

Information-Based Symbolic Regression of Local Closures for RANS Transition Modeling

Javier Capel-Jorquera

j.capel@gmail.com

Universidad Politécnica de Madrid

José I. Cardesa

ONERA

M. Chávez-Módena

Universidad Politécnica de Madrid

E. Valero

Universidad Politécnica de Madrid

L. M. González

Universidad Politécnica de Madrid

Research Article

Keywords: transition, turbulence modeling, information theory, symbolic regression

Posted Date: September 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10535256/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Information-Based Symbolic Regression of Local Closures for RANS Transition Modeling

Javier Capel-Jorquera^{1,2*}, José I. Cardesa³,
M. Chávez-Módena^{1,2}, E. Valero^{1,2}, L. M. González^{1,2}

^{1*}NuMath, CEHINAV, Universidad Politécnica de Madrid, Pl. del Cardenal Cisneros, 3, Moncloa, Madrid, 28040, Spain.

²Center for Computational Simulation, Universidad Politécnica de Madrid, Campus de Montegancedo, Madrid, 28660, Spain.

³DMPE, STAT, ONERA, Centre de Toulouse 2, Avenue Marc Pégélin, Toulouse, 31055, France.

*Corresponding author(s). E-mail(s): j.capel@gmail.com;

Contributing authors: jose.cardesa-duenas@onera.fr; m.chavez@upm.es;
eusebio.valero@upm.es; leo.gonzalez@upm.es;

Abstract

Local-correlation transition models enable laminar–turbulent transition prediction in Reynolds-averaged Navier–Stokes simulations by replacing non-local boundary-layer quantities with local surrogate variables. This work proposes an information-based symbolic-regression framework to construct compact and interpretable surrogates for the Pohlhausen pressure-gradient parameter, λ_θ , and the momentum-thickness Reynolds number, Re_θ . A Falkner–Skan database is used to identify informative local dimensionless features through the IT- π methodology and to derive explicit expressions for λ_{θ_L} and Re_{θ_L} . The resulting closures reconstruct the corresponding integral quantities substantially more accurately than the baseline expressions. Candidate closures are selected according to reconstruction accuracy, algebraic complexity, and wall-normal activation behavior. They are then implemented in a one-equation intermittency model and assessed using the ERCOFTAC T3 transitional flat-plate series. The results are case dependent, with improvements for several bypass and pressure-gradient cases and deterioration for two low-turbulence-intensity cases. Although recalibration of the remaining transition-model correlations may be required, the proposed closures provide a more interpretable representation of the local boundary-layer state and offer a promising route toward compact, data-informed, and physically transparent local-correlation transition models.

Keywords: transition, turbulence modeling, information theory, symbolic regression

1 Introduction

The prediction of laminar–turbulent transition remains one of the major unresolved challenges for the routine use of Reynolds-averaged Navier–Stokes (RANS) simulations in aerodynamic design. Transition affects skin friction, separation, heat transfer and, consequently, global aerodynamic coefficients such as lift, drag, and thrust, where applicable. Its correct prediction is therefore relevant for applications ranging from ship propellers and wind-turbine blades to turbomachinery, rotorcraft dynamics and high-lift configurations. Despite sustained progress in computational fluid dynamics, transition modeling has been repeatedly identified as a critical capability gap for reliable predictive CFD, particularly when simulations are required to remain affordable for industrial design cycles [1].

Within the RANS framework, local-correlation transition models (LCTMs) have become one of the most practical approaches for engineering transition prediction. Their main advantage is that empirical transition correlations are reformulated in terms of local variables, making the models compatible with unstructured grids, parallel solvers and complex geometries [2, 3]. Previous numerical assessments have shown that low-Reynolds-number aerodynamic predictions can be highly sensitive to the selected transition closure [4]. Among the models considered in this class, intermittency-based local-correlation transition models have attracted particular attention because they combine relatively low computational cost with compatibility with general-purpose CFD solvers. The γ – Re_{θ_t} model [5], the one-equation intermittency model [6], and the algebraic intermittency model [7] all rely, in different ways, on local approximations of integral boundary-layer quantities that are not directly available in a RANS computation. These models differ in complexity and in the degree to which transition physics is represented explicitly, as qualitatively illustrated in Fig. 1.

A common feature of the γ – Re_{θ_t} , one-equation γ , and algebraic γ transition models is the need to represent the boundary layer state and to account for pressure-gradient effects in the transition-onset criterion. This is commonly done through a local approximation of the momentum-thickness Reynolds number, Re_{θ} , and through the Pohlhausen pressure-gradient parameter, λ_{θ} .

The difficulty is that those variables are not directly available as local RANS variables in a general CFD solver. As a result, transition models rely on local surrogate expressions, here denoted as Re_{θ_L} and λ_{θ_L} . These expressions are physically motivated but necessarily empirical, and their behavior is coupled to the remaining correlations, limiters and activation functions of the transition model.

This coupling raises an important modeling question. Improving the isolated approximation of a physically meaningful quantity, such as Re_{θ} or λ_{θ} , does not necessarily guarantee a proportional improvement in the final transition prediction. The complete model architecture may amplify, damp or saturate the effect of a modified closure ingredient. Therefore, the assessment of a new local expression should be

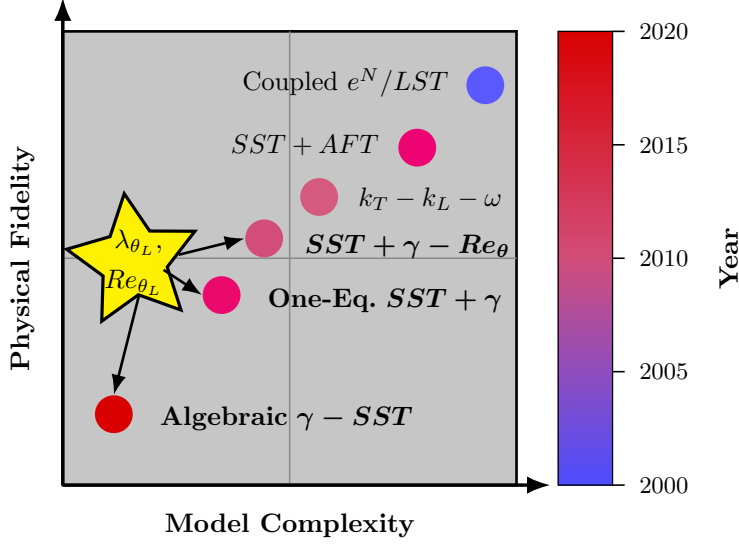


Fig. 1: Qualitative comparison of transition models according to their physical fidelity and model complexity. Models: Coupled e^N/LST [8], $SST + AFT$ [9], $k_T - k_L - \omega$ [10], $SST + \gamma - Re_\theta$ [5], One-Eq. $SST + \gamma$ [6], Algebraic $\gamma - SST$ [7].

performed at two levels: first, by verifying whether the expression provides a more accurate reconstruction of the target physical quantity; and second, by testing whether the improved local reconstruction has a measurable impact after insertion into the complete transition model.

Data-driven methods provide new opportunities to revisit such empirical closures. In turbulence modeling, machine-learning approaches have been used to infer model discrepancies, augment closure terms and construct nonlinear constitutive relations from high-fidelity data [11, 12]. However, many machine-learning closures are difficult to interpret and may be challenging to analyze, constrain, or implement robustly in production CFD solvers [13]. This has motivated increasing interest in sparse and symbolic-regression approaches, where the learned model is expressed as an explicit mathematical equation [14–17]. In transition modeling, related ideas are beginning to appear, including the symbolic-regression-based amplification-factor transport model of Pascal [18]. Such approaches are particularly attractive for transition modeling because existing models already contain physically motivated local correlations; the goal is not necessarily to replace the full model by a black-box predictor, but to improve selected empirical components while preserving interpretability and ease of implementation.

The present work follows this targeted model-enhancement philosophy. Instead of constructing a new transition model from scratch, we focus on the physically-motivated local closures used inside algebraic and transport-equation intermittency models. A two-dimensional, Falkner–Skan boundary-layer database is used as a canonical data

set in which the exact Re_θ and λ_θ variables can be evaluated. Dimensional analysis and information-content variable selection are then used to identify the relevant dimensionless input structure, and symbolic regression is applied to obtain a compact expression for the target variable. The resulting expression is compared against the baseline closure at the data level and then at the complete model level by implementing the expressions in OpenFOAM and assessing how well they perform for the canonical T3 series of flat plate experiments [19].

This controlled replacement allows the influence of the new closures to be isolated. The objective is therefore twofold: to determine whether symbolic regression can improve the local reconstruction of the target variable, and to establish when such an improvement is dynamically active within the complete transition-model architecture.

2 Theoretical and modeling framework

2.1 Local-correlation transition models

Local-correlation transition models (LCTMs) represent laminar–turbulent transition using quantities available locally in a RANS computation. This avoids the explicit evaluation of boundary-layer integral quantities along wall-normal lines, which is generally impractical for unstructured grids, parallel solvers, and complex geometries. Instead, LCTMs introduce local surrogate variables intended to reproduce the relevant integral quantities in the region where transition onset is activated.

The present work focuses on local approximations of the momentum-thickness Reynolds number,

$$Re_\theta = \frac{U_e \theta}{\nu}, \quad (1)$$

and the Pohlhausen pressure-gradient parameter,

$$\lambda_\theta = \frac{\theta^2}{\nu} \frac{dU_e}{ds}, \quad (2)$$

where θ is the momentum thickness, U_e is the boundary-layer edge velocity, ν is the kinematic viscosity, and s is the streamwise coordinate. Since θ and U_e are not generally available as local RANS variables, LCTMs replace these quantities with local surrogate variables, denoted here by Re_{θ_L} and λ_{θ_L} .

Intermittency-based LCTMs introduce an intermittency variable γ , which represents the fraction of time during which the flow at a given location is turbulent [20, 21]. Values close to $\gamma = 0$ and $\gamma = 1$ correspond to predominantly laminar and turbulent flow, respectively. Within the model, γ controls the activation of turbulent kinetic-energy production. Transition onset is therefore determined through local functions that compare Re_{θ_L} with a critical Reynolds number Re_{θ_c} , whose value depends on the local turbulence intensity and pressure-gradient surrogate.

2.1.1 One-equation intermittency model

The one-equation intermittency model of Menter et al. [6] is used as the baseline transition model. It solves the following transport equation for intermittency:

$$\frac{\partial(\rho\gamma)}{\partial t} + \frac{\partial(\rho u_j \gamma)}{\partial x_j} = \frac{\partial}{\partial x_j} \left[\left(\mu + \frac{\mu_t}{\sigma_\gamma} \right) \frac{\partial \gamma}{\partial x_j} \right] + P_\gamma - E_\gamma, \quad (3)$$

where $\sigma_\gamma = 1.0$. The production and destruction terms are

$$P_\gamma = F_{\text{length}} \rho S \gamma (1 - \gamma) F_{\text{onset}}, \quad (4)$$

$$E_\gamma = c_{a2} \rho \Omega \gamma F_{\text{turb}} (c_{e2} \gamma - 1), \quad (5)$$

with $F_{\text{length}} = 100$, $c_{a2} = 0.06$, and $c_{e2} = 50$ in the baseline formulation. Here, $S = \sqrt{2S_{ij}S_{ij}}$ and $\Omega = \sqrt{2\Omega_{ij}\Omega_{ij}}$ are the strain-rate and rotation-rate magnitudes, respectively.

The onset function is primarily controlled by the ratio between the local and critical momentum-thickness Reynolds numbers,

$$\mathcal{A} = \frac{Re_{\theta_L}}{Re_{\theta_c}}, \quad (6)$$

and is defined as

$$F_{\text{onset}} = \max(F_{\text{onset},2} - F_{\text{onset},3}, 0), \quad (7)$$

$$F_{\text{onset},2} = \min(F_{\text{onset},1}, 2), \quad (8)$$

$$F_{\text{onset},3} = \max \left[1 - \left(\frac{Re_T}{3.5} \right)^3, 0 \right], \quad (9)$$

$$F_{\text{onset},1} = \frac{Re_V}{C_{\text{onset},1} Re_{\theta_c}} = \frac{Re_{\theta_L}}{Re_{\theta_c}}, \quad (10)$$

where

$$Re_T = \frac{\rho k}{\mu \omega}, \quad F_{\text{turb}} = \exp \left[- \left(\frac{Re_T}{2} \right)^4 \right]. \quad (11)$$

In the baseline model, the local momentum-thickness Reynolds number is approximated through the vorticity Reynolds number,

$$Re_{\theta_L}^{\text{base}} = \frac{Re_V}{C_{\text{onset},1}}, \quad Re_V = \frac{d_w^2 S}{\nu}, \quad C_{\text{onset},1} = 2.193, \quad (12)$$

where d_w is the wall distance. The use of Re_V is motivated by a simple regression problem using Blasius boundary layer data. However, as presented in [2, 6], this approximation starts to deviate for Falkner-Skan data.

The critical momentum-thickness Reynolds number is expressed as a function of the local turbulence intensity and pressure-gradient surrogate:

$$Re_{\theta_c}(Tu_L, \lambda_{\theta_L}) = C_{TU1} + C_{TU2} \exp[-C_{TU3} Tu_L F_{PG}(\lambda_{\theta_L})], \quad (13)$$

where

$$Tu_L = \min\left(100 \frac{\sqrt{2k/3}}{\omega d_w}, 100\right), \quad (14)$$

and $C_{TU1} = 100$, $C_{TU2} = 1000$, $C_{TU3} = 1$.

The baseline pressure-gradient closure λ_{θ_L} is obtained by approximating λ_θ from Eq. (2) using two-dimensional flow assumptions as

$$\lambda_\theta = \frac{\theta^2}{\nu} \frac{dU_e}{dx} = -\frac{\theta^2}{\nu} \frac{dV_e}{dy}, \quad (15)$$

where V_e is the edge wall normal velocity component, y is the wall-normal coordinate in the freestream. For a general geometry, the wall normal derivative is computed as

$$\frac{dV_e}{dy} \equiv \nabla(\vec{n} \cdot \vec{V}_e) \cdot \vec{n}. \quad (16)$$

Technically speaking, the relation (15) is only correct for incompressible two-dimensional flow and requires components of the edge velocity that must be evaluated outside the boundary layer. However, the expression can serve as a physically motivated template for a local surrogate. This surrogate is obtained by replacing the momentum thickness θ with the wall distance d_w , and the edge-normal velocity gradient is replaced by the local wall-normal velocity gradient. The rationale is that the onset function is activated in the interior of the boundary layer, approximately at half the boundary-layer height, where d_w scales with the local boundary-layer thickness and can be calibrated to reproduce the target Pohlhausen parameter. The resulting local pressure-gradient approximation is

$$\lambda_{\theta_L}^{\text{base}} = -7.57 \cdot 10^{-3} \frac{dv}{dy} \frac{d_w^2}{\nu} + 0.0128. \quad (17)$$

For numerical robustness, λ_{θ_L} is bounded as:

$$\lambda_{\theta_L} \leftarrow \min[\max(\lambda_{\theta_L}, -1.0), 1.0], \quad (18)$$

The pressure-gradient correction appearing in Eq. (13) is

$$F_{PG}(\lambda_{\theta_L}) = \begin{cases} \min(1 + C_{PG1} \lambda_{\theta_L}, C_{PG1}^{\text{lim}}), & \lambda_{\theta_L} \geq 0, \\ \min(1 + C_{PG2} \lambda_{\theta_L} + C_{PG3} \min(\lambda_{\theta_L} + 0.0681, 0), C_{PG2}^{\text{lim}}), & \lambda_{\theta_L} < 0, \end{cases} \quad (19)$$

followed by the limiter

$$F_{PG} \leftarrow \max(F_{PG}, 0). \quad (20)$$

The corresponding constants are: $C_{PG1} = 14.68$, $C_{PG2} = -7.34$, $C_{PG3} = 0$, $C_{PG1}^{\text{lim}} = 1.5$, $C_{PG2}^{\text{lim}} = 3$.

Intermittency is coupled to the SST turbulence model by modifying the production and destruction terms of the turbulent kinetic-energy equation as functions of γ , following Menter et al. [6]. The original SST coupling, additional low-turbulence production limiter, and remaining model constants are retained unchanged and are not repeated here.

The intermittency boundary conditions are

$$\gamma_{\text{inlet}} = 1, \quad \left. \frac{\partial \gamma}{\partial n} \right|_{\text{wall}} = 0. \quad (21)$$

The present work replaces the baseline local closures $Re_{\theta_L}^{\text{base}}$ and $\lambda_{\theta_L}^{\text{base}}$ while retaining the remaining correlations, onset functions, source-term structure, limiters, and SST coupling.

2.2 Falkner–Skan database

The Falkner–Skan family provides canonical self-similar solutions for two-dimensional incompressible laminar boundary layers subject to power-law edge-velocity distributions [22]. They represent the boundary layers that are formed over corners characterized by the angle $\beta(\pi/2)$. These solutions are well suited to the present study because the integral target quantities Re_θ and λ_θ , together with the corresponding local velocity fields and gradients, can be evaluated directly from the similarity solution.

The edge velocity is prescribed as

$$U_e(x) = Ax^m, \quad (22)$$

where A is a constant and m determines the streamwise acceleration of the external flow. The Falkner–Skan parameter β is related to m by

$$\beta = \frac{2m}{m+1}, \quad m = \frac{\beta}{2-\beta}. \quad (23)$$

Positive, zero, and negative values of β correspond to favorable, zero, and adverse pressure gradients, respectively. The Blasius solution is recovered for $\beta = 0$, whereas attached solutions exist down to the separation limit $\beta \approx -0.1988$.

Introducing the similarity coordinate

$$\eta = \frac{y}{x} \sqrt{\frac{U_e x}{\nu}} \sqrt{\frac{1}{2-\beta}}, \quad (24)$$

the Falkner–Skan equation reads

$$\frac{d^3 f}{d\eta^3} + f \frac{d^2 f}{d\eta^2} + \beta \left[1 - \left(\frac{df}{d\eta} \right)^2 \right] = 0, \quad (25)$$

with boundary conditions

$$f(0) = 0, \quad \frac{df}{d\eta}(0) = 0, \quad \frac{df}{d\eta}(\eta \rightarrow \infty) = 1. \quad (26)$$

The corresponding velocity components are

$$u = U_e \frac{df}{d\eta}, \quad (27)$$

and

$$v = \sqrt{\frac{\nu U_e}{(2 - \beta)x}} \left[(1 - \beta)\eta \frac{df}{d\eta} - f \right]. \quad (28)$$

These expressions were used to evaluate the local velocities, gradients, strain-rate magnitude, and other quantities considered in the information-based analysis.

Defining

$$\mathcal{H}(\beta) = \frac{\frac{d^2 f}{d\eta^2}(0) - \beta \lim_{\eta \rightarrow \infty} [\eta - f(\eta)]}{1 + \beta}, \quad (29)$$

the momentum-thickness Reynolds number and Pohlhausen pressure-gradient parameter are

$$Re_\theta = \sqrt{Re_x(2 - \beta)} \mathcal{H}(\beta), \quad Re_x = \frac{U_e x}{\nu}, \quad (30)$$

and

$$\lambda_\theta = \frac{\theta^2}{\nu} \frac{dU_e}{dx} = \beta \mathcal{H}^2(\beta). \quad (31)$$

Consequently, λ_θ depends only on β , whereas Re_θ depends on both β and Re_x . The Falkner–Skan database therefore provides a controlled setting in which to determine whether local dimensionless variables recover the dependencies of the two integral targets. Figure 2 presents representative velocity profiles and the corresponding variations of λ_θ and Re_θ . The database was generated from 200 values of m in the range $m \in [-0.0901, 19.0]$, extending from near the adverse-pressure-gradient separation limit to strongly favorable pressure gradients. The Blasius solution, $m = 0$, was included explicitly. A fourth-order power-law spacing was used to increase the sampling density near the separation limit, where the solutions vary more rapidly.

For each similarity solution, 150 geometrically spaced values of the streamwise Reynolds number $Re_x \in [10^2, 10^6]$ were considered. The solutions were mapped to physical space using $x \in [0.1, 1.0]$ m and $\nu = 1.457 \times 10^{-5}$ m²s⁻¹. The resulting physical-space database contains the integral target quantities and the

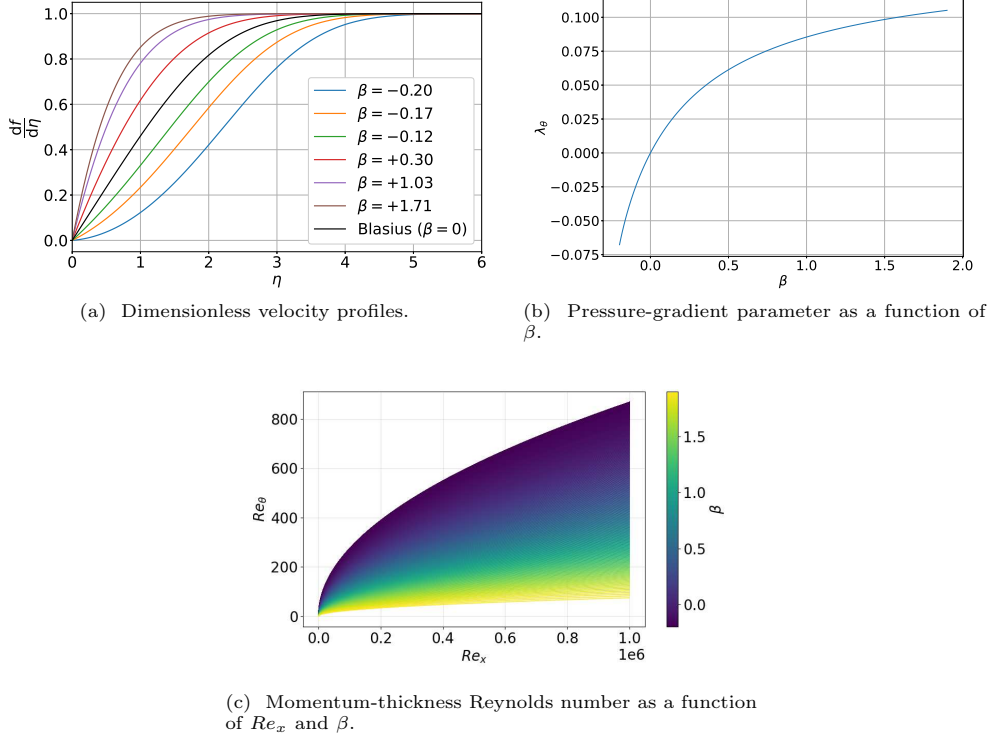


Fig. 2: Falkner–Skan similarity data used to construct the database: representative dimensionless velocity profiles and the corresponding Pohlhausen pressure-gradient parameter λ_θ and momentum-thickness Reynolds number Re_θ .

local variables required for the subsequent information-based feature selection and symbolic-regression analyses.

3 Information-based closure-discovery methodology

The objective of the proposed workflow is to construct compact local closures for the integral boundary-layer quantities λ_θ and Re_θ . Within the LCTM framework, the corresponding local surrogate variables, λ_{θ_L} and Re_{θ_L} , should reproduce the integral targets at a prescribed wall-normal location using only quantities available locally in a RANS computation.

For the Falkner–Skan family, the required functional dependencies are known from the similarity solution. In particular, $\lambda_\theta = \lambda_\theta(\beta)$, as shown in Eq. (31), whereas $Re_\theta = Re_\theta(Re_x, \beta)$, as shown in Eq. (30). The closure-discovery problem is therefore formulated as the search for local dimensionless variables that retain these dependencies, followed by the identification of compact algebraic mappings from the selected variables to the corresponding targets.

The complete workflow is illustrated in Fig. 3. For each Falkner–Skan solution, the candidate local variables are evaluated at prescribed dimensionless wall-normal locations,

$$\eta_w = \frac{d_w}{\delta_{99}}, \quad (32)$$

where d_w is the wall distance and δ_{99} is the boundary-layer thickness. Values in the range $\eta_w \in [0.25, 1.00]$ were considered at increments of 0.05. This wall-normal sweep is used to assess how much information the local variables contain about the integral targets and to identify a common design location for the two closures.

At each wall-normal location, the candidate variables are supplied to the IT- π procedure [23]. This method identifies dimensionless groups that maximize the information shared with the selected target. The resulting informative features are denoted by Π_λ^* and Π_{Re}^* for the pressure-gradient and momentum-thickness closures, respectively.

These features are subsequently used as inputs to PySR [17], which searches for explicit algebraic expressions of the form

$$\lambda_{\theta_L} = g(\Pi_\lambda^*), \quad Re_{\theta_L} = h(\Pi_{Re}^*), \quad (33)$$

where g and h are compact symbolic-regression expressions. The resulting candidates are assessed according to their reconstruction accuracy, algebraic complexity, and activation-coherence behavior away from the prescribed design location.

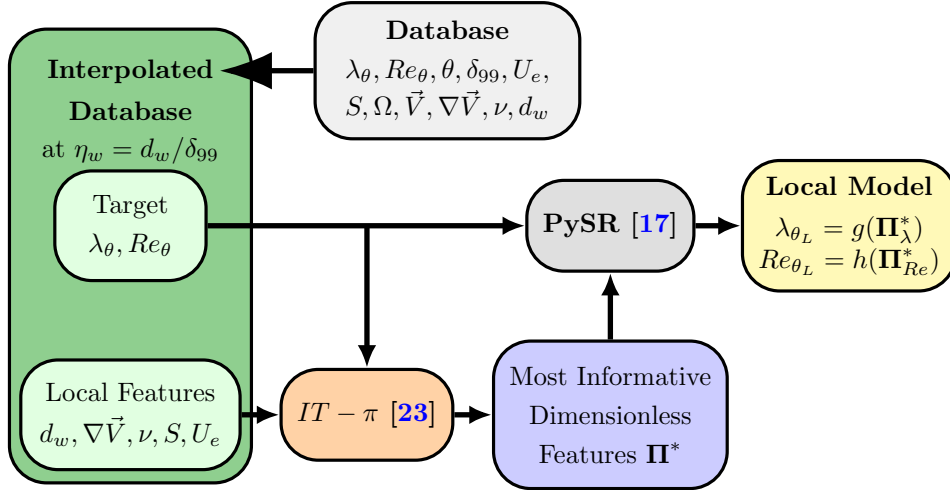


Fig. 3: Information-based symbolic regression workflow for constructing local models for Re_θ and λ_θ .

3.1 Identification of informative dimensionless features

The identification of informative dimensionless features is based on the Vaschy–Buckingham π theorem. For a physical relation involving N dimensional input variables,

$$\mathbf{q} = [q_1, q_2, \dots, q_N], \quad \hat{q}_o = \mathcal{F}(\mathbf{q}), \quad (34)$$

dimensional analysis allows the relation to be expressed as

$$\hat{\Pi}_o = f(\mathbf{\Pi}), \quad (35)$$

where Π_o is a dimensionless target and

$$\mathbf{\Pi} = [\Pi_1, \Pi_2, \dots, \Pi_K] \quad (36)$$

contains $K = N - r$ independent dimensionless input groups, with r denoting the rank of the dimensional matrix.

Each dimensionless group is written as a product of powers of the dimensional variables,

$$\Pi_k(\mathbf{q}) = \prod_{j=1}^N q_j^{\omega_j^{(k)}}, \quad k = 1, \dots, K, \quad (37)$$

where the exponent vector $\boldsymbol{\omega}^{(k)}$ satisfies the dimensional-homogeneity constraints. Although the π theorem determines the number of independent groups, it does not provide a unique basis. Consequently, multiple dimensionally valid combinations of the input variables may exist.

The IT- π method of Yuan and Lozano-Durán [23] addresses this non-uniqueness by selecting the dimensionless groups that carry the largest amount of information about the target. The information shared by the target Π_o and a candidate set of dimensionless groups $\mathbf{\Pi}$ is quantified through the Rényi mutual information of order α , $I_\alpha(\Pi_o; \mathbf{\Pi})$. This quantity is related to an information-theoretic lower bound, ϵ_{LB} , on the error of any model constructed from $\mathbf{\Pi}$, ϵ_f , i.e.,

$$\epsilon_f \geq e^{-I_\alpha(\Pi_o; \mathbf{\Pi})} \cdot c(\alpha, p, h_{\alpha, o}) \equiv \epsilon_{LB}, \quad (38)$$

where p denotes the selected error norm and $h_{\alpha, o}$ represents the information content of the target. Since this bound depends on the selected variables rather than on a particular regression model, it can be used to compare alternative dimensionless representations before performing symbolic regression.

The normalized irreducible error

$$\tilde{\epsilon}_{LB} = e^{-I_\alpha(\Pi_o; \mathbf{\Pi})} \quad (39)$$

is used throughout this work. It is reported as $100\tilde{\epsilon}_{LB}$ and expressed as a percentage. Lower values indicate that the candidate dimensionless groups contain more information about the target.

Because both targets considered here, λ_θ and Re_θ , are already dimensionless, the IT- π problem is restricted to the identification of informative dimensionless input groups:

$$\mathbf{\Pi}^* = \arg \min_{\mathbf{\Pi}} \left[\max_{\alpha} \tilde{\epsilon}_{\text{LB}} \right]. \quad (40)$$

The optimization is performed using the covariance matrix adaptation evolution strategy (CMA-ES) [24]. Bounds are imposed on the exponents to avoid excessively large powers and promote physically interpretable groups. A sparsity penalty is also introduced to discourage small non-zero exponents and favor dimensionless groups involving the fewest possible variables. The resulting objective function is

$$\mathcal{L} = -I_{\alpha}(\Pi_o; \mathbf{\Pi}) + \lambda_s \mathcal{S} \left(\left\{ \omega^{(k)} \right\}_{k=1}^K \right), \quad (41)$$

where

$$\mathcal{S} = \sum_{k=1}^K \sum_{j=1}^N \left[1 - e^{-|\tilde{\omega}_j^{(k)}/\tau|} \right]. \quad (42)$$

Here, $\tilde{\omega}_j^{(k)}$ is the normalized exponent associated with variable q_j in group Π_k , λ_s controls the weight of the sparsity penalty, and τ determines the range over which small non-zero exponents are penalized. The values used in the present analysis are $\lambda_s = 0.3$, and $\tau = 0.072$.

Mutual information is estimated using the k -nearest-neighbor estimator of Kraskov et al. [25], with $k = 5$. Before constructing the power-law groups, magnitudes below 10^{-6} are clipped retaining their sign to avoid numerical singularities during exponentiation.

3.2 Symbolic regression

Once the informative dimensionless features have been identified, symbolic regression is used to construct explicit algebraic expressions for the local closures. The two regression problems are written as

$$\lambda_{\theta_L} = g(\mathbf{\Pi}_{\lambda}^*), \quad Re_{\theta_L} = h(\mathbf{\Pi}_{Re}^*), \quad (43)$$

where $\mathbf{\Pi}_{\lambda}^*$ and $\mathbf{\Pi}_{Re}^*$ denote the informative dimensionless features identified for the pressure-gradient and momentum-thickness closures, respectively.

The expressions are obtained using PySR [17], an open-source symbolic-regression package based on the `SymbolicRegression.jl` backend. PySR applies a multi-population evolutionary search to generate, simplify, and optimize candidate expression trees. The resulting expressions are ranked on an accuracy-complexity Pareto front, allowing compact expressions to be identified without selecting a closure solely on the basis of its minimum training error.

The regression loss is based on the discrepancy between the predicted local surrogate variable and its corresponding integral target in the Falkner–Skan database. The

operator set is deliberately restricted to favor compact expressions that can be implemented robustly in a RANS solver. The binary operators are $\{+, -, \times, \div\}$, and the unary operators are $\{(\cdot)^2, \sqrt{\cdot}\}$.

Each symbolic-regression run uses 10 populations and 100 iterations. The Falkner–Skan data are divided randomly into training and test subsets using an 80–20 split.

The expressions returned by PySR are first ranked according to their position on the accuracy–complexity Pareto front. The final closures are then selected by considering their reconstruction accuracy, algebraic complexity, and activation-coherence behavior away from the prescribed design wall-normal location.

4 Results and discussion

4.1 Informative dimensionless features

The first stage of the analysis identifies local dimensionless groups that retain the information required to reconstruct the integral targets λ_θ and Re_θ . The candidate dimensional variables extracted from the Falkner–Skan database are

$$\mathbf{q} = \left[d_w, \nu, U_e, \frac{\partial u}{\partial y}, \frac{\partial v}{\partial x}, \frac{\partial v}{\partial y}, S \right], \quad (44)$$

where d_w is the wall distance, ν is the kinematic viscosity, U_e is the boundary-layer edge velocity, and S is the strain-rate magnitude. At each wall-normal location, the IT- π optimization is repeated using increasing subsample sizes to assess the robustness of the identified groups.

Two complementary criteria are considered when assessing the resulting local closure variables. The first is a reconstruction-accuracy criterion: at a prescribed design wall-normal location, η_w^* , the local closure should reproduce the corresponding integral target as accurately as possible.

The second is an activation-coherence criterion. Although a local closure is constructed to reproduce its target at η_w^* , it is evaluated throughout the computational domain. Its off-design behavior should therefore not cause the transition-onset criterion to activate at wall-normal locations where the closure is inaccurate. For the one-equation intermittency model, the principal activation ratio is

$$\mathcal{A}(\eta_w) = \frac{Re_{\theta_L}(\eta_w)}{Re_{\theta_c}[Tu_L(\eta_w), \lambda_{\theta_L}(\eta_w)]}. \quad (45)$$

A jointly coherent pair of closures should prevent $\mathcal{A}(\eta_w)$ from exceeding its design-location value at off-design positions, so that the onset threshold is first reached near η_w^* .

When the two closures are examined separately, this requirement can be assessed through proxy conditions. For the momentum-thickness closure, Re_{θ_L} should not be over-predicted away from η_w^* , since this would promote premature activation. For the pressure-gradient closure, off-design values of λ_{θ_L} should not reduce Re_{θ_c} and thereby increase the activation ratio. The latter condition depends on the structure

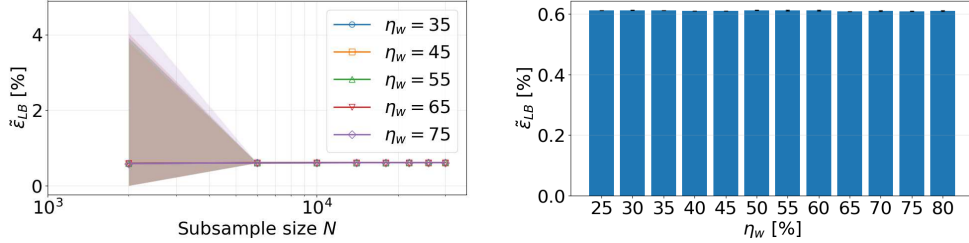
and monotonicity of both F_{PG} and the Re_{θ_c} correlation. Because the baseline correlation does not reproduce the expected pressure-gradient trend over the complete favorable-pressure-gradient range, the activation-coherence behavior of λ_{θ_L} is treated as a diagnostic rather than as a strict selection constraint.

Finally, both closures must satisfy a joint-location criterion: they should reproduce their respective targets at the same design wall-normal location. The activation-coherence assessment performed here is restricted to the sampled boundary-layer region, $\eta_w \leq 1$. The behavior of the transition onset models outside the boundary layer is considered not relevant because at $\eta_w > 1$, $\gamma \approx 1$ and the turbulence model is already fully active. Consequently, the local onset criterion no longer governs the activation of turbulent production in this region.

4.1.1 Feature for the pressure-gradient closure

Within the Falkner–Skan family, the exact Pohlhausen parameter is a one-parameter function, $\lambda_\theta = \lambda_\theta(\beta)$, as shown in Eq. (31). The IT- π procedure is therefore used to identify a single dimensionless local feature, with $\Pi_o = \lambda_\theta$ as the target.

Figure 4 shows the normalized irreducible error as a function of the subsample size and wall-normal location.



(a) Normalized irreducible error as a function of subsample size at selected wall-normal locations. (b) Normalized irreducible error as a function of wall-normal location for the complete dataset.

Fig. 4: Information retained by the dimensionless group $\Pi_\lambda^* = (\partial v / \partial y) d_w^2 / \nu$ about the target $\Pi_o = \lambda_\theta$. The shaded regions in Fig. 4a indicate the estimated uncertainty.

The same dominant dimensional structure is recovered across all considered subsample sizes and wall-normal locations:

$$\Pi_\lambda^* = \frac{\partial v}{\partial y} \frac{d_w^2}{\nu}. \quad (46)$$

For the complete dataset, the corresponding normalized irreducible error is $\tilde{\epsilon}_{\text{LB}} = 0.6\%$, $\Delta\tilde{\epsilon}_{\text{LB}} = 5 \times 10^{-4}\%$. This low value indicates that Π_λ^* retains nearly all the information required to reconstruct λ_θ within the sampled Falkner–Skan domain.

The selected group has the same dimensional structure as the pressure-gradient closure used in the baseline one-equation intermittency model, given in Eq. (17). The information-based analysis therefore recovers the physically motivated structure of the

baseline closure rather than introducing an unrelated combination of variables. The sparsity penalty in Eq. (41) is important in this result: it suppresses dimensionally admissible contributions with small non-zero exponents and promotes the recovery of the same dominant group across the sampled wall-normal locations.

Using the Falkner–Skan similarity solution, Eq. (46) can be written as

$$\Pi_\lambda^* = \frac{\partial v}{\partial y} \frac{d_w^2}{\nu} = -\beta \eta^2 \frac{df}{d\eta} - (\beta - 1) \eta^3 \frac{d^2 f}{d\eta^2}. \quad (47)$$

The group is independent of Re_x . At a prescribed dimensionless wall-normal location, η is fixed and Π_λ^* becomes a function of β only. Consequently, the local feature and the target λ_θ depend on the same underlying Falkner–Skan parameter, which explains the low irreducible error obtained throughout the boundary layer.

Figure 5a shows the relation between λ_θ and Π_λ^* at different wall-normal locations. The collapse onto a one-dimensional mapping confirms that the selected group is a suitable input for symbolic regression. Figure 5b shows its wall-normal evolution. The feature varies non-monotonically under adverse and zero pressure gradients, whereas its behavior is approximately monotonic under favorable pressure gradients. Therefore, no symbolic mapping based solely on Π_λ^* can guarantee activation coherence independently of the downstream F_{PG} and Re_{θ_c} correlations. Its wall-normal behavior is consequently used as a diagnostic when comparing the symbolic-regression candidates, rather than as a strict admissibility condition.

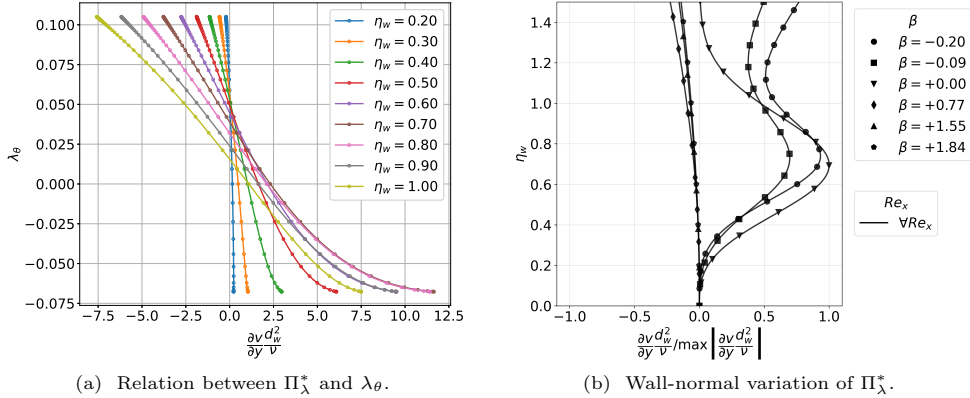


Fig. 5: Relationship between the exact Pohlhausen parameter λ_θ and the selected local dimensionless feature $\Pi_\lambda^* = (\partial v / \partial y) d_w^2 / \nu$. Panel 5a shows the mapping at different fixed values of $\eta_w = d_w / \delta_{99}$, whereas panel 5b shows the wall-normal evolution of the feature for different values of β and Re_x .

4.1.2 Features for the momentum-thickness closure

The exact momentum-thickness Reynolds number depends on two independent Falkner–Skan parameters, $Re_\theta = Re_\theta(Re_x, \beta)$, as shown in Eq. (30). The IT- π procedure is therefore used to identify two independent dimensionless input groups, with $\Pi_o = Re_\theta$ as the target.

When the complete candidate set in Eq. (44) was used, the optimized groups frequently contained variables raised to small non-zero powers, despite the sparsity penalty. To obtain a more interpretable representation, the search was repeated using the reduced dimensional set

$$\mathbf{q}_{Re_\theta} = [U_e, \nu, d_w, S]. \quad (48)$$

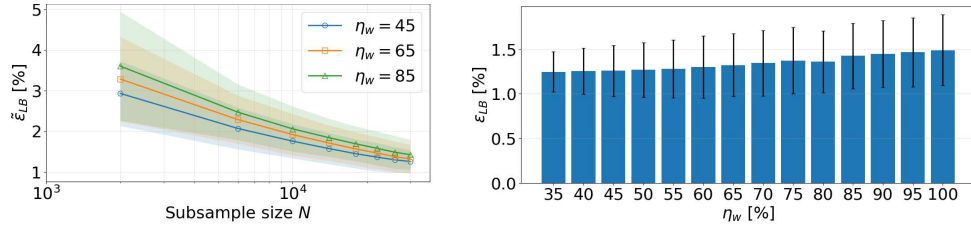
The complete and reduced input sets produced similar values of $\tilde{\epsilon}_{LB}$, indicating that the reduction did not lead to a significant loss of information.

Across the sampled wall-normal locations, three distinct dimensionless groups appeared in the optimized two-group bases:

$$\Pi_{Re}^* = \left[\frac{U_e d_w}{\nu}, \frac{U_e}{\sqrt{S\nu}}, \frac{d_w \sqrt{S}}{\sqrt{\nu}} \right]. \quad (49)$$

The group $d_w \sqrt{S}/\sqrt{\nu}$ was recovered consistently across wall normal locations, but $U_e d_w/\nu$ and $U_e/\sqrt{S\nu}$ were interchanged depending on the dimensionless height studied. For $\eta_w < 0.35$, $U_e/\sqrt{S\nu}$ was recovered, whereas for $\eta_w > 0.35$, $U_e d_w/\nu$ was recovered.

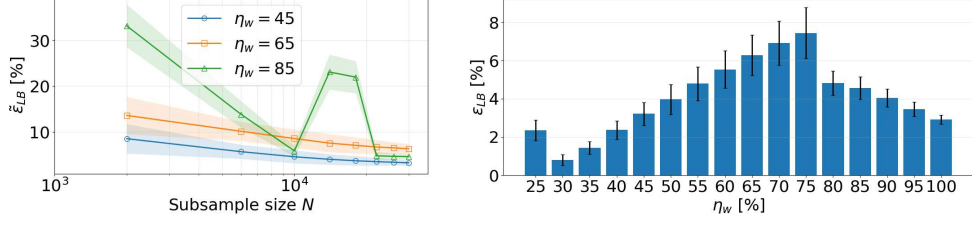
Figures 6 and 7 show the irreducible errors obtained for Re_{U_e} and Re_V , respectively.



(a) Normalized irreducible error as a function of subsample size at selected wall-normal locations. (b) Normalized irreducible error as a function of wall-normal location for the complete dataset.

Fig. 6: Information retained by $Re_{U_e} = U_e d_w/\nu$ about the target $\Pi_o = Re_\theta$. The shaded regions in Fig. 6a indicate the estimated uncertainty.

Both groups provide stable results as the subsample size increases. However, Re_{U_e} exhibits a more uniform irreducible error across the sampled wall-normal locations



(a) Normalized irreducible error as a function of subsample size N at selected wall-normal locations. (b) Normalized irreducible error as a function of wall-normal location for the complete dataset.

Fig. 7: Information retained by $d_w \sqrt{S/\nu} = \sqrt{Re_V}$ about the target $\Pi_o = Re_\theta$. The shaded regions in Fig. 7a indicate the estimated uncertainty.

than $\sqrt{Re_V}$, which is the square root of the variable used in the baseline one-equation intermittency model.

Figure 8 compares the mappings from Re_{U_e} and Re_V to the target Re_θ . The form Re_V is shown instead of $\sqrt{Re_V}$ because the two quantities contain equivalent information and Re_V provides a clearer visual comparison with the target.

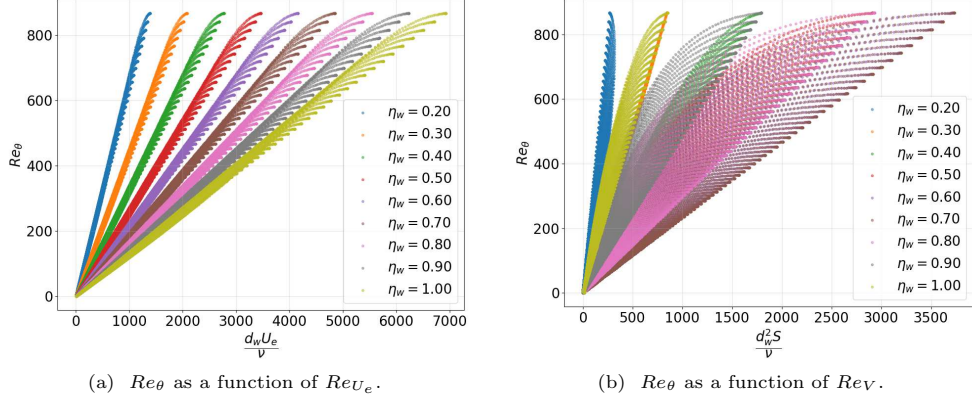


Fig. 8: Relationships between the exact momentum-thickness Reynolds number Re_θ and the candidate local dimensionless features $Re_{U_e} = U_e d_w / \nu$ and $Re_V = d_w^2 S / \nu$ at different wall-normal locations.

Although Re_V provides an almost one-to-one reconstruction of Re_θ near $\eta_w \approx 0.3$, its suitability as a local closure variable must also be assessed from its wall-normal behavior. Figure 9 shows that Re_V has a bell-shaped distribution and may exceed its design-location value over part of the boundary layer. A closure based on Re_V could therefore activate the onset criterion at locations where it no longer provides an accurate reconstruction of Re_θ . It would consequently be difficult for such a closure to simultaneously satisfy the reconstruction-accuracy and activation-coherence criteria.

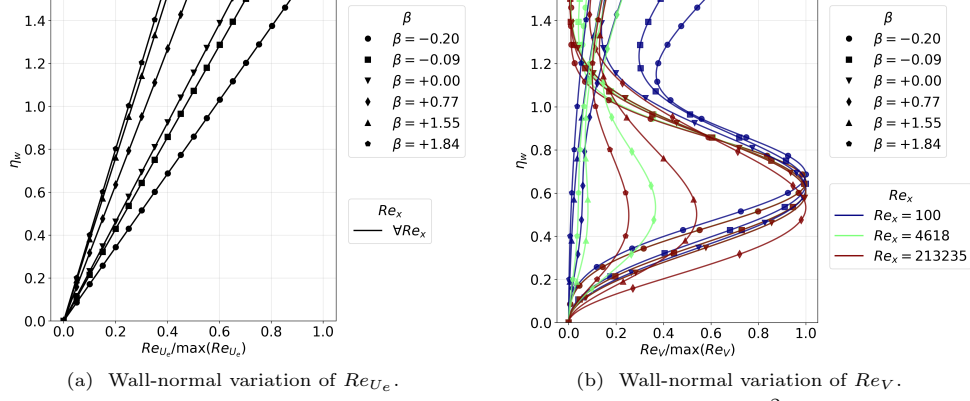


Fig. 9: Wall-normal behavior of $Re_{U_e} = U_e d_w / \nu$ and $Re_V = d_w^2 S / \nu$ for different values of β and Re_x .

By contrast, Re_{U_e} varies approximately linearly with d_w within the Falkner–Skan database, as shown in Fig. 9a, while retaining a strong relationship with Re_θ , as shown in Fig. 8a. Selecting the outer edge of the sampled boundary layer, $\eta_w^* = 1$, as the design location therefore ensures that the corresponding estimate Re_{θ_L} remains smaller at interior wall-normal positions. Within $0 \leq \eta_w \leq 1$, a closure based on Re_{U_e} can consequently satisfy both the reconstruction-accuracy and activation-coherence criteria.

4.2 Symbolic regression of the local closures

The information-content analysis does not identify a preferred wall-normal location for the pressure-gradient closure, since the selected feature retains approximately the same information about λ_θ throughout the sampled boundary layer. The design location is therefore determined from the momentum-thickness closure and imposed on both closures to satisfy the joint-location criterion.

For Re_{θ_L} , the approximately monotonic wall-normal behavior of Re_{U_e} favors the selection of $\eta_w^* = 1$. At this location, the closure can reproduce the integral target while remaining below it at interior wall-normal positions, thereby satisfying the activation-coherence proxy within the sampled boundary layer. Both symbolic-regression problems are consequently formulated at $\eta_w^* = 1$.

Symbolic regression is then applied using the informative dimensionless features identified for each target. The pressure-gradient closure constitutes a one-dimensional regression problem, since λ_{θ_L} is expressed as a function of the single selected feature Π_λ^* . By contrast, the momentum-thickness closure is formulated as a two-dimensional regression problem using Re_{U_e} and $\sqrt{Re_V}$ as candidate inputs for Re_{θ_L} .

4.2.1 Pressure-gradient closure

The baseline pressure-gradient closure and three symbolic-regression candidates of increasing algebraic complexity are compared in Table 1. For compactness, the selected dimensionless feature is denoted by

$$\Pi_\lambda^* = \frac{\partial v}{\partial y} \frac{d_w^2}{\nu}. \quad (50)$$

The coefficient of determination, R^2 , and the root-mean-square error, RMSE, are evaluated at the design wall-normal location, $\eta_w^* = 1$.

Table 1: Baseline and symbolic-regression pressure-gradient closures at $\eta_w^* = 1$, together with their R^2 and RMSE values.

Closure	Expression for λ_{θ_L}	R^2	RMSE ($\times 10^{-4}$)
Baseline [6]	$-7.57 \times 10^{-3} \Pi_\lambda^* + 0.0128$	0.803	288
SR 1	$-12.065 \times 10^{-3} \Pi_\lambda^* + 0.0159$	0.997	34
SR 2	$\Pi_\lambda^* \left[4.6 \times 10^{-5} (\Pi_\lambda^*)^2 - 0.014106 \right] + 0.015696$	0.999	17
SR 3	$\{ \Pi_\lambda^* [-4 \times 10^{-6} \Pi_\lambda^* - 6 \times 10^{-5}] \times (2.7171 - \Pi_\lambda^*) - 0.014251 \} \Pi_\lambda^* + 0.015503$	1.000	1

The baseline closure has a relatively high R^2 value at $\eta_w^* = 1$ despite having originally been calibrated at an interior wall-normal location [6]. Nevertheless, all three symbolic-regression candidates provide substantially more accurate reconstructions at $\eta_w^* = 1$. Even the linear candidate, SR 1, reduces the RMSE by nearly one order of magnitude relative to the baseline closure, while the higher-complexity candidates provide further reductions.

Although the symbolic-regression closures are constructed at $\eta_w^* = 1$, they are also evaluated at off-design wall-normal locations. This assessment is necessary because the closure is evaluated throughout the computational domain when implemented in a RANS solver, rather than exclusively at its design location. Figure 10 shows the reconstruction error as a function of the Falkner–Skan parameter β and the dimensionless wall-normal location η_w .

The baseline closure reaches its minimum reconstruction error near $\eta_w \approx 1.2$. It generally under-predicts λ_θ below this location, except in the adverse-pressure-gradient

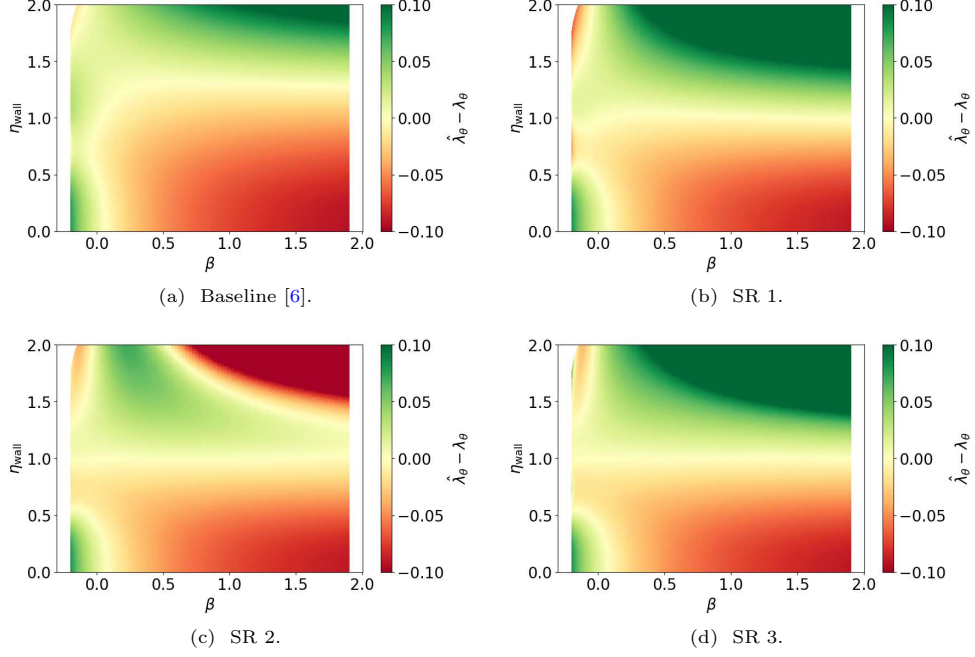


Fig. 10: Reconstruction error of the baseline and symbolic-regression pressure-gradient closures as functions of the dimensionless wall-normal location η_w and the Falkner-Skan parameter β .

range, and over-predicts it farther from the wall. By contrast, the symbolic-regression closures are most accurate near their design location, $\eta_w^* = 1$, and generally under-predict λ_θ at interior wall-normal positions just like the baseline model.

The three symbolic-regression candidates exhibit comparable off-design error distributions within the sampled boundary-layer region, with the main differences appearing outside the boundary layer. Consequently, none of the candidates provides a clear advantage according to the activation-coherence diagnostic. The candidate with the lowest reconstruction error, SR 3, is therefore retained:

$$\lambda_{\theta_L} = \left\{ \Pi_\lambda^* \left[-4 \times 10^{-6} \Pi_\lambda^* - 6 \times 10^{-5} \right] (2.7171 - \Pi_\lambda^*) - 0.014251 \right\} \Pi_\lambda^* + 0.015503. \quad (51)$$

4.2.2 Momentum-thickness Reynolds-number closure

Table 2 compares the baseline momentum-thickness closure with three symbolic-regression candidates of increasing algebraic complexity. The coefficient of determination, R^2 , and the root-mean-square error, RMSE, are evaluated at the design wall-normal location, $\eta_w^* = 1$. The expressions are written in terms of the dimensionless

features

$$Re_{U_e} = \frac{U_e d_w}{\nu}, \quad Re_V = \frac{d_w^2 S}{\nu}. \quad (52)$$

Table 2: Baseline and symbolic-regression momentum-thickness closures at $\eta_w^* = 1$, together with their R^2 and RMSE values.

Closure	Expression for Re_{θ_L}	R^2	RMSE
Baseline [6]	$\frac{Re_V}{2.193}$	0.438	104
SR 1	$\frac{Re_{U_e}}{7.710}$	0.998	6
SR 2	$\left(0.13193 - \frac{0.0374}{\sqrt{Re_V}}\right) Re_{U_e}$	0.998	6
SR 3	$0.13046 \left[\frac{-11.531}{-1.831 + 31.308 Re_V / Re_{U_e}} + Re_{U_e} \right]$	0.999	5

All three symbolic-regression candidates provide substantially more accurate reconstructions of Re_θ at $\eta_w^* = 1$ than the baseline closure, which was originally calibrated to operate at an interior wall-normal location [6]. The linear SR 1 closure already increases R^2 from 0.438 to 0.998 and reduces the RMSE from 104 to 6. SR 2 yields the same reported accuracy while introducing an additional dependence on Re_V , whereas SR 3 provides only a marginal reduction in RMSE at the cost of greater algebraic complexity.

The closures are also evaluated away from the design location to assess their wall-normal behavior. Figure 11 presents the relative reconstruction error as a function of the dimensionless wall-normal location η_w and the Falkner–Skan parameter β .

The baseline closure exhibits a broad region of minimum reconstruction error near $\eta_w \approx 0.5$, consistently with its intended design location. Away from this region, it generally under-predicts Re_θ . The bell-shaped wall-normal behavior of Re_V (see Fig. 9b) therefore provides a degree of activation coherence around the baseline design location, since the estimated Re_{θ_L} decreases on either side of the region where the closure is most accurate. However, its reconstruction accuracy even at its own design height, $\eta_w^* \approx 0.5$, is comparatively poor.

The three symbolic-regression closures exhibit similar wall-normal error distributions. Within the sampled boundary-layer region, $0.25 \leq \eta_w < 1$, they generally under-predict Re_θ and approach the target as η_w increases toward the design location. They therefore satisfy the activation-coherence proxy within this interval by preventing Re_{θ_L} from exceeding its design-location value at interior wall-normal positions. Above $\eta_w = 1$, the closures over-predict the target. This off-design behavior is not considered relevant to the transition-onset prediction because these locations lie outside the boundary layer, where $\gamma \approx 1$ and the turbulence model is already fully active.

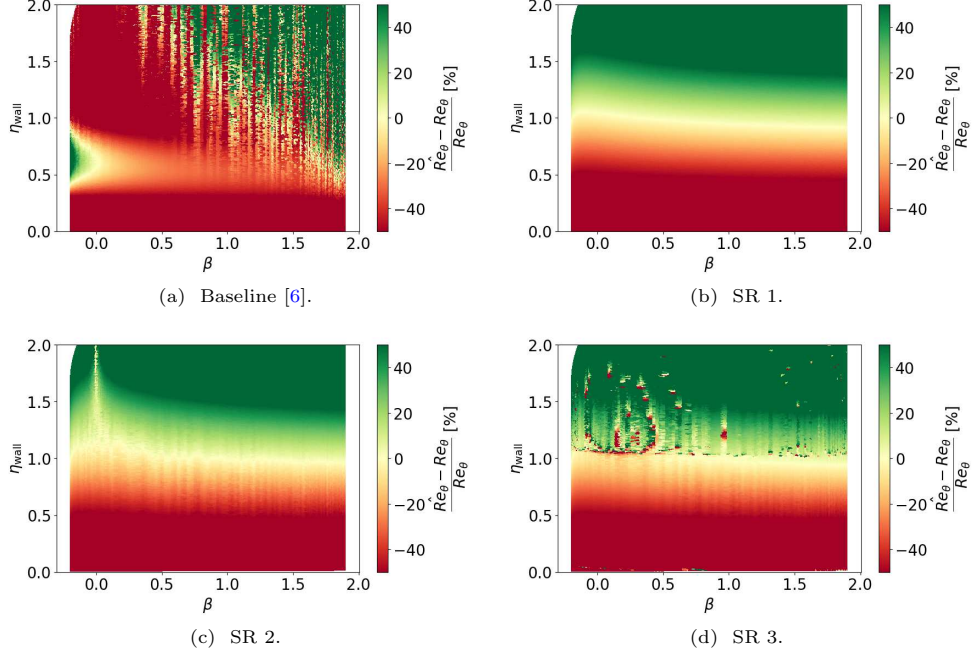


Fig. 11: Relative reconstruction error of the baseline and symbolic-regression momentum-thickness closures as functions of the dimensionless wall-normal location η_w and the Falkner–Skan parameter β .

Consequently, the local onset criterion no longer governs the activation of turbulent production in this region.

Since SR 1, SR 2, and SR 3 provide nearly identical reconstruction accuracy and comparable activation-coherence behavior within the boundary layer, simplicity becomes the determining selection criterion. SR 1 is therefore retained:

$$Re_{\theta_L} = \frac{Re_{U_e}}{7.710} = \frac{d_w U_e}{7.710 \nu}. \quad (53)$$

4.3 A posteriori transition-model assessment

The expressions selected for λ_{θ_L} and Re_{θ_L} in Eqs. (51) and (53), respectively, provide improved local reconstructions of the corresponding Falkner–Skan integral quantities. They should not, however, be interpreted as complete corrections to the transition model. Their selection is based on their reconstruction accuracy, algebraic complexity, and activation-coherence behavior within the sampled Falkner–Skan database. Once inserted into the one-equation intermittency model, their influence on transition onset is mediated by the remaining empirical correlations, limiters, source terms, and turbulence-model coupling.

The pressure-gradient closure does not act directly on the transition location. Instead, λ_{θ_L} enters the pressure-gradient correction F_{PG} , which modifies the critical momentum-thickness Reynolds-number correlation,

$$Re_{\theta_c} = Re_{\theta_c}(Tu_L, \lambda_{\theta_L}). \quad (54)$$

The resulting value of Re_{θ_c} is subsequently used by the onset function that controls intermittency production. This distinction is important because the baseline Re_{θ_c} correlation does not reproduce the same pressure-gradient dependence as classical transition-onset correlations, such as that of Abu-Ghannam and Shaw [26], over the complete favorable-pressure-gradient range. In particular, with the baseline model constants, Eq. (13) predicts a decrease in Re_{θ_c} as positive values of λ_{θ_L} increase. Consequently, a more accurate local reconstruction of the physical Falkner–Skan parameter λ_θ does not necessarily produce the physically expected advancement or delay of transition within the complete model.

The momentum-thickness closure enters the onset mechanism more directly through the ratio between Re_{θ_L} and Re_{θ_c} . Nevertheless, the corresponding onset functions were calibrated using the baseline Re_{θ_L} closure and its original design location. Replacing this closure may therefore require recalibration of the remaining model parameters to account for its different magnitude and wall-normal behavior.

The new momentum-thickness closure also requires an estimate of the boundary-layer edge velocity, which is not directly available as a local RANS variable. Following previous local transition-model formulations, such as the amplification-factor transport model [9], the edge velocity is approximated as

$$\tilde{U}_e^2 = \frac{2\kappa}{\kappa - 1} \left(\frac{P_\infty}{\rho_\infty} - \frac{P}{\rho_e} \right) + U_\infty^2, \quad (55)$$

where

$$\rho_e = \rho_\infty \left(\frac{P}{P_\infty} \right)^{1/\kappa}. \quad (56)$$

Here, P is the local static pressure and κ is the ratio of specific heats. The symbol κ is used to avoid confusion with the intermittency variable γ . For the Falkner–Skan solutions considered in the closure-discovery analysis, Eq. (55) recovers the exact edge velocity U_e . Consequently, the use of U_e during the information-based and symbolic-regression stages is consistent with the local approximation subsequently employed in the RANS implementation. Equation (55) is written for a stationary wall; for moving-wall applications, the velocities should be evaluated relative to the wall velocity to preserve Galilean invariance.

Similarly, the derivative $\partial v / \partial y$ appearing in the pressure-gradient closure represents the derivative of the wall-normal velocity component along the wall-normal direction. For a general geometry, it is evaluated according to Eq. (16). The Cartesian form $\partial v / \partial y$ is recovered for the two-dimensional, wall-aligned Falkner–Skan configuration.

The ERCOFTAC T3 transitional flat-plate series is used as a focused a posteriori assessment of the new closures within the complete one-equation intermittency model.

The closures from Eqs. (51) and (53) were implemented in OpenFOAM v2406. Three model variants are considered:

- **M15**: the baseline one-equation intermittency model of Menter et al. [6];
- **M15 SR**: the baseline model with the new λ_{θ_L} and Re_{θ_L} closures;
- **M15 SR-R**: the modified-closure variant with $F_{\text{length}} = 75$.

The value $F_{\text{length}} = 75$ was selected through a limited manual adjustment against the T3 series. The M15 SR-R variant is included to illustrate the sensitivity of the complete transition model to a minimal recalibration after replacing the local closures. It is therefore not interpreted as an independently validated recalibration.

All simulations were performed using `simpleFoam`. Although the computational meshes were three-dimensional, only one cell was employed in the spanwise direction, making the simulations effectively two-dimensional. The computations were advanced until the integrated force coefficients reached an asymptotic state. Convergence was assessed from the variation of the drag coefficient, C_D , using a target tolerance of $\Delta C_D < 1 \times 10^{-7}$. The pressure equation was solved using the GAMG solver with a Gauss-Seidel smoother. Velocity, turbulent kinetic energy, specific dissipation rate, and intermittency were solved using PBiCGStab with a DILU preconditioner.

Two mesh-sensitivity analyses were performed for each test case: one using the SST $k-\omega$ SST turbulence model [27] and one using the baseline M15 transition model [6]. The selected mesh was the finest mesh for which an additional refinement produced variations below 1% in C_D for both models. The M15 SR and M15 SR-R simulations were then performed using the same selected meshes. This procedure limits the influence of spatial-discretization differences on the comparison among the three transition-model variants.

4.3.1 T3 series of transitional flat plates

The ERCOFTAC T3 transitional flat-plate experiments [19] are used as canonical test cases for natural, bypass, and pressure-gradient-induced transition. The series includes both zero-pressure-gradient (ZPG) and non-zero-pressure-gradient (NZPG) configurations, allowing the a posteriori influence of the modified Re_{θ_L} and λ_{θ_L} closures to be assessed under different transition mechanisms.

The computational domain, boundary conditions, and mesh-generation strategy follow the setup of Capel-Jorquera et al. [28]. As shown in Fig. 12, the domain contains a no-slip flat plate of reference length $L_{\text{ref}} = 2$ m and an upper slip wall. The inlet height is $D = 0.22$ m, and the inlet is located $L_{\text{LE}} = 0.15$ m upstream of the leading edge to prevent the inlet boundary condition from interfering with the development of the plate boundary layer [29]. The upstream extension is treated as a symmetry boundary. For the NZPG cases, the upper-wall contours reported by Suluksna et al. [29] are employed, whereas a horizontal upper wall is used for the ZPG cases.

The inlet velocity, inlet and leading-edge turbulence intensities, inlet turbulent-viscosity ratio, and pressure-gradient classification are summarized in Table 3. The inlet turbulence quantities were selected to reproduce the measured turbulence decay and the experimental turbulence intensity at the plate leading edge.

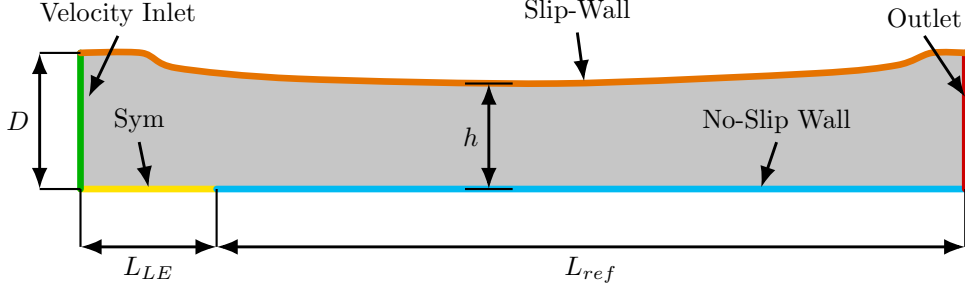


Fig. 12: T3 NZPG computational domain with boundary conditions.

Table 3: Inlet and leading edge conditions for the T3 transitional flat plate cases.

Case	U_{in} [m/s]	Tu_{in} [%]	Tu_{LE} [%]	TVR_{in}	Pressure gradient
T3A	5.40	5.70	3.36	13.3	ZPG
T3B	9.40	8.46	6.09	100	ZPG
T3Am	19.8	1.12	0.94	8.00	ZPG
TSK	50.1	0.03	0.03	1.00	ZPG
T3C1	5.90	7.78	6.60	44.0	NZPG
T3C2	5.00	3.10	3.00	9.00	NZPG
T3C3	3.70	3.10	3.00	6.00	NZPG
T3C4	1.20	3.10	3.00	2.50	NZPG
T3C5	8.40	3.70	3.00	15.0	NZPG

The skin-friction distributions predicted by M15, M15 SR, and M15 SR-R are compared with the experimental measurements in Fig. 13.

For the ZPG cases, both modified variants improve the transition-onset prediction moderately for T3A, while producing only minor changes for T3B. By contrast, the agreement deteriorates for the two low-turbulence-intensity cases, T3Am and TSK. The direction of the onset shift differs between these cases: transition is predicted earlier for T3Am and later for TSK. This opposite response indicates that the effect of the new closures cannot be characterized as a uniform advancement or delay of transition under ZPG conditions.

For the NZPG configurations, the M15 SR variant predicts an earlier transition onset for T3C1, showing that replacing the closures without adjusting the remaining model parameters can alter the original model balance. Reducing F_{length} in M15 SR-R compensates for part of this response. Relative to the baseline M15 model, the recalibrated variant retains a comparable level of agreement for T3C1, T3C3, and T3C4, while improving the predictions for T3C2 and T3C5.

The T3C2 result is particularly noteworthy because M15 SR-R modifies not only the predicted onset location but also the development of the skin-friction coefficient

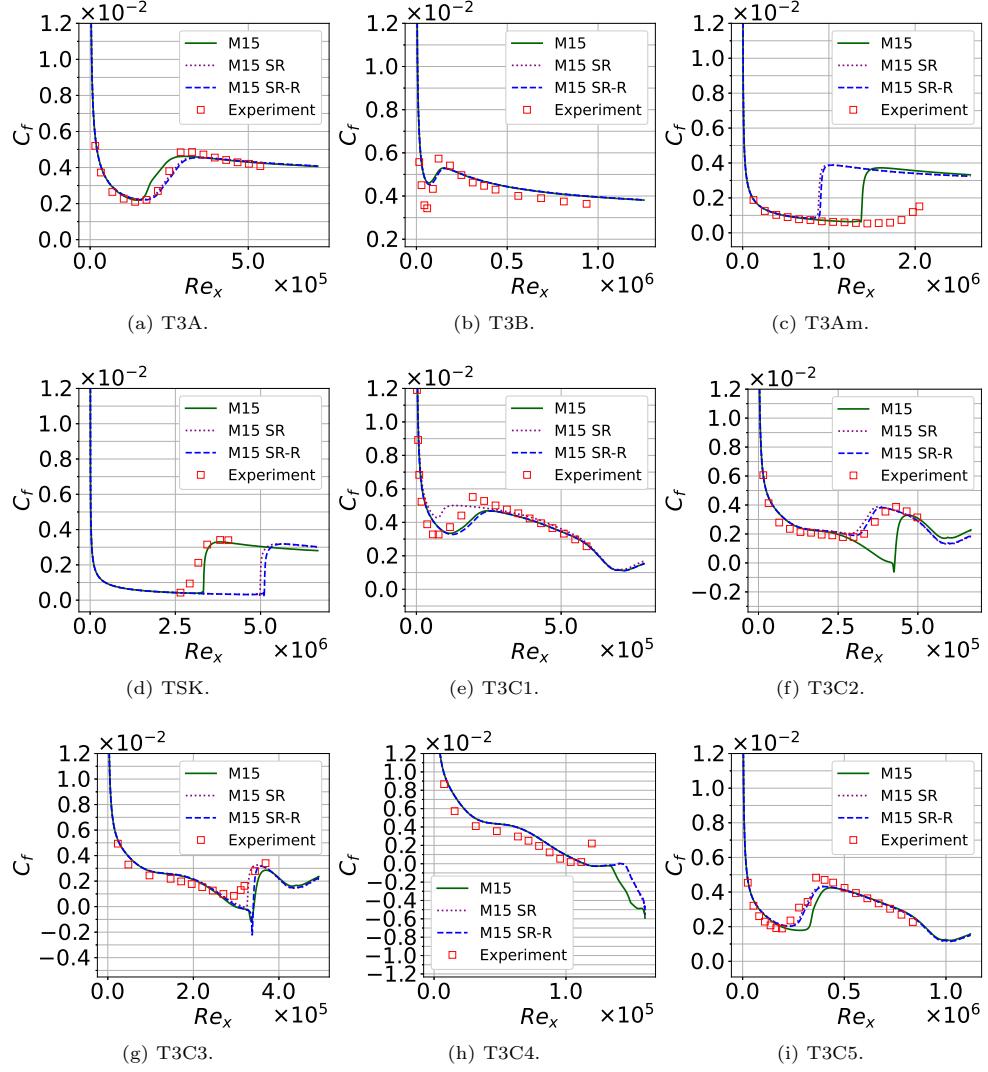


Fig. 13: Skin-friction distributions predicted by the baseline M15, modified-closure M15 SR, and recalibrated M15 SR-R transition-model variants for the ERCOFTAC T3 flat-plate cases.

through the transition region. Previous calibration of the baseline M15 model did not reproduce this behavior [28]. The present result therefore suggests that the combined effect of the new local closures and the limited recalibration can modify aspects of the transition-region evolution that are not readily recovered by adjusting the baseline model constants alone. However, this effect cannot be exclusively attributed to the new closures because M15 SR-R also changes F_{length} .

Overall, the results show that the improved reconstruction of λ_θ and Re_θ within the Falkner–Skan database does not translate into a uniform improvement across the T3 series. The modified variants maintain or improve the agreement for most bypass and NZPG cases, but deteriorate the predictions for T3Am and TSK. This behavior suggests that, under very low free-stream turbulence, the remaining components of the onset mechanism may become more limiting than the local λ_{θ_L} and Re_{θ_L} closures.

One possible source of this limitation is the definition of Tu_L in Eq. (14), which depends on ωd_w and does not directly approximate the free-stream turbulence intensity. A potential alternative is

$$Tu'_L = 100 \frac{\sqrt{2k/3}}{\tilde{U}_e}, \quad (57)$$

where $\tilde{U}_e$ is defined in Eq. (55). This expression would provide a more direct local surrogate of the turbulence intensity based on the estimated boundary-layer edge velocity. Its introduction would, however, alter the input to the Re_{θ_c} correlation and would therefore require a dedicated recalibration. The proposed definition is not assessed in the present work and is retained as a direction for future investigation.

A systematic revision of the Re_{θ_c} correlation would require a database containing consistent information about transition onset, intermittency, free-stream turbulence, and pressure-gradient effects. Such data remain scarce because sufficiently resolved transitional experiments are limited and the experimental determination of intermittency is not fully standardized. High-fidelity simulations of canonical experiments, such as the T3 series [30], may provide the information required to construct such a database. Given suitable data, the information-based symbolic-regression methodology proposed in the present work could also be applied to identify the most informative local dimensionless variables governing Re_{θ_c} and to derive a compact, interpretable replacement for the current critical-Reynolds-number correlation.

5 Conclusions and outlook

An information-based symbolic-regression framework has been developed to construct compact local closures for the pressure-gradient parameter λ_θ and the momentum-thickness Reynolds number Re_θ used in local-correlation transition models. The methodology combines information-based dimensional analysis, symbolic regression, and an assessment of the wall-normal behavior of the resulting closures. It was first applied to a Falkner–Skan database and subsequently evaluated within the one-equation intermittency model using the ERCOFTAC T3 transitional flat-plate cases.

The main conclusions are as follows:

- The IT- π analysis identified

$$\Pi_\lambda^* = \frac{\partial v}{\partial y} \frac{d_w^2}{\nu}$$

as the most informative local dimensionless feature for reconstructing λ_θ within the sampled Falkner–Skan domain. This feature has the same dimensional structure

as the pressure-gradient closure used in the baseline one-equation intermittency model, showing that the information-based analysis recovers its physically motivated variable structure.

- For the reconstruction of Re_θ , the analysis identified

$$Re_{U_e} = \frac{U_e d_w}{\nu}$$

as a more informative and wall-normal-consistent feature than the baseline vorticity Reynolds number,

$$Re_V = \frac{d_w^2 S}{\nu}.$$

In particular, the approximately monotonic variation of Re_{U_e} within the boundary layer makes it suitable for constructing a closure that remains non-activating at interior off-design locations.

- Candidate closures were assessed using three complementary criteria: reconstruction accuracy at the design wall-normal location, activation coherence away from that location, and algebraic parsimony. A common design location of $\eta_w^* = 1$ was selected so that both closures reproduce their corresponding integral targets at the same wall-normal position.
- The symbolic-regression procedure produced substantially more accurate local reconstructions than the baseline closures within the Falkner–Skan database. The selected pressure-gradient closure, given by

$$\lambda_{\theta_L} = \{ \Pi_\lambda^* [-4 \times 10^{-6} \Pi_\lambda^* - 6 \times 10^{-5}] (2.7171 - \Pi_\lambda^*) - 0.014251 \} \Pi_\lambda^* + 0.015503,$$

was chosen because it provided the lowest reconstruction error among candidates with comparable off-design behavior. The selected momentum-thickness closure,

$$Re_{\theta_L} = \frac{Re_{U_e}}{7.710}$$

was selected because it achieved nearly the same accuracy as the more complex candidates while retaining the simplest algebraic form.

- The a posteriori results demonstrate a clear distinction between closure-level accuracy and complete-model response. More accurate reconstructions of λ_θ and Re_θ do not necessarily produce a corresponding improvement in the predicted transition location, because the effect of the closures is mediated by the Re_{θ_c} correlation, F_{PG} , the onset functions, intermittency source terms, model limiters, and the coupling with the SST turbulence model.
- The not re-tuned M15 SR variant produces case-dependent changes across the T3 series, confirming that replacing the local closures alters the balance of the original model architecture. The limited recalibration introduced in M15 SR-R compensates for part of this response. Relative to the baseline model, the recalibrated variant retains comparable agreement for T3B, T3C1, T3C3, and T3C4 and improves the predictions for T3A, T3C2, and T3C5. However, the agreement deteriorates for the low-turbulence-intensity cases T3Am and TSK.

- The deterioration observed for T3Am and TSK suggests that, under very low free-stream turbulence, the remaining transition-onset correlations may become more limiting than the local λ_{θ_L} and Re_{θ_L} closures. In particular, the definitions of Tu_L and Re_{θ_c} require further examination before the benefits of the new closures can be fully exploited.

The present results support the use of information-based symbolic regression as a systematic and interpretable approach for constructing local algebraic closures. Nevertheless, the conclusions are restricted to the equilibrium, two-dimensional Falkner–Skan manifold used for closure discovery and to the T3 cases considered in the a posteriori assessment.

Future work should therefore focus on the following directions:

- Extending the closure-discovery database to include non-equilibrium, separated, compressible, and three-dimensional boundary layers, including crossflow effects. Such databases would make it possible to determine whether additional local variables are required and whether the present closure structures remain valid beyond the Falkner–Skan family. For larger candidate-variable sets, physical screening and sparsity regularization will remain necessary to control dimensional non-uniqueness and promote interpretable groups.
- Assessing the alternative local turbulence-intensity definition

$$Tu'_L = 100 \frac{\sqrt{2k/3}}{\tilde{U}_e}, \quad (58)$$

which provides a more direct approximation of the turbulence intensity based on the estimated edge velocity. Its introduction would require a dedicated recalibration of the Re_{θ_c} correlation.

- Developing a revised Re_{θ_c} correlation from a database containing consistent information about transition onset, intermittency, free-stream turbulence, and pressure-gradient effects. High-fidelity simulations and sufficiently resolved experiments could provide the required data. The information-based symbolic-regression methodology introduced in this work could then be applied to identify the most informative local dimensionless variables governing Re_{θ_c} and to construct a compact and interpretable replacement for the current empirical correlation.

Acknowledgements. This project has received funding from the Spanish Ministry of Defense under grant agreement “Potenciación de Capacidades de Producción de Aeronaves: Servicios de Análisis Aerodinámico para la Ingeniería de Integración de Cargas en Aeronaves del Ejército del Aire”.

This research has received funding from the European Union (ROSAS, project number 101138319). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

The authors gratefully acknowledge the “Centro de Supercomputación y Visualización de Madrid” (CESVIMA) and Universidad Politécnica de Madrid (UPM) for providing computing resources on Magerit Supercomputer.

The authors would like to express their gratitude to Gonzalo Rubio Calzado for the weekly discussions and feedback. In addition, special thanks are given to Antonio Medina-Manuel for insightful discussions on the applicability and coherence of this work.

References

- [1] Slotnick, J.P., Khodadoust, A., Alonso, J.J., Darmofal, D.L., Gropp, W.D., Lurie, E.A., Mavriplis, D.J.: CFD Vision 2030 Study: A Path to Revolutionary Computational Aerosciences. NASA Contractor Report NASA/CR-2014-218178, NASA (March 2014). <https://ntrs.nasa.gov/citations/20140003093>
- [2] Menter, F.R., Langtry, R.B., Likki, S.R., Suzen, Y.B., Huang, P.G., Völker, S.: A Correlation-Based Transition Model Using Local Variables—Part I: Model Formulation. *Journal of Turbomachinery* **128**(3), 413–422 (2006) <https://doi.org/10.1115/1.2184352>
- [3] Langtry, R.B., Menter, F.R., Likki, S.R., Suzen, Y.B., Huang, P.G., Völker, S.: A Correlation-Based Transition Model Using Local Variables—Part II: Test Cases and Industrial Applications. *Journal of Turbomachinery* **128**(3), 423–434 (2006) <https://doi.org/10.1115/1.2184353>
- [4] Lopes, R., Eça, L., Vaz, G.: On the Numerical Behavior of RANS-Based Transition Models. *Journal of Fluids Engineering* **142**(5), 051503 (2020) <https://doi.org/10.1115/1.4045576>
- [5] Langtry, R.B., Menter, F.R.: Correlation-Based Transition Modeling for Unstructured Parallelized Computational Fluid Dynamics Codes. *AIAA Journal* **47**(12), 2894–2906 (2009) <https://doi.org/10.2514/1.42362>
- [6] Menter, F.R., Smirnov, P.E., Liu, T., Avancha, R.: A One-Equation Local Correlation-Based Transition Model. *Flow, Turbulence and Combustion* **95**(4), 583–619 (2015) <https://doi.org/10.1007/s10494-015-9622-4>. Received 19 September 2014; accepted 22 May 2015; published online 5 July 2015; issue date December 2015
- [7] Menter, F., Matyushenko, A., Lechner, R., Stabnikov, A., Garbaruk, A.: An Algebraic LCTM Model for Laminar–Turbulent Transition Prediction. *Flow, Turbulence and Combustion* **109** (2022) <https://doi.org/10.1007/s10494-022-00336-8>
- [8] Malik, M.R.: Numerical Methods for Hypersonic Boundary Layer Stability. *Journal of Computational Physics* **86**(2), 376–413 (1990) [https://doi.org/10.1016/0021-9991\(90\)90106-B](https://doi.org/10.1016/0021-9991(90)90106-B)

- [9] Coder, J.G., Maughmer, M.D.: Application of the Amplification Factor Transport Transition Model to the Shear Stress Transport Model. In: 53rd AIAA Aerospace Sciences Meeting. American Institute of Aeronautics and Astronautics, Reston, VA (2015). <https://doi.org/10.2514/6.2015-0588>
- [10] Walters, D.K., Cokljat, D.: A Three-Equation Eddy-Viscosity Model for Reynolds-Averaged Navier–Stokes Simulations of Transitional Flow. *Journal of Fluids Engineering* **130**(12), 121401 (2008) <https://doi.org/10.1115/1.2979230>
- [11] Duraisamy, K., Iaccarino, G., Xiao, H.: Turbulence Modeling in the Age of Data. *Annual Review of Fluid Mechanics* **51**, 357–377 (2019) <https://doi.org/10.1146/annurev-fluid-010518-040547>
- [12] Ling, J., Kurzwski, A., Templeton, J.: Reynolds Averaged Turbulence Modelling Using Deep Neural Networks with Embedded Invariance. *Journal of Fluid Mechanics* **807**, 155–166 (2016) <https://doi.org/10.1017/jfm.2016.615>
- [13] Spalart, P.: An Old-Fashioned Framework for Machine Learning in Turbulence Modeling (2023). <https://arxiv.org/abs/2308.00837>
- [14] Brunton, S.L., Proctor, J.L., Kutz, J.N.: Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems. *Proceedings of the National Academy of Sciences* **113**(15), 3932–3937 (2016) <https://doi.org/10.1073/pnas.1517384113>
- [15] Schmelzer, M., Dwight, R.P., Cinnella, P.: Discovery of Algebraic Reynolds-Stress Models Using Sparse Symbolic Regression. *Flow, Turbulence and Combustion* **104**(2–3), 579–603 (2020) <https://doi.org/10.1007/s10494-019-00089-x>
- [16] Makke, N., Chawla, S.: Interpretable Scientific Discovery with Symbolic Regression: A Review. *Artificial Intelligence Review* **57**(1), 2 (2024) <https://doi.org/10.1007/s10462-023-10622-0>
- [17] Cranmer, M.: Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl. arXiv preprint arXiv:2305.01582 (2023) [arXiv:2305.01582](https://arxiv.org/abs/2305.01582) [cs.LG]
- [18] Pascal, L.: A Symbolic-Regression-Based Amplification Factor Transport Model for Transition Prediction. In: 60th 3AF International Conference on Applied Aerodynamics - AERO 2026, Paris, France (2026). <https://hal.science/hal-05549042>
- [19] Coupland, J. (ed.): ERCOFTAC Special Interest Group on Laminar to Turbulent Transition and Retrtransition: T3A and T3B Test Cases. A309514 (1990)
- [20] Dhawan, S., Narasimha, R.: Some Properties of Boundary Layer Flow During the Transition from Laminar to Turbulent Motion. *Journal of Fluid Mechanics* **3**(4), 418–436 (1958) <https://doi.org/10.1017/S0022112058000094>

- [21] Narasimha, R.: The Laminar–Turbulent Transition Zone in the Boundary Layer. *Progress in Aerospace Sciences* **22**(1), 29–80 (1985) [https://doi.org/10.1016/0376-0421\(85\)90004-1](https://doi.org/10.1016/0376-0421(85)90004-1)
- [22] Falkneb, V.M., Skan, S.W.: LXXXV. Solutions of the boundary-layer equations. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **12**(80), 865–896 (1931) <https://doi.org/10.1080/14786443109461870>
- [23] Yuan, Y., Lozano-Durán, A.: Dimensionless Learning Based on Information. *Nature Communications* **16**(1), 9171 (2025) <https://doi.org/10.1038/s41467-025-64425-8> . Received 5 March 2025; accepted 15 September 2025; published 16 October 2025
- [24] Hansen, N., Müller, S.D., Koumoutsakos, P.: Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). *Evolutionary Computation* **11**(1), 1–18 (2003) <https://doi.org/10.1162/10635603321828970>
- [25] Kraskov, A., Stögbauer, H., Grassberger, P.: Estimating mutual information. *Physical Review E* **69**(6) (2004) <https://doi.org/10.1103/physreve.69.066138>
- [26] Abu-Ghannam, B.J., Shaw, R.: Natural Transition of Boundary Layers—The Effects of Turbulence, Pressure Gradient, and Flow History. *Journal of Mechanical Engineering Science* **22**(5), 213–228 (1980) https://doi.org/10.1243/JMES_JOUR_1980_022_043_02
- [27] Menter, F.R.: Two-equation eddy-viscosity turbulence models for engineering applications. *AIAA Journal* **32**(8), 1598–1605 (1994) <https://doi.org/10.2514/3.12149>
- [28] Capel-Jorquera, J., Chávez-Modena, M., Valero, E., González, L.M.: Deep Learning for Inverse-Problem-Based Calibration of the γ Transition Model. *AIAA Journal* **63**(11), 4620–4638 (2025) <https://doi.org/10.2514/1.J065391>
- [29] Suluksna, K., Dechaumphai, P., Juntasaro, E.: Correlations for modeling transitional boundary layers under influences of freestream turbulence and pressure gradient. *International Journal of Heat and Fluid Flow* **30**, 66–75 (2008) <https://doi.org/10.1016/j.ijheatfluidflow.2008.09.004>
- [30] Zemmour, L., Gloerfelt, X., Cinnella, P.: A high-fidelity numerical database for free-stream transition. (2026). <https://api.semanticscholar.org/CorpusID:289629230>