A POD-based Autoencoder for Detailed Loss Characterization of Low Pressure Turbine Blades

M Russo[1], **D Petronio**[1], **D Lengani**[1], **D Simoni**[1], **F Bertini**[2]

[1]DIME, Universita` degli Studi di Genova, Genova, IT
[2]Avio Aero, Torino, IT

Keywords: Machine learning, POD, LPT, Loss prediction

Abstract This study introduces a data-driven approach for the tuning of a model capable of identifying the loss sources characterizing different Low Pressure Turbine (LPT) blade geometries. The method involves two main steps: first, Proper Orthogonal Decomposition (POD) is applied to Large Eddy Simulations (LES) of turbine cascades to analyze local entropy production and separate the contribution of coherent flow structures to turbulent kinetic energy. Then, the results of the spatial and modal characterization of losses are used to train a model that predicts loss trends based on blade geometry. The process involves decomposing blade shapes and loss distributions into modes and coefficients, allowing for a low-dimensional representation of both predictors and targeted functions. This enables the creation of an efficient model that correlates the compressed representation of geometry and loss distributions to optimize blade shapes while minimizing overfitting. The application of the POD encoder demonstrates that, despite the simplicity of the model and the limited data available, it provides significant results and allows for an easy interrogation of the profile geometry's design space, thus enabling the identification of the geometric parameters that influence loss mechanisms in different regions of the examined domain.

1. Introduction

The design of Low-Pressure Turbine (LPT) blades has long played a crucial role in the pursuit of improved performance in modern turbomachinery for air propulsion, as their operation profoundly influences overall engine efficiency and fuel consumption. Despite decades of research, the complete understanding of loss mechanisms associated with LPT blade geometries remains challenging due to the complex, unsteady nature of the flow fields involved, which depend on a variety of geometric and operational parameters. Recent advances in computational fluid dynamics, particularly the increasing availability of high-fidelity data from Large Eddy Simulations (LES) and Direct Numerical Simulations (DNS), have provided unprecedented insights into the detailed dynamics of loss generation mechanisms [1, 2, 3, 4]. However, this type of analysis still represents a significant cost at the industrial level, making it a tool that cannot always be systematically employed.

In recent years, machine learning has emerged as a promising approach to the optimization of turbomachinery components, enabling a more efficient allocation of economic and computational resources. It allows identifying more efficient development paths, from the preliminary design of new profiles to final testing [5]. In the scope of design optimization, autoencoders have emerged as a particularly effective technique for compressing and synthesizing high-dimensional information. Autoencoders are artificial neural networks designed to learn compact latent representations of data, allowing for the ex-

traction of essential features for further analysis [6]. In their simplest and most straightforward version, they are usually referred to as linear autoencoders, which rely on simple linear transformations, effectively performing a principal component analysis-like decomposition in a neural network framework. Linear autoencoders are computationally efficient and inherently robust in cases where limited data are available. In such scenarios, their ability to extract meaningful low-dimensional representations without requiring extensive and computationally expensive training data makes them particularly valuable.

Among these, Proper Orthogonal Decomposition (POD) has emerged as a powerful tool for analyzing and simplifying large and complex datasets. Originally introduced by Lumley in 1967 [7], POD can be used for the decomposition of turbulent flow fields into spatial and temporal modes, identifying the dominant coherent structures contributing to loss production. In the context of turbomachinery, POD has been successfully applied to isolate key loss mechanisms, attributing them to specific regions within the examined domain, and guide the optimization of blade geometries [8, 9, 10, 11]. As reported in [12, 13, 14], an orthogonal basis, such as the one spanned by POD modes, can also be effectively used to efficiently parameterize the geometry of blade profiles. The advantage of POD lies in generating a decomposition basis that maximizes the projection of the dataset onto it, allowing for the dimensionality reduction of the parameter space.

Building on this foundation, this paper presents a novel application of POD as a linear autoencoder framework for predicting profile losses in LPT cascades, starting from a limited set of high-fidelity simulations performed on different blade geometries. By leveraging POD properties, we can extract a low-dimensional latent space that efficiently captures the essential geometric variations and their corresponding impact on losses. Then, a functional relationship can be established between the reduced set of geometry descriptors and the reduced set of parameters representing the localization and intensity of loss mechanisms across sub-regions of the examined domain, such as boundary layers, wake, and passage region.

2. LES Database

The present study relies on Large Eddy Simulations (LES) conducted on six different LPT blades using the commercial solver STAR-CCM+. A segregated flow model with bounded-central differencing was employed to ensure accurate convective flux representation. Time integration was performed using a second-order implicit backward differentiation scheme, while the subgrid-scale turbulence was modeled through the WALE model [15]. Additionally, the Synthetic Eddy Model (SEM) [16] was adopted at the domain inlet to generate realistic turbulence structures. Each one of the six simulations was carried out under the incompressible flow assumption. For all cases, the computational setup is 2D-extruded and mimics a repeating-stage LPT environment, extending 25% of the axial chord in the span-wise direction. A polyhedral grid was used, consisting of approximately 25 million cells and refined to ensure wall-resolved conditions with $y^+ < 1$ and $x^+ = z^+ \approx 20$ over the blade surfaces. LES initialization was performed as described in [11]. Instantaneous flow fields were recorded over eight blade-passing periods, storing one snapshot every tenth time step, yielding $D = 400$ snapshots per profile. The operating conditions vary within narrow ranges typical of realistic LPT cascades under normal flight conditions, with the Reynolds number ranging between 7.5×10^4 and 1.1×10^5.

3. Autoencoder Model

3.1. Geometry Parametrization

A turbine cascade is generally identified by a long series of descriptive variables, including inlet and outlet flow angles, thickness distribution, camber line coordinates, and more. This would make searching for a functional relationship between geometry and profile losses a highly complex multidimensional problem, particularly in the case of a very limited dataset. The POD framework can be employed to reduce the dimensionality of the geometry data matrix. Specifically, a blade geometry can be described by a vector containing pairs of coordinate points (x, y).

In this study, a common x-coordinate array was defined and each blade profile was represented by $M = 1600$ y-coordinate values. The normalized coordinates $y(x, p_i)/C_x$ of two adjacent blades (p_i is the representative index of the i-th profile in the dataset) were stored in the column space of the matrix $\boldsymbol{G}$, appropriately centered around the leftmost point near the leading edge of one of the two profiles. Duplicating each geometry array to include two adjacent blades allows for the preservation of information regarding the pitch-to-chord ratio of each cascade. The snapshot POD method, as proposed by Sirovich [17], was applied to the matrix $\boldsymbol{G'}$:

$$\boldsymbol{G'} = \boldsymbol{G} - \overline{\boldsymbol{G}} \tag{3.1}$$

where $\overline{\boldsymbol{G}}$ is the mean of the matrix $\boldsymbol{G}$. The POD generates two complete orthogonal bases, namely the POD modes (φ^k) and the related eigenvectors (x_k), such that:

$$\boldsymbol{G'}(x, p_i) = \sum_k x_{G,k}(p_i)\boldsymbol{\varphi}^k_G(x) \tag{3.2}$$

By truncating the geometry eigenvector matrix $\boldsymbol{\chi}_G$ at an index smaller than k, an approximated reconstruction of the matrix $\boldsymbol{G'}$ is obtained. The degree of approximation can be quantified by computing the mean squared error (MSE) between the original interpolated coordinates and those of the reduced-mode reconstruction. Truncating to the first four POD modes allowed the reconstruction MSE to be limited to 1%. Based on this observation, $k = n_G = 4$ coefficients $x_{G,(1,..,4)}(p_i)$ were selected for each geometry as descriptive parameters.

3.2. Loss Distribution Parametrization

A POD-based post-processing procedure, as described in [10, 11], was employed to split the contribution of individual flow dynamics to the turbulent dissipation for each of the six LES cascades. A more in-depth examination of the contribution of various flow features to overall profile losses was pursued by evaluating the volume integral of the viscous dissipation and the POD-decomposed turbulent dissipation within $Z = 5$ sub-regions limited to the core flow, the boundary layers (BL), and the mixing region downstream of the vane trailing edge, as depicted in figure 1. The resulting cumulative POD-based loss distribution is shown in figure 2 for an exemplary cascade pertaining to the dataset.

A decomposition framework analogous to the one used for the geometry can also be applied to the obtained loss distributions. Similar to the parametrization of the profile geometries, the normalized total dissipation values associated with each coherent flow structure (i.e., each spatial mode into which the turbulent flow has been decomposed) and each integration sub-region were collected in the columns of a matrix $\boldsymbol{L}$, from which the mean was subsequently subtracted. The resulting matrix $\boldsymbol{L'}$ was then

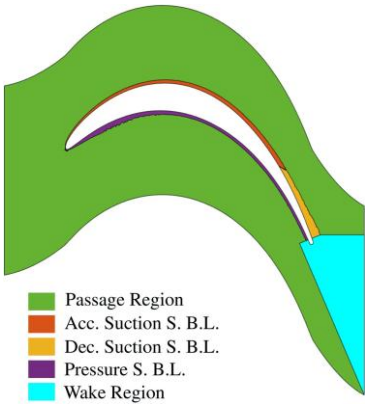

Figure 1. Sub-regions for integration of viscous and turbulent losses.

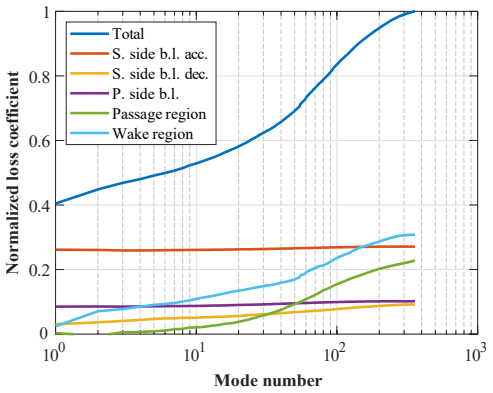

Figure 2. POD mode cumulative contribution to stagnation pressure losses. Data are normalized by the total loss for the present sample blade profile.

decomposed into the corresponding POD modes and eigenvectors:

$$\boldsymbol{L}'(j, p_i) = \sum_k x_{L,k}(p_i)\boldsymbol{\varphi}^k_L(j) \tag{3.3}$$

where j is the coherent flow structure index, replicated for every sub-region. An error metric based on the reconstruction MSE of the loss distribution was considered for selecting the truncation order of the $\boldsymbol{\chi}_L$ matrix. The minimum number of modes required for a reconstruction with an MSE below 1% in all cases was determined to be equal to $k = n_L = 3$. Thus, three coefficients $x_{L,(1,..,3)}(p_i)$ were selected for each profile's loss distribution as descriptive parameters.

3.3. Model Training

The reduced sets of parameters obtained through the encoding procedure of both geometries and loss characteristics were selected as input and output datasets for a low-dimensionality model for loss distribution prediction.

Various predictive models were tested to establish a meaningful relationship between the geometrical parameters and the associated loss features. Instead of relying on a simple linear model, we systematically evaluated all possible combinations of predictors up to the second order. Specifically, for each loss coefficient $x_{L,(1,..,3)}(p_i)$ (representing the objective function), models were constructed using different subsets of four predictors, chosen among the four first-order geometry coefficients $x_{G,(1,..,4)}(p_i)$, their squared values, and the cross-terms involving pairs of coefficients. The maximum order of predictors was restricted to quadratic terms due to the scarcity of training points. For this reason, more complex models, such as deep neural networks or high-order polynomial regressions, were deemed unsuitable, while certainly being better at capturing intricate and/or nonlinear relationships. For each loss coefficient, the best-performing model was selected based on the lowest MSE on the training data. The results showed that the optimal predictor sets for the three loss coefficients consisted of:

- $\left[x_{G,1},\ x_{G,2},\ x_{G,1} * x_{G,4},\ x_{G,2} * x_{G,3}\right]$ for $x_{L,1}$;
- $\left[(x_{G,4})^2,\ x_{G,1} * x_{G,3},\ x_{G,1} * x_{G,4},\ x_{G,3} * x_{G,4}\right]$ for $x_{L,2}$;

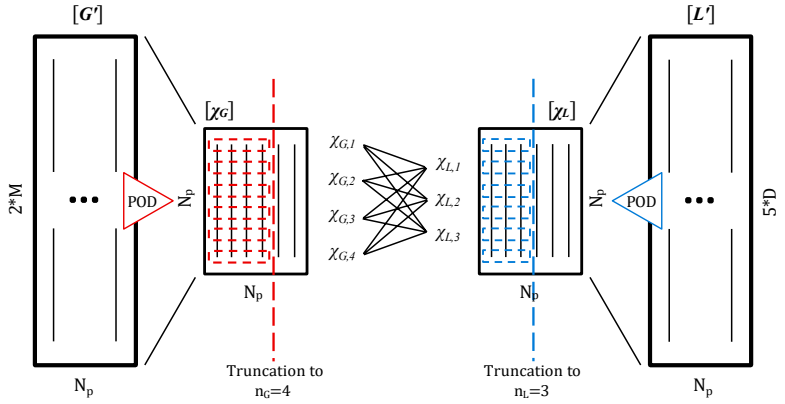

Figure 3. Schematic representation of the POD-encoder model.

- $\left[x_{G,2}\,,\quad (x_{G,1})^2\,,\quad (x_{G,3})^2\,,\quad x_{G,1} * x_{G,4}\right]$ for $x_{L,3}$.

This strategy allowed mitigating the risk of overfitting and ensured that the key underlying relationships between profile geometry and different loss mechanisms (observed in the 5 different spatial sub-regions) could be effectively captured while maintaining computational efficiency and avoiding unnecessary complexity. Figure 3 schematically summarizes the overall structure of the POD-encoder model, highlighting the connection between geometry and loss distribution parametrization.

4. Results and Discussion

The best-fitting functions obtained for each output variable are presented in figure 4 as surfaces in terms of the first two geometry parameters ($x_{G,1}$ and $x_{G,2}$), while the other two parameters ($x_{G,3}$ and $x_{G,4}$) are fixed at mean values. In the plots, the training points are also shown as red dots, providing a visual reference for the distribution of the original dataset in the chosen subspace. The differences in the predictor sets across the three models highlight the varying sensitivity of each loss parameter to different geometric features.

The results of the model training were analyzed by generating a dataset of 10,000 synthetic geometries as input. This dataset was constructed by systematically varying each of the four geometry coefficients ($x_{G,(1,..,4)}(p_i)$) over a discrete set of 10 values, spanning the full range observed across the six

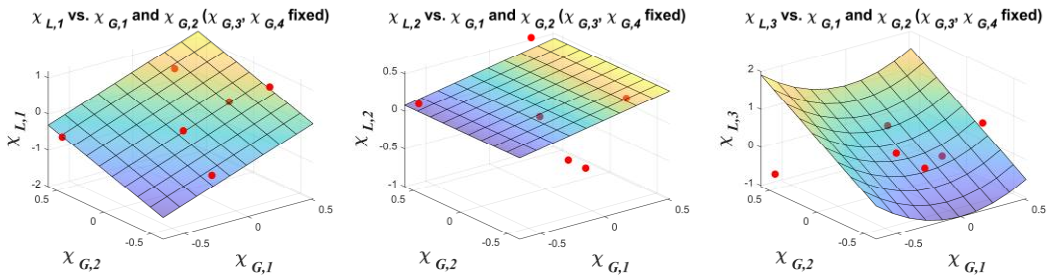

Figure 4. Best-fitting surfaces for the three loss coefficients as functions of the first two geometric parameters, with the remaining parameters fixed at constant values. Training points are also shown in red.

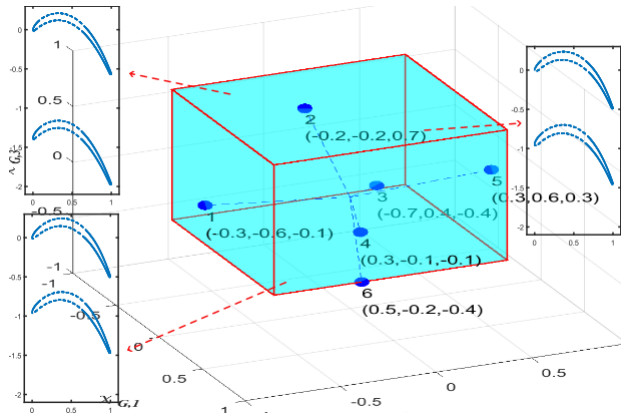

Figure 5. Representation of the geometry parameter space in the ($x_{G,1}$, $x_{G,2}$, $x_{G,3}$) plane.

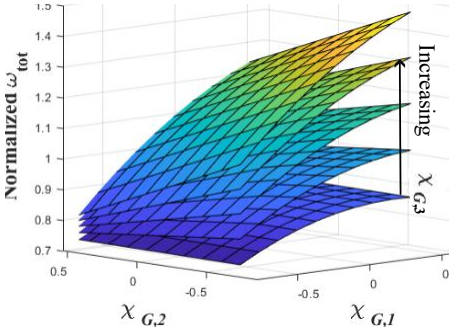

Figure 6. Multidimensional response surface obtained from the hypercube parameter space as a function of $x_{G,1}$ and $x_{G,2}$, for discrete values of $x_{G,3}$ and the mean value of $x_{G,4}$.

original blade profiles. Figure 5 provides a graphical representation of the parameter space, illustrating the hypercube defined by the first three geometry coefficients while keeping the fourth coefficient at its mean value. The geometries corresponding to the parameter combinations at three sample edges of the cube are also depicted, highlighting the extent of variability captured in the generated dataset. By evaluating the trained regression model on this dataset, a multidimensional response surface was obtained, mapping the geometry parameter variations to the three loss-related output coefficients ($x_{L,(1,..,3)}(p_i)$).

The response surface is shown in figure 6, as derived by computing the total pressure loss coefficient ω_{tot} for each geometry sampled within the hypercube parameter space. The value of ω_{tot} is derived as normalization of the viscous dissipation and TKE production associated with each coherent flow structure over the entire computational domain, reconstructed from the corresponding loss parameters predicted by the model:

$$\omega_{tot} = \frac{\int \frac{\overline{Dp_t}}{Dt} dV}{\dot{V} \cdot p_{in}} \quad \text{with} \quad \frac{\overline{Dp_t}}{Dt} = -\overline{\tau}_{ij}\frac{\partial \overline{u}_i}{\partial x_j} - -\rho\overline{u_i'u_j'}\frac{\partial \overline{u}_i}{\partial x_j} \tag{4.1}$$

In Eq. (4.1), V is the volume of the domain, $\dot{V}$ is the volumetric flow rate through it, and p_{din} is a reference dynamic pressure.

To further illustrate the capabilities of the model and the utility of its predictions, figure 7 presents the total pressure loss curves across the different sub-regions for the profile within the hypercube that exhibits the lowest loss coefficient within the maximum confidence region of the parameter space. Data are normalized with the global loss coefficient of the profile previously introduced as an example from the dataset (fig. 2), which is now set as the reference geometry. As expected, the profile features viscous losses mainly concentrated in the blade boundary layer, showing minimal dependence on mode order in the accelerating suction side (red curve) and pressure side (violet curve) regions. The most significant contribution to loss production is attributed to the decelerating suction side region (yellow curve), where turbulence-induced losses are more pronounced. The same can be observed for the wake region (cyan curve) and, albeit with minimal contribution, for the passage region (green curve). Overall, a 20% reduction in the total loss coefficient is achieved compared to the reference. The selected profile geometry,

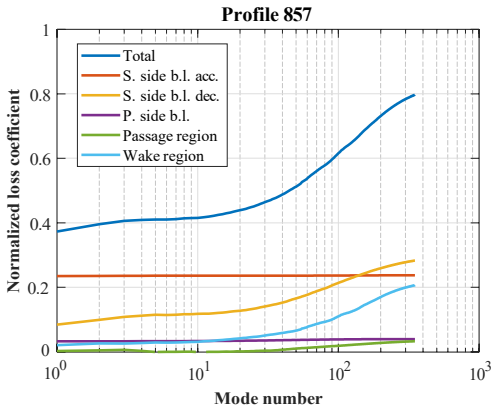

Figure 7. POD mode cumulative contribution to stagnation pressure losses for the selected profile.

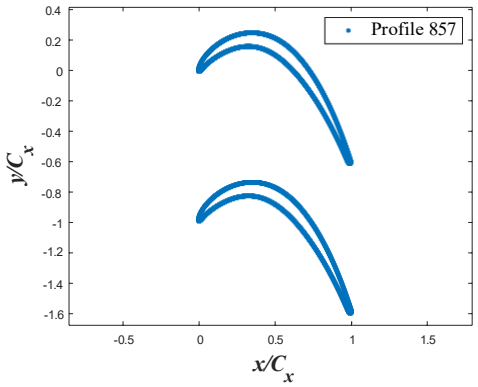

Figure 8. Geometry of the cascade with the lowest predicted total pressure loss within the maximum confidence region.

described by the latent geometry array $(0.12,\ 0.05,\ 0.62,\ -0.33)$, is depicted in figure 8. It could serve as a starting point for further detailed analyses aimed at verifying the accuracy of the prediction and refining the design accordingly.

5. Conclusions

This study presented a data-driven approach to modeling the spatial and dynamic decomposition of total pressure losses in LPT blade profiles using Proper Orthogonal Decomposition. By encoding blade geometries and related loss distributions (obtained from post-processing of high-fidelity simulations) into a reduced set of parameters, a predictive model can be trained, capturing key loss trends with minimal computational cost. This enables an initial exploration of the design space, allowing designers to gain preliminary insights into how losses and their generation mechanisms vary with geometry. With this information, high-fidelity simulations can then be directed more precisely toward a narrower, more promising subset of blade geometries, improving computational efficiency and design optimization. The results demonstrate that, by leveraging the strength of the described encoding procedure, even a simple regression framework can effectively map geometric variations to loss behavior. Future work will explore the integration of new training data points and more complex models to further refine predictive accuracy.

References

[1] S. Sarkar and P.R. Voke. “Large-eddy simulation of unsteady surface pressure over a low-pressure turbine blade due to interactions of passing wakes and inflexional boundary layer”. In: *Journal of Turbomachinery* 128.2 (2006), pp. 221–231.

[2] V. Michelassi et al. “Compressible direct numerical simulation of low-pressure turbines—part II: effect of inflow disturbances”. In: *Journal of Turbomachinery* 137.7 (2015), p. 071005. DOI: 10 . 1115 / 1 . 4029126.

[3] D. Lengani et al. “Coherent structures formation during wake-boundary layer interaction on a LP turbine blade”. In: *Flow, Turbulence and Combustion* 98 (2017), pp. 57–81.

[4] M.A. Mendez et al. *Data-driven fluid mechanics: combining first principles and machine learning*. Cambridge University Press, 2023. DOI: 10.1017/9781108896214.

[5] J. Chen et al. “Knowledge-based turbomachinery design system via a deep neural network and multi-output Gaussian process”. In: *Knowledge-Based Systems* 252 (2022), p. 109352. DOI: 10.1016/j.knosys.2022.109352.

[6] J. Bellucci et al. “A Real Time AI-Based Strategy for the Design of a Low-Pressure Turbine Profile”. In: *Turbo Expo: Power for Land, Sea, and Air*. Vol. 88087. American Society of Mechanical Engineers. 2024, V12DT34A026.

[7] J. L. Lumley. “The structure of inhomogeneous turbulent flows”. In: *Atmospheric turbulence and radio wave propagation* (1967), pp. 166–178.

[8] Z. Liu, R.J. Adrian, and T.J. Hanratty. “Large-scale modes of turbulent channel flow: transport and structure”. In: *Journal of Fluid Mechanics* 448 (2001), pp. 53–80.

[9] E. Canepa et al. “Flow coefficient and reduced frequency effects on low pressure turbine unsteady losses”. In: *Journal of Propulsion and Power* 38.1 (2022), pp. 18–29.

[10] M. Dellacasagrande et al. “Analysis of Unsteady Loss Sensitivity to Incidence Angle Variation in Low Pressure Turbine”. In: *Turbo Expo: Power for Land, Sea, and Air*. American Society of Mechanical Engineers, paper No. GT2023-103770. 2023.

[11] M. Russo et al. “Optimization of Low-Pressure Turbine Blade by Means of Fine Inspection of Loss Production Mechanisms”. In: *Journal of Turbomachinery* 147.4 (Oct. 2024), p. 041005. DOI: 10.1115/1.4066818.

[12] G.M. Robinson and A.J. Keane. “Concise orthogonal representation of supercritical airfoils”. In: *Journal of Aircraft* 38.3 (2001), pp. 580–583.

[13] D.J.J. Toal et al. “Geometric Filtration Using Proper Orthogonal Decomposition for Aerodynamic Design Optimization”. In: *AIAA Journal* 48.5 (2010), pp. 916–928. DOI: 10.2514/1.41420.

[14] D. Petronio et al. “A ML strategy for the identification of optimal LPT design region and related blade shape”. In: *Aerospace Science and Technology* 148 (2024), p. 109118. DOI: 10.1016/j.ast.2024.109118.

[15] F. Nicoud and F. Ducros. “Subgrid-scale stress modelling based on the square of the velocity gradient tensor”. In: *Flow, turbulence and Combustion* 62.3 (1999), pp. 183–200.

[16] S. Benhamadouche et al. “Synthetic turbulent inflow conditions based on a vortex method for large-eddy simulation”. In: *Progress in Computational Fluid Dynamics, an International Journal* 6.1-3 (2006), pp. 50–57.

[17] L. Sirovich. “Turbulence and the dynamics of coherent structures, Parts I, II and III”. In: *Quart. Appl. Math.* (1987), pp. 561–590.