDER REDUCTION FOR THE AUTOMOTIVE INDUSTRY

Eiximeno, Benet[1,2], Miró, Arnau[1,2], Rodríguez, Ivette[2] and Lehmkuhl, Oriol[1]

[1] Barcelona Supercomputing Center, 08034 Barcelona, Spain
[2] Turbulence and Aerodynamics Research Group. Universitat Politècnica de Catalunya (UPC), 08221 Terrassa, Spain

Key words: Reduced order models (ROM), deep learning, aerodynamics, variational autoencoders

Summary. This manuscript compares the model order reduction capacity of POD with the one of more modern techniques as β variational autoencoders in the wake of the Windsor body at $Re_L = 2.9 \times 10^6$. A β VAE with 20 latent dimensions which are 89.72% orthogonal between themselves recovers 96.75% of thw flow energy while 20 POD modes barely recover the 40%.

1 INTRODUCTION

When a car is exposed to lateral wind gusts, the yaw angle of the incident velocity increases, leading to an intensification of the suction at the vehicle's rear and elevated drag forces. This effect is particularly relevant in square-back vehicles, as their drag force is completely driven by the pressure on their back face. This work focuses on building a surrogate model for the changes in the aerodynamic performance with the yaw angle of a simplified square-back vehicle, the Windsor body.

Surrogate models are data-driven computational methods widely used across various scientific and engineering disciplines to approximate complex systems or functions. These models act as simplified substitutes for both experimental procedures and computationally intensive simulations, offering faster yet sufficiently accurate results [23]. Surrogate models are primarily employed to estimate optimal solutions or serve as essential tools for performance evaluation in the early stages of development, as they significantly reduce the resources needed for design exploration [18, 26].

Given the complexity and high dimensionality of the original model, surrogate models are typically constructed in a reduced space [26]. Dimensionality reduction can be achieved through algebraic methods such as proper orthogonal decomposition (POD) [20] and its variants [24], or by employing deep learning techniques. Examples of autoencoder-based applications in the field of fluid dynamics include [1, 5, 6, 10, 14, 21, 22, 25, 27]. The rising popularity of autoencoders for model order reduction is due to their ability to capture the non-linear behavior of dynamical systems with a higher compression capacity than any POD-based methodology [4, 5, 22, 25].

The aim of this study is to develop a similar CNN-based β variational autoencoder for model order reduction to the one implemented by Eivazi et al. [5] in a simplified urban environment to prove the capabilites of this methodology in such a common flow in the automotive industry. The rest of the manuscript includes a section describing the methodology, including the dataset

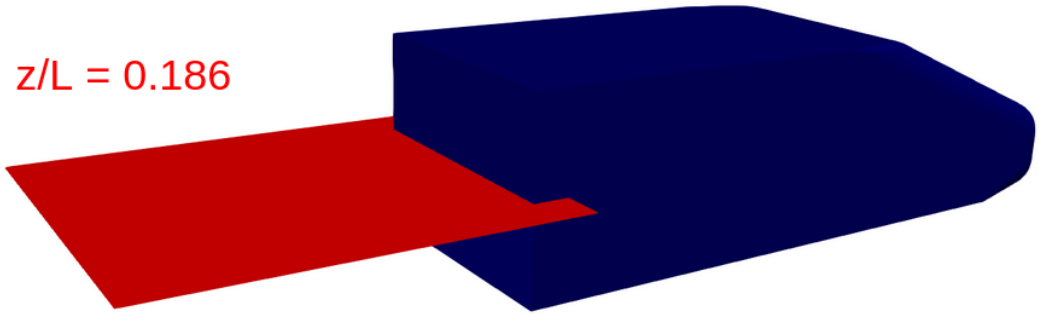


Figure 1: Geometry of the Windsor body and the working plane where the data is interpolated to develop the model. The plane is perpendicular to the vertical axis and is located at $z/L = 0.186$.

description, a numerical definition of POD and the architecture used for the β VAE, and finally the comparison between the compression capacity of the β VAE and POD for that case.

2 METHODOLOGY

This section describes the methods used in the manuscript, including the Windsor body dataset employed to test the methodology, a mathematical description of POD and the selection of the most significant modes, together with a description of the model used to add the energy from the truncated modes.

Dataset description

The test dataset consists of the turbulent wake behind the Windsor body, a simplified square-back vehicle illustrated in Figure 1, at a Reynolds number of $Re_L = U_\infty L/\nu = 2.9 \times 10^6$. Here, U_∞ is the free-stream velocity magnitude, L is the model length, and ν denotes the fluid's kinematic viscosity. The data was generated through wall-modeled large eddy simulations conducted at yaw angle of $\delta = 10°$. The simulations involved numerically solving the spatially filtered incompressible Navier-Stokes equations (Equation 1 and Equation 2) using SOD2D (Spectral high-Order coDe 2 solve partial Differential equations) [11,12], a spectral element method (SEM) code designed for low-dissipation.

$$\frac{\partial \overline{u}_i}{\partial x_i} = 0 \tag{1}$$

$$\frac{\partial \overline{u}_i}{\partial t} + \frac{\partial \overline{u}_i \overline{u}_j}{\partial x_j} - \nu \frac{\partial^2 \overline{u}_i}{\partial x_j \partial x_j} + \rho^{-1} \frac{\partial \overline{p}}{\partial x_i} = -\frac{\partial \mathcal{T}_{ij}}{\partial x_j} \tag{2}$$

In these equations, x_i represent the spatial coordinates (i.e., x, y, and z), u_i (i.e., u, v, and w) are the velocity components, and p is the pressure. The kinematic viscosity and fluid density are denoted by ν and ρ, respectively. Filtered variables are represented by an overline ($\overline{\cdot}$). The right-hand side of Equation 2 represents the sub-grid scale (SGS) stresses, with its anisotropic part given by:

$$\mathcal{T}_{ij} - \frac{1}{3}\mathcal{T}_{kk}\delta_{ij} = -2\nu_{sgs}\overline{\mathcal{S}}_{ij} \tag{3}$$

where the large-scale rate-of-strain tensor $\overline{\mathcal{S}}_{ij}$ is calculated as $\overline{\mathcal{S}}_{ij} = \frac{1}{2}\left(g_{ij} + g_{ji}\right)$, with $g_{ij} = \partial\overline{u}_i/\partial x_j$ and δ_{ij} being the Kronecker delta. The equations are closed using the local formulation of the integral length-scale approximation (ILSA) [19] as the SGS viscosity (ν_{sgs}).

The run was extended for 60 convective time units, $t = 60L/U_\infty$, after the initial transient phase was completed. A total of 660 snapshots were collected during this period. The data used for model evaluation was interpolated onto the red plane shown in Figure 1. This plane is perpendicular to the vertical axis and positioned at $z/L = 0.186$. For a detailed description of the numerical methodology, domain, and grid used in the simulations, as well as validation, the reader is referred to previous work by Eiximeno et al. [9].

Proper Orthogonal Decomposition (POD)

In this work, Proper Orthogonal Decomposition (POD) is employed as the baseline dimensionality reduction technique due to its efficiency in capturing an infinite-dimensional process with a finite number of modes [15]. POD is based on identifying a set of deterministic functions that represent the dominant features of the system, allowing for the decomposition of a field $F(X,t)$. This decomposition can be expressed as:

$$F(X,t) = \sum_{i=1}^{N} a_i(t)\Phi_i(X), \tag{4}$$

where N denotes the number of functions used to decompose the field. POD requires the spatial mode basis to be orthonormal, i.e.,

$$\int_{\mathbf{X}} \Phi_{i_1}(X)\Phi_{i_2}(X)\mathrm{d}x = \begin{cases} 1 \text{ if } i_1 = i_2 \\ 0 \text{ otherwise} \end{cases} \tag{5}$$

and optimal, meaning that the first N_r vectors are those that reconstruct the dataset with the minimum possible error.

In this study, Singular Value Decomposition (SVD) is the chosen method for performing POD. SVD decomposes the initial snapshot matrix, $\mathcal{X}$, into the left singular vectors, U, the singular values, S, and the right singular vectors, V, as follows:

$$\mathcal{X} = USV^T. \tag{6}$$

Each column of U contains a spatial mode, $\Phi_i(X)$, and each column of V provides the temporal evolution of the corresponding mode's coefficient, $a_i(t)$. The singular values, organized in a diagonal matrix, represent the energy contribution of each mode in descending order. The larger the singular value, the more energy the mode contains. The POD analysis in this work was conducted using pyLOM [7], a high-performance computing (HPC) enabled reduced order modeling code that implements a parallel and scalable algorithm for singular value decomposition [8].

Variational autoencoders (VAE)

Kingma and Welling [17] introduced an extension to standard autoencoders (AEs), known as variational autoencoders (VAEs), which allows for the mapping of input data into a smooth and informative latent space distribution. This approach addresses the limitations of standard AEs, particularly their fixed and non-informative latent space, which hinders the generation of new data. In this study, a Gaussian distribution is employed, where the input data is projected via $\mathcal{E}$ into a mean vector $\boldsymbol{\mu}$ and a standard deviation vector $\boldsymbol{\sigma}$. The latent space is sampled using a normally distributed random variable $\boldsymbol{\varepsilon}$, such that $\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\varepsilon}$. The loss function for a variational autoencoder includes a penalization term from the latent space, calculated using the Kullback-Leibler (KL) divergence, D_{KL}:

$$\mathcal{L}(\mathbf{x}) = \mathcal{L}_{rec} - \frac{\beta}{2} \sum_{i=1}^{d} \left(1 + \log(\sigma_i^2) - \mu_i^2 - \sigma_i^2\right) \tag{7}$$

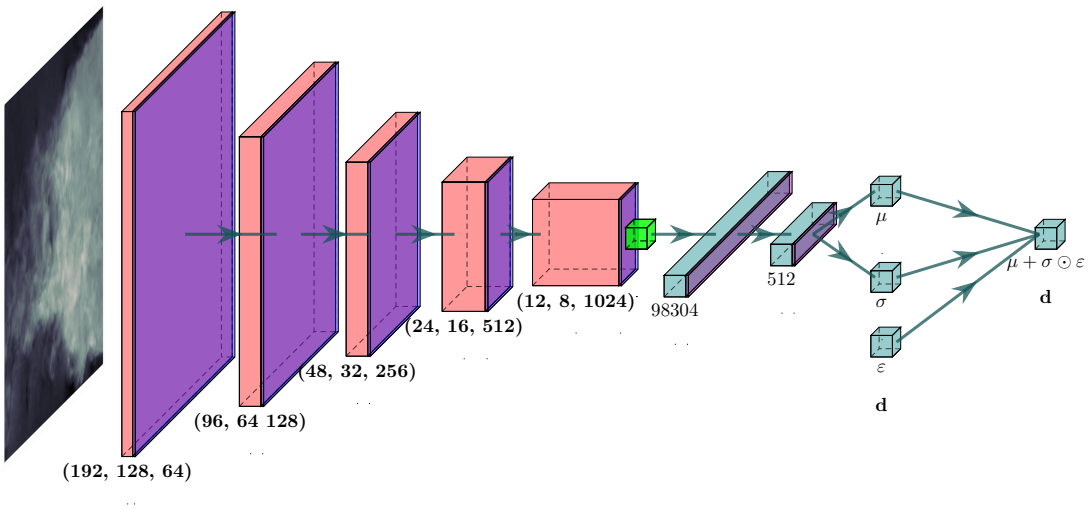
The term β is served as a regularization for the KL divergence used to disentangle the obtained latent vectors. In this work $\beta = 1 \times 10^{-3}$. The VAE architecture used in this study is shown in Figure 2. The input data consists of snapshots of streamwise velocity in a plane located at $z/L = 0.186$, with $n_x = 384$ points in the horizontal direction and $n_y = 256$ points in the vertical direction. The baseline architecture is adapted from previous works on autoencoders designed to compress fluid flow data [2, 5, 10, 25]. All of these studies utilized five convolutional layers to capture flow features. For the present application, using a filter with a kernel size that reduces the input data dimensions proved to yield more robust results than pooling layers for dimensionality reduction. Therefore, to halve the output data dimensions at each layer, a kernel size of (4×4) with a padding of 1 and a (2×2) stride is applied. The first layer employs a filter size of 16, which doubles at each subsequent layer to retain data patterns. After the final convolutional layer, the extracted features are flattened and fed into a fully-connected layer of length $l = 512$ to facilitate a smooth dimensionality reduction. The final step of the encoder involves mapping the features into two parallel fully-connected layers of dimension $\mathbf{d}$, representing the mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}$ of the latent distribution. In this work, $\mathbf{d}$ is set to $\mathbf{d} = 2$. The sizes of the filter, the fully-connected layer, and the latent space are chosen as the minimum required to achieve coherent reconstructions of the streamwise velocity.

The decoder is symmetrically designed, with fully-connected layers receiving input from the latent distribution. The data is then reshaped to match the output of the encoder's final convolutional layer. Five transposed convolutional layers are employed to progressively reconstruct the original data by increasing the spatial dimensions. In this study, the hyperbolic tangent (tanh) function is used, as it performed best with the data under consideration. All layers in the model are initialized using the Xavier uniform initialization scheme [13].

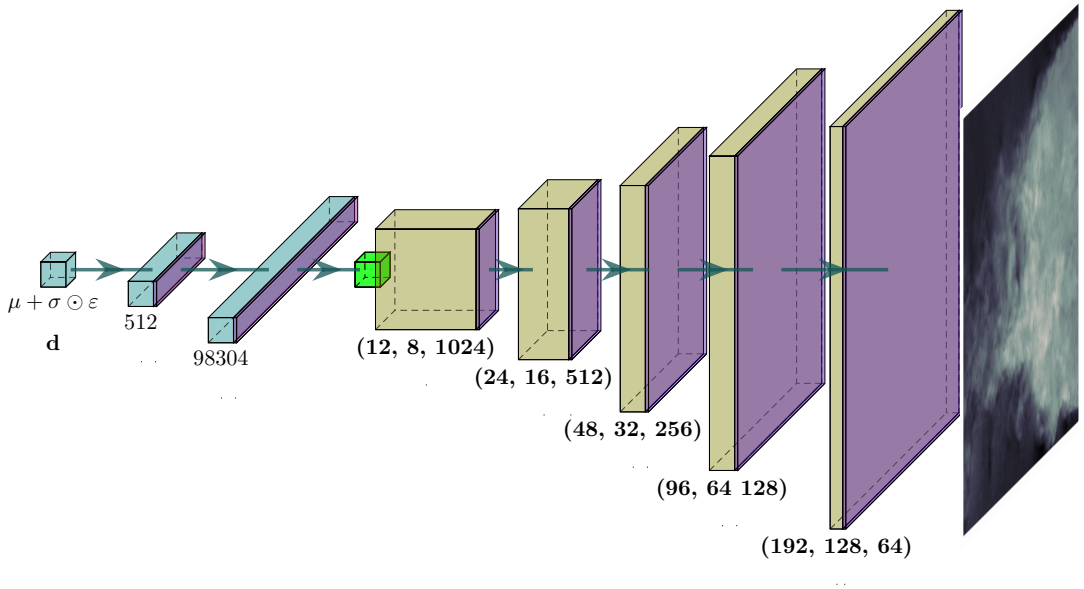
The reconstruction loss $\mathcal{L}_{\text{rec}}$ is defined as the mean-squared error (MSE), and the model parameters $\mathbf{w}$ are updated using the Adam optimization algorithm [16]. To ensure convergence of the loss function, the model is trained for 300 epochs with a learning rate lr, which follows a decay schedule starting from 5×10^{-4}:

$$lr_i = \frac{lr_{i-1}}{1 + 1 \times 10^{-4} \cdot i} \tag{8}$$

where i denotes the current epoch number. To prevent overfitting of $\mathbf{w}$, an early stopping



(a)



(b)

Figure 2: Encoder (a) and decoder (b) architectures. The output size of each layer is indicated under the corresponding block in the format (Height × Width × Channels). The color coding for each layer is as follows: 2D-convolution layer, 2D-transpose convolution layer, fully-connected layer, reshape, and tanh activation function.

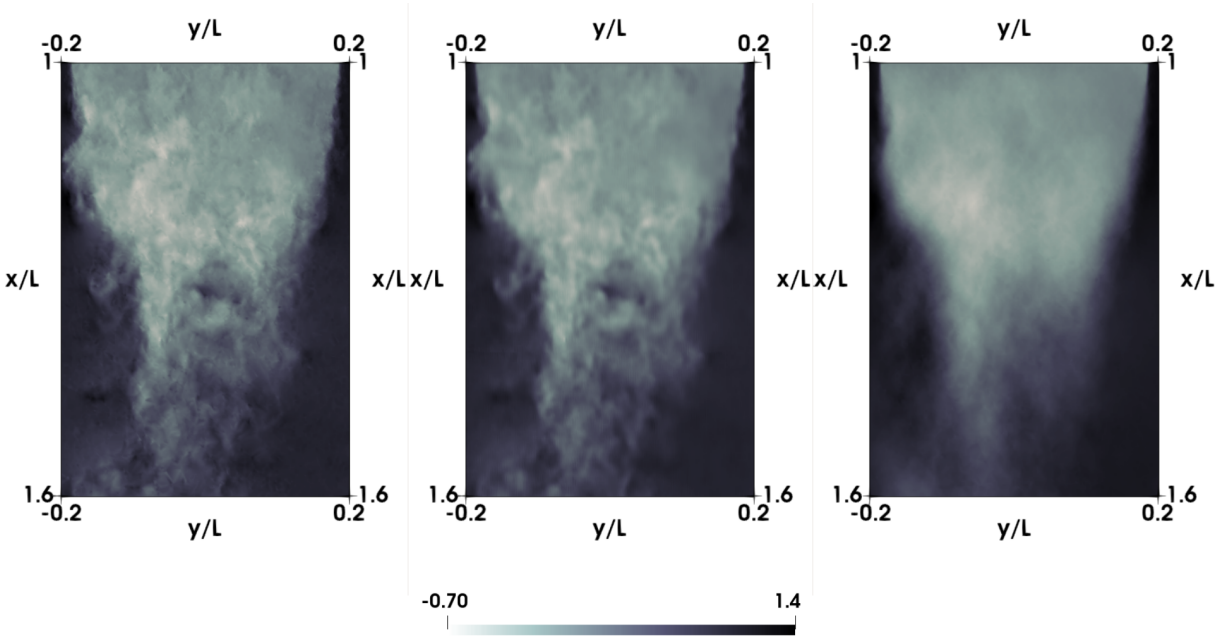


Figure 3: Comparison between the original flow (left), the autoencoder reconstruction (center) and POD reconstruction (right)

mechanism is applied based on the total loss $\mathcal{L}(\mathbf{x})$, halting training when no further improvement is observed [3].

Eighty percent of the snapshots collected were used to train the autoencoder, while the remaining 20% served as a validation set to evaluate potential overfitting to the training data.

3 RESULTS

Figure 3 compares the reconstruction of one snapshot of the streamwise velocity with the original dataset using the autoencoder with 20 latent dimensions and the reconstruction with 20 POD modes, effectively showcasing the high compression capacity of the variational autoencoder. The 20 latent dimensions of the variational autoencoder recover up to 96.75% of the total flow energy while the 20 POD modes barely reach the 40% of the flow energy.

Thanks to the regularization parameter β it was possible to obtain a near-disentangled latent representation of the fluid flow as the determinant of the correlation matrix of the latent vectors is of 89.72%.

4 CONCLUSIONS

This work provides additional evidence of the capacity of variational autoencoders for model order reduction. The model can be further extended for more flow conditions to create a surrogate model between them. However, one should note that the core of this methodology is the usage of spatial convolutional layers and is restricted to flows that can be meshed with a regular grid. The future work by the authors in this topic will be directed on methodologies that avoid the need of having a regular grid and can extend the model to additional flow conditions.

REFERENCES

[1] N. Akkari, F. Casenave, Elie Hachem, and D. Ryckelynck. A bayesian nonlinear reduced order modeling using variational autoencoders. *Fluids*, 7(10), 2022.

[2] Rossella Arcucci, Dunhui Xiao, Fangxin Fang, Ionel Michael Navon, Pin Wu, Christopher C. Pain, and Yi-Ke Guo. A reduced order with data assimilation model: Theory and practice. *Computers & Fluids*, 257:105862, 2023.

[3] Yingbin Bai, Erkun Yang, Bo Han, Yanhua Yang, Jiatong Li, Yinian Mao, Gang Niu, and Tongliang Liu. Understanding and improving early stopping for learning with noisy labels. *arXiv preprint arXiv:2106.15853*, 2021.

[4] Steven L. Brunton, Bernd R. Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, 52(1):477–508, 2020.

[5] H. Eivazi, S. L. C. Martínez, S. Hoyas, and R. Vinuesa. Towards extraction of orthogonal and parsimonious non-linear modes from turbulent flows. *Expert Systems with Applications*, 202:117038, 2022.

[6] H. Eivazi, H. Veisi, M. H. Naderi, and V. Esfahanian. Deep neural networks for nonlinear model order reduction of unsteady flows. *Physics of Fluids*, 32:Article 105104, 2020.

[7] Benet Eiximeno, Beka Begiashvili, Arnau Miro, Eusebio Valero, and Oriol Lehmkuhl. pylom: Low order modelling in python, 2024.

[8] Benet Eiximeno, Arnau Miró, Beka Begiashvili, Eusebio Valero, Ivette Rodriguez, and Oriol Lehmkuhl. pyLOM: A HPC open source reduced order model suite for fluid dynamics applications. *arXiv preprint arXiv:2405.15529*, 2024.

[9] Benet Eiximeno, Arnau Miró, Ivette Rodríguez, and Oriol Lehmkuhl. Toward the usage of deep learning surrogate models in ground vehicle aerodynamics. *Mathematics*, 12(7):998, 2024.

[10] K. Fukami, T. Nakamura, and K. Fukagata. Convolutional neural network based hierarchical autoencoder for nonlinear mode decomposition of fluid field data. *Physics of Fluids*, 32:Article 095110, 2020.

[11] L. Gasparino, F. Spiga, and O. Lehmkuhl. Sod2d: A gpu-enabled spectral finite elements method for compressible scale-resolving simulations. *Computer Physics Communications*, 297:109067, 2024.

[12] Lucas Gasparino, Jordi Muela, and Oriol Lehmkuhl. Sod2d repository. `https://gitlab.com/LucasBSC21387/sod2d_gitlab/-/tags`.

[13] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9, pages 249–256, Chia Laguna Resort, Sardinia, Italy, 2010. PMLR.

[14] Derrick Hines and P. Bekemeyer. Graph neural networks for the prediction of aircraft surface pressure distributions. *Aerospace Science and Technology*, null:108268, 2023.

[15] Philip J. Holmes, John L. Lumley, Gal Berkooz, Jonathan C. Mattingly, and Ralf W. Wittenberg. Low-dimensional models of coherent structures in turbulence. *Physics Reports*, 287(4):337–384, 1997.

[16] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2017.

[17] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *arXiv*, 2013.

[18] Yuichi Kuya, Kenji Takeda, Xin Zhang, and Alexander I. J. Forrester. Multifidelity surrogate modeling of experimental and computational aerodynamic data sets. *AIAA Journal*, 49(2):289–298, 2011.

[19] Oriol Lehmkuhl, Ugo Piomelli, and Guillaume Houzeaux. On the extension of the integral length-scale approximation model to complex geometries. *International Journal of Heat and Fluid Flow*, 78:108422, 2019.

[20] J. L. Lumley. Rational Approach to Relations between Motions of Differing Scales in Turbulent Flows. *Physics of Fluids*, 10(7):1405, 1981.

[21] Takaaki Murata, Kai Fukami, and K. Fukagata. Nonlinear mode decomposition with convolutional neural networks for fluid dynamics. *Journal of Fluid Mechanics*, null(822):null, 2019.

[22] Alberto Solera-Rico, Carlos Sanmiguel Vila, Miguel Gómez-López, Yuning Wang, Abdulrahman Almashjary, Scott TM Dawson, and Ricardo Vinuesa. β-variational autoencoders and transformers for reduced-order modelling of fluid flows. *Nature Communications*, 15(1):1361, 2024.

[23] Gang Sun and Shuyue Wang. A review of the artificial neural network surrogate modeling in aerodynamic design. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 233(16):5863–5872, 2019.

[24] Aaron Towne, Oliver T Schmidt, and Tim Colonius. Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis. *Journal of Fluid Mechanics*, 847:821–867, 2018.

[25] Yuning Wang, Alberto Solera-Rico, Carlos Sanmiguel Vila, and Ricardo Vinuesa. Towards optimal beta-variational autoencoders combined with transformers for reduced-order modelling of turbulent flows. *International Journal of Heat and Fluid Flow*, 105:109254, 2024.

[26] Raul Yondo, Esther Andrés, and Eusebio Valero. A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. *Progress in Aerospace Sciences*, 96:23–61, 2018.

[27] Bo Zhang. Nonlinear mode decomposition via physics-assimilated convolutional autoencoder for unsteady flows over an airfoil. *Physics of Fluids*, 5(0164250), 2023.