Proceedings of the 1st international Symposium on AI and Fluid Mechanics
Paper No [leave blank]

Data-driven skin friction estimation for UAV wings in subsonic flows

Chris Pliakos[1,2], Giorgos Efrem[1,2], Dimitrios Terzis[1,2], Pericles Panagiotou[1,2, *]

[1] Laboratory of Fluid Mechanics and Turbomachinery, Department of Mechanical Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

[2] UAV Integrated Research Center (UAV-iRC), Center for Interdisciplinary Research and Innovation (CIRI), Aristotle University of Thessaloniki, 57001 Thessaloniki, Greece

Keywords: skin friction coefficient, UAV, RANS, symbolic regression, correlation

Abstract. Accurate estimation of the skin friction coefficient (C_f) is essential for estimating the wall shear stresses (τ_w) and ultimately the first-layer cell height (y) in wall-resolved RANS simulations of wings, where turbulence models are used, demanding a specific grid resolution near walls (primarily the y^+_{target}). Conventional flat-plate correlations often fail to account for the three-dimensional nature of real wing flows, introducing uncertainties in C_f predictions and leading to multiple CFD analyses and mesh refinements to meet the targets. In this work, we propose a machine learning – based approach – exploring symbolic regression, to derive a model that correlates wing-specific parameters (e.g., Reynolds number, angle of attack, thickness-to-chord ratio, wing sweep angle) with C_f at the Mean Aerodynamic Chord (MAC). Data are acquired from an in-house database of over 5,000 RANS simulations for UAV wings operating in the low subsonic regime, covering a wide design space, all conducted following best-practice CFD guidelines to ensure high fidelity. These analyses are performed at various flow conditions covering Reynolds numbers from 10^5 to 10^7, and include the complete drag polar for each wing. The proposed correlation provides improved agreement with CFD data and enables more accurate y^+ estimations. Validation on different wing geometries, including the ONERA M6 and in-house UAV wings, confirmed the robustness of the model, which improves boundary-layer resolution with only a marginal (~2%) increase in total mesh size, while achieving an R^2 of 0.68 with negligible computational inference cost. This explicit, data-driven equation offers an efficient method for streamlining mesh generation in aerodynamic simulations.

1. Introduction

In external aerodynamics, accurately predicting skin friction coefficient (C_f) is crucial for understanding viscous flow effects and aerodynamic performance. Skin friction arises due to shear stress at the wall, directly contributing to total drag on an aircraft wing, particularly in high Reynolds number flows where turbulent boundary layers dominate [1,2]. A critical aspect of computational fluid dynamics (CFD) simulations for external aerodynamic applications is the appropriate resolution of the near-wall region. To achieve this, the y^+ is employed to quantify the relative positioning of the first computational cell adjacent to the wall. The selection of an appropriate first-layer height is vital for ensuring accurate turbulence modelling [3]. Specifically, resolving the viscous sublayer ($y^+ < 1$) is required for wall-resolved Reynolds-Averaged Navier-Stokes (RANS) simulations using low-Reynolds-number turbulence models like k-ω SST. These turbulence models are the industry and research standard for external aerodynamics [4]. Improper selection of y^+ values can introduce numerical errors, leading to inaccurate wall shear stress predictions and ultimately affecting lift and drag.

Traditionally, engineers during mesh generation estimate the skin friction coefficient using semi-empirical correlations derived from canonical boundary layer flows. Flat-plate solutions, such as the Blasius relation for laminar boundary layers and power-law correlations (e.g., Prandtl, Schlichting, and

Kármán-Schoenherr) for turbulent regimes [5,6], are widely employed to approximate C_f and determine an appropriate first-layer height. However, these correlations assume simplified, two-dimensional flow conditions and often fail to capture the complex three-dimensional effects present in real aircraft configurations. As a result, engineers must perform iterative CFD refinements, re-meshing geometries multiple times to achieve the target y^+ distribution along wall surfaces (i.e. wings), leading to increased computational cost, labor and time.

Given these limitations, recent advancements in data-driven modeling have sought to enhance C_f predictions through machine learning (ML) techniques. For instance, authors in [7] developed a neural network-based wall model that predicts the velocity profile in turbulent boundary layers, effectively reducing the need for fine near-wall meshing while maintaining accuracy. Another notable work by Zhao et al., proposed a hybrid physics-informed ML approach, combining Euler solutions with a limited number of RANS simulations to predict skin friction distributions with high generalizability, this approach significantly reduced computational cost [8]. While these approaches show promise, they often face challenges in interpretability and generalization across diverse aerodynamic conditions. In contrast, symbolic regression offers a balance between accuracy and interpretability, deriving explicit mathematical expressions that can be readily used in CFD preprocessing.

This study introduces a symbolic regression-based correlation for predicting skin friction at the Mean Aerodynamic Chord (MAC) of UAV wings. A high-fidelity dataset of RANS simulations across various UAV configurations is used to train and validate the model. By incorporating key geometric and flow parameters the derived equation accounts for three-dimensional aerodynamic effects often overlooked in traditional flat-plate correlations, thereby improving skin friction predictions across a broader range of UAV configurations. The resulting formulation aims to streamline mesh generation by providing a more accurate estimation of y^+, reducing the need for repeated mesh refinement and improving the efficiency of CFD simulations in UAV aerodynamic design.

2. Methodology

In this study, an in-house automated framework for high-fidelity RANS simulations of UAV wings is employed to acquire the necessary aerodynamic data [9]. This framework minimizes human intervention by automating the key stages: (i) geometry generation, (ii) mesh creation, (iii) solution setup, and (iv) solution execution. The turbulence model selected is the low-Reynolds-number $k-\omega\ SST$, which has been validated in previous work against experimental data and is widely used for external aerodynamic applications [4,10]. The design space explored includes a diverse set of UAV configurations covering a range from Micro UAVs (<2 kg MTOW) to MALE & HALE Strike UAVs (>600 kg MTOW) as classified by NATO [11].

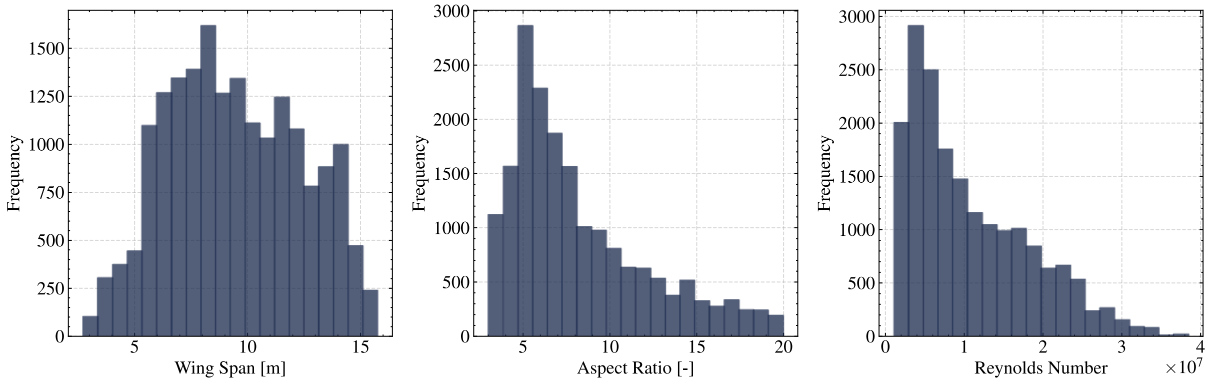


Figure 1: Distribution of key geometric and operational parameters of the UAVs in the dataset.

In total, 5,150 RANS CFD simulations were selected, corresponding to approximately 900 distinct wing planforms evaluated at various angles of attack. This broad parameter space ensures that the database reflects a wide spectrum of flow conditions, Reynolds numbers, and geometry variations relevant to UAV mission requirements. The generation of computational meshes for aerospace applications, particularly in resolving the boundary layer near the wing surface, requires calculating an appropriately small first-layer height to achieve a specific non-dimensional wall distance (y^+) dictated by the turbulence model. First, the Reynolds number is computed with the Mean Aerodynamic Chord (MAC) as the characteristic length. Next, the skin friction coefficient (C_f) is determined using one of the correlations outlined in Introduction. The wall shear stress (τ_w) is then calculated as a function of the skin friction coefficient and dynamic pressure. From this, the friction velocity (u^*) is derived, which is subsequently used to estimate the required wall distance to achieve the desired y^+ value [Equations (1-2)].

$$\tau_w = C_f \times \frac{1}{2} \rho V_\infty^2 \tag{1}$$

$$y = \frac{y^+_{target} \mu}{\rho\, u^*} = \frac{y^+_{target} \mu}{\rho \sqrt{\tau_w / \rho}} \tag{2}$$

For each geometry, the MAC position relative to the wing root was computed using the methodology described by Gudmundsson [12] for two-section wings. A normal plane slice was created in PyFluent [13] to extract geometric characteristics of the MAC and the skin friction distribution along its surface. The selected CFD analyses were filtered from the dataset according to the rules presented in Equation (3), in order to ensure adequate boundary layer resolution according to best CFD practices and prior validation [3]. Additionally, all selected simulations had to converge on a steady solution based on the criteria outlined in [9]

$$y^+_{avg} < 0.60 \; and \; y^+_{max} < 1.50 \tag{3}$$

The geometric characteristics that are retrieved from the airfoil at the MAC of each wing include: maximum thickness (t_{max}), maximum camber (y_{camber}), maximum thickness position (x_{tmax}), maximum camber position (x_{camber}) and leading edge radius (r_{LE}). Moreover, Reynolds number, angle of attack (AoA), wing sweep angle (Λ) and wing aspect ratio (AR) conclude the candidate features. The average skin friction coefficient (${C_f}_{avg}$) along the MAC surface, is chosen as the dependent variable for the machine learning model to predict. It is important to note, that the average value is chosen over the maximum one (${C_f}_{\max}$), since the latter can lead to significant small size for the first – layer computational cells, increasing drastically the number of subsequent cells that comprise the inflation layer near the wall. Data gathered are pre-processed with the IQR technique for the detection of outliers [14]. Also, a correlation analysis is executed taking into consideration Pearson and Spearman coefficients [15]. This leads to four (4) main features considered for the Symbolic Regression task, which are: $AoA, Reynolds, t_{max}\; and\; \Lambda$.

Symbolic regression (SR) is a machine learning technique that derives explicit and possibly interpretable mathematical expressions to map dependent to the independent variables. Unlike traditional regression techniques, it does not assume a predefined functional form but instead discovers the optimal equation using evolutionary algorithms (e.g. genetic algorithms). For this study, the PySR module [16] is employed to establish a data-driven correlation between the selected features and $C_{f_{avg}}$. PySR optimizes symbolic expressions by iteratively combining predefined mathematical operators, as shown in Table 1, while balancing accuracy and complexity.

Table 1: Mathematical operators used in the symbolic regression process.

Operator type	Operator
Unary operators	$\exp(x)\,, \log_{10} x\,, \sin(x)\,, \cos(x)\,, abs(x)$
Binary operators	$+, -, \times, /, ^\wedge$

Accuracy in this context refers to the ability of the derived equation to match the CFD-computed C_f values by calculating the Mean Squared Error (MSE) of the predictions to the true values, while complexity measures the equation's length and number of operations. In this work, we do not strictly constrain complexity to a low number. Given that any analytical expression is computationally inexpensive, allowing for more complex formulations may yield better $C_{f_{avg}}$ predictions, ultimately improving the meshing process in CFD simulations. Other than the MSE, the Mean Absolute Error (MAE) is also calculated since it is more interpretable, along with the coefficient of determination R^2.

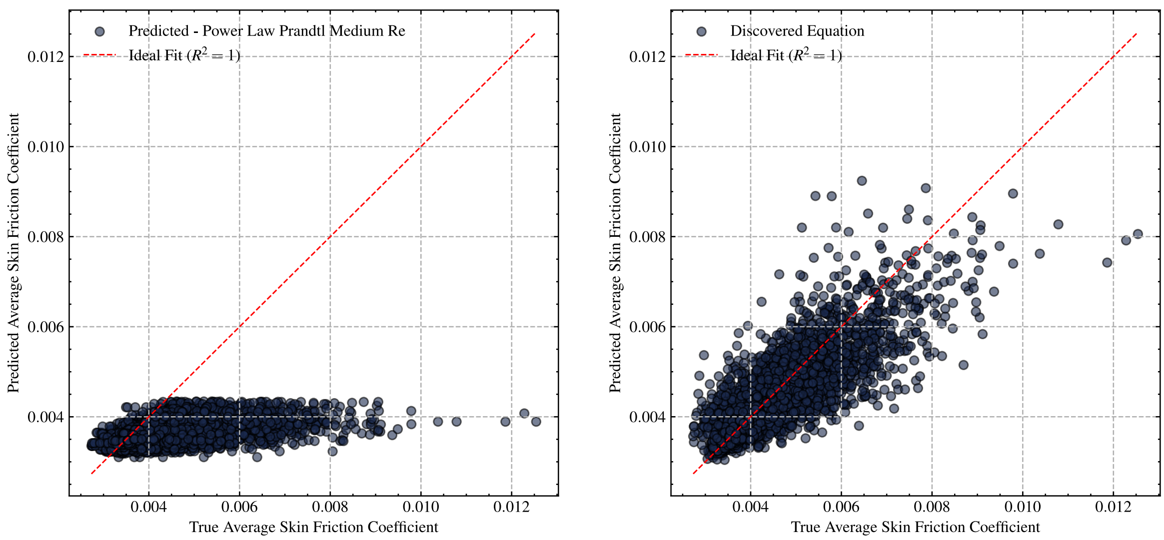
3. Results

PySR generated equations with complexity values ranging from 3 to 40 through 20,000 iterations. An equation with a complexity of 19 is selected, as further increase in complexity results in negligible improvements in predictive accuracy. Table 2 presents the selected equation alongside classical flat-plate correlations. Traditional correlations exhibit significantly lower agreement with the high-fidelity CFD data, with negative R^2 values, indicating that they perform worse than a simple mean prediction. This suggests that flat-plate-based approaches, which are primarily developed for canonical boundary-layer flows, fail to account for the complex three-dimensional effects present in real wing flows. In contrast, the symbolic regression model captures these effects more effectively, providing a notable improvement in predictive capability.

Table 2: Typical correlations and the discovered equation for the prediction of C_f.

Equation	**Complexity**	$\boldsymbol{R^2}$	**MAE**	**MSE**	**Derived From**
$C_f = (3.46 \log_{10} Re - 5.6)^{-2}$	6	-2.19	$1.01e^{-03}$	$1.84e^{-06}$	Explicit Schoenherr
$C_f = (4.13 \log_{10} Re\, C_f)^{-2}$	6	-2.04	$9.48e^{-04}$	$1.72e^{-06}$	Implicit Karman
$C_f = 0.072 \cdot Re^{-1/5}$	4	-1.96	$9.25e^{-04}$	$1.66e^{-06}$	Power Law Blasius
$C_f = 0.0725 \cdot Re^{-1/5}$	4	-1.91	$9.02e^{-04}$	$1.61e^{-06}$	Power Law Prandtl
$C_f = 0.455 \cdot \log_{10}(Re)^{-2.58}$	5	-1.89	$9.08e^{-04}$	$1.60e^{-06}$	Schlichting
$\boldsymbol{C_f = \dfrac{(0.9036^{AoA} \cdot 0.02228) + (AoA \cdot 0.00214)}{((Re \cdot (t_{max} + (\Lambda \cdot 0.00322))^{1/7.71}}}$	**19**	**0.68**	$\boldsymbol{3.95e^{-04}}$	$\boldsymbol{3.46e^{-07}}$	**Proposed model**

Additionally, Figure 2 provides a visual comparison of predicted $C_{f\,avg}$ versus true CFD values for the flat-plate Prandtl Law correlation (Figure 2a) and the symbolic regression model (Figure 2b).



(a) **(b)**

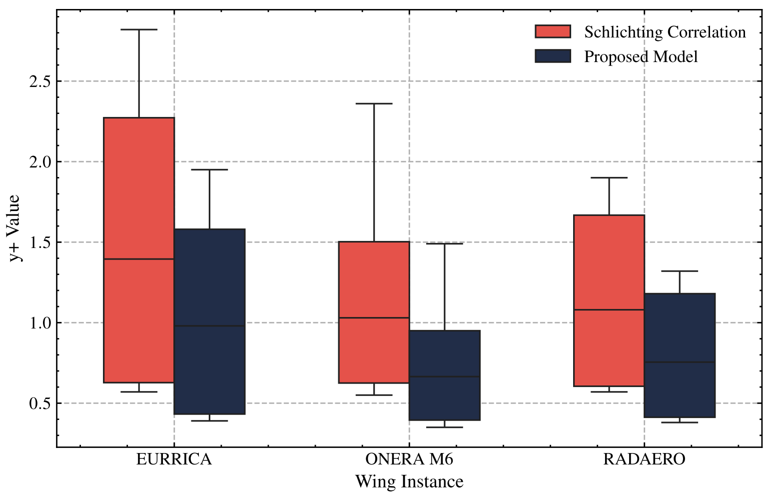
Figure 2: Predictions vs CFD of $C_{f_{avg}}$ at MAC. (a): Prandtl Law Medium Re model, (b) SR model.

The flat-plate correlation systematically underpredicts C_f, clustering most predictions around a narrow range and failing to capture the variation present in the dataset. In contrast, the symbolic regression model exhibits a broader, more accurate spread of predictions, aligning more closely with the ideal fit. However, some underprediction remains, particularly at higher C_f values.

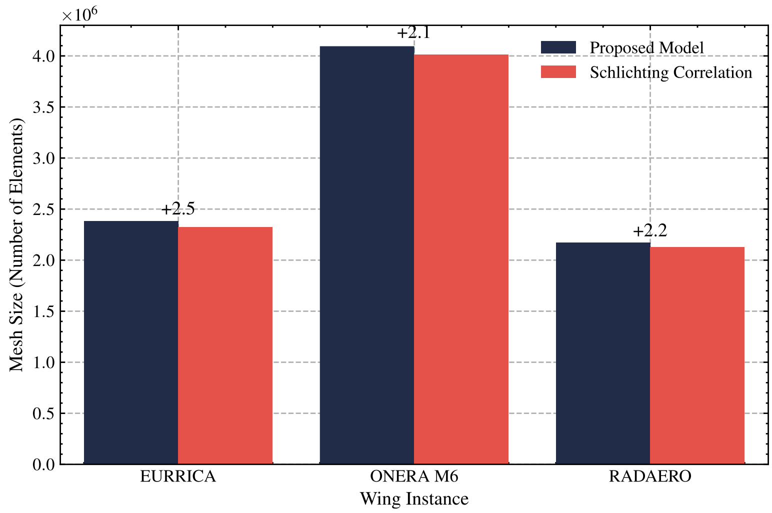
One key consideration when applying the derived equation is the inclusion of angle of attack (AoA) as an input feature. In practice, CFD meshes are typically generated once and used for the entire drag polar, rather than being re-meshed for each AoA. To account for this, a recommended approach is to estimate C_f at an AoA corresponding to 75% of the expected AoA range for the analysis (e.g. 12° for $AoA \in [0, 16]$). This "safety factor" helps prevent an excessively small first-layer height, ensuring a well-balanced mesh without unnecessary refinement. As observed in Figure 2b, the model slightly underpredicts C_f, meaning that this safety factor helps shift predictions closer to the $Y = X$ line.

To further illustrate the effectiveness of the proposed model, three distinct wing geometries were analysed: the ONERA M6 wing and two wings from in-house designed UAV, specifically the Blended-Wing-Body EURRICA and conventional RADAERO configurations [9]. Each wing was meshed in two distinct approaches: the first utilized the Schlichting correlation to estimate the skin friction coefficient (C_f), while the second applied the derived model showcased in Table 2. Both meshing approaches aimed for a target y^+ value of 1, consistent with best practices for the k-ω SST turbulence model [3].

Steady-state RANS simulations were performed for each wing under identical flight conditions (25 m/s at sea level conditions), covering the full drag polar range. Simulations were considered converged when the change in aerodynamic forces fell below 0.1% over 100 iterations. Figure 3a presents the boxplot distributions of the actual y^+ values obtained on the wing surfaces across the entire drag polar for each configuration. The results clearly indicate that the mesh derived from the proposed model achieved consistently lower y^+ values compared to the flat plate correlation. This outcome confirms the superior accuracy of the proposed approach in predicting C_f. Especially for the EURRICA case, the classic approach would lead to a mesh refinement and a repetition of the CFD run ($y^+ > 2.5$).



(a)



(b)

Figure 3: Comparison of (a) achieved y^+ values and (b) mesh sizes using Schlichting correlation vs. proposed model.

Moreover, Figure 3b compares the total number of mesh elements generated by both methodologies. Despite providing more accurate C_f predictions and consequently refined mesh layers near the wall, the new grids exhibit only a marginal increase in total element count—approximately 2% more—than those generated using the Schlichting correlation. Thus, the proposed data-driven correlation achieves significant improvements in boundary layer resolution without substantial computational overhead.

4. Conclusions

This study demonstrates a data-driven approach for improving the prediction of the skin friction coefficient (C_f) in subsonic wing flows, emphasizing the importance of high-fidelity aerodynamic data. Using symbolic regression, an analytical equation was derived that correlates key aerodynamic parameters – Reynolds number, angle of attack, maximum thickness-to-chord ratio, and sweep angle – to the average C_f at the Mean Aerodynamic Chord. Unlike conventional flat-plate correlations, which neglect three-dimensional flow effects, the proposed equation results in consistently lower and more accurate y^+ values in CFD simulations, improving boundary-layer resolution without significantly increasing computational cost. The impact on total mesh size remains minimal, with an average increase of only 2%. A key advantage of this approach is the derivation of an explicit analytical equation with virtually zero computational cost, making it highly suitable for practical applications in aerodynamic design. Although derived from a UAV-focused dataset, the equation generalizes well to other aerodynamic configurations, including the ONERA M6 wing, suggesting broader applicability in aircraft design. The results highlight the power of symbolic regression in generating interpretable models that improve aerodynamic predictions while avoiding the limitations of black-box machine learning methods. Future work could explore further simplifications of the equation, improve predictive accuracy, and extend this methodology to estimating additional boundary-layer properties, further optimizing computational grid generation in CFD-based aerodynamic analysis.

5. References

[1] Anderson, J., "Fundamentals of Aerodynamics."

[2] Wilcox, D. C., "Turbulence Modeling for CFD," *DCW industries, La Canada*, 1998.

[3] Falk Goetten, D. Felix Finger, Matthew Marino, and Cees BIl, "A Review of Guidelines and Best Practices for Subsonic Aerodynamic Simulations Using RANS CFD," 2019.

[4] Menter, F., Sechner, R., Germany GmbH Matyushenko, A. A., and Petersburg, S., "Best Practice: RANS Turbulence Modeling in Ansys CFD."

[5] Schlichting, H., and Gersten, K., "Boundary-Layer Theory," *Boundary-Layer Theory*, 2016, pp. 1–799. https://doi.org/10.1007/978-3-662-52919-5/COVER

[6] Lazauskas, L. V., "Hydrodynamics of Advanced High-Speed Sealift Vessels," University of Adelaide, Australia, 2005.

[7] Romanelli, M., Beneddine, S., Mary, I., Beaugendre, H., Bergmann, M., Sipp, D., and Hélo¨ıse Beaugendre, H., "Data-Driven Wall Models for Reynolds Averaged Navier-Stokes Simulations," *International Journal of Heat and Fluid Flow*, Vol. 99, 2023. https://doi.org/10.1016/j.ijheatfluidflow.2022.109097ï

[8] Shule, Z., and Weiwei, Z., "CORRELATION MODELLING OF SURFACE INVISCID FLOW CHARACTERISTICS AND FRICTION DISTRIBUTION BASED ON MACHINE LEARNING."

[9] Pliakos, C., Efrem, G., Terzis, D., and Panagiotou, P., "An Automated Framework for Streamlined CFD-Based Design and Optimization of Fixed-Wing UAV Wings," *Algorithms*, 2025.

[10] Dimopoulos, T., Paliaikos, D., Christou, V., Kaparos-Tsafos, P., and Panagiotou, P., "Experimental and Computational Investigation of the Vortical Structures Generated from a Blended-Wing-Body UAV Model," *Aerospace Science and Technology*, Vol. 139, 2023, p. 108377. https://doi.org/10.1016/j.ast.2023.108377

[11] Mitridis, D., Kapsalis, S., Terzis, D., and Panagiotou, P., "An Evaluation of Fixed-Wing Unmanned Aerial Vehicle Trends and Correlations with Respect to NATO Classification, Region, EIS Date and Operational Specifications," *Aerospace*, Vol. 10, No. 4, 2023. https://doi.org/10.3390/aerospace10040382

[12] Gudmundsson, S., "General Aviation Aircraft Design: Applied Methods and Procedures," 2022.

[13] "PyFluent Documentation 0.21.0 — PyFluent." Retrieved 30 December 2024. https://fluent.docs.pyansys.com/version/0.21/

[14] Beyer, H., "Tukey, John W.: Exploratory Data Analysis. Addison-Wesley Publishing Company Reading, Mass. — Menlo Park, Cal., London, Amsterdam, Don Mills, Ontario, Sydney 1977, XVI, 688 S.," *Biometrical Journal*, Vol. 23, No. 4, 1981, pp. 413–414. https://doi.org/10.1002/BIMJ.4710230408

[15] Schober, P., and Schwarte, L. A., "Correlation Coefficients: Appropriate Use and Interpretation," *Anesthesia and analgesia*, Vol. 126, No. 5, 2018, pp. 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

[16] Cranmer, M., "Interpretable Machine Learning for Science with PySR and SymbolicRegression.Jl," 2023. http://arxiv.org/abs/2305.01582