Improving the Spalart-Allmaras turbulence model for separated flows using field inversion and symbolic regression

Paul Bataillie,[1,2] Maxime Casanova,[1] and Pedro Stefanin Volpiani[1,*]
[1]*ONERA, The French Aerospace Lab. 8 Rue des Vertugadins, 92190 Meudon, France*
[2]*Aix Marseille Univ, CNRS, Centrale Med, M2P2, 13013 Marseille, France*

(Received 27 October 2025; accepted 10 June 2026; published 16 July 2026)

Reynolds-averaged Navier-Stokes (RANS) turbulence models rely on several simplifying assumptions that limit their accuracy, particularly in predicting separated flows. To address these limitations, significant efforts have focused on augmenting RANS models through field inversion and machine learning. In the present study, data assimilation and symbolic regression are used to formulate an analytical correction to the Spalart-Allmaras model, addressing local deficiencies in its production term. A key objective of this work is to develop a correction applicable to a wide range of flow configurations, improving predictions for separated flows without degrading the performance of the baseline model for wall-attached flows. To this end, symbolic regression is performed using assimilated data from two-dimensional separated-flow cases (converging-diverging channel, bump H42, and NASA wall-mounted hump) together with a reference case where the Spalart-Allmaras model remains accurate (flat plate). This methodology enforces the desired correction in separated flows while minimizing the impact on the model's accuracy for wall-attached flows. The symbolic correction accurately reproduces the assimilated quantities and its performance on unseen flows demonstrates that the resulting expression remains applicable across a broad range of flow configurations and Reynolds numbers (bump H38, square cylinder, and periodic hill). The correction is also assessed on three-dimensional cases, namely, the FAITH hill and the Ahmed body. While improvements are observed for some configurations, these tests also reveal limitations that warrant further investigation.

DOI: 10.1103/dk9r-td14

I. INTRODUCTION

Turbulence modeling is an active topic of research given its complexity and importance for fluid dynamics simulations. The prediction of turbulent flows represents one of the most persistent challenges in physics and engineering. The Reynolds-averaged Navier-Stokes (RANS) framework remains the most widely used approach for modeling turbulent flows. Although continuous advances in computational power are making large eddy simulation (LES) and direct numerical simulation (DNS) increasingly accessible, these methods are still impractical for most real-world applications. As a result, the RANS framework is expected to remain the prevailing paradigm for the foreseeable future [1].

The main limitation of RANS simulations lies in the closure problem, which arises from the averaging process in their derivation. To address this issue, closure models are introduced to provide the missing information and enable the solution of the RANS equations. While decades of development have led to widely used formulations such as the k-ϵ [2], k-ω [3], and Spalart-Allmaras (SA) [4] models, their predictive capabilities often struggle to accurately capture separated regimes.

[*]Contact author: pedro.stefanin_volpiani@onera.fr

Therefore, enhancing their performance in such conditions is essential to significantly broaden their applicability across a wide range of engineering problems.

Some attempts have been made to improve turbulence models with analytical laws based on physical insights. For instance, Rumsey *et al.* [5] introduced an *ad hoc* term defined as a function of the local ratio of turbulent production to dissipation. This *ad hoc* term was used to multiply the ω-destruction term, increasing the eddy viscosity in separated regions. The authors highlighted that the proposed "separation fix" yielded dramatic improvements in certain cases, but also acknowledged that the correction was not universal. This shows that deriving broadly applicable models or corrections based solely on physical insight and fundamental laws is extremely challenging.

With an ever-increasing volume of high-fidelity datasets now available, data-driven turbulence modeling is emerging as a powerful approach to overcome the predictive limitations of classical RANS models. Duraisamy [6] discussed several approaches to leverage data from experiments or high-fidelity numerical simulations. The most straightforward approach consists of training machine-learning models on DNS data to reconstruct the missing information. Following this strategy, Ling *et al.* [7] trained a deep neural network to learn the Reynolds stress anisotropy tensor from high-fidelity data before incorporating these predictions into a RANS framework. Similarly, Wang *et al.* [8] tried to reconstruct the discrepancies of the RANS modeled Reynolds stresses by using random forests. The discrepancy functions were trained directly on DNS databases and then used to predict Reynolds stress discrepancies in different flows where data were not available. More recently, Volpiani *et al.* [9,10] developed a machine-learning oriented turbulence model that directly estimates the eddy viscosity correction from LES data of turbulent flows over parametric bumps. Their methodology, using artificial neural networks (ANN) and random forests (RF), was proved to be robust even in predicting flows with strong separation where the eddy-viscosity assumption lacks accuracy.

Another approach consists in solving an inverse problem. It supposes the inference of a corrective term added to the baseline model that minimizes the discrepancies between the enhanced model and the reference data. Parish *et al.* [11] introduced the field inversion and machine learning (FIML) paradigm and applied it to the prediction of a turbulent channel flow. Their approach consisted of modifying the source term in the turbulence transport equations. Solving this inverse problem yields a spatially varying correction field. This allows model-form errors to be addressed rather than only parametric ones. Once the correction is obtained, machine learning can be used to establish the mapping between the input features and the expected corrective field.

Several studies have employed the FIML method to address inherent deficiencies in RANS turbulence models. Singh *et al.* [12] used experimental lift coefficients to infer corrections for the SA production term. These corrections were then used to train a neural network to correct the prediction of turbulent flows over airfoils involving flow separation. This approach was also adopted by Ferrero *et al.* [13] in the case of low-pressure gas turbine cascades. More recently, Yan *et al.* [14] used this framework to improve the SA model on three cases, the S809 airfoil, the periodic hill and the GLC305 airfoil with ice formation. While these efforts focused on recalibrating the production term, other research has explored alternative corrective formulations. Cato *et al.* [15] evaluated the performance of several corrective terms in the SA model using dense velocity measurements, and showed that the choice of correction strongly affects the inverse optimization process. They observed that introducing corrections directly in the momentum equations could lead to a good reconstruction of the flow field. Looking for a perfect flow reconstruction, Volpiani *et al.* [16] introduced a vectorial source term into the momentum equation to improve the SA predictions over periodic hills. However, this type of correction might be difficult to generalize across different types of flows. Applying a corrective factor to the production term of the SA transport equation is particularly attractive from a FIML perspective, as it corresponds to a nondimensional scalar quantity. This property facilitates the learning process and potentially improve the robustness of the resulting model. Nevertheless, because the correction remains constrained by the Boussinesq hypothesis, the achievable level of flow reconstruction is inherently limited.

In the literature, many studies focused on improving turbulence models with machine learning for specific flow configurations: separated flows over airfoils [12,17,18], periodic hills [16,19], square duct [20], or bidimensional bumps [9,10]. While the literature almost unanimously acknowledges the success of machine-learning-based corrections for specific types of flows, the generality of such improved models remains an open question. Rumsey *et al.* [21] incorporated diverse flow configurations into the training dataset, including a flat-plate flow, the NASA hump, and the NACA 0012 airfoil, in an effort to improve model generalization. However, the resulting improvements remained limited, and robust generalization was not demonstrated. The authors concluded that a consistent and broadly applicable data-driven turbulence model for separated flows has not yet been achieved.

Another fundamental issue in turbulence modeling concerns model interpretability. Machine-learning tools such as neural networks are often regarded as black boxes due to their layered structure (characterized by the superposition of hidden layers and nonlinear activation functions), which makes physical interpretation difficult. In this context, symbolic regression (SR) offers a compelling alternative, as it aims to derive explicit analytical expressions directly from data. SR is a type of regression analysis that tries to combine operators and functionals to find the model that best fits a given dataset, both in terms of accuracy and simplicity. The recent development of open-source libraries such as *PySR* [22] and *PhySO* [23], provided the flexibility needed to discover symbolic expressions. To derive analytical models from data, these frameworks rely on advanced search strategies, such as genetic algorithms (GA), as implemented in *PySR*, or deep reinforcement learning (DRL), as employed in *PhySO*. Because the outputs take the form of explicit mathematical expressions, they can be easily analyzed and understood by modelers. This explicit structure also enables the formal enforcement of physical constraints. For example, *PhySO* was specifically designed to generate expressions that are dimensionally consistent by construction [23]. However, this interpretability is maintained only when the complexity of the expression remains relatively small. If an expression consists of multiple nested operators (a common occurrence when regressing complex data), it quickly becomes impossible to interpret. Despite this challenge, symbolic regression effectively solves the "portability" problem associated with models based on neural networks for instance. While black-box models are often difficult to interface with different flow solvers [24], analytical expressions can be easily shared and incorporated into new numerical frameworks.

Several research groups are currently using symbolic regression to improve the predictive accuracy of the SST turbulence model. Weatheritt *et al.* [25] symbolically regressed algebraic forms of the Reynolds stress anisotropy tensor using data of rectangular ducts. When tested on an asymmetric diffuser, their new models demonstrated significant improvements over the baseline linear model. Schmelzer *et al.* [26] introduced sparse regression of turbulent stress anisotropy (SpaRTA), a deterministic symbolic regression method. SpaRTA was used to infer algebraic stress models for the closure of RANS equations directly from high-fidelity LES or DNS data. Following this idea, Buchanan *et al.* [27] derived a new zonally augmented SST model by applying a Relative Importance Term Analysis (RITA) methodology to keep the baseline performance in fully attached flows. Although the model was regressed on two-dimensional configurations, it demonstrated significant improvements in predicting three-dimensional separated flows, such as the Ahmed body and a wall-mounted hill named FAITH (fundamental aeronautics investigates the hill). Focusing on the transport equations themselves, Wu *et al.* [28] applied symbolic regression to derive a corrective factor that adjusts the destruction term in the ω equation. Although the regression was performed only on a 2D curved backward-facing step, the resulting corrections proved effective for two-dimensional bumps, periodic hills and on a three-dimensional Ahmed body. In a subsequent study, Wu *et al.* [29] coupled symbolic regression with a conditioned field inversion and expanded the training set to include a two-dimensional hump. The corrective field applied to the destruction term in the ω equation was multiplied by a shielding function f_d, which is inactive within the boundary layer and active elsewhere. This conditioned field inversion allowed to improve separated flows predictions while maintaining the accuracy of the original model for wall-attached flows. An *et al.* [30] employed the field inversion and symbolic regression (FISR) methodology to incorporate streamline curvature effects into the SST model. Their approach modifies the eddy viscosity by introducing a correction that accounts for directional flow effects, thereby enhancing the baseline SST formulation for this class of flows.

Studies focused on improving the SA turbulence model by combining data assimilation with symbolic regression are more scarce. In fact, these studies have focused on deriving models for a specific flow scenario [18,31]. He *et al.* [18] and Liao *et al.* [31] applied symbolic regression to model incompressible flows over airfoils. The novelty of the present study lies in the application of the FISR methodology to improve the predictions of the Spalart-Allmaras turbulence model for a significantly broader scope of flow configurations. Unlike previous studies confined to specific flows, this work improves the baseline model across a diverse set of separated flows spanning a large range of Reynolds numbers. To achieve this, the regression is performed simultaneously on separated flows, where a modification is required, and on a reference case representative of the baseline SA model calibration. This dual-target approach yields a symbolic expression that locally adjusts the production term in the SA equation. As a result, the derived correction provides the necessary adjustments for nonequilibrium flows while minimally affecting the baseline behavior of the SA model in attached-flow regimes. The outline of this study is as follows. First, the RANS framework and the numerical discretization setup are detailed (Sec. II). We then present the data assimilation procedure and the corrective fields inferred by field inversion (Sec. III). In Sec. IV, we explain the strategy employed to determine an analytical correction through symbolic regression. The resulting SR expression and its performance across both regression and unseen cases are provided in Secs. IV C–IV E. Finally, a discussion is provided in Sec. V.

II. REYNOLDS-AVERAGED NAVIER-STOKES EQUATIONS AND NUMERICAL DISCRETIZATION

A. RANS equations and turbulence closure

Following the Reynolds decomposition, the flow variables are written as the sum of mean and fluctuating components, here formalized as $\mathbf{u}(\mathbf{x},t) = \overline{\mathbf{u}}(\mathbf{x}) + \mathbf{u}'(\mathbf{x},t)$ for the velocity and $p(\mathbf{x},t) = \overline{p}(\mathbf{x}) + p'(\mathbf{x},t)$ for the pressure. The incompressible steady-state RANS equations for the mean flow variables are given by

$$\nabla \cdot \overline{\mathbf{u}} = 0, \tag{1}$$

$$(\overline{\mathbf{u}} \cdot \nabla)\overline{\mathbf{u}} = -\nabla \overline{p} + \nabla \cdot (2\nu \mathbf{S}) - \nabla \cdot (\overline{\mathbf{u}' \otimes \mathbf{u}'}), \tag{2}$$

where $\mathbf{S} = \frac{1}{2}(\nabla\, \overline{\mathbf{u}} + \nabla\, \overline{\mathbf{u}}^{\,T})$ is the symmetric component of the mean velocity gradient, ν is the kinematic viscosity, $\overline{\mathbf{u}' \otimes \mathbf{u}'}$ is the Reynolds stress tensor, and the operator $\otimes$ refers to the dyadic product. In the context of RANS modeling, the Reynolds stresses are commonly approximated using the Boussinesq hypothesis:

$$-\overline{\mathbf{u}' \otimes \mathbf{u}'} \approx 2\nu_t \mathbf{S} - \tfrac{2}{3}k\mathbf{I}, \tag{3}$$

where ν_t is the turbulent viscosity and $k = \frac{1}{2}(\overline{u'^2} + \overline{v'^2} + \overline{w'^2})$ is the turbulent kinetic energy. The turbulent viscosity ν_t is determined by solving the transport equation for $\tilde{\nu}$ in the Spalart-Allmaras turbulence model described in Ref. [32]. The *RANS-SA* equations, describing the modeled mean flow variables $\overline{\mathbf{q}} := \overline{\mathbf{q}}(\mathbf{x}) = [\overline{\mathbf{u}}, \overline{p}, \tilde{\nu}]^T$, are thus defined as

$$\nabla \cdot \overline{\mathbf{u}} = 0, \tag{4}$$

$$(\overline{\mathbf{u}} \cdot \nabla)\overline{\mathbf{u}} = -\nabla \overline{p} + \nabla \cdot (2\nu \mathbf{S}) + \nabla \cdot [2\nu_t(\tilde{\nu})\mathbf{S}], \tag{5}$$

$$\overline{\mathbf{u}} \cdot \nabla \tilde{\nu} - \nabla \cdot [\eta(\tilde{\nu}) \nabla \tilde{\nu}] = \underbrace{P(\tilde{\nu}, \nabla \overline{\mathbf{u}}) - D(\tilde{\nu}, \nabla \overline{\mathbf{u}}) + C(\nabla \tilde{\nu})}_{s(\tilde{\nu}, \nabla \tilde{\nu}, \nabla \overline{\mathbf{u}})}, \tag{6}$$

while the turbulent viscosity $\nu_t(\tilde{\nu})$ and diffusion coefficient $\eta(\tilde{\nu})$ are modeled as

$$\nu_t(\tilde{\nu}) = \begin{cases} \tilde{\nu} f_{v_1}, & \tilde{\nu} \geqslant 0 \\ 0, & \tilde{\nu} <, 0 \end{cases}, \quad \eta(\tilde{\nu}) = \begin{cases} \sigma^{-1} \nu (1 + \chi), & \chi \geqslant 0 \\ \sigma^{-1} \nu \big(1 + \chi + \frac{1}{2} \chi^2\big), & \chi < 0 \end{cases},$$

$$f_{v_1} = \frac{\chi^3}{c_{v_1}^3 + \chi^3}, \quad \chi = \frac{\tilde{\nu}}{\nu}.$$

The source term $s(\tilde{\nu}, \nabla \tilde{\nu}, \nabla \overline{\mathbf{u}})$ is composed of a production, destruction and cross-diffusion term:

$$P(\tilde{\nu}, \nabla \overline{\mathbf{u}}) = \begin{cases} c_{b_1} \tilde{\nu} \tilde{S}, & \chi \geqslant 0 \\ c_{b_1} \tilde{\nu} S g_n, & \chi < 0 \end{cases}, \quad D(\tilde{\nu}, \nabla \overline{\mathbf{u}}) = \begin{cases} c_{w_1} f_w \big(\frac{\tilde{\nu}}{d}\big)^2, & \chi \geqslant 0 \\ -c_{w_1} \big(\frac{\tilde{\nu}}{d}\big)^2, & \chi < 0 \end{cases},$$

$$C(\nabla \tilde{\nu}) = \frac{c_{b_2}}{\sigma} \nabla \tilde{\nu} \cdot \nabla \tilde{\nu},$$

where

$$\tilde{S} = S + \frac{\tilde{\nu} f_{v_2}}{\kappa^2 d^2}, \quad S = \sqrt{|\nabla \times \overline{\mathbf{u}}|^2 + M^2} - M, \quad f_{v_2} = 1 - \frac{\chi}{1 + \chi f_{v_1}}, \quad g_n = 1 - 1000 \frac{\chi^2}{1 + \chi^2},$$

$$f_w = g \left[\frac{1 + c_{w_3}^6}{g^6 + c_{w_3}^6} \right]^{\frac{1}{6}}, \quad g = r + c_{w2}(r^6 - r), \quad r = \begin{cases} 0, & r' < 0, \\ r', & 0 \leqslant r' \leqslant 10, \\ 10, & r' > 10, \end{cases} \quad r' = \frac{\tilde{\nu}}{\tilde{S} \kappa^2 d^2}.$$

Here d is the distance to the nearest wall, $c_{v_1} = 7.1$, $c_{b_1} = 0.1355$, $c_{b_2} = 0.622$, $\sigma = 2/3$, $\kappa = 0.41$, $c_{w_1} = c_{b_1}/\kappa^2 + (1 + c_{b_2})/\sigma$, $c_{w_2} = 0.3$, and $c_{w_3} = 2$. We also introduced the regularizing constant $M = 10^{-5}$ to avoid differentiability issues of S.

B. Finite-element formulation

The discretization of the problem is performed using the finite-element method (FEM). To this extent, a finite-dimensional scalar component $\overline{w}$ of the full state $\overline{\mathbf{q}}$ can be modeled as a finite linear combination of degrees of freedoms $\{w_i\}^{n_w}$ and associated Lagrange basis functions $\phi_{w_i} = \phi_{w_i}(\mathbf{x})$ as

$$\overline{w} = \sum_{i=1}^{N} w_i \phi_{w_i}. \tag{7}$$

In the scope of the study, $\overline{w}$ can be either the velocity $\overline{\mathbf{u}}$, the pressure $\overline{p}$, or the Spalart-Allmaras variable $\tilde{\nu}$.

To ensure a numerically stable solution, first-order Lagrange elements P^1 are used for the pressure $\overline{p}$ and the Spalart-Allmaras variable $\tilde{\nu}$ while first-order Lagrange elements enriched with bubble functions P_b^1 are employed for the velocity $\overline{\mathbf{u}}$.

The *RANS-SA* equations, Eqs. (4)–(6), are solved numerically with *Firedrake* [33] which is an open source library to solve partial differential equations using the FEM. *Firedrake* uses the nonlinear parallel solvers of *PETSc* and relies on the *unified form language (UFL)* [34] from the *FEniCS* Project. *Firedrake* is able to discretize automatically the weak formulation of the *RANS-SA* residual. Thus, the variational formulation corresponding to the nonlinear *RANS-SA* residual reads

$$\begin{aligned}\mathcal{R}_{\text{RANS-SA}}([\overline{\mathbf{u}},\overline{p},\tilde{\nu}],[\boldsymbol{\phi}_{\overline{\mathbf{u}}},\phi_{\overline{p}},\phi_{\tilde{\nu}}]) = &-\int_\Omega \overline{p}\,[\nabla\cdot\boldsymbol{\phi}_{\overline{\mathbf{u}}}]\,\mathrm{d}\Omega + \int_\Omega 2[\nu+\nu_t(\tilde{\nu})][\mathbf{S}:\nabla\boldsymbol{\phi}_{\overline{\mathbf{u}}}]\,\mathrm{d}\Omega\\ &+\int_\Omega [\nabla\overline{\mathbf{u}}\cdot\overline{\mathbf{u}}]\cdot\boldsymbol{\phi}_{\overline{\mathbf{u}}}\,\mathrm{d}\Omega - \int_\Omega [\nabla\cdot\overline{\mathbf{u}}]\,\phi_{\overline{p}}\,\mathrm{d}\Omega\\ &+\int_\Omega \eta(\tilde{\nu})[\nabla\tilde{\nu}\cdot\nabla\phi_{\tilde{\nu}}]\,\mathrm{d}\Omega + \int_\Omega [\overline{\mathbf{u}}\cdot\nabla\tilde{\nu}]\phi_{\tilde{\nu}}\,\mathrm{d}\Omega\\ &-\int_\Omega P\phi_{\tilde{\nu}}\,\mathrm{d}\Omega + \int_\Omega D\phi_{\tilde{\nu}}\,\mathrm{d}\Omega - \int_\Omega C\phi_{\tilde{\nu}}\,\mathrm{d}\Omega,\end{aligned}$$

where ϕ denotes the basis functions chosen as test functions and the operator : refers to the Frobenius inner product.

Due to the numerical instability of the FEM for flows at high Reynolds number, a stabilization scheme is necessary. Simplified versions of the streamline-upwind Petrov-Galerkin (SUPG) [35] and GRAD-DIV [36] stabilization schemes have been used. The variational formulation of the stabilization scheme residual reads

$$\begin{aligned}\mathcal{R}_{\text{STAB}}([\overline{\mathbf{u}},\overline{p},\tilde{\nu}],[\boldsymbol{\phi}_{\overline{\mathbf{u}}},\phi_{\overline{p}},\phi_{\tilde{\nu}}]) = &\int_\Omega \tau^{\text{SUPG}}[\nabla\overline{\mathbf{u}}\cdot\overline{\mathbf{u}}]\cdot[\nabla\boldsymbol{\phi}_{\overline{\mathbf{u}}}\cdot\overline{\mathbf{u}}]\,\mathrm{d}\Omega\\ &+\int_\Omega \tau^{\text{SUPG}}[\nabla\tilde{\nu}\cdot\overline{\mathbf{u}}][\nabla\phi_{\tilde{\nu}}\cdot\overline{\mathbf{u}}]\,\mathrm{d}\Omega\\ &+\int_\Omega \tau^{\text{GRAD-DIV}}[\nabla\cdot\overline{\mathbf{u}}][\nabla\cdot\boldsymbol{\phi}_{\overline{\mathbf{u}}}]\,\mathrm{d}\Omega,\end{aligned}$$

where τ^{SUPG} and $\tau^{\text{GRAD-DIV}}$ are defined as follows:

$$\tau^{\text{SUPG}} = \frac{h_t}{2||\overline{\mathbf{u}}||}\xi(\mathrm{Re}_h) \quad \text{with} \quad \xi(\mathrm{Re}_h) = \begin{cases}\frac{\mathrm{Re}_h}{3}, & \mathrm{Re}_h \leqslant 3\\ 1, & \mathrm{Re}_h > 3\end{cases}, \quad \mathrm{Re}_h = \frac{h_t||\overline{\mathbf{u}}||}{2\nu}, \tag{8}$$

$$\tau^{\text{GRAD-DIV}} = \frac{h_t}{2||\overline{\mathbf{u}}||}\psi(\mathrm{Re}_h) \quad \text{with} \quad \psi(\mathrm{Re}_h) = \begin{cases}\frac{1}{3}||\overline{\mathbf{u}}||^2\mathrm{Re}_h, & \mathrm{Re}_h \leqslant 3\\ ||\overline{\mathbf{u}}||^2, & \mathrm{Re}_h > 3\end{cases}, \quad \mathrm{Re}_h = \frac{h_t||\overline{\mathbf{u}}||}{2\nu}, \tag{9}$$

with h_t the local element size taken as the diameter of the cell (i.e., the maximum distance of two points in the cell).

Finally, the variational formulation of the stabilized steady nonlinear residual $\mathcal{R}$ which is solved by the FEM is given as follows:

$$\mathcal{R}([\overline{\mathbf{u}},\overline{p},\tilde{\nu}],[\boldsymbol{\phi}_{\overline{\mathbf{u}}},\phi_{\overline{p}},\phi_{\tilde{\nu}}]) = \mathcal{R}_{\text{RANS-SA}}([\overline{\mathbf{u}},\overline{p},\tilde{\nu}],[\boldsymbol{\phi}_{\overline{\mathbf{u}}},\phi_{\overline{p}},\phi_{\tilde{\nu}}]) + \mathcal{R}_{\text{STAB}}([\overline{\mathbf{u}},\overline{p},\tilde{\nu}],[\boldsymbol{\phi}_{\overline{\mathbf{u}}},\phi_{\overline{p}},\phi_{\tilde{\nu}}]). \tag{10}$$

A Newton solver is used to solve the nonlinear problem. This implies the solution of a sequence of linear systems, which are solved via the multifrontal massively parallel sparse direct solver (MUMPS) [37,38]. The meshes for the flow configurations presented in the following have been constructed with a wall refinement strategy ensuring $y^+ \leqslant 1$. A grid convergence study allowed to validate the refinement of all meshes used further in this study.

Note that in the following, we solve the nondimensional form of the Navier-Stokes equations using a characteristic length scale L_{ref} and a reference velocity scale U_{ref}. Consequently, all geometries are nondimensionalized based on their respective reference length.

III. FIELD INVERSION

The field inversion step supposes the inference of a corrective term added to the baseline model that minimizes the discrepancies between the enhanced model and the reference data. This section provides a discussion about the choice of the corrective term (Sec. III A) and gives details of how the inverse problem is posed (Sec. III B). The adjoint solver and the optimization algorithm used are presented in Sec. III C. Finally, the results for the data assimilation step are provided in Sec. III D for all flow configurations considered.

A. Corrective term

The main goal of this study is to improve the predictive performance of turbulence models for separated flows, where traditional models often show notable inaccuracies. This is accomplished by using data-assimilation techniques to integrate high-fidelity information into the simulations.

Cato *et al.* [15] compared the performance of several correction terms to match a full mean-flow velocity field, provided by averaged DNS simulations. They analyzed the pros and cons of each of these correction terms when used in the FIML framework. In the present study, all corrective terms in the momentum equations have been excluded because we intend to preserve the baseline behavior of the SA model and the physical arguments on which it has been derived. Such corrections have been proven very efficient to correct a specific type of flow [15] but might be detrimental for generalization. Inside the turbulent SA equation, we focused only on nondimensional corrective terms to make the symbolic regression step more robust. Deriving an analytical correction with dimensional units using symbolic regression is not straightforward.

The corrective field β added in front of the production term in the SA transport equation, was deemed relevant to improve separated flows. This corrective factor has the ability to act directly on the production of $\tilde{\nu}$ and consequently allows the turbulent viscosity to attain higher or lower values. This correction has been extensively used to improve the Spalart-Allmaras model with field inversion and machine learning with overall positive results [11–14]. Outside the scope of FIML, Shur *et al.* [39] attempted to account for rotation and curvature effects in the SA model by multiplying the production term in the $\tilde{\nu}$ equation by a function f_{r_1}.

Therefore, a spatially varying correction, β, is introduced as a multiplier to the production term $P(\tilde{\nu}, \nabla\overline{\mathbf{u}})$:

$$\overline{\mathbf{u}} \cdot \nabla\tilde{\nu} - \nabla \cdot [\eta(\tilde{\nu})\nabla\tilde{\nu}] = (1+\beta)P(\tilde{\nu}, \nabla\overline{\mathbf{u}}) - D(\tilde{\nu}, \nabla\overline{\mathbf{u}}) + C(\nabla\tilde{\nu}). \tag{11}$$

The reason for the $(1+\beta)$ form of the correction is only for practicality. In this way, when β is equal to zero, the baseline production is recovered.

B. Inverse problem

Inferring the appropriate β field to correct the baseline Spalart-Allmaras model leads to an inverse problem, often solved by data assimilation techniques. The reference data taken in the present study is the velocity vector field $\overline{\mathbf{u}}_{\mathrm{ref}}$ coming from LES or DNS. The choice regarding the reference data is motivated by the level of correction intended. Fanizza *et al.* [40] noticed that when abundant data were used in the data assimilation step, it was possible to get accurate correction of wall variables such as the skin-friction coefficient C_f and pressure coefficient C_p while ensuring a fairly good correction of the flow field. However, when only wall measurements were considered, the skin-friction coefficient could not be accurately corrected. They also raised the need for careful regularization techniques in the case of strongly intrusive corrections based on sparse measurements.

By using a variational approach, it can be shown that the inverse problem leads to a constrained optimization problem, which aims to minimize the cost functional $\mathcal{J}(\overline{\mathbf{q}}, \beta)$. Let $\mathcal{R}(\overline{\mathbf{q}}, \beta)$ be the residual vector of the stabilized nonlinear *RANS-SA* equations, with $\overline{\mathbf{q}}$ the vector containing the mean state variables and β the corrective field. The cost functional $\mathcal{J}(\overline{\mathbf{q}}, \beta)$ is function of both the mean state variables $\overline{\mathbf{q}}$ and the corrective term β. The mean state variables $\overline{\mathbf{q}}$

depend implicitly on the corrective field β as follows: $\overline{\mathbf{q}} = \overline{\mathbf{q}}(\beta)$. Thus, the constrained optimization problem to be solved can be expressed as

$$\begin{aligned} \min \quad & \mathcal{J}(\overline{\mathbf{q}}(\beta), \beta) \\ \text{s.t.} \quad & \mathcal{R}(\overline{\mathbf{q}}(\beta), \beta) = 0, \end{aligned} \tag{12}$$

with

$$\mathcal{J}(\overline{\mathbf{q}}(\beta), \beta) = \frac{1}{2}\|w_d[\overline{\mathbf{u}}(\beta) - \overline{\mathbf{u}}_{\text{ref}}]\|^2_{\mathrm{L}^2} + \frac{\lambda}{2}\|\beta\|^2_{\mathrm{L}^2}, \quad \text{with} \quad w_d = \frac{1}{d^* + 0.01} \quad \text{and} \quad d^* = \frac{d}{\frac{1}{2}\delta_0}, \tag{13}$$

where d is the wall distance used in the SA model, δ_0 is the height of the boundary layer taken at the inlet of the domain, and w_d is a geometrical weight to emphasize the importance of flow characteristics near the wall, as suggested by Fanizza *et al.* [40]. When the constant 0.01 in the expression for w_d is replaced by a different value, the contribution of the near-wall region to the cost functional is correspondingly increased or decreased. For example, without this weighting, the skin-friction coefficient could be poorly captured due to the low velocity magnitude near the walls. Adjusting the constant therefore allows better correction of wall quantities during the data-assimilation step. In practice, values in the range 0.01–0.1 yield similar reconstructions.

The first term in the cost functional $\mathcal{J}$ refers to the error between the modeled mean velocity field (provided by the RANS simulation) and the reference mean velocity field (coming from LES or DNS). The second term represents the deviation of β from its baseline value of zero, also known as the Tikhonov regularization. This term helps to constrain the magnitude of the correction applied to the model, by promoting a magnitude of β as small as possible. Smoother corrective fields can be obtained with such regularization techniques.

The choice of the regularization factor λ results in a tradeoff between ensuring a satisfactory drop in the cost functional $\mathcal{J}$ and reducing the deviation of β from its baseline value. Keeping β as small as possible reduces the impact of the correction on the baseline model, which is often desirable to preserve its original behavior and, consequently, the physical arguments on which it has been derived. In the present study, the value of λ was obtained by assessing this tradeoff for all cases used in the data assimilation step. $\lambda = 10^{-3}$ was deemed to provide a satisfactory tradeoff for the intended corrections.

C. Adjoint solver and optimization algorithm

The constrained optimization problem equivalent to Eq. (12), can be formalized through the Lagrangian $\mathcal{L}(\overline{\mathbf{q}}, \beta, \mathbf{Z})$, which is constructed with the help of Lagrange multipliers or adjoint states $\mathbf{Z}$:

$$\mathcal{L}(\overline{\mathbf{q}}(\beta), \beta, \mathbf{Z}) = \mathcal{J}(\overline{\mathbf{q}}(\beta), \beta) - \mathbf{Z}^T \mathcal{R}(\overline{\mathbf{q}}(\beta), \beta). \tag{14}$$

The stationary points of $\mathcal{L}(\overline{\mathbf{q}}(\beta), \beta, \mathbf{Z})$ are given by

$$\frac{\partial \mathcal{L}}{\partial \mathbf{Z}} = 0 \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \overline{\mathbf{q}}} = 0. \tag{15}$$

The first equality gives the constraint relative to the optimization problem:

$$\mathcal{R}(\overline{\mathbf{q}}(\beta), \beta) = 0. \tag{16}$$

The second equality yields the adjoint system:

$$\frac{\partial \mathcal{J}}{\partial \overline{\mathbf{q}}} - \mathbf{Z}^T \frac{\partial \mathcal{R}}{\partial \overline{\mathbf{q}}} = 0, \tag{17}$$

$$\left(\frac{\partial \mathcal{R}}{\partial \overline{\mathbf{q}}}\right)^T \mathbf{Z} = \left(\frac{\partial \mathcal{J}}{\partial \overline{\mathbf{q}}}\right)^T. \tag{18}$$

Applying the chain rule to obtain the total derivative $\mathrm{d}\mathcal{L}/\mathrm{d}\beta$ yields

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\beta} = \frac{\partial \mathcal{J}}{\partial \beta} + \frac{\partial \mathcal{J}}{\partial \overline{\mathbf{q}}}\frac{\mathrm{d}\overline{\mathbf{q}}}{\mathrm{d}\beta} - \mathbf{Z}^T\left[\frac{\partial \mathcal{R}}{\partial \beta} + \frac{\partial \mathcal{R}}{\partial \overline{\mathbf{q}}}\frac{\mathrm{d}\overline{\mathbf{q}}}{\mathrm{d}\beta}\right], \tag{19}$$

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\beta} = \frac{\partial \mathcal{J}}{\partial \beta} - \mathbf{Z}^T\frac{\partial \mathcal{R}}{\partial \beta} + \underbrace{\left[\frac{\partial \mathcal{J}}{\partial \overline{\mathbf{q}}} - \mathbf{Z}^T\frac{\partial \mathcal{R}}{\partial \overline{\mathbf{q}}}\right]}_{=0}\frac{\mathrm{d}\overline{\mathbf{q}}}{\mathrm{d}\beta}, \tag{20}$$

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\beta} = \frac{\partial \mathcal{J}}{\partial \beta} - \mathbf{Z}^T\frac{\partial \mathcal{R}}{\partial \beta}, \tag{21}$$

where $\mathbf{Z}$ is the solution of the adjoint system defined in Eq. (18).

Solving for the adjoint equations allows us to remove the dependence on $\mathrm{d}\overline{\mathbf{q}}/\mathrm{d}\beta$ which is computationally expensive to obtain. This makes the adjoint approach a remarkably efficient method in terms of computational effort [41].

The gradient $\mathrm{d}\mathcal{L}/\mathrm{d}\beta$ is provided by *dolfin-adjoint* which is the adjoint implementation used in *Firedrake*. *dolfin-adjoint* [42] is a generic operator-overloading based algorithmic differentiation framework for Python. It is employed for the automatic adjoint capabilities of *Firedrake*. Based on the weak form of the prescribed PDEs, the gradient $\mathrm{d}\mathcal{L}/\mathrm{d}\beta$ is calculated automatically.

To solve the inverse problem, the limited-memory Broyden-Fletcher-Goldfarb-Shanno bounded (L-BFGS-B) algorithm was used. This algorithm is suitable to solve nonlinear optimization in which information on the Hessian matrix is difficult to obtain, or for large dense problems [43]. An open-source implementation is available in the SciPy library. The L-BFGS-B algorithm only requires the gradient, it does not require the knowledge of second derivatives and approximates the Hessian matrix [44]. The gradient $\mathrm{d}\mathcal{L}/\mathrm{d}\beta$ computed by *dolfin-adjoint* is thus forwarded to the optimization algorithm to solve the problem.

The gradients provided by *dolfin-adjoint* are defined with respect to the scalar-product $\langle f, g\rangle_{\mathrm{L}^2(\Omega)} = \int_\Omega f(\mathbf{x})g(\mathbf{x})d\Omega$, where Ω refers to the volume of the computational domain. In the FEM, the integral of the basis functions $\int_\Omega \phi_i(\mathbf{x})\phi_j(\mathbf{x})d\Omega$ defines the components of the mass matrix $\mathbf{M}$. Therefore, the discrete L^2 scalar-product is calculated as: $\langle \boldsymbol{f}, \boldsymbol{g}\rangle_{\mathbf{M}} = \boldsymbol{f}^T\mathbf{M}\boldsymbol{g}$. However, the L-BFGS-B optimizer is restricted to the Euclidean scalar product (defined as $\langle \boldsymbol{f}, \boldsymbol{g}\rangle_{\ell^2} = \boldsymbol{f}^T\boldsymbol{g}$) and not the L^2 used in FEM. Following Refs. [40,45], the optimization is thus performed with respect to β^Q for which the degrees of freedom vector has been rescaled as $\beta^Q = Q\,\beta$ to be consistent with the L-BFGS-B implementation. The diagonal matrix Q is computed from a Cholesky factorization of a lumped mass matrix $\mathbf{M}_{\mathrm{Lump}}$ with $\mathbf{M}_{\mathrm{Lump}} = QQ^T$. The mass lumping technique has been used to handle instantly the Cholesky factorization of the mass matrix $\mathbf{M}$.

D. Assimilation results

The data-assimilation methodology was performed on the flow configurations presented in Table I. These cases exhibit flow separation of varying intensity, from mild to strong, and are selected to highlight the limitations of the baseline SA model in predicting separated flows. The objective of this study is to improve the SA model through a correction applicable to such flows. To this end, considering reference cases across a wide range of Reynolds numbers is essential. Moreover, the diversity of separation types encountered in these cases may further support the development of a broadly applicable correction.

TABLE I. Flow configurations used for field inversion. The Reynolds numbers are based on the reference lengths and reference velocities relative to each configuration.

Flow configurations	Reynolds number	High-fidelity data	$\mathcal{J}/\mathcal{J}_0$
Converging-diverging channel	$\mathrm{Re}_{h/2} = 12\,600$	DNS [46]	0.093
Bump H42	$\mathrm{Re}_c = 205\,000$	LES [47]	0.278
NASA wall-mounted hump	$\mathrm{Re}_c = 936\,000$	LES [48]	0.345

The corrective fields obtained by data assimilation serve as reference data for the machine learning step. In this way, they represent the best corrections that an expression obtained with symbolic regression can reach.

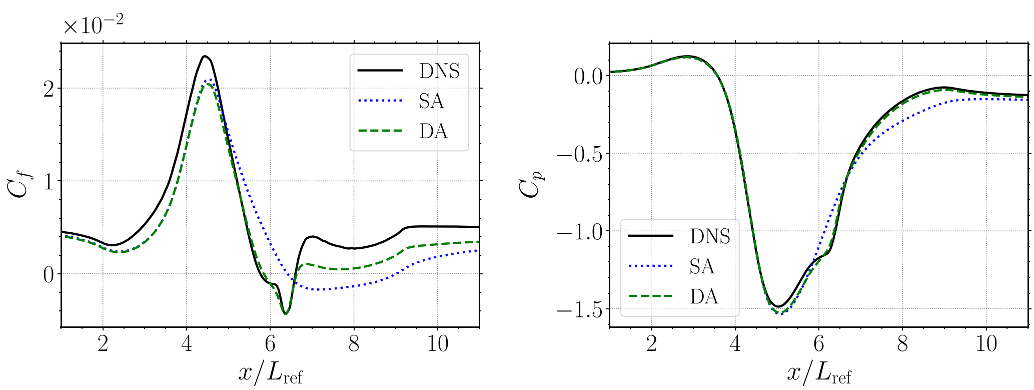
The cost functional ratios $\mathcal{J}/\mathcal{J}_0$ obtained after field inversion are provided in Table I for all cases considered.

1. Converging-diverging channel

Marquille *et al.* [46] performed a direct numerical simulation of a turbulent channel flow with a lower curved wall. The flow slightly separates over the bottom wall. This behavior appears to be due to the low Reynolds number investigated, which is based on the half-channel height and the mean velocity at the inlet. They mentioned the absence of any flow separation in experimental data measured at much higher Reynolds number. The authors also pointed out a strong production of turbulent kinetic energy at the location where the pressure gradient takes its maximum value, on the upper and lower walls.

The original SA model completely fails to identify the small recirculation exhibited by this flow configuration as shown by the skin friction coefficient C_f in Fig. 1(a). The separation and reattachment points are incorrectly predicted. However, the data assimilation provides a corrective field β [Fig. 4(a)] allowing an accurate prediction of the recirculation. Wall quantities are overall well corrected after field inversion [Figs. 1(a) and 1(b)].

However, the C_f further downstream still exhibits some discrepancies that are mostly due to the architecture of the underlying model or the correction term considered. The results obtained for this configuration are in agreement with those by Cato *et al.* [15].



(a) **(b)**

FIG. 1. Wall quantities corrected after field inversion (FI) on the converging-diverging channel, (a) skin-friction coefficient C_f and (b) pressure coefficient C_p.

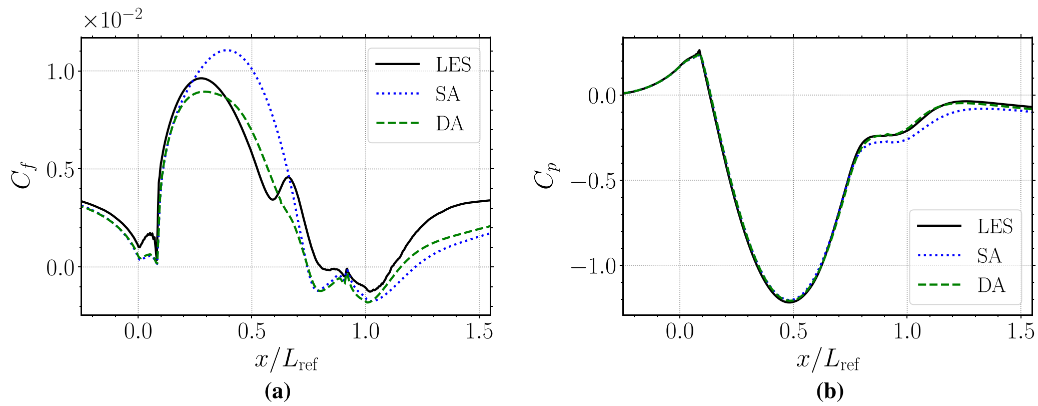


FIG. 2. Wall quantities corrected after field inversion (FI) on the bump H42. (a) Skin-friction coefficient C_f and (b) pressure coefficient C_p.

2. Bump H42

Matai *et al.* [47] investigated turbulent flows over a series of two-dimensional bumps of varying heights using wall-resolved large-eddy simulations. The Reynolds number, based on the bump length and freestream velocity, was kept constant across all configurations. The flow over the highest bump (H42) presents a recirculation region on the bottom surface. In this configuration, the flow experiences a favorable pressure gradient along the front half of the bump, followed by the formation of a region of high turbulent kinetic energy in the lee of the bump that extends into the wake. The flow appears to be significantly influenced by this region of high turbulent kinetic energy, which originates before the flow separation.

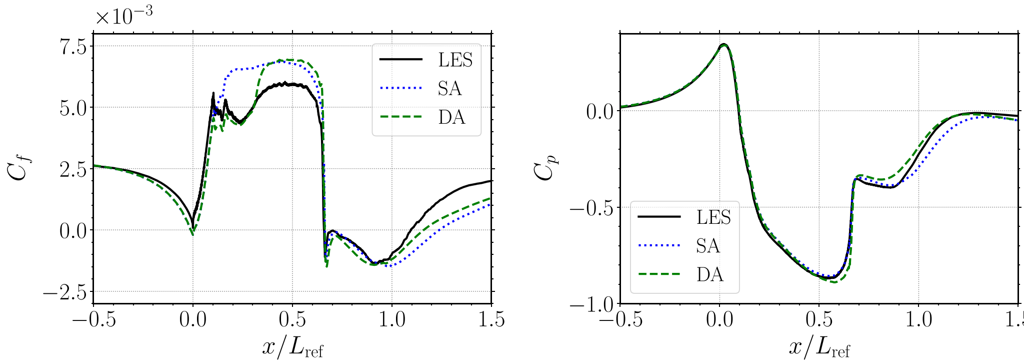
Two aspects of the flow are missing in the baseline predictions. The skin friction coefficient over the windward side of the bump is over-predicted and fails to replicate the oscillating pattern observed in the reference data [Fig. 2(a)]. Moreover, the length of the recirculation region is once again over-predicted by the baseline SA model. The data-assimilation produces a correction allowing to predict more accurately the reattachment point. The skin-friction and pressure coefficients are better estimated after field inversion [Figs. 2(a) and 2(b)].

As previously observed, the data assimilation is not able to provide a field for β [Fig. 5(a)] allowing to fully correct the baseline model. This illustrates inadequacies both in the model structure and corrective term. The pressure coefficient, however, is correctly predicted by the data-assimilated (DA) approach [Fig. 2(b)].

3. NASA wall-mounted hump

Uzun *et al.* [48] conducted a wall-resolved large-eddy simulation to investigate the flow separation over the NASA wall-mounted hump. Over the front part of the hump, the flow presents a relaminarization observed due to a strong favorable pressure gradient. Then, an adverse pressure gradient appears on the aft region of the hump and causes the flow to separate. Representative of many industrial cases, the main challenge is to accurately predict the reattachment point. Similarly to the bump, the Reynolds number for this flow is based on the length of the hump and on the freestream velocity.

The assimilated correction [Fig. 6(a)] tries to account for the flow relaminarization as shown in Fig. 3(a). The reattachment point, in particular, is predicted with higher accuracy once the field inversion is applied [Fig. 3(a)]. However, the improvements on the pressure coefficient are less significant [Fig. 3(b)].



(a) (b)

FIG. 3. Wall quantities corrected after field inversion (FI) on the NASA wall-mounted hump. (a) Skin-friction coefficient C_f and (b) pressure coefficient C_p.

IV. SYMBOLIC REGRESSION

The data assimilation step provided optimized correction fields β_{DA} for multiple flow configurations. For these corrections to be implemented in the SA turbulence model, a mapping between the local flow features $\boldsymbol{\eta}$ and the correction β needs to be established. To this purpose, symbolic regression is used to determine an analytical expression for $\beta_{\mathrm{SR}}(\boldsymbol{\eta})$ based on the corrective fields previously inferred by field inversion. Contrary to classical black-box corrections performed with traditional machine learning techniques (artificial neural networks, random forests, etc.), an analytical expression is expected. This section presents the overall setup used for symbolic regression (Sec. IV A) and provides a discussion about the input features and their nondimensionalization (Sec. IV B). The final SR expression is provided in Sec. IV C. Finally, results for both in-sample and out-of-sample cases are discussed in Secs. IV D and IV E, respectively.

A. PySR

PySR is a Python library developed by Cranmer [22], which has been used in this study to obtain analytical correction through symbolic regression. Its search strategy relies on a multi-population evolutionary algorithm, which consists of a unique evolve-simplify-optimize loop [22]. Multiple settings (operators, complexity, cost functional, etc.) can be used to constrain the search algorithm of PySR and to promote expressions that are physically meaningful and numerically compatible with RANS solvers.

1. Operators

Operators and elementary functions usually used in turbulence modeling have been selected. They are presented in Table II. However, some constraints have been established to ensure numerically stable solutions. Nests of a particular operator are prohibited (an exception has been made for tanh(tanh(·))), meaning that $((\cdot)^2)^2$ is forbidden, however, $\tanh(\cdot)^2$ is possible. All division

TABLE II. List of operators considered for the regression.

Operators type	Operators
Unary	$\tanh(\cdot)$, $\sqrt{\cdot}$, $(\cdot)^2$, $(\cdot)^3$, $\min(\cdot)$, $\max(\cdot)$
Binary	$+, -, \times$

operators $(\cdot)/(\cdot)$ or $1/(\cdot)$ have been removed after some regression attempts where they led to the divergence of the solver depending on the input features involved in the expressions. An input feature which is equal to zero at the wall for instance, cannot be in the denominator of a division. Another custom division operator $1/(1+\cdot)$ which could prevent this situation has been considered but with no major improvement on the results. Note that the selection of suitable operators is mainly guided by trial-and-error even if some physical and mathematical insights might be favorable for this exercise.

2. Cost functional

The cost functional prescribed in PySR is the weighted mean-square error (WMSE) which consists of the sum of the squared difference between β_{DA}, the correction found during the data assimilation step and $\beta_{\mathrm{SR}}(\boldsymbol{\eta}_{\mathrm{DA}})$, the corrective field predicted by the SR expression based on the input features $\boldsymbol{\eta}_{\mathrm{DA}}$. An additional weight α has been introduced in the cost functional $\mathcal{J}_{\mathrm{SR}}$:

$$\mathcal{J}_{\mathrm{SR}} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i\left[\beta_{\mathrm{SR}_i}(\boldsymbol{\eta}_{\mathrm{DA}}) - \beta_{\mathrm{DA}_i}\right]^2, \tag{22}$$

where N is the total number of sample and $\boldsymbol{\eta}_{\mathrm{DA}}$, the input features corresponding to the data assimilation solution.

Since input features from the data assimilation solution, $\boldsymbol{\eta}_{\mathrm{DA}}$, are used in the cost functional $\mathcal{J}_{\mathrm{SR}}$, the resulting regression expressions are expected to yield corrective fields consistent with the assimilated state. However, when the expressions are plugged into the RANS simulation, nothing ensures that the correction will effectively converge toward the same solution as obtained for the data assimilation.

3. Regression parameters

For a given regression case, the PySR algorithm is run for $n_p = 200$ populations, each containing $n_e = 40$ individuals (mathematical expressions). At each iteration, the most accurate expressions in each population and for each complexity are saved. Then, the most accurate individuals among all the populations are saved for each complexity. The individuals evolve inside each population through mutation and crossover, happening at each iteration. Repeating this procedure for $n_i = 250$ iterations gives the set of most accurate expressions ranked by complexity. The overall complexity of an expression is formed by summing the complexities associated with its constants γ_{const}, variables γ_{var}, and operators γ_{oper}. The maximum equation complexity has been set to $C(E) \leqslant 30$ while a unitary complexity has been applied for constants, variables and operators: $\gamma_{\mathrm{const}} = \gamma_{\mathrm{var}} = \gamma_{\mathrm{oper}} = 1$:

$$C(E) = \gamma_{\mathrm{const}}N_{\mathrm{const}} + \gamma_{\mathrm{var}}N_{\mathrm{var}} + \gamma_{\mathrm{oper}}N_{\mathrm{oper}} = N_{\mathrm{const}} + N_{\mathrm{var}} + N_{\mathrm{oper}}, \tag{23}$$

where $C(E)$ is the complexity of a given equation and N the occurrence of either a constant, a variable or an operator. All parameters related to mutations, migrations, and tournament selection were kept at their default values. More information can be found in Ref. [22].

4. Downsampling

In PySR, symbolic regression relies on genetic algorithms which is essentially a brute force approach scaling exponentially with the number of input variables and operators [49]. The correction field β_{DA} inferred by field inversion contains as many points as the meshes used for the RANS simulations, representing approximately around $\sim 10^5$ points. The efficiency of symbolic regression depends on the size of the dataset used to obtain the equations. Cranmer [22] argues that generally $\sim 10^4$ sample points is enough to find a functional form with symbolic regression, more datapoints will drastically lower the search speed. Consequently, the dataset containing the correction field β_{DA} and the input features $\boldsymbol{\eta}_{\mathrm{DA}}$ should be downsampled. The samples constituting the inferred β_{DA} field are split into two categories: trivial and nontrivial samples. A sample will be considered as nontrivial if the following condition is verified: $|\beta_{\mathrm{DA}}| > 0.1$. The nontrivial samples are likely to be the datapoints defining the regions of the flow where a correction is needed. However, many samples still satisfy this condition $|\beta_{\mathrm{DA}}| > 0.1$, thus only a random subset $N_{\mathrm{nontrivial}}$ of these datapoints will be effectively taken as nontrivial samples. The same reasoning approach applies to trivial samples, only a random subset N_{trivial} among all samples verifying the condition $|\beta_{\mathrm{DA}}| \leqslant 0.1$ will be taken as such. The maximum number of sample per flow configuration is fixed at $N_{\mathrm{max/flow}} = N_{\mathrm{trivial}} + N_{\mathrm{nontrivial}} = 5 \times 10^3$, with the number of trivial samples being equal to the number of nontrivial samples $N_{\mathrm{trivial}} = N_{\mathrm{nontrivial}}$.

B. Input features

1. Guidelines

The selection of input features on which the analytical correction is based is a critical step in turbulence modeling. There is a broad consensus in the literature on the general guidelines that should be followed.

The intended correction is applied through a nondimensional term β multiplying the production term in the $\tilde{\nu}$ transport equation. Consequently, the inputs selected to obtain such correction must be nondimensional. Duraisamy *et al.* [6] argued that this nondimensionalization should be done locally rather than globally because it offers more possibility to generalize across different flow configurations.

In the field of turbulence modeling, invariance properties (i.e., Galilean, rotational, reflectional, and frame invariance) must be enforced in the models [50]. This should apply to both the selected input features $\boldsymbol{\eta}$ and quantities used for global or local nondimensionalization [51]. Spalart [50] considers that if the velocity is included in the general formulation of a model, then it is "not to be a model," because Galilean invariance is violated.

2. Selection and nondimensionalization

The set of input features $\boldsymbol{\eta}$ which has been considered is shown in Table III. Only a subset of the inputs presented in Table III have effectively been used by the symbolic regression. The three damping functions f_{v_1}, f_{v_2}, and f_w have been taken as inputs for their role in identifying the different regions of the boundary layers. These functions have been calibrated to ensure proper near-wall behavior in the baseline SA model. The physical knowledge incorporated in these functions might be favorable to obtain a correction for separated flows which does not harm the predictions for wall-attached flows. The shielding function f_d' which is a modified version of f_d initially introduced by Spalart *et al.* [53] to identify the boundary layer, has also been included in the set of input features. Ferrero *et al.* [13] argued that f_d' better represents the near-wall flow features while f_d tends to compress the information and does not allow to distinguish the different structures.

The intended correction β should be active in separated regions. We thus need some inputs to relate the correction to these regions.

The norms of the strain rate tensor S or rotational rate tensor Ω are widely used to correct separated flows but they need to be nondimensionalized in our case. Wu *et al.* [29] and Srivastava *et al.* [55], used $\mathrm{Re}_\Omega = (d^2\Omega)/\nu$ to identify regions of high vorticity away from the wall. This nondimensionalization procedure with d^2/ν is appealing since it is local (even if the locality of d can be discussed [50]) and mostly, it verifies all invariance properties. However, due to the poor properties of d^2/ν used to nondimensionalize Ω, no satisfactory results have been obtained by including Re_Ω in the set of input features. Tracey *et al.* [51] acknowledged the poor properties of nondimensionalizers such as $d/(\nu + \nu_t)$, $d^2/(\nu + \nu_t)$, or $d^2/(\nu + \nu_t)^2$ as they often hide the small region of relevant differences in input values (especially when d decreases). Another problem with these nondimensionalizers is the high probability to encounter extreme outliers (as d increases) which might be detrimental for unseen flow configurations.

TABLE III. List of input features $\boldsymbol{\eta} = [\eta_1, \cdots, \eta_7]$ considered for symbolic regression.

Name	Quantity	Expression	Description
η_1	f_{v_1}	$\frac{\chi^3}{c_{v_1}^3+\chi^3}$	Damping function in the Spalart-Allmaras model, originally from Herring *et al.* [52]. Allows us to retrieve the right behavior of ν_t in the viscous region of the boundary layer.
η_2	f_{v_2}	$1-\frac{\chi}{1+\chi f_{v_1}}$	Damping function in the Spalart-Allmaras model, constructed for $\tilde{S}$ to maintain its log-layer behavior, $\tilde{S} = u_\tau/(\kappa y)$, all the way to the wall.
η_3	f_w	$g\left[\frac{1+c_{w_3}^6}{g^6+c_{w_3}^6}\right]^{\frac{1}{6}}$	Damping function in the Spalart-Allmaras model, allowing to obtain the right behavior for the destruction in the outer region of the boundary layer.
η_4	f'_d	$\tanh(\sqrt{r'_d})$ with $r'_d = \frac{\nu+\nu_t}{\lVert\nabla\bar{\mathbf{u}}\rVert\kappa^2 d^2}$	Analogous to the shielding function modified by Ferrero *et al.* [13] to identify flow features close to the wall but initially introduced by Spalart *et al.* [53].
η_5	Re_Ω	$\frac{d^2\Omega}{\nu}$	Reynolds based on the norm of the rotation rate tensor Ω. Used to identify high vorticity region away from the wall.
η_6	$\mathrm{Re}_{k_{\mathrm{qcr}}}$	$\frac{d\sqrt{k_{\mathrm{qcr}}}}{\nu}$	Reynolds based on the quadratic constitutive relation (QCR) form of the turbulent kinetic energy k_{qcr}, introduced in Ref. [54]. Used to identify regions away from the wall characterized by high turbulent kinetic energy.
η_7	$\frac{k_{\mathrm{qcr}}}{U_{\mathrm{ref}}^2}$	$\frac{3}{2}\frac{c_{r_2}\nu_t}{U_{\mathrm{ref}}^2}\sqrt{2S_{ij}S_{ij}}$	Quadratic constitutive relation (QCR) form of the turbulent kinetic energy k_{qcr}, introduced in Ref. [54]. It aims to be analog to the Turbulent Kinetic Energy for the SA model (where there is no transport equation for k). Nondimensionalized by the square of the reference velocity U_{ref}.

However, Marquillie *et al.* [46] and Matai *et al.* [47] mentioned the importance of the turbulent kinetic energy for flows on which data assimilation has been performed, including flows exhibiting strong disequilibrium. Consequently, the turbulent kinetic energy k appears to be a relevant input to identify these regions of strong disequilibrium. Similarly, Volpiani [10] mentioned the significant relevance of the turbulent kinetic energy in the estimation of the corrected eddy viscosity with random forest and artificial neural networks. Fang *et al.* [56] also considered the turbulent kinetic energy as a relevant input feature to correct free-shear flows.

The 2013 version of the quadratic constitutive relation (QCR) for the SA model [54] contains the k_{qcr} term, which is supposed to be analogous to the turbulent kinetic energy (not solved in the SA model). k_{qcr} is a dimensional quantity ($\mathrm{m}^2\,\mathrm{s}^{-2}$) and consequently its nondimensionalization is necessary to be considered as an input feature. As mentioned previously, we can again imagine a nondimensionalization by constructing a Reynolds number based on k_{qcr} with d/ν or $d/(\nu+\nu_t)$. However, such nondimensionalization did not yield satisfactory results for the same reasons as already discussed. Another possibility to nondimensionalize k_{qcr} is by using the global quantity U_{ref}. However, the major problem with using U_{ref} is the violation of the Galilean invariance, which must be enforced when developing a model. Using such a quantity to nondimensionalize k_{qcr} might be detrimental for the ability of the model to satisfy Galilean invariance. Nevertheless, the improvement in the model predictions justifies its use as shown in the following.

It should be noted that other models likewise do not satisfy Galilean invariance. Langtry and Menter [57] violated Galilean invariance in the derivation of a correlation-based transition model. All turbulence models based on correlations with the turbulence intensity T_u inevitably violate

TABLE IV. Flow configurations used for the symbolic regression. The Reynolds numbers are based on the reference lengths and reference velocities relative to each configuration.

Flow configurations	Reynolds number	Corrective term β	Weight α
Converging-diverging channel	$\mathrm{Re}_{h/2} = 12\,600$	$\beta = \beta_{\mathrm{DA}}$	$\alpha = 1$
Bump H42	$\mathrm{Re}_c = 205\,000$	$\beta = \beta_{\mathrm{DA}}$	$\alpha = 1$
NASA wall-mounted hump	$\mathrm{Re}_c = 936\,000$	$\beta = \beta_{\mathrm{DA}}$	$\alpha = 1$
Flat plate	$\mathrm{Re}_L = 5 \times 10^6$	$\beta = 0$	$\alpha = 10$

Galilean invariance because of the relationship between T_u and the local velocity. In this case, Langtry and Menter considered the violation of the Galilean invariance "not problematic" since it serves the purpose of the model. They also acknowledged that multiple moving walls in one domain would require additional information.

He *et al.* [18] tried to enhance the prediction accuracy of the SA model for airfoil stall using field inversion and symbolic regression. However, they used a length scale which depends on the chord of the airfoil to define the nondimensional strain $\hat{S}$ and rotation tensor $\hat{\Omega}$. The use of such length scale prevents any application to other geometries. We consider that using the reference velocity to nondimensionalize the inputs can lead to better generalization properties than by using a length scale relative to the geometry for instance. Furthermore, this lack of Galilean invariance is not detrimental for the wall-bounded flow configurations under consideration in this study, as the wall velocity can be considered as the zero-velocity reference.

This discussion illustrates the difficulty in selecting inputs which are nondimensional, local and verifying all invariance properties. In the case of the Spalart-Allmaras turbulence model, there is, in fact, not many quantities that could be used to nondimensionalize relevant input features for separated flows. Note that this problem concerns mainly the SA turbulence model as only one additional transport equation is solved compared to the $k - \epsilon$ or $k - \omega$ turbulence models where two additional transport equations are solved, offering more possibilities for the nondimensionalization of the input features.

C. Symbolic correction

1. Regression set

The symbolic regression was performed on a set containing four different flow configurations: the three cases for which a corrective field β has been inferred through field inversion (converging-diverging channel, bump h42, and NASA wall-mounted hump) and an additional case where no correction should be applied. More details are available in Table IV.

The flat plate has been selected for this additional case for which we set $\beta = 0$ as reference data. By including the flat-plate in the regression set, we intend to keep the behavior of the baseline model for wall-attached flows. The weight α in the cost functional used to guide the symbolic regression has been set to a higher value for the flat plate case. This weighting strategy penalizes expressions that do not predict $\beta = 0$ in cases where no correction is required. As a result, the learned correction is expected to have a limited impact on the baseline model calibration in attached-flow regimes.

2. Symbolic expression

The best analytical correction β_{SR} provided by SR for the set of input features $\boldsymbol{\eta}$ reads

$$\beta_{\mathrm{SR}}(\boldsymbol{\eta}_{\mathrm{SR}}) = c_1\eta_7 + \tanh[f(\boldsymbol{\eta}_{\mathrm{SR}}) - c_2\eta_7], \tag{24}$$

$$f(\boldsymbol{\eta}_{\mathrm{SR}}) = c_3\eta_2\big[\eta_7\big(\eta_2^2 + \eta_4\big)[\tanh(c_4\eta_3) - c_5]\big]^2, \tag{25}$$

where

$$\boldsymbol{\eta}_{\mathrm{SR}} = [\eta_2, \eta_3, \eta_4, \eta_7],$$

$$\eta_2 = f_{v_2}, \quad \eta_3 = f_w, \quad \eta_4 = f_d', \quad \eta_7 = \frac{k_{\mathrm{qcr}}}{U_{\mathrm{ref}}^2},$$

$$c_1 = 22.0, \quad c_2 = 22.1326, \quad c_3 = 703381.9214, \quad c_4 = 1.7723, \quad c_5 = 0.9823.$$

The expression β_{SR} is based on the input features included in $\boldsymbol{\eta}_{\mathrm{SR}}$ (subset of $\boldsymbol{\eta}$): the damping functions f_{v_2} and f_w, the shielding function f_d' and the turbulent kinetic energy k_{qcr} nondimensionalized by the square of the reference velocity U_{ref}^2. The correction yields a direct relationship between the corrective term β and the turbulent kinetic energy k_{qcr}, promoting a correction of the SA model in regions where k_{qcr} takes significant values. This type of correction has the ability to improve the predictions of the SA model for detached flows by increasing locally the production of $\tilde{\nu}$, leading to higher values for the turbulent viscosity ν_t in separated regions.

The structure of the equation $c_1\eta_7 + \tanh[f(\boldsymbol{\eta}_{\mathrm{SR}}) - c_2\eta_7]$ is simple. However, we acknowledge that the form of $f(\boldsymbol{\eta}_{\mathrm{SR}})$ is more complex. According to the complexity rules introduced in Sec. IV A, the current expression has an overall complexity of 28 which is high.

The expression given in Eq. (24) is unbounded. To avoid regions of the flow where the correction could lead to an excessive amplification of the SA production term, a limiter is introduced. In the present study, the increase of the production term is restricted to a maximum of 3.5 times its baseline value. Accordingly, the symbolic correction is clipped at a value of 2.5. This clipping is also important to prevent unphysical values of the correction factor β (for instance near sharp corners) and to improve solver convergence. All results reported in this work were obtained using this limiter value.

D. Results on regression cases

In this section, the SR correction is validated on the cases included in the regression set. A quick clarification is given concerning the notations used in the following. SA refers to the baseline Spalart-Allmaras model with no correction (i.e., $\beta = 0$). DA refers to the Spalart-Allmaras model enhanced with the correction inferred via data assimilation, representing the optimal performance achievable within the constraints of the original model formulation. SR refers to the Spalart-Allmaras model improved with the analytical correction found through symbolic regression.

1. Converging-diverging channel

The symbolic correction β_{SR} allows us to fairly reconstruct the corrective field inferred by data assimilation β_{DA} as shown in Figs. 4(a) and 4(b). The region in the divergent part requiring high values of β is well captured by the expression. This increase in the turbulent viscosity ν_t forces the flow to reattach earlier than the predictions of the baseline Spalart-Allmaras model. Figures 4(c) and 4(d) show good agreement between the skin-friction and wall-pressure coefficient predictions obtained from data assimilation and those computed using the symbolic expression (24). The SR model yields overall better predictions than the baseline SA model.

It should be noticed that the analytical correction β_{SR} found by symbolic regression should theoretically not surpass the predictions provided by the field inversion procedure. This expected behavior can be confirmed here, our symbolic correction β_{SR}, which we expect to be applicable over different flow configurations, reveals slight discrepancies compared to the data-assimilation, especially where the flow separates.

2. Bump H42

As observed for the converging-diverging channel, the SR expression reconstructs the corrective field inferred by data assimilation well [Figs. 5(a) and 5(b)]. The SR correction is concentrated

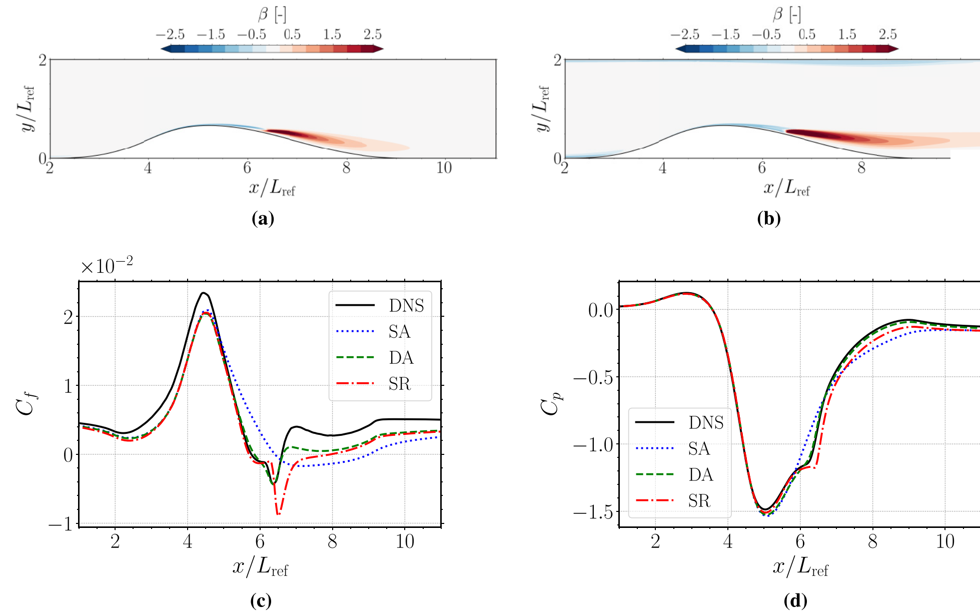


FIG. 4. Field inversion (FI) and symbolic regression (SR) procedure on the converging-diverging channel. (a) Corrective field β_{DA}, (b) symbolic corrective field β_{SR} (c) skin-friction coefficient C_f, and (d) pressure coefficient C_p.

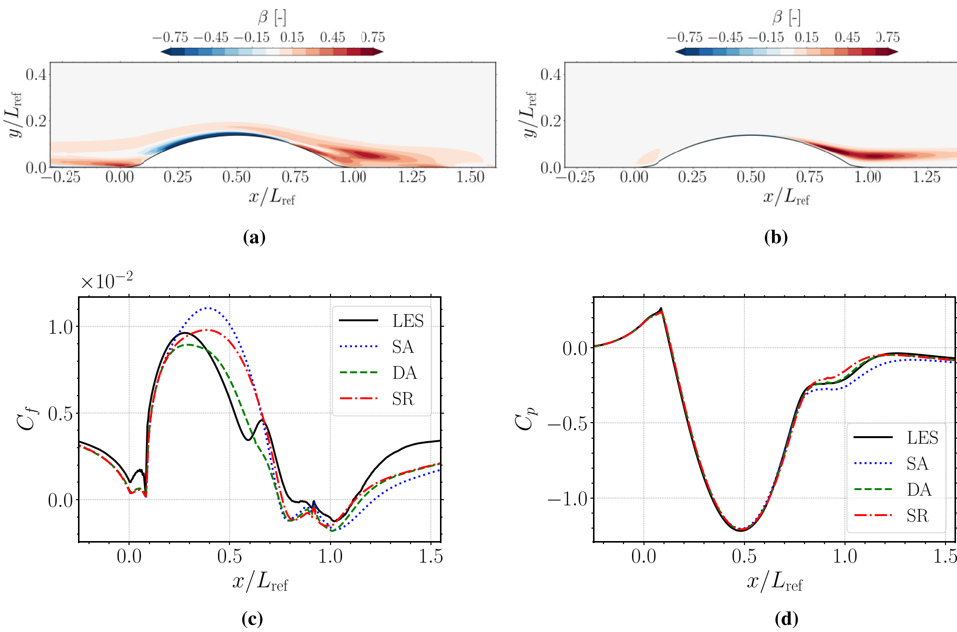


FIG. 5. Field inversion (FI) and symbolic regression (SR) procedure on the bump H42. (a) Corrective field β_{DA}, (b) symbolic corrective field β_{SR} (c) skin-friction coefficient C_f, and (d) pressure coefficient C_p.

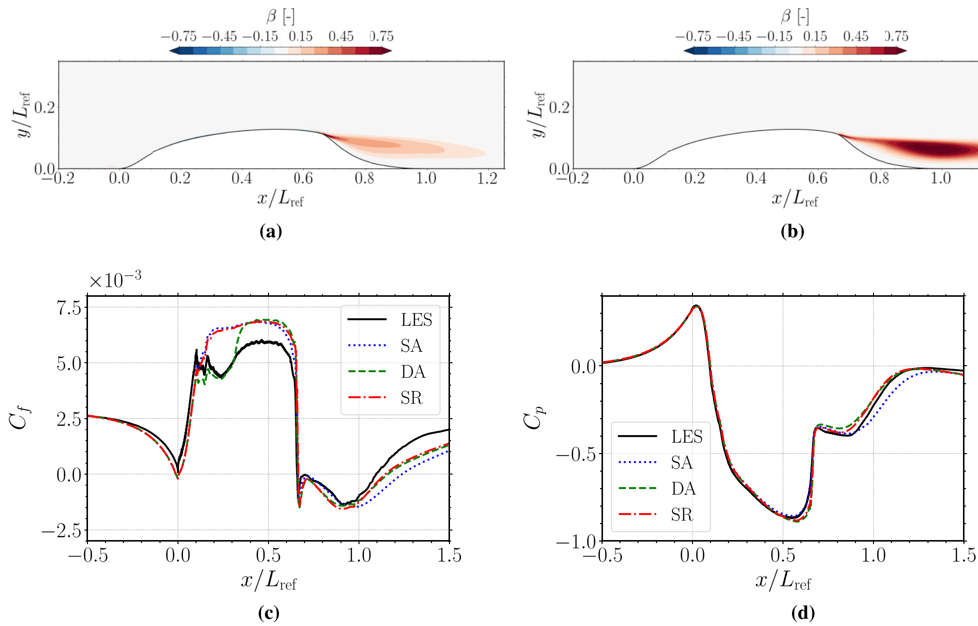


FIG. 6. Field inversion (FI) and symbolic regression (SR) procedure on the NASA wall-mounted hump. (a) Corrective field β_{DA}, (b) symbolic corrective field β_{SR}, (c) skin-friction coefficient C_f, and (d) pressure coefficient C_p.

behind the bump in the recirculation region of the flow, while the field inversion predicted a correction across the full domain. This is mainly due to the type of correction sought in the present study, which is intended to act in flow-separation regions. Because the SR expression was derived from all cases in the regression dataset, it primarily captures the corrections that are common to the different flow fields. Consequently, the symbolic regression model embeds the shared characteristics of the corrective fields across all cases.

Figures 5(c) and 5(d) demonstrate that the wall quantities, including skin-friction and pressure coefficients, are well captured by the model. The SR model provides a much better prediction of the length associated to the recirculation region compared to the baseline SA model [Fig. 5(c)]. The increase in the C_f over the first half of the bump is also more accurately predicted, yielding a maximum value closer to the LES estimations. Despite the good estimation of the C_p by the baseline SA model, the SR correction still provides slight improvements [Fig. 5(d)].

3. NASA wall-mounted hump

As in the previous cases, the SR expression reproduces well the corrective field inferred by data assimilation for the flow over the NASA wall-mounted hump, especially near the aft of the hump [Figs. 6(a) and 6(b)]. The SR expression predicts a stronger patch of production downstream of the hump. The relaminarization on the windward side of the hump captured by the data assimilation has not been well embedded in the SR correction, as expected. The symbolic regression, due to the simplicity of the expression, mostly captures features common to all DA corrections considered (as discussed for the h42 bump).

The skin-friction coefficient corrected by the SR model exactly follows the data assimilation in the separated region. The reattachment point is estimated better with the SR model than with the baseline SA model [Fig. 6(c)]. The wall pressure is also in agreement with the reference data for the SR-based model [Fig. 6(d)].

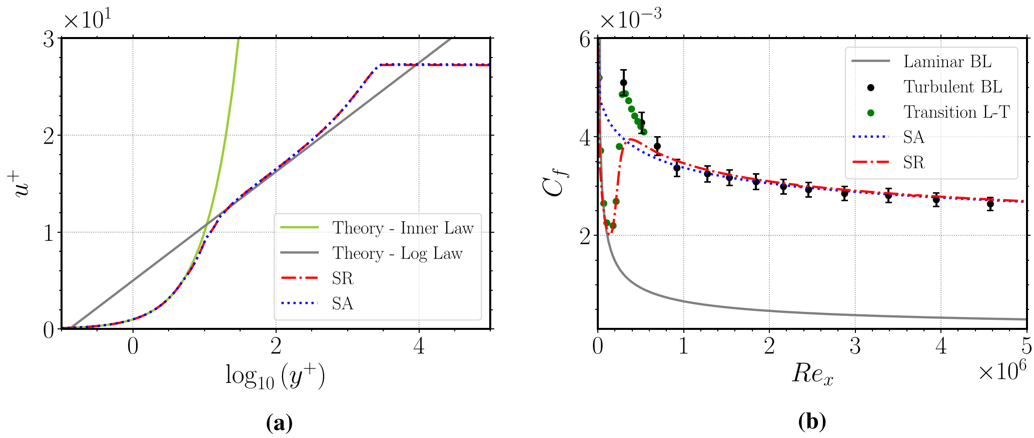


FIG. 7. Velocity Profile u^+ versus y^+ and Skin-Friction Coefficient C_f for a flow over a flat plate with no pressure gradient. Experimental data for turbulent boundary layer from Wieghardt *et al.* [58]. Experimental data relative to the transition from laminar to turbulent states from Savill *et al.* [59], (a) $\mathrm{Re}_x = 4.85.4 \times 10^6$, (b) skin-friction coefficient C_f.

4. *Zero pressure gradient flat plate*

The zero pressure gradient flat plate was also included in the regression set to keep the good behavior of the baseline SA model for wall-attached flows. The regression algorithm was guided to favor expressions that preserve the baseline turbulent boundary layer predictions (i.e., $\beta = 0$) through a strong penalization in the cost functional. The velocity profiles taken at $\mathrm{Re}_x = 4.8504 \times 10^6$ reveal perfect agreement between the SR and SA predictions [Fig. 7(a)]. At this location, the boundary layer is fully developed and fully turbulent.

The SR predictions for the skin-friction coefficient [Fig. 7(b)] also show overall agreement with the baseline SA predictions for turbulent boundary layers. Experimental results from Wieghardt *et al.* [58] have been included as reference. They investigated a turbulent boundary layer developing over a flat plate with no pressure gradient. In the experiments, the flow has been tripped at $\mathrm{Re}_x = 0$ to ensure a fully turbulent boundary layer [58]. Both SR and SA predictions lie inside the 5% error bar associated to the experiments. A threshold of 5% difference between the reference and the predictions is often used in the literature for the C_f of the turbulent flat plate [21,60].

Nevertheless, near the leading edge of the flat-plate, right after $\mathrm{Re}_x = 0$, the skin-friction coefficient [Fig. 7(b)] reveals surprising results. Curiously, the SR model predicts a C_f that appears to match the behavior of a laminar boundary layer, followed by a transition toward the turbulent regime. This observation was supported by the Blasius solution of a laminar boundary layer $C_f = 0.664/\sqrt{\mathrm{Re}_x}$ and the experimental results of Savill [59] for a flow over a zero pressure gradient flat-plate. Savill [59] did not trip the boundary layer contrary to Wieghardt *et al.* [58] and focused on the transition between laminar and turbulent states.

The Spalart-Allmaras model considered assumes a fully turbulent boundary layer and is therefore unable to model the transition from laminar to turbulent states without some additional terms or equations. Moreover, no correction β_{DA} was inferred by data assimilation for a laminar boundary layer transitioning to a turbulent state. The prediction of this laminar flow by the SR model was therefore not expected. However, Rumsey *et al.* [21] already reported a similar observation regarding a correction on β embedded within a neural network for the Spalart-Allmaras model.

E. Results on unseen scenarios

The SR correction was further assessed on seven unseen flow configurations, including five two-dimensional cases (bump H38, square cylinder, periodic hill, backward-facing steps both standard

TABLE V. Flow Configurations used to test the expression β_{SR} found by Symbolic Regression. The Reynolds numbers are based on the reference lengths and reference velocities relative to each configuration.

Flow configurations	Reynolds number	High-fidelity data
Bump H38	$\mathrm{Re}_c = 205\,000$	LES [47]
Square cylinder	$\mathrm{Re}_h = 22\,000$	DNS [61]
Periodic hill	$\mathrm{Re}_h = 10\,595$	DNS [62]
Backward-facing step	$\mathrm{Re}_h = 36\,000$	EXP [63]
Curved backward-facing step	$\mathrm{Re}_h = 13\,700$	LES [64]
FAITH hill	$\mathrm{Re}_h = 500\,000$	EXP [65]
Ahmed body	$\mathrm{Re}_h = 768\,000$	EXP [66]

and curved) and two three-dimensional configurations (FAITH hill and Ahmed body). The high-fidelity data considered in this section are provided in Table V. Particular care was taken to include flows exhibiting turbulent behavior markedly different from those used to derive the SR correction. The resulting SR model shows good agreement with the reference data across a broad range of configurations. Nevertheless, its limitations become apparent in certain cases where the correction proves less effective.

1. Bump H38

Despite its similarity to the h42 bump used in the regression set, the h38 bump is characterized by a different level of curvature. Consequently, the pressure gradient and the flow separation are also different. The skin-friction coefficient C_f is well predicted by the SR model, with a reattachment point closely matching the reference data [Fig. 8(a)]. The pressure coefficient is also in agreement with the LES predictions [Fig. 8(b)]. This test case confirms the ability of the SR correction to improve the predictions for geometries similar to those used in the regression set.

Other levels of curvature were available in the LES data from Matai *et al.* [47]. The results were not included in the present study; however, a good agreement was obtained between the SR predictions and the reference data. For cases without flow separation, the SR model produces results comparable to those of the baseline SA model (i.e., the correction is inactive).

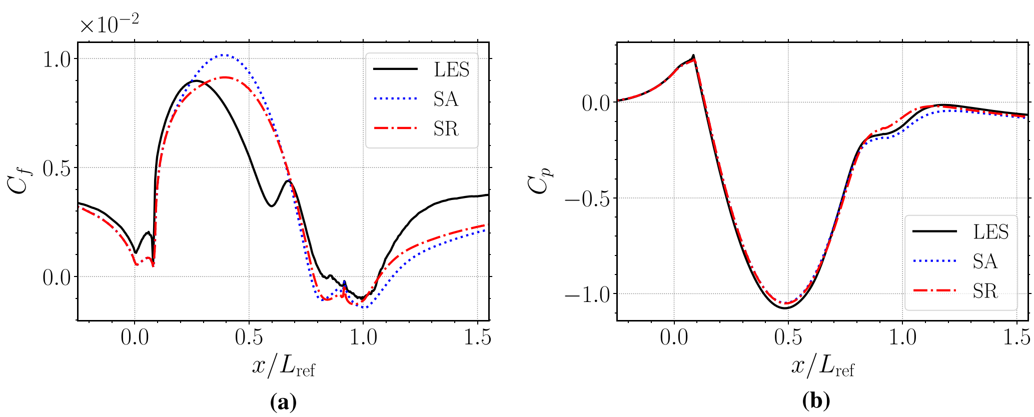


FIG. 8. Predictions with the symbolic expression β_{SR} on the bump H38. (a) Skin-friction coefficient C_f and (b) pressure coefficient C_p.

2. Square cylinder

The flow past a square cylinder is remarkably different from the flows included in the regression set. The flow exhibits a massive recirculation region downstream of the square cylinder caused by a strong separation occurring at its front corners. The estimation of the flow field around a bluff body is known to be challenging for RANS turbulence models [67]. The Reynolds number is based on the height of the square cylinder and on the freestream velocity.

The baseline SA model poorly predicts the recirculation region behind the square cylinder, producing a massive separation that extends far downstream, as shown by the streamlines in Figs. 9(a) and 9(b). The SR expression predicts the length of the recirculation region behind the square cylinder with good accuracy, although a small discrepancy with the reference data remains [Fig. 9(c)]. This nevertheless represents a significant improvement over the baseline SA model predictions. The velocity profiles presented in Figs. 9(d) and 9(e) further confirm this trend, showing closer agreement with the reference data than those obtained using the baseline SA model.

3. Periodic hill

The periodic-hill configuration considered in this study is characterized by a Reynolds number defined based on the hill height and the mean streamwise velocity at the hill crest. The flow exhibits a large recirculation region, for which linear eddy-viscosity models are known to suffer from significant inaccuracies. In particular, such models tend to overpredict the extent of the recirculation bubble. The velocity profiles are shown in Figs. 10(a) and 10(b). The SR-corrected predictions display improved agreement with the DNS results, particularly through a more accurate representation of the recirculation region. Figure 10(c) also highlights improvements in the skin-friction coefficient, especially within the region $3.0 < x/L_{\mathrm{ref}} < 7.0$.

4. Backward-facing step

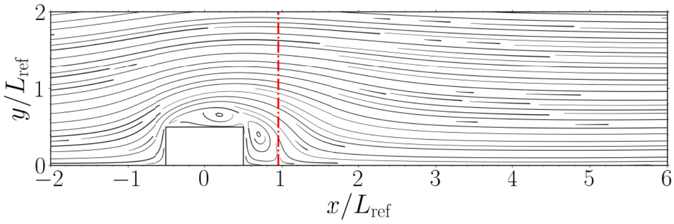
For the backward-facing step configuration, the Reynolds number is defined based on the step height and the freestream velocity. Predictions for the wall quantities demonstrate only minor variations between the baseline and the augmented model [Figs. 11(a) and 11(b)]. Although the baseline SA model yields predictions slightly closer to the experimental data, the proposed correction still performs comparably to the baseline.

The curved backward-facing step configuration was also examined, representing another challenging case where the baseline SA model fails to accurately predict the flow separation location. Similar to the standard backward-facing step, the SR correction exhibits only limited improvement for the curved configuration, indicating that the correction is less effective for this class of flows. Nevertheless, the SR expression slightly improves the predicted separation point [Fig. 12(a)] as well as the pressure coefficient distribution [Fig. 12(b)].

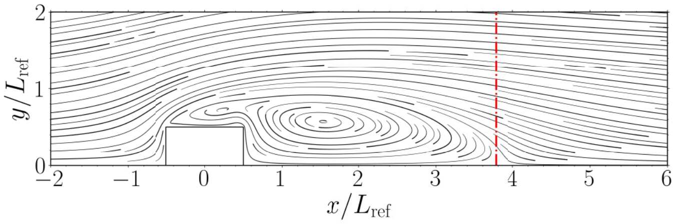
5. FAITH Hill

Moving beyond two-dimensional configurations, three-dimensional flow cases are now considered. The FAITH hill case involves a separated flow developing over a three-dimensional hill, with the Reynolds number defined based on the hill height and the freestream velocity. The mesh and OpenFoam case were taken from the *Benchmark challenge for machine learning in RANS turbulence modeling* [68]. This configuration represents a particularly demanding test for the SR correction, which was originally derived using two-dimensional cases. Three-dimensional flows inherently exhibit more complex turbulent structures, for which the correction was not specifically optimized.

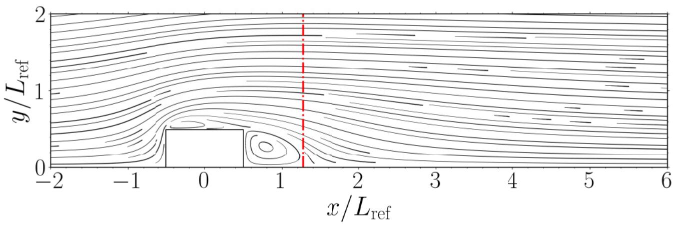
The velocity profiles are compared with available experimental data in Figs. 13(a) and 13(b). The SR correction is primarily active within the separated region downstream of the hill. In this region, the velocity profiles predicted with the SR correction show better agreement with the experimental data than those obtained with the baseline SA model, particularly for the vertical velocity component [Fig. 13(b)].



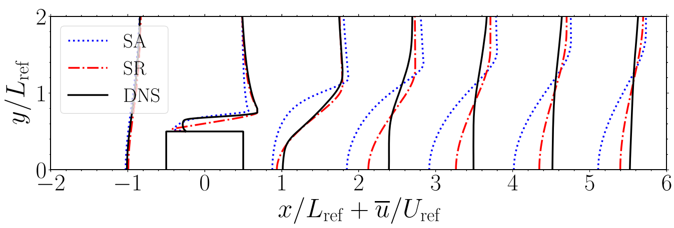
(a)



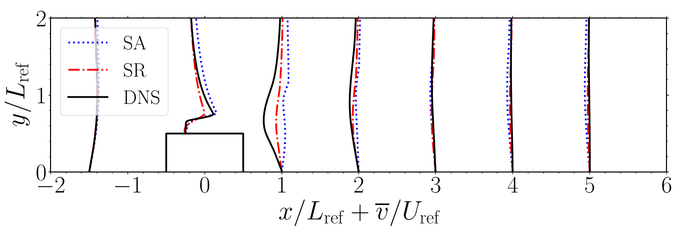
(b)



(c)



(d)



(e)

FIG. 9. Predictions with the symbolic expression β_{SR} on the square cylinder. (a) Streamlines-DNS [61], (b) streamlines-Spalart-Allmaras, (c) streamlines-symbolic expression β_{SR}, (d) $\bar{u}$-velocity profiles, and (e) $\bar{v}$-velocity profiles.

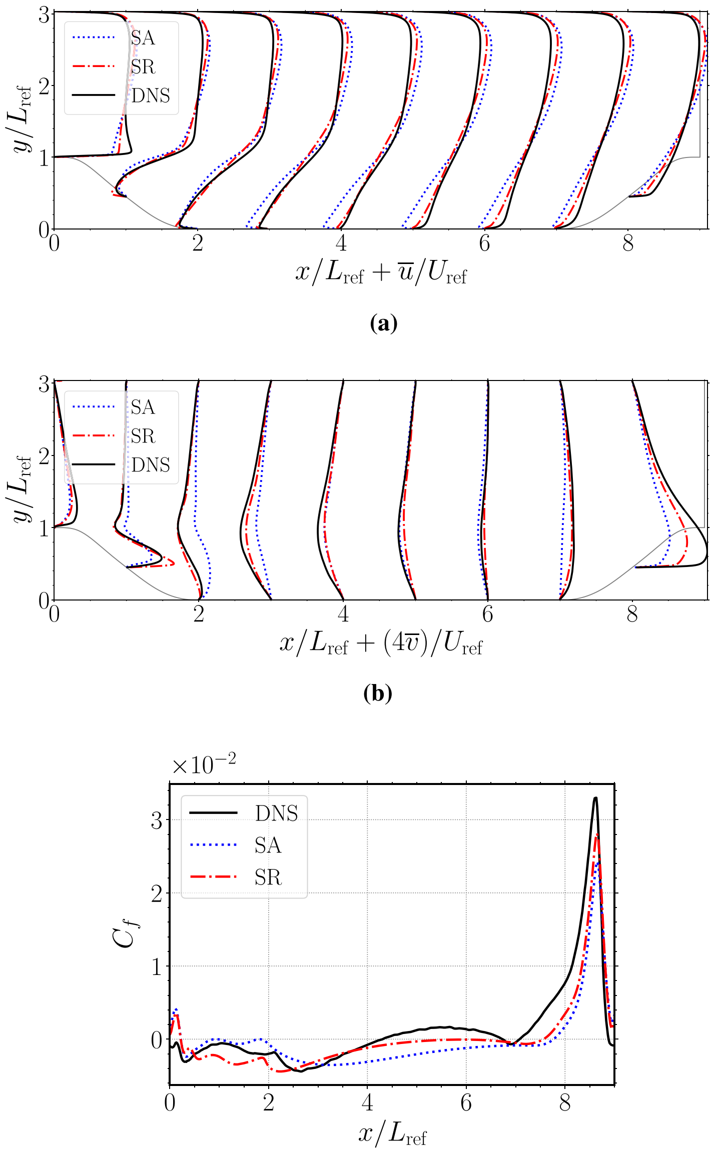


FIG. 10. Predictions with the symbolic expression β_{SR} on the periodic hill ($\mathrm{Re}_h = 10\,595$). (a) $\bar{u}$-velocity profiles, (b) $\bar{v}$-velocity profiles, and (c) skin-friction coefficient C_f.

The streamwise velocity contours in the symmetry plane are shown in Figs. 13(c), 13(d), and 13(e) for the experiment, the SA model, and the SR model, respectively. The recirculation region, indicated by the white contour lines, is noticeably smaller with the SR correction than with the baseline SA model. Although experimental data are not available near the wall, the SR correction clearly reduces the extent of the recirculation zone, which is overpredicted by the baseline model. Nevertheless, this reduction appears slightly excessive compared to the experimental observations.

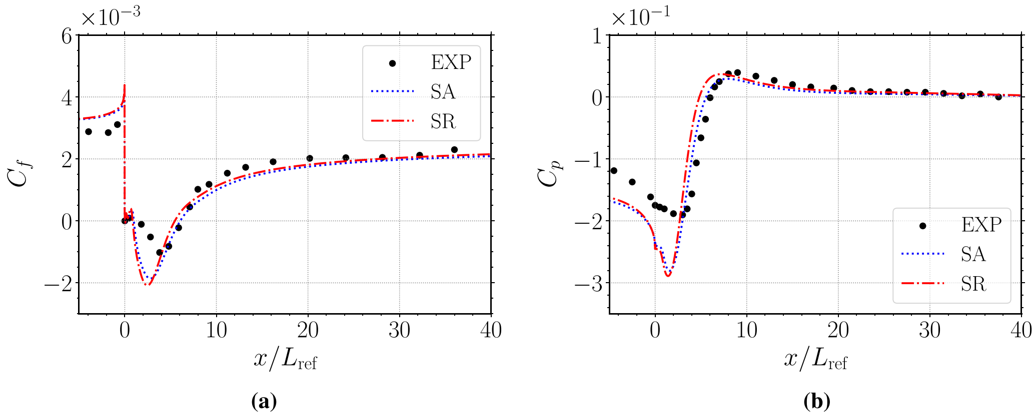


FIG. 11. Predictions with the symbolic expression β_{SR} on the backward-facing step. (a) Skin-friction coefficient C_f and (b) pressure coefficient C_p.

6. Ahmed body

The SR correction was also evaluated on another three-dimensional case: the Ahmed body. For this configuration, the Reynolds number is based on the model height and the freestream velocity. This test case represents a simplified car model that generates complex turbulent structures, including vortices shed from the upper corners of the slant and a large recirculation region developing behind the geometry. The slant of the Ahmed body is set at 25 degrees. The mesh and OpenFoam case were also taken from the *Benchmark challenge for machine learning in RANS turbulence modeling* [68].

The velocity profiles along the slant and in the downstream region of the model are presented in Figs. 14(a) and 14(b). The baseline SA model does not predict any flow separation over the slant, whereas other RANS models indicate a separated region [69]. Although experimental data in the near-wall region are limited, the SR correction appears to slightly degrade the predictions in this area. In the wake region, assessing the effectiveness of the correction based solely on the velocity profiles is difficult. At some measurement stations, the SA model aligns more closely with the experimental data, while at others, the SR correction provides better agreement, for both streamwise and vertical velocity components.

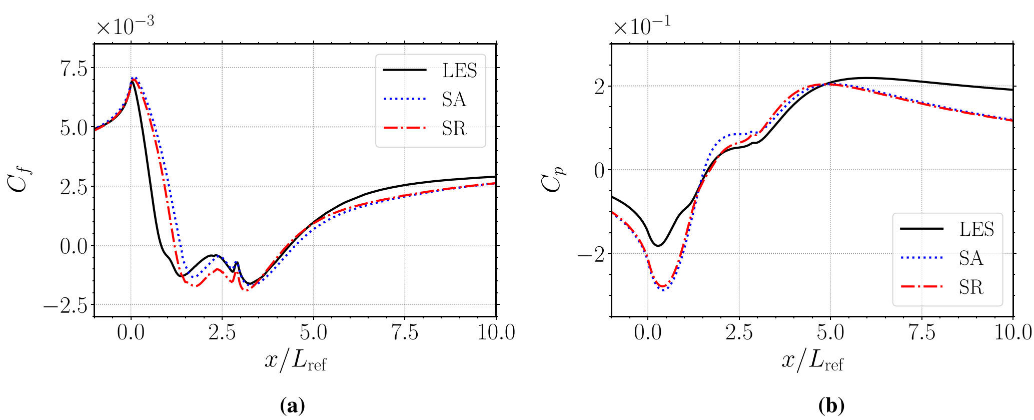


FIG. 12. Predictions with the symbolic expression β_{SR} on the curved backward-facing step. (a) Skin-friction coefficient C_f and (b) pressure coefficient C_p.

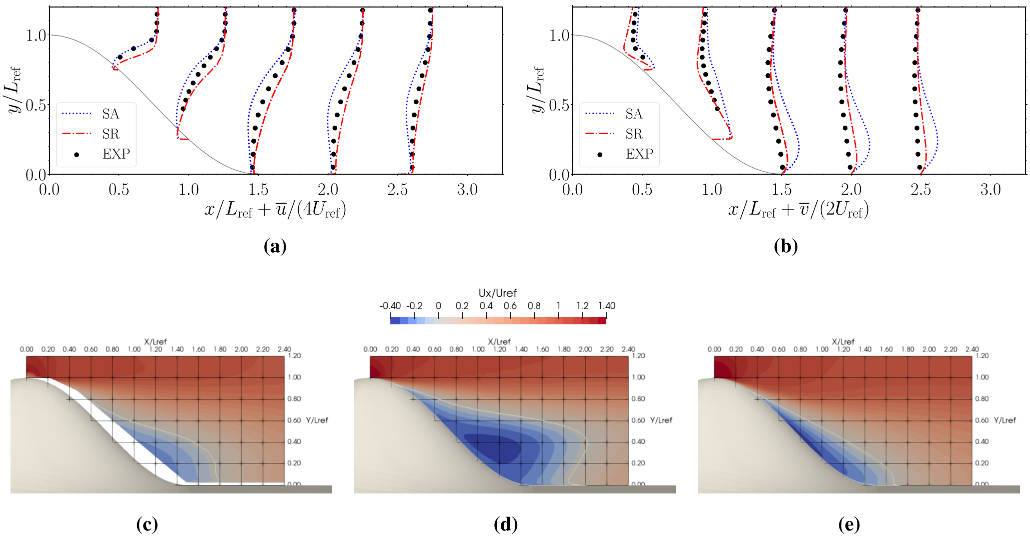


FIG. 13. Predictions with the Symbolic Expression β_{SR} on the FAITH hill. Streamwise velocity extracted on the *xy* Plane at $z/L_{\mathrm{ref}} = 0$. White contour lines indicate where the streamwise velocity is zero, identifying the recirculation region. (a) $\bar{u}$-velocity profiles, (b) $\bar{v}$-velocity profiles (c) $\bar{u}$-velocity-Exp. [65], (d) $\bar{u}$-velocity-SA, and (e) $\bar{u}$-velocity-SR.

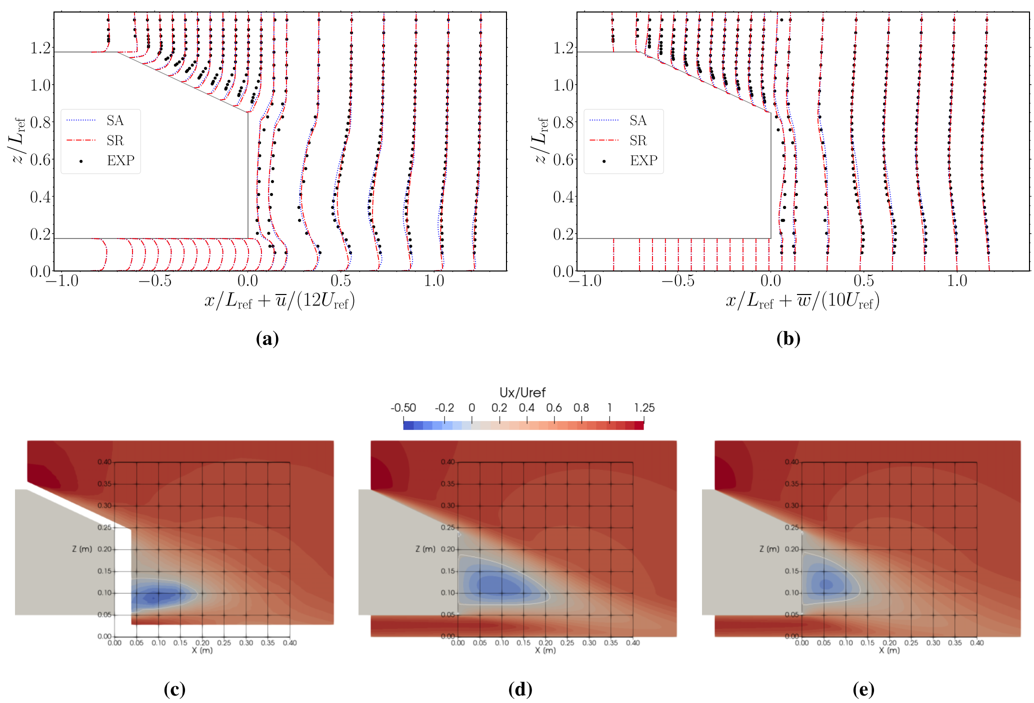


FIG. 14. Predictions with the symbolic expression β_{SR} on the Ahmed body. Streamwise velocity extracted on the *xz* Plane at $y/L_{\mathrm{ref}} = 0$. White contour lines indicate where the streamwise velocity is zero, identifying the recirculation region. (a) $\bar{u}$-velocity profiles, (b) $\bar{w}$-velocity profies (c) $\bar{u}$-velocity-Exp. [66], (d) $\bar{u}$-velocity-SA, and (e) $\bar{u}$-velocity-SR.

TABLE VI. Reduction in error achieved by the DA/SR predictions (two-dimensional flows for which high-fidelity numerical data were available) compared to the baseline SA model error. The errors are computed according to the reference data listed in Table V. The column UV refers to the reduction in L^2 error computed on the velocity field. The C_f and C_p columns refer to the reduction in root-mean-square error computed on the skin friction coefficient and on the pressure coefficient, respectively.

	$\|\bar{\mathbf{u}} - \bar{\mathbf{u}}_{\text{ref}}\|_{L^2}$		C_f		C_p	
	DA	SR	DA	SR	DA	SR
CDC	−59.98%	−26.92%	−47.88%	−19.47%	−69.62%	−17.03%
Bump H42	−60.56%	−39.66%	−43.40%	−22.50%	−63.63%	−32.13%
Hump	−6.07%	−19.84%	−27.32%	−8.96%	+0.60%	−18.44%
Bump H38		−10.00%		−17.67%		−11.54%
Square cylinder		−60.41%				
Periodic hill		−49.35%		−27.78%		
CBFS		−10.38%		−18.35%		−6.22%

Figures 14(c), 14(d), and 14(e) present the streamwise velocity contours in the symmetry plane for the experiment, the SA model, and the SR model, respectively. As in the FAITH hill case, the recirculation region behind the Ahmed Body is indicated by the white contour lines. The baseline SA model predicts a separated region that extends slightly too far downstream compared to the experiments, whereas the SR model significantly shortens the recirculation region. As a result, the SA model appears to align slightly better with the experimental data for this specific configuration. However, it should be emphasized that three-dimensional flows were not included in the training set, and incorporating such cases could potentially improve the SR model's predictions for this type of flow.

V. DISCUSSION

The SR correction was evaluated across different configurations, from two-dimensional flows similar to those used in the regression to three-dimensional flows that differ significantly from the regression dataset. An error analysis for the two-dimensional flows studied in the previous section is presented in Table VI. The error assessment was restricted to cases for which high-fidelity numerical data were available. The values reported correspond to the relative variation of the error with respect to the baseline SA model, defined as

$$\Delta E = \frac{E_{\text{model}} - E_{\text{SA}}}{E_{\text{SA}}} \times 100, \tag{26}$$

where E_{model} denotes the error obtained with either the DA or SR corrected model, and E_{SA} is the error associated with the baseline SA prediction. The error E corresponds to the L^2 norm of the velocity field, and to the root-mean-square error for the skin-friction and pressure coefficients C_f and C_p. Negative values therefore indicate a reduction of the error relative to the baseline SA model, while positive values correspond to a degradation of the prediction.

Overall, the results clearly show that the SR correction improves the SA predictions for most of the two-dimensional configurations considered. The correction is particularly effective for the cases included in the regression set, especially the converging-diverging channel (CDC), which is known to be a challenging configuration for RANS models [26,70]. For this case, the SR model reduces the velocity-field error by approximately 27% and significantly improves the prediction of wall quantities, with error reductions of about 19% for C_f and 17% for C_p. Similar improvements are observed for the bump h42 case and the NASA wall-mounted hump.

Interestingly, the SR correction also performs well for flows that were not included in the training set. For instance, the square cylinder case exhibits a reduction of more than 60% in the velocity-field error, while the periodic-hill case shows an error reduction of approximately 50%. These configurations differ substantially from those used to optimize the correction, involving sharp-edge separation and large recirculation regions. This indicates that the correction captures flow mechanisms that are relevant across a broad class of separated flows.

Finally, the standard and curved backward-facing steps illustrate cases where the correction does not significantly improve the baseline model. For the standard backward-facing step, the baseline SA model is slightly closer to experimental measurements; nevertheless, the proposed correction does not significantly degrade performance. For the curved backward-facing step, the SR model reduces the velocity error by about 10%, and yields improvements of roughly 18% and 6% for C_f and C_p, respectively. Although the gains are smaller than in other cases, the correction still leads to a consistent improvement over the baseline SA model for the curved configuration of the backward-facing step.

Regarding three-dimensional flows, the impact of the correction varies depending on the configuration considered. For the FAITH hill geometry, the correction reduces the extent of the recirculation region which was initially overpredicted by the baseline model. In particular, the modified model provides a more accurate representation of the velocity profiles in the wake. In contrast, for the Ahmed body configuration, the correction is significantly stronger and leads to an excessive reduction of the recirculation region. While the baseline SA model predicts a recirculation bubble that extends too far downstream, the SR correction shortens this region more than observed experimentally. Nevertheless, for both three-dimensional flows considered, the correction acts in the physically appropriate direction, as it consistently tends to reduce the size of the recirculation region predicted by the baseline model. It should be emphasized that the correction was not optimized for such configurations, since the regression dataset used during the training phase did not include any three-dimensional cases. Incorporating such configurations in future training datasets could therefore improve the robustness and predictive capability of the correction for complex three-dimensional separated flows.

VI. CONCLUSION

In this study, field inversion and symbolic regression were combined to derive an analytical correction to the Spalart-Allmaras production term, improving the model's predictive accuracy for separated flows. First, data assimilation was used to infer corrective fields relative to a wide range of cases presenting flow separation. Then, the analytical correction was determined by means of symbolic regression according to constraints on the operators and functionals used. The simplicity of an analytical expression facilitates its implementation in any existing CFD solver. A regression set, composed of the converging-diverging channel, the h42 bump, the NASA wall-mounted hump, and a flat-plate flow, was used to derive a symbolic correction that improves the baseline model's predictions without compromising its accuracy for wall-attached flows. Tests on additional two-dimensional flows, such as the h38 bump, square cylinder and periodic hill, confirm that combining field inversion with symbolic regression can generate an expression capable of addressing deficiencies in existing turbulence models across a broad range of flows. To further evaluate the performance of the SR correction in separated flows, three-dimensional test cases were also considered, including the FAITH hill and the Ahmed body. Depending on the configuration, the correction proved effective in improving predictions, while also highlighting certain limitations. Considerable effort was devoted to regress expressions using dimensionless input features while ensuring that all invariance properties were satisfied. The QCR form of the turbulent kinetic energy, k_{qcr}, was found to be a relevant input to identify regions where a correction to the production term in the SA equation is needed. However, the difficulty of finding a relevant quantity for its nondimensionalization led us to choose the square of a reference velocity, U_{ref}^2, thus violating Galilean invariance. We acknowledge that the applicability of the model is therefore restricted to wall-bounded flows. We believe that the improvements in the model predictions (for the intended flow configurations) justify its use. The study revealed a fundamental tradeoff between a correction tailored to a specific type of flow, which can achieve high accuracy, and a fully general, invariance-preserving correction, which offers broader applicability but delivers only marginal improvements for the flows considered. We prioritized a correction that is effective across a wide range of flows, rather than pursuing a fully general expression that provides little practical benefit. Nevertheless, the goal of developing an expression based on dimensionless inputs and satisfying all invariance properties remains important and should continue to guide future work. Because the Spalart-Allmaras model is based on just one equation for a viscosity-like quantity, achieving corrections that are both fully invariant and broadly applicable is particularly difficult. Access to additional quantities, such as the turbulent kinetic energy k and its dissipation ϵ, could enable new nondimensionalization strategies. In this context, the methodology that has been shown to be effective in the present study could be applied to other turbulence models (e.g., k-ϵ or k-ω) to enhance the generality of the corrections for both wall-attached and separated flows.

ACKNOWLEDGMENT

This work was supported by ONERA through the internal project TURBINA (Modélisation de la turbulence assistée par intelligence artificielle).

DATA AVAILABILITY

The proposed model implemented in OpenFoam will be shared on the author's GitHub page [71].

---

[1] A. W. Cary, J. Chawner, E. P. Duque, W. Gropp, W. L. Kleb, R. M. Kolonay, E. Nielsen, and B. Smith, CFD vision 2030 road map: Progress and perspectives, in *Proceedings of the AIAA Aviation Forum* (American Institute of Aeronautics and Astronautics, Reston, VA, 2021).

[2] B. E. Launder and D. B. Spalding, The numerical computation of turbulent flows, Comput. Methods Appl. Mech. Eng. **3**, 269 (1974).

[3] D. C. Wilcox, Formulation of the k-w turbulence model revisited, AIAA J. **46**, 2823 (2008).

[4] P. Spalart and S. Allmaras, A one-equation turbulence model for aerodynamic flows, in *Proceedings of the 30th Aerospace Sciences Meeting and Exhibit* (American Institute of Aeronautics and Astronautics, Reston, VA, 1992).

[5] C. L. Rumsey, Exploring a method for improving turbulent separated-flow predictions with k-omega models, NASA Technical report, NASA, Washington, D.C., 2009.

[6] K. Duraisamy, Perspectives on machine learning-augmented Reynolds-averaged and large eddy simulation models of turbulence, Phys. Rev. Fluids **6**, 050504 (2021).

[7] J. Ling, A. Kurzawski, and J. Templeton, Reynolds averaged turbulence modelling using deep neural networks with embedded invariance, J. Fluid Mech. **807**, 155 (2016).

[8] J.-X. Wang, J.-L. Wu, and H. Xiao, Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data, Phys. Rev. Fluids **2**, 034603 (2017).

[9] P. S. Volpiani, R. F. Bernardini, and L. Franceschini, Neural network-based eddy-viscosity correction for RANS simulations of flows over bi-dimensional bumps, Int. J. Heat Fluid Flow **97**, 109034 (2022).

[10] P. S. Volpiani, Are random forests better suited than neural networks to augment RANS turbulence models? Int. J. Heat Fluid Flow **107**, 109348 (2024).

[11] E. J. Parish and K. Duraisamy, A paradigm for data-driven predictive modeling using field inversion and machine learning, J. Comput. Phys. **305**, 758 (2016).

[12] A. P. Singh, S. Medida, and K. Duraisamy, Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils, AIAA J. **55**, 2215 (2017).

[13] A. Ferrero, A. Iollo, and F. Larocca, Field inversion for data-augmented RANS modelling in turbomachinery flows, Comput. Fluids **201**, 104474 (2020).

[14] C. Yan, H. Li, Y. Zhang, and H. Chen, Data-driven turbulence modeling in separated flows considering physical mechanism analysis, Int. J. Heat Fluid Flow **96**, 109004 (2022).

[15] A. S. Cato, P. S. Volpiani, V. Mons, O. Marquet, and D. Sipp, Comparison of different data-assimilation approaches to augment RANS turbulence models, Comput. Fluids **266**, 106054 (2023).

[16] P. S. Volpiani, M. Meyer, L. Franceschini, J. Dandois, F. Renac, E. Martin, O. Marquet, and D. Sipp, Machine learning-augmented turbulence modeling for RANS simulations of massively separated flows, Phys. Rev. Fluids **6**, 064607 (2021).

[17] F. Jäckel, A closed-form correction for the Spalart-Allmaras turbulence model for separated flows, AIAA J. **61**, 2319 (2023).

[18] Z. He, C. Wu, and Y. Zhang, A field inversion and symbolic regression enhanced Spalart-Allmaras model for airfoil stall prediction, Phys. Fluids **36**, 065111 (2024).

[19] R. McConkey, E. Yee, and F. S. Lien, Deep structured neural networks for turbulence closure modeling, Phys. Fluids **34**, 035110 (2022).

[20] M. A. Cruz, R. L. Thompson, L. E. Sampaio, and R. D. Bacchi, The use of the Reynolds force vector in a physics informed machine learning approach for predictive turbulence modeling, Comput. Fluids **192**, 104258 (2019).

[21] C. L. Rumsey, G. N. Coleman, and L. Wang, In search of data-driven improvements to RANS models applied to separated flows, in *Proceedings of the AIAA SCITECH Forum* (American Institute of Aeronautics and Astronautics, Reston, VA, 2021).

[22] M. Cranmer, Interpretable machine learning for science with PySR and SymbolicRegression.jl, arXiv:2305.01582v3.

[23] W. Tenachi, R. Ibata, and F. I. Diakogiannis, Deep symbolic regression for physics guided by units constraints: Toward the automated discovery of physical laws, Astrophys. J. **959**, 99 (2023).

[24] P. Spalart, An old-fashioned framework for machine learning in turbulence modeling, arXiv:2308.00837.

[25] J. Weatheritt and R. Sandberg, The development of algebraic stress models using a novel evolutionary algorithm, Int. J. Heat Fluid Flow **68**, 298 (2017).

[26] M. Schmelzer, R. P. Dwight, and P. Cinnella, Discovery of algebraic Reynolds-stress models using sparse symbolic regression, Flow Turbul. Combust. **104**, 579 (2020).

[27] T. Buchanan, M. Lăcătuş, A. West, and R. P. Dwight, Data-driven RANS closures using a relative importance term analysis based classifier for 2D and 3D separated flows, Comput. Fluids **305**, 106899 (2026).

[28] C. Wu and Y. Zhang, Enhancing the shear-stress-transport turbulence model with symbolic regression: A generalizable and interpretable data-driven approach, Phys. Rev. Fluids **8**, 084604 (2023).

[29] C. Wu, S. Zhang, and Y. Zhang, Development of a generalizable data-driven turbulence model: Conditioned field inversion and symbolic regression, AIAA J. **63**, 687 (2025).

[30] B. An, Z. Jiang, C. Yi, M. Han, and C. Yan, Improvement of the SST model for streamline curvature effect based on field inversion and symbolic regression, Aerosp. Sci. Technol. **157**, 109828 (2025).

[31] H. Liao, X. Sun, Y. Liu, and W. Zhang, Data-driven turbulence modeling: A mutually coupled framework for symbolic regression and data assimilation, Phys. Fluids **37**, 075211 (2025).

[32] A. Crivellini, V. D'Alessandro, and F. Bassi, A Spalart-Allmaras turbulence model implementation in a discontinuous Galerkin solver for incompressible flows, J. Comput. Phys. **241**, 388 (2013).

[33] D. A. Ham, P. H. Kelly, L. Mitchell, C. Cotter, R. C. Kirby, K. Sagiyama, N. Bouziani, S. Vorderwuelbecke, T. Gregory, J. Betteridge, D. R. Shapero, R. Nixon-Hill, C. Ward, P. E. Farrell, P. D. Brubeck, I. Marsden, T. H. Gibson, M. Homolya, T. Sun, A. T. McRae *et al.*, Firedrake User Manual, Imperial College London and University of Oxford and Baylor University and University of Washington, 1st ed., 2023.

[34] M. S. Alnæs, A. Logg, K. B. Ølgaard, M. E. Rognes, and G. N. Wells, Unified form language: A domain-specific language for weak formulations of partial differential equations, ACM Trans. Math. Softw. **40**, 1 (2014).

[35] Y. Bao, D. Zhou, C. Huang, Q. Wu, and X. Chen, Numerical prediction of aerodynamic characteristics of prismatic cylinder by finite element method with Spalart-Allmaras turbulence model, Comput. Struct. **89**, 325 (2011).

[36] M. Olshanskii, G. Lube, T. Heister, and J. Löwe, GRAD-DIV stabilization and subgrid pressure models for the incompressible Navier-Stokes equations, Comput. Methods Appl. Mech. Eng. **198**, 3975 (2009).

[37] P. R. Amestoy, I. S. Duff, J.-Y. L'Excellent, and J. Koster, A fully asynchronous multifrontal solver using distributed dynamic scheduling, SIAM J. Matrix Anal. Appl. **23**, 15 (2001).

[38] P. R. Amestoy, A. Buttari, J.-Y. L'Excellent, and T. Mary, Performance and scalability of the block low-rank multifrontal factorization on multicore architectures, ACM Trans. Math. Softw. **45**, 1 (2019).

[39] M. L. Shur, M. K. Strelets, A. K. Travin, and P. R. Spalart, Turbulence modeling in rotating and curved channels: Assessing the Spalart-Shur correction, AIAA J. **38**, 784 (2000).

[40] B. Fanizza, P. Stefanin Volpiani, F. Renac, E. Martin, and D. Sipp, Adjoint-based optimization for non-linear inverse problems with high-order discretization of the compressible RANS equations, Appl. Math. Modell. **142**, 115984 (2025).

[41] G. K. Kenway, C. A. Mader, P. He, and J. R. Martins, Effective adjoint approaches for computational fluid dynamics, Prog. Aerosp. Sci. **110**, 100542 (2019).

[42] S. K. Mitusch, S. W. Funke, and J. S. Dokken, dolfin-adjoint 2018.1: Automated adjoints for FEniCS and Firedrake, J. Open Source Software **4**, 1292 (2019).

[43] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization, ACM Trans. Math. Softw. **23**, 550 (1997).

[44] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Comput. **16**, 1190 (1995).

[45] L. Franceschini, D. Sipp, and O. Marquet, Mean-flow data assimilation based on minimal correction of turbulence models: Application to turbulent high Reynolds number backward-facing step, Phys. Rev. Fluids **5**, 094603 (2020).

[46] M. Marquillie, U. Ehrenstein, and J.-P Laval, Instability of streaks in wall turbulence with adverse pressure gradient, J. Fluid Mech. **681**, 205 (2011).

[47] R. Matai and P. Durbin, Large-eddy simulation of turbulent flow over a parametric set of bumps, J. Fluid Mech. **866**, 503 (2019).

[48] A. Uzun and M. R. Malik, Large-eddy simulation of flow over a wall-mounted hump with separation and reattachment, AIAA J. **56**, 715 (2018).

[49] M. Cranmer, A. Sanchez-Gonzalez, P. Battaglia, R. Xu, K. Cranmer, D. Spergel, and S. Ho, Discovering symbolic models from deep learning with inductive biases, arXiv:2006.11287.

[50] P. R. Spalart, Philosophies and fallacies in turbulence modeling, Prog. Aerosp. Sci. **74**, 1 (2015).

[51] B. D. Tracey, K. Duraisamy, and J. J. Alonso, A machine learning strategy to assist turbulence model development, in *Proceedings of the 53rd AIAA Aerospace Sciences Meeting* (American Institute of Aeronautics and Astronautics, Kissimmee, FL, 2015).

[52] H. J. Herring and G. L. Mellor, Two methods of calculating turbulent boundary layer behavior based on numerical solutions of the equations of motion (Stanford, CA 1969), https://ntrs.nasa.gov/citations/19700036233.

[53] P. R. Spalart, S. Deck, M. L. Shur, K. D. Squires, M. K. Strelets, and A. Travin, A new version of detached-eddy simulation, resistant to ambiguous grid densities, Theor. Comput. Fluid Dyn. **20**, 181 (2006).

[54] M. Mani, D. Babcock, C. Winkler, and P. Spalart, Predictions of a supersonic turbulent flow in a square duct, in *Proceedings of the 51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition* (AIAA, Reston, VA, 2013).

[55] V. Srivastava, C. L. Rumsey, G. N. Coleman, and L. Wang, On generalizably improving RANS predictions of flow separation and reattachment, in *Proceedings of the AIAA SCITECH Forum* (American Institute of Aeronautics and Astronautics, Orlando, FL, 2024).

[56] Y. Fang, Y. Zhao, F. Waschkowski, A. S. H. Ooi, and R. D. Sandberg, Toward more general turbulence models via multicase computational-fluid-dynamics-driven training, AIAA J. **61**, 2100 (2023).

[57] R. B. Langtry and F. R. Menter, Correlation-based transition modeling for unstructured parallelized computational fluid dynamics codes, AIAA J. **47**, 2894 (2009).

[58] K. Wieghardt and W. Tillmann, On the turbulent friction layer for rising pressure, Technical Memorandum No. 1314 (National Advisory Committee for Aeronautics, Washington, 1951).

[59] A. M. Savill, in *Further Progress in the Turbulence Modelling of By-Pass Transition, Engineering Turbulence Modelling and Experiments*, edited by W. Rodi and F. Martelli (Elsevier, Oxford, UK, 1993), p. 583.

[60] C. Wu and Y. Zhang, Field inversion and machine learning based on the rubber-band Spalart-Allmaras model, Theor. Appl. Mech. Lett. **15**, 100564 (2025).

[61] L. Franceschini, D. Sipp, O. Marquet, J. Moulin, and J. Dandois, Identification and reconstruction of high-frequency fluctuations evolving on a low-frequency periodic limit cycle: Application to turbulent cylinder flow, J. Fluid Mech. **942**, A28 (2022).

[62] X. Gloerfelt and P. Cinnella, Large eddy simulation requirements for the flow over periodic hills, Flow Turbul. Combust. **103**, 55 (2019).

[63] D. M. Driver and H. L. Seegmiller, Features of a reattaching turbulent shear layer in divergent channel flow, AIAA J. **23**, 163 (1985).

[64] Y. Bentaleb, S. Lardeau, and M. A. Leschziner, Large-eddy simulation of turbulent boundary layer separation from a rounded step, J. Turbul. **13**, N4 (2012).

[65] J. Bell, J. Heineck, G. Zilliac, R. Mehta, and K. Long, Surface and flow field measurements on the FAITH hill model, in *Proceedings of the 50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition* (American Institute of Aeronautics and Astronautics, Reston, VA, 2012).

[66] H. Lienhart, C. Stoots, and S. Becker, in *Flow and Turbulence Structures in the Wake of a Simplified Car Model (Ahmed Modell)*, edited by S. Wagner, U. Rist, H.-J. Heinemann, and R. Hilbig, New Results in Numerical and Experimental Fluid Mechanics III Vol. 77 (Springer, Berlin, Heidelberg, 2002), p. 323.

[67] J. P. Huijing, R. P. Dwight, and M. Schmelzer, Data-driven RANS closures for three-dimensional flows around bluff bodies, Comput. Fluids **225**, 104997 (2021).

[68] R. McConkey, T. Buchanan, T. Smidt, A. Bodner, R. Dwight, and P. Cinnella, The closure challenge: A benchmark task for machine learning in turbulence modelling, arXiv:2603.28884.

[69] E. Guilmineau, Computational study of flow around a simplified car body, J. Wind Eng. Ind. Aerodyn. **96**, 1207 (2008).

[70] M. Oulghelou, S. Cherroud, X. Merle, and P. Cinnella, Machine-learning-assisted blending of data-driven turbulence models, Flow Turbul. Combust. **115**, 1095 (2025).

[71] https://github.com/p-bataillie/2026-PRF-SA-FISR.