Wall model based on a Mixture Density Network to predict the wall shear stress distribution for turbulent separated flows

Margaux Boxho[1,4*], Thomas Toulorge[1], Michel Rasquin[1], Grégoire Winckelmans[2], Grégory Dergham[3], Koen Hillewaert[1,4]

[1*]Cenaero, Charleroi, 6041, Belgium.
[2]Institute of Mechanics, Materials and Civil Engineering, UCLouvain, Louvain-la-Neuve, 1348, Belgium.
[3]Safran Tech, Châteaufort, 78114, Magny-les-Hameaux, France.
[4]Department of Aerospace and Mechanics, ULiege, Liège, 4000, Belgium.

*Corresponding author(s). E-mail(s): margaux.boxho@cenaero.be;

Abstract

Most wall shear stress models assume the boundary layer to be fully turbulent, at equilibrium, and attached. Under these strong assumptions, that are often not verified in industrial applications, these models predict an *averaged behavior*. To address the instantaneous and non-equilibrium phenomenon of separation, the Mixture Density Network (MDN), the neural network implementation of a Gaussian Mixture Model, initially deployed for uncertainty prediction, is employed as a wall shear stress model in the context of wall-modeled Large Eddy Simulations (wmLES) of turbulent separated flows. The MDN is trained to estimate the conditional probability $\boldsymbol{p(\tau_w|\mathbf{x})}$, knowing certain entries $\mathbf{x}$, to better predict the instantaneous wall shear stress $\boldsymbol{\tau_w}$ (which is then sampled from the distribution). In this work, an MDN is trained on a turbulent channel flow at the friction Reynolds number $\boldsymbol{Re_\tau}$ of 1,000 and on the two-dimensional periodic hill at the bulk Reynolds number of 10,595. The latter test case is known to feature a massive separation from the hill crest. By construction, the model outputs the probability distribution of the two wall-parallel components of the wall shear stress, conditioned by the model inputs: the instantaneous velocity field, the instantaneous and mean pressure gradients, and the wall curvature. Generalizability is ensured by carefully non-dimensionalizing databases with the kinematic viscosity and wall-model height. The relevance of the MDN model is evaluated *a posteriori* by performing wmLES using the in-house high-order Discontinuous Galerkin (DG) flow solver, named Argo-DG, on a turbulent channel flow at $\boldsymbol{Re_\tau = 2{,}000}$ and on the same periodic hill flow. The data-driven WSS model significantly improves the prediction of the wall shear stress on both the upper and lower walls of the periodic hill compared to quasi-analytical WSS models.

Keywords: Wall-modeled Large Eddy Simulation, Wall shear stress model, Machine Learning, Turbulent boundary layer, Separated flows, Conditional probability distribution

1 Introduction

Large Eddy Simulation (LES) is an effective tool for simulating turbulent flows. LES lies at the border between Direct Numerical Simulation (DNS), which resolves all scales, and Reynolds Averaged Navier-Stokes (RANS), which models all turbulent scales. However, LES is still computationally unfeasible for real engineering applications due to the rapid increase in resolution required to accurately resolve the small structures near the wall at high Reynolds numbers as estimated by Choi and Moin (2012). To reduce the computational cost, the energetic scales of turbulence in the innermost (i.e., the near-wall region) 10-20% of the boundary layer are modelled rather than resolved. Consequently, in a wall-modeled LES (wmLES), the mesh is coarsened near the solid walls, and a model is defined to relate the external physics to the wall. In the wmLES framework, models based on the law of the wall, which is valid for attached flow at moderate pressure gradients, are restricted in their ability to address complex, unsteady and non-equilibrium flow features, including secondary flows and separation. Other effects relevant to separation, such as pressure gradients, are not included in these models.

In an effort to address these issues, the Two-Layer Model (TLM), proposed by Balaras *et al.* (1996), and the Detached Eddy Simulation (DES), introduced by Spalart *et al.* (1997), have been developed as hybrid RANS/LES approaches. The main feature of DES and its modification, Delayed DES (DDES) (Spalart *et al.*, 2006b), is that a significant part of the boundary layer is treated by Reynolds-Averaged Navier-Stokes (RANS), while the separated flow regions are treated by LES. However, both TLM and DES have their own limitations. Although TLM has been successfully applied to simulate various scenarios, such as square ducts, rotating channels, and backward-facing steps, it still encounters two main issues: the log-layer mismatch and the resolved Reynolds stresses inflow. Additionally, TLM struggles with managing strong pressure gradients and flow separation effectively. Breuer *et al.* (2007) used the approach of artificial viscosity to build a simple analytical model if an appropriate definition of the relative thickness of the viscous sublayer is obtained. Encouraging results were obtained on the two-dimensional periodic hill at $Re_b = 10595$. Cadieux *et al.* (2016) addressed the separation problem using an integral wall model for LES with additional non-equilibrium terms. An analytically tractable integral formulation was obtained and successfully applied to a flat plate subjected to an adverse pressure gradient. Krank *et al.* (2019) used a turbulent boundary layer velocity profile model to enrich the Discontinuous Galerkin (DG) solution in the near wall region. Such a model has been applied to turbulent channel and periodic hill flows with encouraging results. Although much progress has been achieved in this field, most existing equilibrium wall models (Piomelli, 2008; Bose and Park, 2018) are still unable to predict flow separation and reattachment (e.g., boundary layer misalignment). The use of Deep Learning (DL) techniques offers a great opportunity to avoid making strong assumptions about the data, to incorporate the effects of the pressure gradient, and to predict the wall-parallel components of the wall shear stress (i.e., to cope with boundary layer misalignment), allowing non-equilibrium conditions such as turbulent separated flows to be addressed.

The advent of new hardware, such as GPUs and TPUs, combined with the exponential generation and accumulation of data, has made it possible to train deeper neural networks to support the generation of new engineering models. A deep neural network consists of a large number of parameters (up to $\approx$ 100 billion for the largest neural network trained in 2024 and used in the field of large language models) combined with non-linear activation functions that enable it to learn complex relationships between its inputs and outputs and to recognize features automatically through data assimilation.

Since the birth of the first neural network by Frank Rosenblatt in 1957, significant efforts have been made to train neural networks efficiently and to develop new architectures, such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long-Short Term Memory (LSTM), Residual Networks (ResNet), and UNet, which are specifically designed to perform dedicated tasks (Zhang *et al.*, 2023). In recent years, the fluid dynamics community has increasingly adopted ML and DL techniques for dimensionality reduction, closure models (e.g., RANS closures, wall models, and subgrid-scale models), and flow control, uncertainty quantification, and optimization (Duraisamy *et al.*, 2019; Brunton *et al.*, 2020). This work focuses on the use of neural networks to support the development of new wall shear stress models.

Wall models can be classified into two broad categories: hybrid RANS/LES (e.g., Detached Eddy Simulations Spalart *et al.* (2006a)) and wall stress modeling approaches. Each of these categories can be further subdivided into *standard* and *data-driven* approaches. The interested reader is referred to Heinz (2020) and references therein.

The rough principle of a wall shear stress model is to establish the local non-linear relationship between a set of entries (e.g., the velocity) and the wall shear stress. Most wall shear stress models take as input a single point located at a certain height h_{wm} in the wall-normal direction, or slightly shifted downstream, as proposed by Piomelli *et al.* (1989). Subsequently, the predicted wall shear stress is then imposed as a Neumann boundary condition. The reader is referred to the literature reviews by Piomelli and Balaras (2002), Piomelli (2008), and Larsson *et al.* (2016) for more details on the standard wall shear stress modeling approaches.

The last decade has seen the emergence of data-driven wall models due to the increase in computational resources for the generation of high-quality databases and the training of ML and DL techniques. Research was initially conducted on fully turbulent boundary layers for comparison with existing wall models, which have already proven successful under such equilibrium conditions. Yang *et al.* (2019) developed a physics-informed data-driven wall model for the channel flows by training MLPs on filtered DNS data of the channel flow at a friction Reynolds number Re_τ (defined as $Re_\tau = u_\tau h/\nu$, where u_τ is the friction velocity, h is the channel half height and ν the kinematic viscosity) of 1,000. The non-dimensionalization is inspired by the vertically integrated thin boundary layer equations and the eddy population density scaling to ensure a less aggressive extrapolation when increasing Re_τ in the *a posteriori* test. Radhakrishnan *et al.* (2021) focus on a wall shear stress model for channel flows using Gradient Boosted decision trees. Their model is trained on dimensionless features extracted from two channel flows ($Re_\tau = 180$ and $Re_\tau = 1{,}000$) and data synthetically generated by rescaling the instantaneous channel velocity fields. Their model performs similarly to the Equilibrium Wall Model (EQWM) on a channel at $Re_\tau = 2{,}005$ and fails to predict the streamwise velocity and stress profiles on a wall-mounted hump. Jamaat and Hattori (2023) *a priori* assess the performance of convolutional neural networks (CNN) for predicting the wall shear stress in channel flows. The model predicted the wall shear stress with high correlation coefficients for small y^+ values (defined as $y^+ = yu_\tau/\nu$, where y is the wall-normal distance), making the model unsuitable for realistic configurations. Furthermore, the non-local approach is not the most appropriate for massively parallel computations.

These wall shear stress models target the channel flow configuration and attempt to match or exceed the standard Law-of-the-wall (LOTW). Recent research also aims to overcome the existing weaknesses of standard wall models using ML and DL techniques.

Zhou *et al.* (2021) targets the turbulent separation by training a feed-forward neural network on the DNS of multiple geometries of the periodic hill (i.e., modification of the streamwise length). Although they obtained satisfactory *a priori* results, the *a posteriori* test fails in predicting the mean velocity profile on the nominal geometry at $Re_b = 10{,}595$. The work of Lozano-Durán and Bae (2023) is based on the hypothesis that any complex flow can be decomposed as a non-linear combination of simpler flows, called *building-block flows*. They developed a wall-flux-based wall model for LES using a self-critical machine-learning approach, which was successfully trained on DNS data (e.g., turbulent channel flows, turbulent Poiseuille–Couette flow with a strong adverse mean pressure gradient, or turbulent channel flow suddenly subjected to a spanwise pressure gradient). However, when applied to the NASA Juncture Flow, the model fails to predict the separation correctly. The lack of separated flows in the training data may be the origin of this failure. Dupuy *et al.* (2023a) focuses on data-driven wall models for separated flows by training a multi-layer perceptron (MLP) on filtered and sampled (according to Leonard's LES formalism) DNS data of two turbulent channel flows and the Stanford three-dimensional diffuser. The model is made invariant to the Mach number, Galilean invariant, and rotationally invariant. By increasing the spatial information (i.e., the input stencil size), the model can discriminate between separated and non-separated flow regions. *A posteriori*, their model performs better than the standard LOTW on the backward step, but there is still a misprediction in the recirculation bubble.

Dupuy *et al.* (2023b) continue their work on developing new wall shear stress (WSS) models by using Graph Neural Networks (GNN), which are appropriate tools for encoding unstructured data. The DNS data of three channel flows, a three-dimensional diffuser, a backward-facing step, an adverse pressure gradient test case, and a NACA blade at two angles of incidence are filtered onto coarser meshes (i.e., representative of LES meshes). The model is made orthogonal invariant, equivariant under rotation and reflection, independent of the coordinate system, Galilean invariant, and Mach number invariant. The authors observe that message passing steps (i.e., the capability of a GNN to exchange and aggregate information among the nodes) $N \geq 3$ are necessary to correctly predict wall shear stress and discriminate among the various flow configurations. The backward-facing step is simulated with their GNN-WSS model. Although the predictions are improved by increasing the message passing steps, the size of the recirculation bubble is still underestimated. Finally, their model is applied to the NACA65-009 blade at 7° of the angle of incidence and shows a clear improvement over to the standard LOTW. Lee *et al.* (2023) took a different direction in defining the input data for the model. The input variable used by standard LOTW can lead to poor performance when the configuration has complex flow physics. To circumvent this observation, they used the Fukagata-Iwamoto-Kasagi (FIK) identity to predict the skin friction coefficient coupled with an artificial neural network (ANN). *A posteriori*, the mean velocity profile agrees well with the DNS results at the lowest Re_τ. At higher Reynolds numbers, a log-layer mismatch is observed, which becomes even more pronounced when the mesh is coarsened. The model is also applied to the separated turbulent boundary layer flow in a periodic domain, where the upper wall is subjected to blowing and suction. The model shows relatively good agreement with the DNS reference for the separation and reattachment points. Zhideng *et al.* (2023) trained an MLP on DNS data of the two-dimensional periodic hill and synthetic data extracted from the LOTW to predict the two-wall parallel components of the wall shear stress. Although the model is trained on separated flows, it is only tested *a posteriori* on turbulent channel flows ($Re_\tau = 10^3$ to 1.2×10^8). They showed that as a standard wall model, the behavior of a data-driven wall model is also influenced by the numerical method (e.g., subgrid-scale model).

Existing data-driven wall models do not address the statistical aspect of the turbulent wall shear stress. These models are treated as a regression problem, where the network is trained using Mean Square Error (MSE) loss. However, most MSE-trained neural networks make a strong assumption about the conditional distribution of the output. Consequently, many authors have observed that the predicted WSS exhibits a *lower variance* than the actual or filtered DNS values. To address this issue, our work incorporates the prediction of the probability density function (PDF) of the wall shear stress instead of predicting a single-point estimate. For this purpose, a Mixture Density Network (MDN) is trained to predict the probability distribution as a linear combination of K Gaussian distributions. To the authors' knowledge, this work is the first attempt to develop a statistically-based wall shear stress model.

The development of novel data-driven WSS models into a high-order flow solver introduces a number of design constraints (or limitations). Firstly, the model must be instantaneous, i.e., the prediction of the wall shear stress at time t is based only on the volume data extracted at the same time t. This constraint prevents the storage of the flow history, which can be expensive in higher-order flow solvers. Secondly, the model must be local. Only the immediate vicinity of the prediction point can be used to infer the wall shear stress. This design constraint aims to reduce the amount of communication between partitions in highly scalable flow solvers. Thirdly, an important consideration related to the wall model robustness is its ability to predict the two wall-parallel components of τ_w, rather than making the prediction dependent on the velocity direction. This constraint is of interest for skewed and separated boundary layers. Fourthly, the model must be height-independent. By definition, the wall shear stress should remain constant regardless of the height at which the input data is measured in the boundary layer. Fifthly, a final constraint is that the model should be independent of the flow solver used to generate the database. This final constraint dictates that the model should not learn the numerical error. This last constraint is very difficult to satisfy because both DNS and wall-resolved LES (wrLES) are not perfect. Since this constraint is very strong, it is first relaxed to the constraint that the model should be independent of the polynomial order p. This last constraint is validated in Section 3.2.

The remainder of the paper is structured as follows. Section 2 describes the methodology employed to train the neural network and is subdivided into Section 2.1, which describes the generation and preprocessing of the database, and Section 2.2, which presents the selected neural network. Section 3 presents the *a priori* (Section 3.1) and *a posteriori* (Section 3.2) tests on both turbulent channel flows and the turbulent separated flow.

2 Methodology

This section discusses the development of a data-driven wall model, from data generation to neural network training. Emphasis is placed on the Mixture Density Network architecture and its loss function, which differs from the standard Mean Square Error loss.

2.1 Databases

Database generation is a crucial step in training data-driven WSS models. It directly impacts the model's ability to discriminate between different flow physics. The challenge is to identify test cases that best represent the complexity of fluid flows, which may feature laminar and turbulent boundary layers, transition, shock, and separation. The present objective focuses on the phenomenon of separation. For this purpose, a database of four test cases has been constructed, including both turbulent boundary layers (at

equilibrium and subjected to moderate adverse pressure gradients) and separated regions. Each test case provides three-dimensional time-dependent data, including the velocity, pressure gradient,, and wall shear stress. Three are obtained using the Argo-DG flow solver, and one is downloaded from the Johns Hopkins Turbulence Databases (JHTDB) website (Perlman *et al.*, 2007; Li *et al.*, 2008).

2.1.1 Simulation details

The turbulent channel flow and the two-dimensional periodic hill are simulated with the in-house code Argo-DG, developed at Cenaero. This code is a high-order Discontinuous Galerkin (DG) flow solver that solves the compressible Navier-Stokes equations. This solver (Hillewaert, February 2013) implements the Discontinuous Galerkin method with the Symmetric Interior Penalty (SIP) method for the discretization of diffusive terms. The code can operate on large cases due to the high scalability of the DG method. It implements a hybrid parallelism based on Message Passing Interface (MPI) and Open Multi-Processing (OpenMP). DG methods (DGM) are a specific class of the Galerkin finite element method (FEM) in which the shape functions defined in each element are not required to be continuous at the interfaces, leading to a compact and local set of discretized equations on the element. On the element interfaces, a Riemann solver treats the convective terms, while the SIP method controls the diffusive terms. Regarding time discretization, the BDF2 scheme, which is an implicit method, is employed. At each time step, a non-linear problem resulting from the implicit integration is solved by a Newton/GMRES method, preconditioned with an elementwise block-Jacobi technique. In terms of turbulence modeling, Argo-DG relies on the implicit LES (ILES) approach, for which the numerical dissipation of the underlying high-order DG scheme spectrally acts similarly to the explicit subgrid-scale models traditionally used in classical LES methods (Carton de Wiart *et al.*, 2014).

2.1.2 Test cases

The development of a novel WSS model is classified as supervised learning. In this category, it is necessary to create an explicitly labeled database containing pairs of inputs and outputs. Consider a joint probability distribution $p_{X,Y}$, where X represents the input (i.e., usually a vector of features or descriptors $\in \mathbb{R}^q$) and Y is the output (i.e., a real value or a category). The database is drawn from this distribution as,

$$(\boldsymbol{x}_i, y_i) \sim p_{X,Y}$$

where $x_i \in \mathcal{X}$, $y_i \in \mathcal{Y}$, $i = 1, \ldots, N$. The training data $\mathbf{d} = \{(\boldsymbol{x}_i, y_i) | i = 1, \ldots, N\}$ are generated identically and independently distributed (i.i.d.) with a finite size N. It is important to note that in practice, there is no prior information available about this joint probability. In the present case, the process behind this joint probability is the wall-resolved LES (wrLES) using Argo-DG, which approximates the solutions of the compressible Navier-Stokes equations for various test cases.

Turbulent channel flow at $Re_\tau = 950$. The geometry of the turbulent channel flow is periodic and homogeneous in the streamwise and spanwise directions. The flow evolves between two non-slip adiabatic walls separated by a distance 2δ (Hoyas and Jimenez, 2008). The computational domain sizes are $L_x/\delta = 2\pi$ and $L_z/\delta = \pi$ in the streamwise and spanwise directions, respectively. A uniform pressure gradient drives the flow to impose the exact friction Reynolds number. To ensure a fair comparison with the existing literature, the simulation is conducted at a low Mach number of $M = 0.1$, i.e., it can be considered almost incompressible. The simulation is performed with a mesh

resolution comparable to a wrLES: $\Delta x^+ \simeq 90$ and $\Delta z^+ \simeq 46$, where the superscript $+$ denotes the wall unit normalization (i.e., $\Delta x^+ = u_\tau \Delta x/\nu$). The Lagrange polynomial order p of the DG scheme is set to 3, resulting in an effective resolution of $\Delta x^+ \simeq 30$ and $\Delta z^+ \simeq 15$. Close to the wall, the effective resolution is set such that $\Delta y^+ \simeq 1$. After the numerical transient has been evacuated, the statistics are accumulated for approximately $13.9\,t^+$, where t^+ is defined as $t\,u_\tau^2/\nu$. It corresponds to 45.8 flow-through time (t_c), defined as $t\,u_b/L_x$, where u_b is the bulk velocity. The database is accumulated over a structured grid of probes of size $(N_x \times N_y \times N_z) = (192 \times 6 \times 192)$. The six wall-normal positions are located at $y^+ = \{10, 20, 50, 100, 150, 200\}$. This test case is subject to further analysis and validation in Boxho *et al.* (2022).

Turbulent channel flow at $Re_\tau = 1{,}000$ (JHTD, 2019). This test case is very similar to the previous one, but was obtained from a different flow solver. This test case has two objectives. The first is to get a wider range of y^+ values and the second is to briefly evaluate the last constraint discussed in Section 1 regarding the independence of the model on the flow solver. This test case is part of the Johns Hopkins Turbulence Databases (Graham *et al.*, 2016). A DNS of turbulent channel flow is performed in a domain of size $8\pi h \times 2h \times 3\pi h$. The incompressible Navier-Stokes equations are solved with a pseudo-spectral method in the two periodic and homogeneous directions (x and z), while a 7^{th} order B-spline collocation method is applied in the wall-normal direction. The simulation has required $2048 \times 512 \times 1536$ nodes. The channel flow is first driven by a bulk velocity of 1 and then switched to impose a uniform and constant pressure gradient. After accumulating the statistics, the friction velocity reaches the value of 0.0499. It corresponds to a friction Reynolds number of 999.35 with a kinematic viscosity of 5×10^{-5} [m^2/s]. From this online database, 16 snapshots of size $112 \times 35 \times 15$ of the velocity and the wall shear stress are extracted on both the upper and lower walls, for a total of 1,881,600 input/output pairs. Compared to the previous test case, the number of wall-normal positions has been extended to cover a wider range of y^+ values. The wall shear stress is scaled with the kinematic viscosity to get a friction velocity of 1 as follows,

$$\tau_w^\dagger = \left(\frac{\nu_{new}}{\nu}\right)^2 \tau_w \,, \tag{1}$$

where $\nu_{new} = 1/Re_\tau$. The velocity is scaled accordingly as $u^\dagger = (\nu_{new}/\nu)u$.

Two-dimensional periodic hill at $Re_b = 10{,}595$. This flow consists of a bi-periodic flow between two walls with a streamwise constriction, as shown in Figure 1. This test case is carefully designed to allow for a flow separation from the hill crest, followed by a massive recirculation bubble and the reattachment of the free shear layer on the flat bottom surface. The streamwise periodicity is fixed to $L_x/h = 9$ with a first hill located at $x/h = 0$ and a second at $x/h = L_x$. The spanwise periodicity is fixed to $L_z/h = 4.50$ to ensure sufficient decorrelation between the two periodic planes (Mellen *et al.*, 2000). A uniform pressure gradient whose magnitude is controlled to match the bulk Reynolds number ($Re_b = u_b h/\nu$) drives the flow. This procedure was initially proposed by Benocci and Pinelli (1990), but minor modifications were introduced by Carton de Wiart *et al.* (2015) to account for compressibility effects. This test case is the UFR 3-30 reference in the ERCOFTAC KB Wiki (ERCOFTAC, 2010) and has been extensively studied both numerically (Gloerfelt and Cinnella, 2015; Breuer *et al.*, 2009), and experimentally (Song and Eaton, 2004). The reader is referred to Fröhlich *et al.* (2005) and Gloerfelt and Cinnella (2015) for a detailed description of the 3D extruded geometry and a comprehensive analysis of the flow behavior.

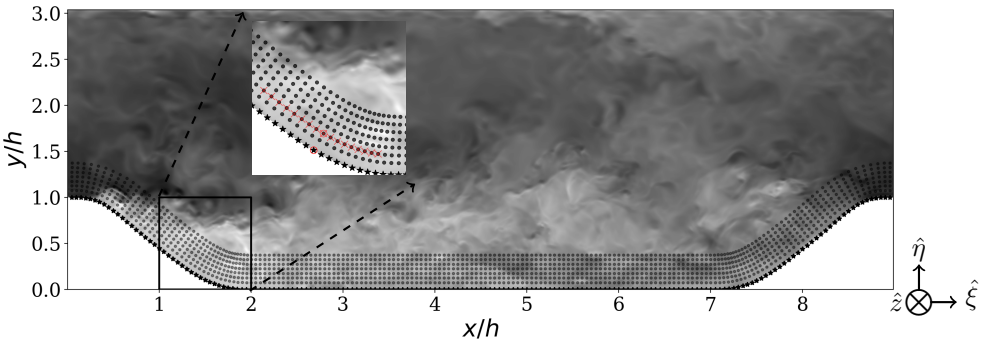


Fig. 1: Visualization of the instantaneous velocity field for the two-dimensional periodic hill generated with a dedicated plugin for the visualization of high-order fields with ParaView Rasquin *et al.* (2019); the hyperbolic probe grid used to interpolate the three-dimensional time-dependent flow fields is represented by the black dots; the corresponding surface probe grid, on which the wall shear stress is interpolated, is shown as black stars.

In the near-wall regions, where no-slip wall boundary conditions are imposed, representing approximately 10% of the hill height, the mesh is structured with a geometric progression to enforce a first cell size of $y^+ = 1$. The averaged grid sizes are discussed and shown in Figure 6b in Bose and Moin (2014). The rest of the mesh is unstructured with a refinement region located near the separation and in the free shear layer. The mesh is, therefore, composed of 445,000 hexahedra. The simulation is carried out using a polynomial Finite Element space of order $p = 3$, leading to 28 million degrees of freedom (dof) per equation. Considering the bulk velocity u_b, the spatial resolution of the mesh near the separation, the polynomial order, and the imposed time step, the convective CFL is maintained at about 0.3, which ensures that turbulence-related time structures are properly resolved. The flow was simulated at a low bulk Mach number of $M_b = 0.1$, and statistical data have been accumulated over about 40 flow-through times t_c (defined as $t_c = u_b/L_x$). This test case has been subjected to further analysis in Boxho *et al.* (2022).

This test case can be divided into two databases: (i) the lower wall and (ii) the upper wall. The lower wall is subjected to a strong separation leading to the development of a free shear layer which reattaches further downstream at $x/h \simeq 4.2$. This separation creates a pressure gradient that acts on the upper wall. The upper wall can therefore be considered as a turbulent boundary layer subject to a moderate pressure gradient. The database is extracted on a structured grid of probes composed of $(N_\xi \times N_\eta \times N_z) = (180, 40, 100)$, as illustrated in Figure 1. The lower wall grid is generated by a hyperbolic grid generator instead of just propagating straight lines normal to the wall. The velocity field and pressure gradient are interpolated on this grid and projected in the local reference frame (i.e., in curvilinear coordinates). The inset graph of Figure 1 shows a closer view of the hyperbolic probe grid. The volume fields are extracted on the black points, while the two wall-parallel components of the wall shear stress $\tau_{w,\xi}$ and $\tau_{w,z}$ are interpolated on the black stars obtained from the projection of the probe grid onto the surface.

Table 1 lists all the test cases. Some of the test cases are used both for training and the *a posteriori* tests. It is erroneous to assume that satisfactory training on a test case will necessarily lead to accurate predictions in the subsequent *a posteriori* test on the same

Table 1: Summary of the simulations employed for the training and validation, as well as for both the *a priori* and *a posteriori* tests.

Test cases	Training/Validation	A piori tests	A posteriori tests
TC950	✗	✓	✗
TC1000	✓	✗	✗
TC2000	✗	✗	✓
PHL10595	✓	✗	✓
PHU10595	✓	✗	✓

test case. Indeed, the *a posteriori* environment (i.e., a high-order flow solver) is very different from the training environment. In the *a posteriori* environment, the wall model begins to produce new values that will interact with the flow solver.

2.1.3 Preprocessing of the database

Database preprocessing is the set of procedures that prepares the three-dimensional time-dependent flow fields for the subsequent machine learning process. This step includes a number of operations, including filtering, cleaning, non-dimensional procedures, data structuring into an appropriate format, and data augmentation. The preprocessing stage will influence the capacity of the trained neural network to generalise to previously unseen configurations. It can also bring desired invariances into the data (such as Mach number invariance and rotational invariance, to name a few). The filtering process is frequently employed to ensure that the flow field resembles the one observed during the wmLES (Dupuy *et al.*, 2023b). However, the data is not explicitly filtered in this work, as the interpolation onto the probe grid already acts as a filtering process. Even if filtering is implicitly applied, the non-dimensionalization is explicitly applied to both the inputs and the outputs.

In fluid dynamics, non-dimensional variables that use integral or non-local flow quantities (e.g., the boundary layer height, the freestream velocity, etc.) are frequently utilized. Unfortunately, these quantities may not be the most relevant for generalization purposes and are not always available during the computation. The inner scaling of the boundary layer is an excellent choice for the non-dimensionalization. Nonetheless, the wall shear stress is not available to scale the input. However, the result of the product y^+u^+ is independent of u_τ and can be used to scale the input data. Moreover, the relation between y^+u^+ and y^+ is almost linear for a turbulent boundary layer. Frère (2018) employed a similar trick to tabulate the Reichardt LOTW and get a faster evaluation of the model in the high-order flow solver Argo-DG. The product y^+u^+ is nothing else than a local Reynolds number. This non-dimensionalization only requires the kinematic viscosity ν [m^2/s] and the wall model height h_{wm} [m], two quantities that are directly available for the wall model. Hence, h_{wm} and ν serve as a velocity scale (i.e., ν/h_{wm}) to scale the velocity field accordingly. The scaled velocity $\boldsymbol{u}^*$ is defined as,

$$\boldsymbol{u}^* = \frac{\boldsymbol{u}\, h}{\nu}\,, \tag{2}$$

where $\boldsymbol{u}$ is the velocity extracted at h_{wm}, and projected into the curvilinear coordinates (ξ, η, z) of the probe grid. The wall-model height is not used as a model input because it is implicitly defined in the non-dimensional velocity. Regarding the pressure gradient,

we first define it as a velocity $\boldsymbol{u}_p$ and then, this velocity is scaled as in Equation 2 to get,

$$\boldsymbol{u}_p^* = \frac{\boldsymbol{u}_p\, h}{\nu} \quad \text{where} \quad \boldsymbol{u}_p = \left(\frac{\nu}{\rho}\nabla p\right)^{1/3} . \tag{3}$$

Concerning the outputs, the two-wall parallel components of wall shear stresses are scaled as a signed version of y^+,

$$\boldsymbol{\tau}_w^* = \text{sign}\,(\boldsymbol{\tau}_w)\,\frac{y}{\nu}\sqrt{\frac{|\boldsymbol{\tau}_w|}{\rho}} . \tag{4}$$

The interesting fact behind using a square root for the non-dimensionalization of the wall shear stress is that this variable transformation reduces the skewness of the distribution. Although a Gaussian mixture network can recover the skewness of the underlying distribution, it may struggle to correctly predict this third statistical moment if the skewness is large. It is therefore of interest to reduce the skewness before training.

In addition to the non-dimensionalization, the data must be scaled for the training. The scaling constrains a given feature ϕ to lie between 0 and 1 by using the minimum and the maximum values of that feature as $\phi^\ddagger = \frac{\phi - \min(\phi)}{\max(\phi) - \min(\phi)}$. Assuming that the feature ϕ follows a normal distribution, it can be standardized using its mean μ and its standard deviation σ to obtain a rescaled distribution with zero mean and unit variance.

2.2 Neural Network Architecture

Artificial Neural Networks (ANNs) are a subset of ML and the core of deep learning algorithms. These learning algorithms aim to produce a function $f : \mathcal{X} \to \mathcal{Y}$ that maps the inputs to the outputs using a set of non-linear transformations (Zhang *et al.*, 2023). The present task of predicting the wall-parallel components of the shear stress, knowing a set of inputs, can be categorized as a regression problem. In this context, inference involves estimating the conditional probability $p(Y = y|X = \boldsymbol{x})$ for any new pair $(\boldsymbol{x}, y)$.

In the case of standard regression problems, this conditional probability is modeled by a Gaussian distribution with a fixed variance σ^2:

$$p(y|\boldsymbol{x}, f) = \mathcal{N}(y|f(\boldsymbol{x}), \sigma^2), \tag{5}$$

where $f(\boldsymbol{x})$ describes the expected value function. The prediction of this function/model f can be evaluated through a loss function $\mathcal{L} : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}^+$ so that $\mathcal{L}(y, f(\boldsymbol{x})) \geq 0$ measures how close the predictions $f(\boldsymbol{x})$ are to the actual value y. Under Equation 5, the well-known Mean Square Error (MSE) loss is obtained by maximizing the likelihood of the data over f.

Instead of considering a fixed variance, the variance can be, as the mean, a function of the input data such that,

$$p(y|\boldsymbol{x}, \mu, \sigma) = \mathcal{N}(y|\mu(\boldsymbol{x}), \sigma^2(\boldsymbol{x})), \tag{6}$$

where $\mu(\boldsymbol{x})$ and $\sigma^2(\boldsymbol{x})$ are parametric functions to be learned. For each input, the model does not predict a point estimate but a distribution of the output. The objective is to maximize the likelihood of the data over μ and σ:

$$\arg\max_{\mu,\sigma} p(\mathbf{d}|\mu, \sigma) = \arg\max_{\mu,\sigma} \prod_{\mathbf{x}_i, y_i \in \mathbf{d}} p(y_i|\mathbf{x}_i, \mu, \sigma)$$

$$
\begin{aligned}
&= \arg\max_{\mu,\sigma} \prod_{\mathbf{x}_i, y_i \in \mathbf{d}} \mathcal{N}(y_i | \mu(\mathbf{x}_i), \mathbf{x}_i(\sigma)) \\
&= \arg\max_{\mu,\sigma} \prod_{\mathbf{x}_i, y_i \in \mathbf{d}} \frac{1}{\sqrt{2\pi}\sigma(\mathbf{x}_i)} \exp\left(-\frac{(y_i - \mu(\mathbf{x}_i))^2}{2\sigma^2(\mathbf{x}_i)} \right) \\
&= \arg\min_{\mu,\sigma} \sum_{\mathbf{x}_i, y_i \in \mathbf{d}} \frac{(y_i - \mu(\mathbf{x}_i))^2}{2\sigma^2(\mathbf{x}_i)} + \log(\sigma(\mathbf{x}_i)) + \text{const.}
\end{aligned}
$$

In this last expression, a trade-off exists between the first and second terms. If the estimated value $\mu(\mathbf{x}_i)$ is far from the true value y_i, the difference in the numerator will be large and this term will drive the loss. To reduce the loss, the variance appearing in the denominator may increase. The second term serves to prevent an infinite variance and to maintain it within acceptable limits.

It can be assumed that any general distribution can be approximated by a mixture of simpler distributions. Consequently, the conditional distribution $p(y|\boldsymbol{x})$ can be modelled as a mixture of K Gaussian components (i.e., multi-modal Gaussian). In general, any probability distribution may be employed. However, the Gaussian distribution is frequently adopted due to its favorable mathematical properties and computational performance. Under this assumption, $p(y|\boldsymbol{x})$ is written as,

$$
p(y|\boldsymbol{x}, \mu_k, \sigma_k) = \sum_{k=1}^{K} \pi_k \mathcal{N}(y|\mu_k(\boldsymbol{x}), \sigma_k^2(\boldsymbol{x})), \tag{7}
$$

where $0 \leq \pi_k \leq 1$ for all k and $\sum_{k=1}^{K} \pi_k = 1$. A mixture density network is the neural network implementation of the Gaussian mixture model, and it is schematized in Figure 2. This figure shows a network composed of one head of K components, each producing a mean μ_k, a standard deviation σ_k, and a mixture coefficient π_k.

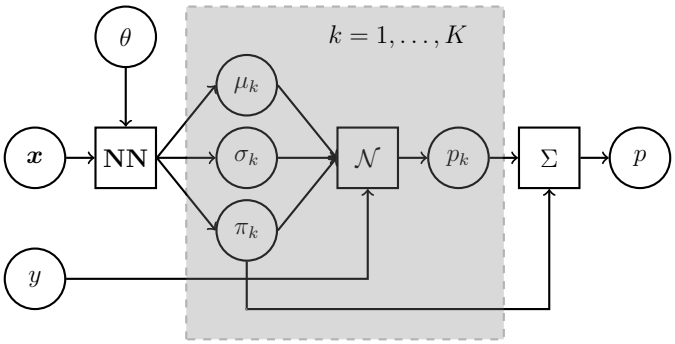


Fig. 2: Schematic of the architecture of a Mixture Density Network with K modes.

The aforementioned head can be connected to any neural network (denoted as **NN** in Figure 2), such as a multi-layer perceptron (MLP) or a convolutional neural network (CNN), to name a few. Following space-time correlations analysis (Boxho *et al.*, 2022) of turbulent and separated boundary layers, the input stencil has to be enlarged in the streamwise direction to capture the shifted correlation between the input features and the wall shear stress. Instead of just a single point, a series of streamwise locations will

be used to predict τ_w. The insert graph of Figure 1 illustrates a possible enlargement of the input stencil. The red cross indicates where the points are extracted at a given wall-normal location. The positioning of the points in the stencil is important and needs to be encoded in the network. The whole network is trained end-to-end using backpropagation applied to the Negative Log Likelihood.

Table 2 summarizes the parameters associated to the CNN-skip-GMH network. The CNN is composed of convolutional layers combined with residual blocks (i.e., one implementation of skip connections, originally proposed by He *et al.* (2015) in their ResNet), allowing for smoother gradients and better training. The network is consists of three convolutional blocks, two residual blocks, a linear block to switch from the convolutional part to the MDN, and two Gaussian Mixture Heads. One GMH is dedicated exclusively to the prediction of the conditional distribution of $\tau_{w,\xi}$ and one for $\tau_{w,z}$, respectively. Such construction imposes the independence between $\tau_{w,\xi}$ and $\tau_{w,z}$. Although no correlation between $\tau_{w,\xi}$ and $\tau_{w,z}$ have been observed (Boxho *et al.*, 2022), this assumption could still be debated for 2D extruded configurations. However, it may fail for more realistic three-dimensional configurations. One potential solution is to join the two GMHs by employing bivariate Gaussian distributions.

Table 2: Neural network hyperparameters (databases, pairs of input/output, size of the input stencil along the streamwise direction, optimizer, learning rate, activation function, batch size, number of Gaussian per head, etc.).

Architecture	**CNN-skip-GMH**
Train/Test data	Tab.1
Inputs	u^*, v^*, w^*, u_p^*, v_p^*, w_p^*, $\overline{u_p^*}$, $\overline{v_p^*}$, $\overline{w_p^*}$, $\mathcal{K}^*$
Input size	$[-5:10]$
Nb. of h_{wm}	5,5,35
Outputs	$\tau_{w,\xi}^*$, $\tau_{w,z}^*$
Batch size	512
Learning rate	10^{-3}
Weight regularization	10^{-5}
Number of epochs	10^3
Training size	$\simeq 4\,10^6$
Validation size	$\simeq 4\,10^5$
Optimizer	`torch.optim.Adam()`
Activation fct.	Sigmoid
Nb. of Gaussian/head	$K = 2$
Tot. nb. params	10,212

The model is trained on a stencil size of 16 with 5 points taken upstream and 10 points taken downstream, plus the current point at index 0. The analysis of the space-time correlations Boxho *et al.* (2022) has shown that the instantaneous and local wall shear stress is more correlated with the downstream velocity than with the upstream velocity at the same time step. Therefore, the input stencil is enlarged in the streamwise direction. Although this choice may seem counter-intuitive, it is not based on the concept of causality. This observation is also consistent with the work of Dupuy *et al.* (2023b),

where the prediction improved as the size of the message passing increased. With more points, the model can better discriminate between different flow physics.

3 Results

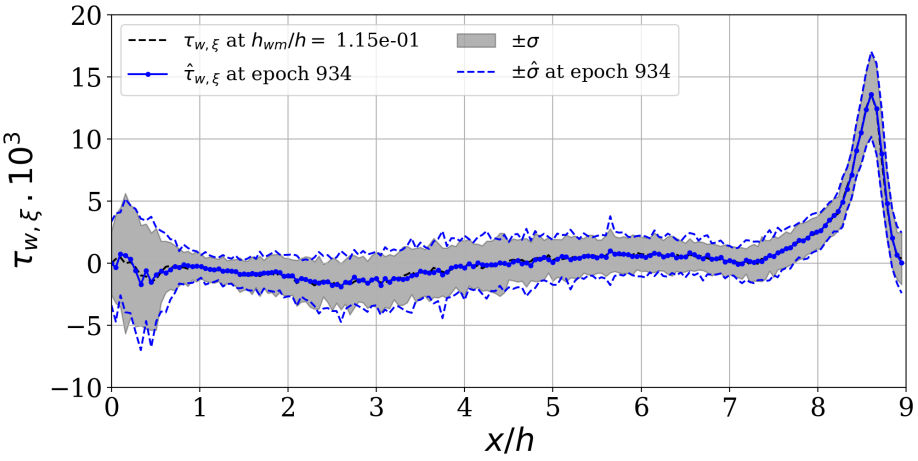
The CNN-skip-GMH network is trained for a predefined number of epochs and stopped at the minimum of the validation loss (i.e., early stop). At epoch 934, the validation loss (not shown here) starts to increase, indicating that the model is overfitting the training data, reducing its ability to generalize to the validation set. Therefore, the model parameters are frozen at epoch 934 and used for both the *a priori* (Section 3.1) and *a posteriori* (Section 3.2) tests, where the capabilities of the model once implemented in a high-order flow solver are evaluated.

3.1 *A priori* tests on the CNN-skip-GMH model

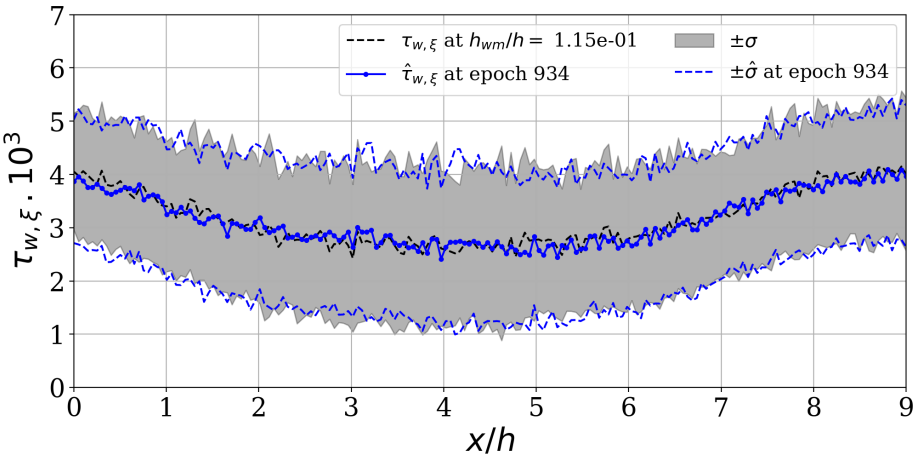
***A priori* evaluation on the lower wall.** The model is first validated on the lower wall of the two-dimensional periodic hill. Figure 3a shows the mean and standard deviation of the prediction at one height ($h_{wm}/h = 0.115$). The other four heights give similar results and are not shown here for the sake of brevity. Since the validation set is small, the target values and their corresponding predictions appear slightly noisy. The mean and the variance of the true wall shear stress are accurately captured by the model. The model is also capable to predict the large variance at locations $x/h \in [0, 0.7]$, which suggests that it is able to recognize that the separation point can vary within these specified bounds. The graph (y^+, u^+), shown in Figure 4a, is plotted based on the instantaneous wrLES data and the corresponding predictions. The four colored regions in green, cyan, pink, and yellow represent the distribution of (y^+, u^+) for the separation, recirculation, recovery, and acceleration regions, respectively. The recirculation region is located below Reichardt's profile with negative values, while the recovery region is above. These two regions appear to mirror each other. The separation is located in the highest y^+, corresponding to the part of the flow that accelerated strongly on the hill. Many points reach high u^+ values because the friction velocity tends to zero faster. The accelerated region slightly coincides with the separation and barely aligns with Reichardt's mean velocity profile. This graph reinforces the observation about the capability of the model to successfully predict the variance and that a WSS model based on the Reichardt's profile would have a hard time correctly predicting the separation location.

***A priori* evaluation on the upper wall.** Figure 3b shows the mean and standard deviation of the prediction on the upper wall. Predicting the wall shear stress on the upper wall turns out to be difficult. The wall shear stress $\tau_{w,\xi}$ is more correlated with the pressure gradient than with the velocity field. Therefore, the pressure gradient input greatly helps the network to closely match the ground truth for all streamwise positions. Figure 4b shows the graph (y^+, u^+) for the upper wall. The trend is different from a turbulent channel flow due to the impact of a moderate pressure gradient. The instantaneous pairs (y^+, u^+) are asymmetrically distributed around Reichardt's profile, with a trail of points extending over larger u^+ values. The data-driven model successfully captures the instantaneous trend compared to a quasi-analytical WSS model.

***A priori* test on TC950.** The model is trained on the test case TC1000 and is evaluated on TC950 obtained with Argo-DG at three heights: 100, 150, and 200 in wall units. Although the two test cases are almost identical, the objective is to evaluate *a priori* the independence of the model with respect to the flow solver (and therefore to some extent to the numerical errors). Table 3 summarizes the first three statistical moments



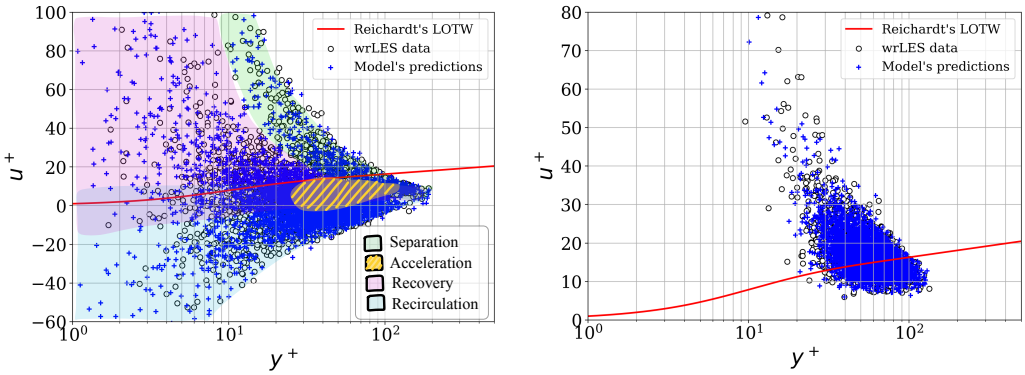
(a) PHL10595



(b) PHU10595

Fig. 3: *A priori* validation on two dimensional periodic hill at $Re_b = 10{,}595$. The target wall shear stress is plotted as a dashed black line, while the blue line represents the predictions averaged along the time and spanwise samples of the validation set; the gray area corresponds to one standard deviation and the predicted σ is plotted in dashed blue.

of the predicted wall shear stress. The predicted streamwise wall shear stress is in good agreement with the wrLES one. Although the model is not explicitly trained to match the skewness, this third moment of the distribution is also recovered. Predicting the spanwise wall shear stress appears more challenging for the network. It is possible that the non-zero spanwise wall shear stress is the result of a potential bias in the training database. This indicates the need for careful attention to the generation of the database. Although the database was generated over a sufficiently large number of flow-through times to allow the statistics to converge, subsampling could introduce bias into the statistics over the subset. This observation is a potential area for improvement.



(a) PHL10595 (b) PHU10595

Fig. 4: *A priori* validation: Instantaneous representation of the (y^+, u^+) graph, where the empty black circles correspond to the wall-resolved data and the blue '+' signs correspond to the (y^+, u^+) computed based on the predicted wall shear stress; Reichardt's mean velocity profile is shown in red. The left graph also identifies the different physics encountered on the lower wall (i.e., separation, acceleration, recovery and recirculation).

Table 3: Three first statistical moments of the wrLES $(\tau_{w,\xi}, \tau_{w,z})$ and predicted $(\hat{\tau}_{w,\xi}, \hat{\tau}_{w,z})$ wall parallel components of the wall shear stress. The unit of measurement of the data is the Pascal.

Statistics	$\tau_{w,\xi}$	$\hat{\tau}_{w,\xi}$	$\tau_{w,z}$	$\hat{\tau}_{w,z}$
μ	1.000	0.965	0.000	0.013
σ	0.424	0.423	0.284	0.352
S	1.008	0.936	-0.067	0.285

3.2 A posteriori tests on the CNN-skip-GMH model

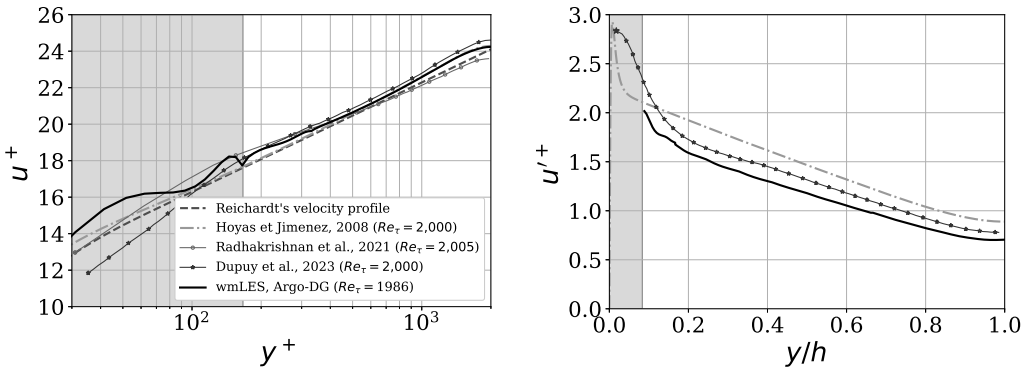
The new Mixture Density Network wall model is assessed *a posteriori* by implementing the CNN-skip-GMH network described in Section 2.2 in the Argo-DG flow solver presented in Section 2.1.1. The two-dimensional periodic hill at $Re_b = 10{,}595$ and the turbulent channel flow at $Re_\tau = 2{,}000$ are simulated with the new wall model.

3.2.1 Turbulent channel flow at $Re_\tau = 2{,}000$

This friction Reynolds number is selected because it has a logarithmic layer $135 \leq y^+ 300$ included in the training y^+ values (Marusic *et al.*, 2013). The domain size is the same as the channel flow at $Re_\tau = 950$, described in Section 2.1. The mesh is regular and uniform and composed of $(N_x, N_y, N_z) = (20, 24, 20)$ cells. The wmLES is performed at a polynomial degree $p = 4$. The effective resolution in the three directions is thus $(\delta x^+, \Delta y^+, \Delta z^+) \simeq (160, 42, 80)$. The instantaneous data are extracted at $y^+ = 180$ which is located in the second grid cell. The time step is fixed at $dt(u_\tau/h) = 1 \times 10^{-3}$ to match a CFL number of 0.28.

After evacuating the initial transient, statistical data is collected over about $30t_c$. The mean and Reynolds stress profiles are shown in Figures 5a and 5b, respectively. They are compared with the DNS data of Hoyas and Jimenez (2008) and to the wmLES of Dupuy

et al. (2023b) and of Radhakrishnan *et al.* (2021). The former used a Graph Neural Network to predict the wall shear stress, and the latter used Gradient Boosted Decision Tree. Dupuy *et al.* (2023b) used the same wall unit mesh refinement as for $Re_\tau = 950$. Therefore, the input velocity is extracted at $h_{wm}^+ \simeq 110$. Our results are compared with the finer mesh of Radhakrishnan *et al.* (2021) with a wall-normal height of $h_{wm}^+ = 31$.



(a) Mean streamwise velocity profile u^+

(b) Reynolds stresses profile u'^+

Fig. 5: *A posteriori* test on the turbulent channel flow at $Re_\tau = 2{,}000$; in plain black line, the data-driven wall model; in gray dash-dotted line, the DNS of Hoyas and Jimenez (2008); in circle symbols, the wmLES of Radhakrishnan *et al.* (2021); in dark star symbols, the wmLES of Dupuy *et al.* (2023b); and in dashed line, the Reichardt's mean velocity profile; the gray area indicates the size of the first grid cell.

Mean velocity profile. The mean profile agrees with the DNS results, except in the second cell ($166 \leq y^+ \leq 333$), where a relative error of 2% is measured. The oscillations of the mean velocity profile in the first cell (i.e., the under-resolved region) are directly linked to the DG method, as discussed in the thesis of Frère (2018). The mean profile of Radhakrishnan *et al.* (2021) is overestimated near the wall and underestimates the wake, while the one of Dupuy *et al.* (2023b) overestimates the DNS at every wall-normal coordinate. Regarding the mean velocity profile, our WSS model demonstrates competitive performance against existing data-driven models.

Mean Reynolds stresses. The Reynolds stress profile u'^+ is underestimated. However, such an underestimation has already been reported by Frère (2018) when using the WSS model based on Reichardt's velocity profile. The u'^+ profile of Dupuy *et al.* (2023b) underestimates the DNS with a smaller offset. Firstly, their wmLES grid is finer than ours. Secondly, their input velocity is extracted closer to the wall. Therefore, their boundary layer is better resolved, resulting in a better fit of u'^+ close to the wall.

3.2.2 Two-dimensional periodic hill flow at $Re_b = 10{,}595$

The CNN-skip-GMH model is *a posteriori* assessed on the two-dimensional periodic hill at the same bulk Reynolds number ($Re_b = 10{,}595$). At first sight, this task appears straightforward, since the model was trained on this specific test case. However, the production environment differs from the training environment as the wall model will interact with the resolved volume data. According to the best practices section of UFR 3-30[1], the flow physics is highly sensitive to the grid resolution and the numerical scheme,

[1] https://kbwiki.ercoftac.org/w/index.php/UFR_3-30_Best_Practice_Advice

which should not be too dissipative to ensure appropriate resolution of the scales and not to damp them. Moreover, the reattachment location is very sensitive to small errors. Accurate resolution of the free shear layer is of high importance. The non-equilibrium boundary layer within the recirculation bubble necessitates the accurate resolution of all terms in the Navier-Stokes equations to yield correct results. For these reasons, the wmLES mesh is very similar to the wrLES one, except cells within $0 \leq \eta/h \leq 0.1$ that are removed on both the upper and lower walls. The lower half of the mesh is refined to accurately capture the free shear layer, which is a region of high gradients.

Several experiments have been conducted and are summarized in Table 4. The parameters changed from one experiment to another are: (i) the WSS model, (ii) the alignment of the predicted wall shear stress, (iii) the polynomial order, and (iv) the wall model height. The objective is to verify that the constructed model satisfies the constraints listed in Section 1, and that the model produces satisfactory results in terms of mean wall shear stress, mean velocity, and Reynolds stress profiles.

(i) **The WSS model.** The CNN-skip-GMH data-driven wall model is compared to the Analytical WSS model based on Reichardt's velocity profile (AWSSR).

(ii) **The alignment of the predicted wall shear stress to the velocity.** The AWSSR predicts the magnitude of the wall shear stress and aligns it with the velocity extracted at h_{wm}. Our neural network predicts the wall-parallel components of the wall shear stress. Therefore, there is no explicit alignment of the predicted wall shear stress. For the test `DD-A-512-p4`, $\tau_{w,\xi}$ and $\tau_{w,z}$ are sampled from their respective distributions, the magnitude is evaluated, and then the wall shear stress is aligned with the velocity projected in the local frame of reference made by the solid wall.

(iii) **The polynomial order.** According to Frère (2018), the WSS model in Argo-DG gives better results for even polynomial degrees. Since the model was trained on data collected at $p3$, the data-driven wmLES is run at both $p3$ and $p4$. The $p4$ simulation is run on a slightly coarser mesh which allows an equivalent number of dof.

(iv) **The wall model height.** To ensure that the model is independent of the matching location, two h_{wm}/h are tested. A change in the matching location results in an adjustment of the first cell size.

Since the model has not been trained to capture the numerical transient, the simulation is restarted from a coarse wrLES of the same test case. The simulation `Rcht-A-512-p4` is also restarted from the same coarse wrLES for a fair comparison. For each wmLES, the time step is fixed at $\mathrm{d}t(u_b/h) = 5 \times 10^{-3}$. The Mach number is maintained at the low value of 0.1, and a pressure gradient is imposed to ensure the bulk Reynolds number, as previously mentioned in Section 2.1.

Table 4: Summary of the numerical experiments carried out on the two-dimensional periodic hill at $Re_b = 10{,}595$, where **NA** stands for **N**o **A**lignment, and **A** stands for **A**lignment; t_c corresponds to the number of flow through time.

	WSS model	**Align.**	p	**DOF**	h_{wm}/h	**Accum.**
`DD-NA-512-p3`	CNN-skip-GMH	✗	3	25,473,600	0.1	$\sim 36\,t_c$
`DD-NA-512-p4`	CNN-skip-GMH	✗	4	28,788,750	0.08	$\sim 25\,t_c$
`DD-A-512-p4`	CNN-skip-GMH	✓	4	28,788,750	0.1	$\sim 19\,t_c$
`Rcht-A-512-p4`	AWSSR	✓	4	28,788,750	0.1	$\sim 15\,t_c$

Mean friction coefficient. The four wmLES are compared to the wrLES reference used to train the data-driven wall model. Figure 6a shows the wall shear stress obtained on the lower wall. The mean separation location (Table 5) is correctly predicted by `DD-NA-512-p3` and `DD-NA-512-p4`. As also observed in the work of Krank *et al.* (2019) and Zhou *et al.* (2022), the friction is overestimated in the separation vicinity. The friction peak and the tiny separation before the windward foot of the next hill are also well captured. These two simulations show a good agreement with the reference, except near the reattachment. By imposing the alignment, `DD-A-512-p4` and `Rcht-A-512-p4` result in a delay of the mean separation line. `DD-A-512-p4` better predicts the friction peak while `Rcht-A-512-p4` misses the strong acceleration on the hill and underpredicts the peak. The mean reattachment location is underestimated by the four wmLES (Table 5), as also reported by Dupuy *et al.* (2023b) on the backward-facing step. The faster reattachment allows the flow to recover over a larger part of the flat bottom surface, generating a higher friction for $x/h \in [4, 7]$, at least for `DD-NA-512-p3` and `DD-NA-512-p4`.

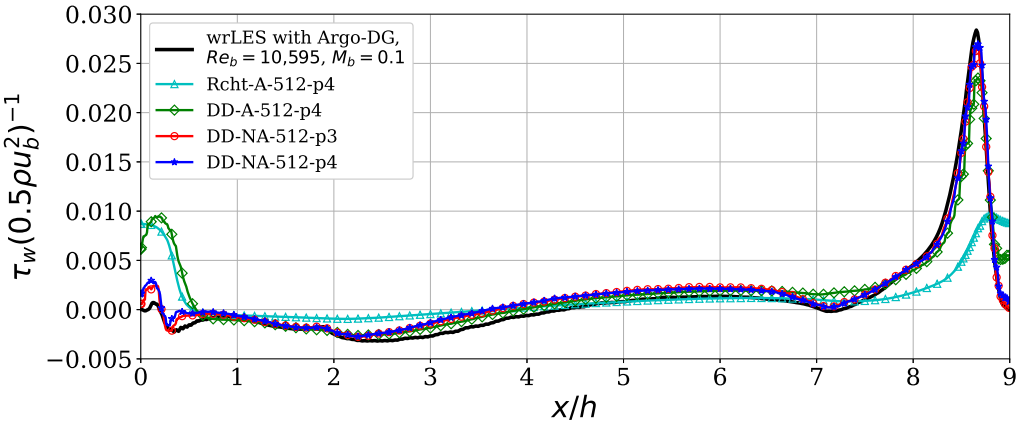
Table 5: Mean position of the separation and reattachment lines for the two-dimensional periodic hill at $Re_b = 10{,}595$.

	`DD-NA-512-p3`	`DD-NA-512-p4`	`DD-A-512-p4`	`Rcht-A-512-p4`	**Expe.**
x_{sep}/h	0.217	0.216	0.573	0.538	0.19
x_{reatt}/h	3.690	3.589	3.935	3.652	4.21

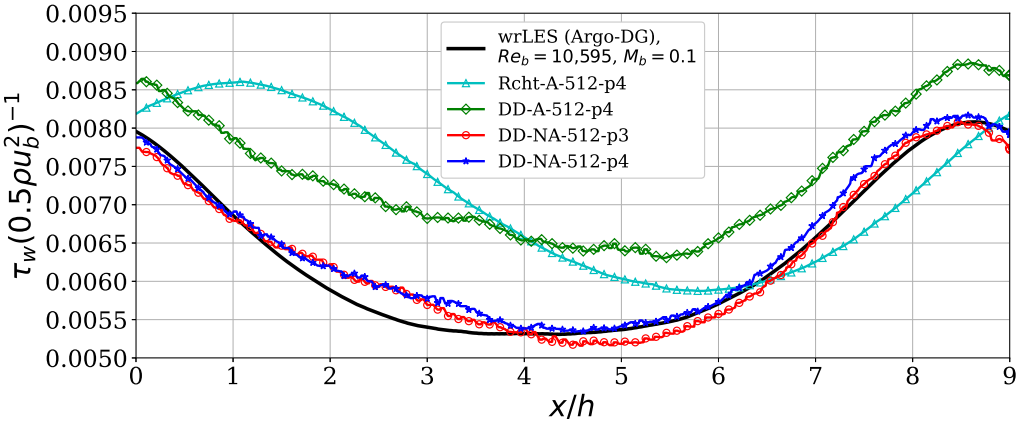
Figure 6b shows the friction on the upper wall of the periodic hill. Since the friction on this wall is more correlated with the pressure gradient than with the velocity, the friction imposed by `Rcht-A-512-p4` does not match the reference. The predicted friction of `DD-NA-512-p3` and `DD-NA-512-p4` is in good agreement with the reference. However, `DD-A-512-p4` overestimates the friction on the upper wall at every location.

Mean velocity profiles. The mean velocity profiles are shown in Figure 7 at ten stations ($x/h = 0.05$, 0.5, 1, 2, 3, 4, 5, 6, 7, 8). Overall, good agreement with the reference is observed for the four wmLES. `DD-NA-512-p3` and `DD-NA-512-p4` accurately predict the mean velocity profile at the separation, even in the first cell (the grey area in Figure 7), compared to `DD-A-512-p4` and `Rcht-A-512-p4` which underestimate the acceleration at the hill top. Other minor discrepancies were identified due to the early reattachment of the free shear layer for `DD-NA-512-p3` and `DD-NA-512-p4`. The faster reattachment allows the flow to recover over a larger portion of the flat bottom surface, as shown by the slight overestimation of the velocity for $3 \leq x/h \leq 7$. Conversely, the velocity profiles near the reattachment are better captured by `DD-A-512-p4` and `Rcht-A-512-p4`, although $\overline{u}$ is overestimated on the upper wall, which is consistent with the overestimation of the friction coefficient in Figure 6b.

Mean Reynolds stresses. The most significant discrepancy in the Reynolds stress $\overline{u'u'}$ profiles (Figure 8) is within the recirculation bubble ($1 \leq x/h \leq 4$). `DD-NA-512-p3` and `DD-NA-512-p4` overestimate $\overline{u'u'}$, while `DD-A-512-p4` underestimates this component of the Reynold stress. Although `Rcht-A-512-p4` mispredicts the friction on both the lower and upper walls, it accurately predicts $\overline{u'u'}$ throughout the domain except in the near-wall region on the lower wall where it overpredicts this component. Regarding the covariance $\overline{u'v'}$ in Figure 9, all simulations are in agreement with the



(a) PHL10595



(b) PHU10595

Fig. 6: Mean streamwise wall shear stress measured on the solid walls of the two-dimensional periodic hill at $Re_b = 10{,}595$; in black solid line, the reference wrLES obtained with Argo-DG; in cyan triangle symbols, the wmLES using the WSS model based on Reichardt's velocity profile; in green diamonds, the `DD-A-512-p4`; in red circles, the `DD-NA-512-p3`; in blue stars, the `DD-NA-512-p4`.

reference, except inside the recirculation bubble where a slight deviation is observed.

Summary. It is difficult to determine the origin of the discrepancy in the predicted reattachment location in this specific configuration. Similarly, it is unclear why `Rcht-A-512-p4` produces better results for the mean velocity profiles and Reynolds stresses even though it mispredicts the wall shear stress. Compensation errors may exist between different terms that allow for such results, but these have not yet been quantified. Nevertheless, the principal distinction between (`DD-NA-512-p3`, `DD-NA-512-p4`) and (`DD-A-512-p4`, `Rcht-A-512-p4`) is the imposition of the alignment with the velocity. Many standard WSS models in the literature naturally align the wall shear stress with the velocity extracted at the matching location. At the separation, there is a strong misalignment between τ_w and u. In wall models, the magnitude of the wall shear stress is not the only relevant factor, and misalignment issues should be treated differently.

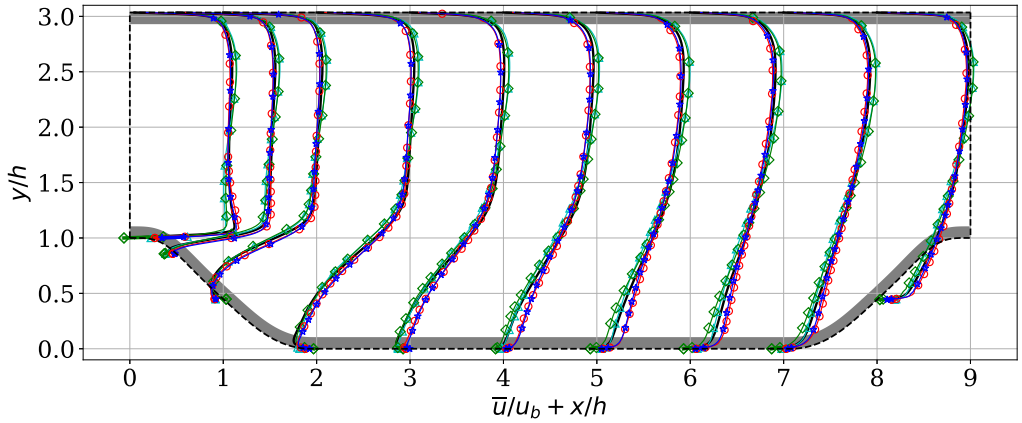


Fig. 7: Mean velocity profiles of the two-dimensional periodic hill at $Re_b = 10{,}595$ (see legend of Figure 6a).

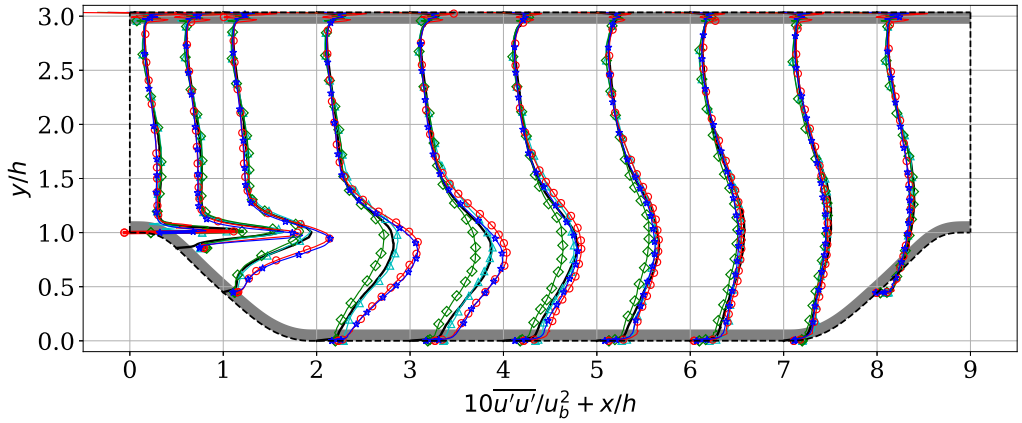


Fig. 8: Reynolds stress $\overline{u'u'}$ profiles of the two-dimensional periodic hill at $Re_b = 10{,}595$ (see legend of Figure 6a).

4 Conclusion and perspectives

To the authors' knowledge, this work is the first attempt to create a statistically based WSS model to address separated flows. The model consists of a CNN connected to a Mixture Density Network to predict the wall shear stress distribution as a linear combination of K Gaussians. This network satisfies the design constraints stated in Section 1 and predicts the conditional probability distribution of the wall-parallel component of the wall shear stress. The model was trained on the turbulent channel flow at $Re_\tau = 1{,}000$ and data extracted from the lower and upper walls of the two-dimensional periodic hill at $Re_b = 10{,}595$, two computationally feasible test cases with turbulent boundary layers, one at equilibrium, one subjected to a moderate pressure gradient, and one featuring a massive separation. The data collected during the scale-resolved simulations of these test cases is not fed directly into the network. The preprocessing of the databases and their in-depth analysis (Boxho *et al.*, 2022) constituted a crucial step towards a new data-driven WSS model. For this purpose, the input/output pairs are non-dimensionalized using local quantities directly available to the model.

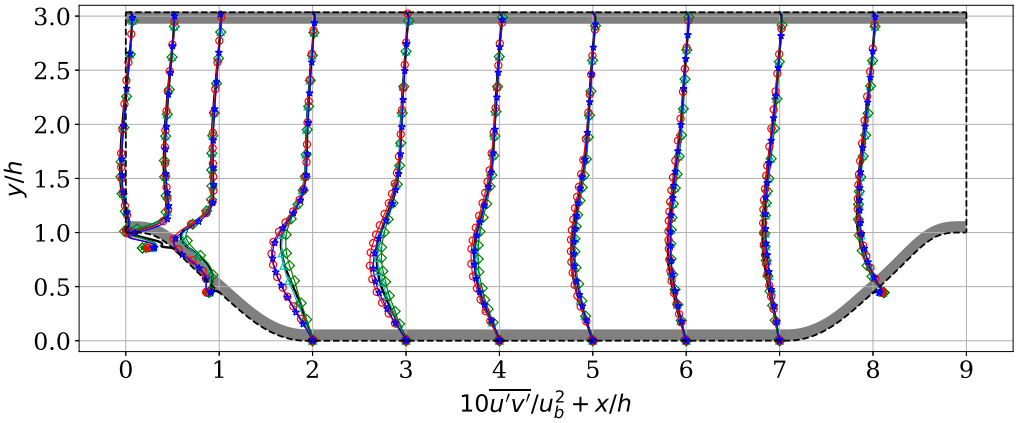


Fig. 9: Covariance between u and v of the two-dimensional periodic hill at $Re_b = 10{,}595$ (see legend of Figure 6a).

The model is *a posteriori* tested on two configurations: a turbulent channel flow at $Re_\tau = 2{,}000$ and the periodic hill at $Re_b = 10{,}595$. The results for both configurations are promising. Regarding the periodic hill, a clear improvement over the analytical WSS model based on Reichardt's velocity profile is observed in the prediction of the wall shear stress on the upper and lower walls. Although the separation point is accurately predicted, the reattachment location is shifted upwards compared to the DNS predictions. Such an underestimation of the recirculation bubble size is also observed in other studies applying data-driven wall shear stress models to separated flows (Zhou *et al.*, 2021; Lozano-Durán and Bae, 2021; Dupuy *et al.*, 2023b).

Perspectives. A first short-term perspective is to combine the two Gaussian heads into a single bivariate Gaussian head. This approach avoids the strong assumption of independence between the two wall-parallel components of the wall shear stress, discussed in Section 2.2. A second short-term perspective is to extend the input stencil in the spanwise direction. The input is then transformed into a two-dimensional image. This extension is consistent with the spanwise space-time correlations presented in Boxho *et al.* (2022). A third perspective is the use of Graph Neural Networks (GNNs), a powerful type of neural network designed to encode unstructured data. The use of GNNs eliminates the need for an additional probe grid on which to interpolate the input data by directly exploiting the connectivity of the computational mesh. Dupuy *et al.* (2023b) have already trained this type of network, but they still observed an underestimation of the recirculation bubble size, suggesting that further improvement is possible.

Acknowledgments. The funding of M. Boxho's PhD project by Safran Tech is gratefully acknowledged. The present research benefited from computational resources made available on the Tier-1 supercomputer Zenobe of the Fédération Wallonie-Bruxelles, infrastructure funded by the Walloon Region under the grant agreement n°1117545, and from computational resources made available on Lucia, the new Tier-1 supercomputer of the Walloon Region, infrastructure funded by the Walloon Region under the grant agreement n°1910247.

Declarations

Some journals require declarations to be submitted in a standardised format. Please check the Instructions for Authors of the journal to which you are submitting to see if you need to complete this section. If yes, your manuscript must contain the following sections under the heading 'Declarations':

- Funding: Safran Tech and computational resources made available on the Tier-1 supercomputer of the Fédération Wallonie-Bruxelles, infrastructure funded by the Walloon Region under the grant agreement n°1117545.

- Conflict of interest/Competing interests: Not applicable
- Ethics approval: Not applicable
- Consent to participate: Not applicable
- Consent for publication: Not applicable
- Availability of data and materials: Data can be made available on specific and motivated demand.
- Code availability: Proprietary code that may be made available free of charge under a specific Software License Agreement for academic research purposes.
- Authors' contributions:
  - Conceptualization: [Koen Hillewaert, Grégoire Winckelmans, Thomas Toulorge, Margaux Boxho];
  - Data curation: [Margaux Boxho];
  - Formal analysis: [Margaux Boxho, Koen Hillewaert];
  - Funding acquisition: [Koen Hillewaert];
  - Investigation: [Margaux Boxho];
  - Methodology: [Margaux Boxho, Koen Hillewaert, Grégoire Winckelmans];
  - Project administration: [Koen Hillewaert, Grégoire Winckelmans, Thomas Toulorge, Grégory Dergham];
  - Resources: [Cenaero, Safran Tech];
  - Software: [Margaux Boxho, Koen Hillewaert, Michel Rasquin, Thomas Toulorge, Grégory Dergham];
  - Supervision: [Koen Hillewaert, Grégoire Winckelmans, Thomas Toulorge, Michel Rasquin, Grégory Dergham];
  - Validation: [Margaux Boxho, Koen Hillewaert, Grégoire Winckelmans, Thomas Toulorge, Michel Rasquin, Grégory Dergham];
  - Visualization: [Margaux Boxho];
  - Writing - original draft preparation: [Margaux Boxho];
  - Writing - review and editing: [Margaux Boxho, Koen Hillewaert, Grégoire Winckelmans, Thomas Toulorge, Michel Rasquin, Grégory Dergham];

References

Balaras E, Benocci C, Piomelli U (1996) Two-layer approximate boundary conditions for large-eddy simulations. AIAA J 34(6):1111—1119. https://doi.org/10.2514/3.13200

Benocci C, Pinelli A (1990) The role of the forcing term in the large eddy simulation of equilibrium channel flow. In: International Symposium On Engineering Turbulence Modelling And Measurements. Elsevier, URL https://api.semanticscholar.org/CorpusID:129356123

Bose ST, Moin P (2014) A dynamic slip boundary condition for wall-modeled large-eddy simulation. Physics of Fluids 26(1):015104. https://doi.org/10.1063/1.4849535

Bose ST, Park GI (2018) Wall-modeled large-eddy simulation for complex turbulent flows. Annual Review of Fluid Mechanics 50:535–561. https://doi.org/10.1146/annurev-fluid-122316-045241

Boxho M, Rasquin M, Toulorge T, *et al.* (2022) Analysis of Space-Time Correlations to Support the Development of Wall-Modeled LES. Flow, Turbulence and Combustion 109(4):1081–1109. https://doi.org/10.1007/s10494-022-00365-3

Breuer M, Kniazev B, Abel M (2007) Development of wall models for LES of separated flows using statistical evaluations. Computers & Fluids 36(5):817–837. https://doi.org/10.1016/j.compfluid.2006.09.001

Breuer M, Peller N, Rapp C, *et al.* (2009) Flow over periodic hills – numerical and experimental study in a wide range of Reynolds numbers. Computers & Fluids 38(2):433–457. https://doi.org/10.1016/j.compfluid.2008.05.002

Brunton SL, Noack BR, Koumoutsakos P (2020) Machine learning for fluid mechanics. Annual Review of Fluid Mechanics 52:477—-508. https://doi.org/10.1146/annurev-fluid-010719-060214

Cadieux F, Sadique J, Yang XI, *et al.* (2016) Wall-Modeled Large Eddy Simulation of Laminar and Turbulent Separation Bubble Flows. In: 46th AIAA Fluid Dynamics Conference. AIAA, pp 2016–3189, https://doi.org/10.2514/6.2016-3189

Choi H, Moin P (2012) Grid-point requirements for large eddy simulation: Chapman's estimates revisited. Phys Fluids 24(1):011702. https://doi.org/10.1063/1.3676783

Dupuy D, Odier N, Lapeyre C (2023a) Data-driven wall modeling for turbulent separated flows. Journal of Computational Physics 487:112173. https://doi.org/10.1016/j.jcp.2023.112173, URL https://www.sciencedirect.com/science/article/pii/S0021999123002681

Dupuy D, Odier N, Lapeyre C, *et al.* (2023b) Modeling the wall shear stress in large-eddy simulation using graph neural networks. Data-Centric Engineering 4:e7. https://doi.org/10.1017/dce.2023.2, URL https://www.cambridge.org/core/product/identifier/S2632673623000023/type/journal_article

Duraisamy K, Iaccarino G, Xiao H (2019) Turbulence modeling in the age of data. Annual Review of Fluid Mechanics 51:357—377. https://doi.org/10.1146/annurev-fluid-010518-040547

ERCOFTAC (2010) URF 3-30 References. https://www.kbwiki.ercoftac.org/w/index.php/UFR_3-30_References

Frère A (2018) Towards wall-modeled Large-Eddy Simulations of high Reynolds number airfoils using a discontinuous Galerkin method. PhD thesis, UCL - Université Catholique de Louvain

Fröhlich J, Mellen CP, Rodi W, *et al.* (2005) Highly resolved large-eddy simulation of separated flow in a channel with streamwise periodic constrictions. J Fluid Mech 526:19–66. https://doi.org/10.1017/S0022112004002812

Gloerfelt X, Cinnella P (2015) Investigation of the flow dynamics in a channel constricted by periodic hills. In: 45th AIAA Fluid Dynamics Conference. AIAA

Graham J, Kanov K, Yang X, *et al.* (2016) A web services-accessible database of turbulent channel flow and its use for testing a new integral wall model for les. Journal of Turbulence 17(2):181–215. https://doi.org/http://www.tandfonline.com/doi/pdf/10.1080/14685248.2015.1088656

He K, Zhang X, Ren S, *et al.* (2015) Deep residual learning for image recognition. https://doi.org/10.48550/arXiv.1512.03385, URL http://arxiv.org/abs/1512.03385, 1512.03385[cs]

Heinz S (2020) A review of hybrid RANS-LES methods for turbulent flows: Concepts and applications. Progress in Aerospace Sciences 114:100597. https://doi.org/10.1016/j.paerosci.2019.100597

Hillewaert K (February 2013) Development of the Discontinuous Galerkin method for high-resolution, large scale CFD and acoustics in industrial geometries. PhD thesis, Ecole polytechnique de Louvain/iMMC

Hoyas S, Jimenez J (2008) Reynolds number effects on the Reynolds-stress budgets in turbulent channels. Physics of Fluids 20:101511. https://doi.org/10.1063/1.3005862

Jamaat GT, Hattori Y (2023) *A priori* assessment of nonlocal data-driven wall modeling in large eddy simulation. Physics of Fluids 35(5):055117. https://doi.org/10.1063/5.0146770, URL https://pubs.aip.org/pof/article/35/5/055117/2888149/A-priori-assessment-of-nonlocal-data-driven-wall

JHTD (2019) Turbulent channel flow. https://doi.org/10.7281/T10K26QW

Krank B, Kronbichler M, Wall WA (2019) A multiscale approach to hybrid RANS/LES wall modeling within a high-order Discontinuous Galerkin scheme using function enrichment. Int J Numer Meth Fluids 90(2):81–113. https://doi.org/10.1002/fld.4712

Larsson J, Kawai S, Bodart J, *et al.* (2016) Large eddy simulation with modeled wall-stress: recent progress and future directions. Mechanical Engineering Reviews 3(1):15–00418. https://doi.org/10.1299/mer.15-00418

Lee YM, Lee JH, Lee J (2023) Artificial neural network-based wall-modeled large-eddy simulations of turbulent channel and separated boundary layer flows. Aerospace Science and Technology 132:108014. https://doi.org/10.1016/j.ast.2022.108014

Li Y, Perlman E, Wan M, *et al.* (2008) A public turbulence database cluster and applications to study lagrangian evolution of velocity increments in turbulence. Journal of Turbulence 9(31)

Lozano-Durán A, Bae HJ (2021) Self-critical machine-learning wall-modeled LES for external aerodynamics. arXiv:201210005 [physics] URL http://arxiv.org/abs/2012.10005.

Lozano-Durán A, Bae HJ (2023) Machine learning building-block-flow wall model for large-eddy simulation. Journal of Fluid Mechanics 963:A35. https://doi.org/10.1017/jfm.2023.331

Marusic I, Monty JP, Hultmark M, *et al.* (2013) On the logarithmic region in wall turbulence. Journal of Fluid Mechanics 716(R3). https://doi.org/10.1017/jfm.2012.511

Mellen CP, Fröhlich J, Rodi W (2000) Large-eddy simulation of the flow over periodic hills. In: 16th IMACS World Congress, Lausanne, Switzerland

Perlman E, Burns R, Li Y, *et al.* (2007) Data exploration of turbulence simulations using a database cluster. Supercomputing SC07, ACM, IEEE

Piomelli U (2008) Wall-layer models for large-eddy simulations. Progress in Aerospace Sciences 44(6):437–446. https://doi.org/10.1016/j.paerosci.2008.06.001

Piomelli U, Balaras E (2002) Wall-Layer Models for large Eddy Simulations. Annual Review of Fluid Mechanics 34:349–374. https://doi.org/10.1146/annurev.fluid.34.082901.144919

Piomelli U, Moin P, Ferziger JH, *et al.* (1989) New approximate boundary conditions for large-eddy simulations of wall-bounded flows. Physics of Fluids A: Fluid Dynamics 1 pp 1061—-1068. https://doi.org/10.1063/1.857397

Radhakrishnan S, Gyamfi LA, Miró A, *et al.* (2021) A Data-Driven Wall-Shear Stress Model for LES Using Gradient Boosted Decision Trees. In: Jagode H, Anzt H, Ltaief H, *et al.* (eds) High Performance Computing, vol 12761. Springer International Publishing, p 105–121, https://doi.org/10.1007/978-3-030-90539-2_7, URL https://link.springer.com/10.1007/978-3-030-90539-2_7, series Title: Lecture Notes in Computer Science

Rasquin M, Bauer A, Hillewaert K (2019) Scientific post hoc and in situ visualisation of high-order polynomial solutions from massively parallel simulations. International Journal of Computational Fluid Dynamics 4(22):171–180. https://doi.org/10.1080/10618562.2019.1618453

Song S, Eaton J (2004) Flow structures of a separating, reattaching, and recovering boundary layer for a large range of Reynolds number. Experiments in Fluids 36:642—-653. https://doi.org/10.1007/s00348-003-0762-2

Spalart P, Deck S, Shur M, *et al.* (2006a) A New Version of Detached-Eddy Simulation, Resistant to Ambiguous Grid Densities. Theoretical and Computational Fluid Dynamics 20:181–195. https://doi.org/10.1007/s00162-006-0015-0

Spalart PR, H. JW, M. S, *et al.* (1997) Comments on the feasibility of LES for wings and on a hybrid RANS/LES approach. Advances in DNS/LES: Direct numerical simulation and large eddy simulation pp 137–148

Spalart PR, Deck S, Shur ML, *et al.* (2006b) Approximate wall boundary conditions for large eddy simulations. Theor Comput Fluid Dyn 20(3):181–195

Carton de Wiart C, Hillewaert K, Bricteux L, *et al.* (2014) Implicit LES of free and wall bounded turbulent flows based on the discontinuous Galerkin/symmetric interior penalty method. International Journal of Numerical Methods in Fluids 78(6):335—-354. https://doi.org/10.1002/fld.4021

Carton de Wiart C, Hillewaert K, Bricteux L, *et al.* (2015) LES using a Discontinuous Galerkin method: Isotropic turbulence, channel flow and periodic hill flow. In: Fröhlich, J, Kuerten, H, Geurts, B, Armenio, V (eds) Direct and Large-Eddy Simulation IX ERCOFTAC Series, vol 20 Springer, Cham https://doi.org/10.1007/978-3-319-14448-1$_$13

Yang XIA, S. Z, J.-X. W, *et al.* (2019) Predictive large-eddy-simulation wall modeling via physics-informed neural networks. Physical Review Fluids 4:034602. https://doi.org/10.1103/PhysRevFluids.4.034602

Zhang A, Lipton ZC, Li M, *et al.* (2023) Dive into Deep Learning. Cambridge University Press, https://D2L.ai

Zhideng Z, Xiang XIA, Fengshun Z, *et al.* (2023) A wall model learned from the periodic hill data and the law of the wall. Physics of Fluids 35(5):055108. https://doi.org/10.1063/5.0143650

Zhou D, Whitmore MP, Griffin KP, *et al.* (2022) Multi-agent reinforcement learning for wall modeling in LES of flow over periodic hills. https://doi.org/10.48550/arXiv.2211.16427, URL http://arxiv.org/abs/2211.16427, 2211.16427[physics]

Zhou Z, He G, X. Y (2021) Wall model based on neural networks for LES of turbulent flows over periodic hills. Physical Review Fluids 6:054610. https://doi.org/10.1103/PhysRevFluids.6.054610