JFM PAPERS

Machine-learning wall-model large-eddy simulation accounting for isotropic roughness under local equilibrium

Rong Ma[1] and **Adrián Lozano-Durán**[1,2]

[1]Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[2]Graduate Aerospace Laboratories, California Institute of Technology, Pasadena, CA 91125, USA
Corresponding author: Rong Ma, rongma@mit.edu

We introduce a wall model for large-eddy simulation (WMLES) applicable to rough surfaces with Gaussian and non-Gaussian distributions for both the transitionally and fully rough regimes. The model is applicable to arbitrary complex geometries where roughness elements are assumed to be underresolved, i.e. subgrid-scale roughness. The wall model is implemented using a multi-hidden-layer feedforward neural network, with the mean geometric properties of the roughness topology and near-wall flow quantities serving as input. The optimal set of non-dimensional input features is identified using information theory, selecting variables that maximize information about the output while minimizing redundancy among inputs. The model also incorporates a confidence score based on Gaussian process modelling, enabling the detection of potentially low model performance for untrained rough surfaces. The model is trained using a direct numerical simulation (DNS) roughness database comprising approximately 200 cases. The roughness geometries for the database are selected from a large repository through active learning. This approach ensures that the rough surfaces incorporated into the database are the most informative, achieving higher model performance with fewer DNS cases compared with passive learning techniques. The performance of the model is evaluated both *apriori* and *aposteriori* in WMLES of turbulent channel flows with rough walls. Over 550 channel flow cases are considered, including untrained roughness geometries, roughness Reynolds numbers and grid resolutions for both transitionally and fully rough regimes. Our rough-wall model offers higher accuracy than existing models, generally predicting wall shear stress within an accuracy range of 1 %–15 %. The performance of the model is also assessed on a high-pressure turbine blade with two different rough surfaces. We show that the new wall model predicts the skin friction and the mean velocity deficit

1007 A17-1

induced by the rough surface on the blade within 1 %–10 % accuracy except the region with transition or shock waves. This work extends the building-block flow wall model (BFWM) introduced by Lozano-Durán & Bae (2023. *J. Fluid Mech.* **963**, A35) for smooth walls, expanding the BFWM framework to account for rough-wall scenarios.

Key words: turbulent boundary layers, turbulence simulation, turbulence modelling

1. Introduction

Turbulent boundary layers (TBL) over rough surfaces are prevalent in engineering applications. Examples include the deposition of fuel and airborne contaminants, as well as erosion in turbomachinery applications contributing to the formation of roughness on turbine blades (Bons *et al.* 2001). Another example is biofouling, resulting from the accumulation of living organisms, which generates multiscale roughness geometries on the immersed surfaces of marine vessels (Munk *et al.* 2009). These rough surfaces significantly increase hydrodynamic drag and reduce the overall efficiency of engineering systems (Bons 2010; Kirschner & Brennan 2012). Therefore, modelling the effects of roughness and predicting drag in rough-wall flows are essential tasks in engineering. Computational fluid dynamics complements experimental investigations by enabling more cost-effective exploration under various operating conditions and reducing the need for extensive physical testing. While roughness-resolved simulations such as direct numerical simulation (DNS) and wall-resolved large-eddy simulation (WRLES) provide valuable insights into the underlying physics and aid in the development of models, their practical applicability to high Reynolds number flows is limited due to their high computational cost. Recently, wall-modelled large-eddy simulation (WMLES) has emerged as a competitive approach for modelling the effects of roughness on the outer flow without resolving the small-scale flow and roughness details in the near-wall region. In this work, we develop a rough-wall model for WMLES that is applicable to various roughness geometries and flow conditions.

A surface is considered rough when its topographical features are large enough to disrupt the near-wall eddies, resulting in increased drag and momentum deficit across the TBL (Raupach *et al.* 1991; Jiménez 2004; Chung *et al.* 2021). In incompressible zero-pressure-gradient TBL, the velocity deficit caused by roughness is quantified by the roughness function $\triangle U^+ = \triangle U / u_\tau$, where $\triangle U$ represents the downward shift of the mean velocity profiles in the logarithmic layer, and u_τ denotes the mean friction velocity. In the fully rough regime, the momentum deficit is primarily caused by form drag. This scenario is typically easier to investigate as wall friction becomes independent of Reynolds number. In contrast, both form drag and viscous drag contribute to wall friction in the transitionally rough regime. In these cases, drag is highly sensitive to Reynolds number and roughness topographies, making the search for universal scaling laws and models for drag challenging tasks. For zero-pressure-gradient TBL, the effect of roughness in the fully rough regime is generally characterized by the equivalent sand-grain roughness height k_s. This hydraulic roughness scale, proposed by Nikuradse (1933), represents the size of uniformly packed sand-grain roughness that produces the same frictional drag as the actual roughness geometry. In the fully rough regime, k_s quantifies hydrodynamic drag through a logarithmic relationship with the roughness function.

Models with varying fidelity have been devised to account for wall-roughness effects, ranging from empirical correlations based on Moody charts to wall functions for the Reynolds-averaged Navier–Stokes equations (RANS), WRLES and WMLES. Many of these rough-wall models are formulated in terms of equivalent sand-grain roughness height, making the prediction of k_s the main goal. The first, lower-fidelity family of models aims to establish correlations between k_s and other roughness geometrical parameters without explicitly resolving the flow motion (e.g. Bons 2002; Flack & Schultz 2010; Forooghi *et al.* 2017; Chung *et al.* 2021). Further details about this topic can be found in the reviews by Bons (2002); Flack & Schultz (2010) and Forooghi *et al.* (2017). More recently, data-driven methods have also been leveraged to enhance the prediction of k_s without resolving the flow. Jouybari *et al.* (2021) used machine-learning methods to predict k_s based on a large set of roughness parameters and demonstrated more accurate results than previous empirical correlations. Ma *et al.* (2023) proposed a particle swarm optimized backpropagation method to estimate k_s and showed better performance in evaluation metrics compared with both existing roughness correlation formulae and the traditional backpropagation model. Yang *et al.* (2023) utilized ensemble neural networks to predict k_s based on the roughness height probability density function (PDF) and power spectrum. Other methods do not rely on the use of k_s. For example, Yang *et al.* (2016) proposed an analytical roughness model based on the exponential velocity profile within the roughness layer for rectangular-prism roughness elements and demonstrated good predictions of mean velocity and drag forces for this type of roughness where the flow separation point is easily identified.

The second family of rough-wall models incorporates surface roughness effects directly into RANS, WRLES or WMLES. These methods usually adopt one of the following approaches.

(i) In the first approach, wall roughness is represented using a closure model for RANS simulations. Cebeci & Chang (1978) adapted the mixing-length formulation of eddy viscosity near rough walls by introducing an effective wall displacement as a function of k_s. Feiereisen & Acharya (1986) further refined the model proposed by Cebeci & Chang (1978) by directly incorporating measurable roughness parameters instead of relying solely on k_s. Durbin *et al.* (2001) extended the two-layer k–ϵ model to rough walls by modifying the calculation of the eddy viscosity based on k_s. Aupoix & Spalart (2003) proposed two extensions of the Spalart–Allmaras model to account for roughness effects using the value of k_s as key parameter. Knopp *et al.* (2009) presented an extension for k–ω type turbulence models to account for surface roughness based on k_s and the rough-wall logarithmic law, demonstrating its capability of predicting the aerodynamic effects of surface roughness on the flow past an airfoil. Brereton & Yuan (2018) proposed a model of equivalent shear force for the wall-roughness eddy viscosity, demonstrating good agreement with experimental data for zero and favourable pressure gradient TBL over fully rough surfaces.

(ii) The second approach consists of imposing the fluxes as wall boundary condition obtained from analytical wall functions or rough-wall models that account for roughness effects. Wilcox (1998) incorporated roughness effects into the boundary condition for the ω equation of the k–ω turbulence closure model for RANS by introducing a functional dependence with k_s. Suga *et al.* (2006) derived an analytical wall function accounting for the effects of fine-grain surface roughness on turbulence and heat transfer in RANS simulations. In the context of WMLES, the logarithmic law for rough walls as a function of k_s has been used to capture the downward shift of velocity profiles (Yang *et al.* 2017; Li & Yang 2021). Li *et al.* (2022) provided

a systematic assessment of the predictive capability of the logarithmic law rough-wall model for WMLES and demonstrated good predictions of the mean velocity against DNS data. The logarithmic law rough-wall model has also been widely used to develop morphological models for flows over urban-like surfaces (Theurer 1993; Macdonald *et al*. 1998; Grimmond & Oke 1999; Hanna & Britter 2010).

(iii) The third approach involves introducing a body force term to the Navier–Stokes equations as a drag model representing roughness effects. The rationale for the forcing term was discussed by Stripf *et al.* (2009). The body-force drag model has been employed in both RANS equations (Aupoix 2016; Chedevergne & Forooghi 2020) and eddy-resolving simulations (Shaw & Schumann 1992; Busse & Sandham 2012). Busse & Sandham (2012) used a body-force term model and showed good agreement for the mean flow and Reynolds stresses everywhere except in the immediate vicinity of the rough surface. However, they noticed that the model parameters need to be calibrated against experiments or DNS to be successfully applied in a simulation setting. Anderson & Meneveau (2011) developed a dynamic roughness model for large-eddy simulation (LES) applicable to multiscale, fractal-like roughness by decomposing the surface into resolved and subgrid-scale (SGS) height contributions. The unresolved height fluctuations were modelled using the equilibrium logarithmic law, and the SGS roughness parameter was dynamically estimated to achieve resolution-independent mean velocity profiles.

The roughness modelling approaches presented above have greatly facilitated the prediction of surface roughness effects on turbulent flows. However, current rough-wall models still face important limitations:

1) Many wall models for WMLES lack true predictability, as they require the specification of the non-trivial hydrodynamic property k_s which is often empirically measured rather than derived from a model. This reliance on a prescribed k_s hinders their ability to provide true predictions.

2) Although k_s is effective for predicting drag in fully rough flows, its utility diminishes in transitionally rough flows. This limitation stems from the fact that the logarithmic relationship between k_s and $\triangle U^+$ is valid only within the fully rough regime.

3) The assumptions underlying many wall models are rooted in 'equilibrium' turbulence, i.e. the presence of wall-attached, statistically steady turbulence under zero-pressure-gradient. Consequently, these wall models can accurately predict outcomes only for a limited number of cases and cannot be generalized to complex scenarios (e.g. adverse/favourable mean pressure gradient and separated flows) which are of significant interest in practical applications.

4) Many models are tailored for specific roughness geometries. The challenge remains to develop a rough wall model that can accommodate a broad spectrum of surface topologies without sacrificing prediction accuracy.

5) On some occasions, the rough-wall models are, by construction, only applicable to simple flow configurations such as channel flows and flat plates. This limitation may be due to assumptions of flat walls, periodic boundary conditions or the need for global flow quantities (e.g. turbulent channel height) that might not be well-defined in other scenarios. As a result, these models are unsuitable for the complex geometries typical in real-world engineering applications.

The reader is referred to the recent work by Durbin (2023) for a discussion on the strengths and limitations of different approaches to formulate rough-wall models.

Addressing current limitations is crucial for developing accurate and robust wall models that capture the effects of roughness on turbulent flows across diverse conditions, geometries and flow regimes. To overcome these limitations, the present work aims to develop a wall model by utilizing the flow over rough walls in minimal turbulent channels.

This effort builds upon the concept of a building-block-flow wall model (BFWM) for WMLES of smooth-wall flows, initially introduced by our group (Lozano-Durán & Bae 2023). The core assumption of the model is that the essential physics required to predict wall stress in complex scenarios can be captured by a finite set of simpler canonical flows, referred to as building-block flows. Seven types of building-block flows, based on turbulent Poiseuille–Couette flows, were used to train the BFWM. Model inputs include local flow quantities (e.g. velocities, density, viscosity), while outputs consist of wall shear stress and the angle between wall shear stress and wall-parallel velocity at the first control volume. The model also provides the probability of flow association with each building-block category and a confidence score for the predictions. Lozano-Durán & Bae (2023) demonstrated that the BFWM approach successfully accounts for multiple flow regimes (e.g. zero, adverse and favourable mean pressure gradients, as well as separation) within a unified model, outperforming predictions by an equilibrium wall model.

In this work, we aim to extend this framework to incorporate wall roughness, referred to as BFWM-rough. Our long-term goal is to develop a rough-wall model for WMLES that accommodates multiple roughness geometries and flow regimes, including both zero and non-zero mean pressure gradient effects and flow separation. Towards this goal, the primary objective of this first work is to develop the initial version of BFWM-rough for WMLES under near-wall equilibrium assumptions, applicable to transitionally and fully rough regimes with both Gaussian and non-Gaussian roughness geometries. Efforts are already underway to incorporate additional rough surface geometries, non-equilibrium effects and compressibility effects (Ma *et al.* 2024).

In this work, we build a database of rough-wall turbulent flows using DNS cases selected through active learning (AL). The resulting data are utilized to train the machine-learning-based wall model, which is implemented in both structured and unstructured WMLES solvers. To evaluate the performance of the model, BFWM-rough is assessed across a wide range of cases, spanning from canonical turbulent channel flows to high-pressure turbine (HPT) blade configurations. The manuscript is organized as follows. The roughness database is introduced in § 2. The formulation of the newly proposed rough-wall model is discussed in § 3. The model evaluation is presented in § 4. Finally, the conclusions are offered in § 5.

2. Database of turbulence over rough surfaces

We generate a database of turbulent channel flows over rough surfaces to train the wall model. The streamwise, wall-normal and spanwise directions are denoted by x, y and z, respectively, and occasionally referred to as x_1, x_2 and x_3. The friction Reynolds number is $Re_\tau = u_\tau \delta / \nu$, where u_τ is the friction velocity, ν is the kinematic viscosity and δ is the channel half-height. The database is constructed in three steps. First, we create a repository containing various irregular rough surfaces. Second, we apply an AL framework to identify and select the most informative rough surfaces from the repository. Third, we conduct DNS of turbulent channel flows with the rough walls selected in the previous step. The resulting database is used to train the wall model, as described in § 3.

2.1. *Roughness repository*

The roughness repository is a collection of rough surfaces designed for generating the DNS database, which is utilized to train and validate the wall model. The repository includes irregular rough surfaces characterized by different PDFs and power spectra, resembling the realistic roughness encountered in engineering applications. These irregular rough surfaces are created from the PDF and power spectra using a rough surface generator (Pérez-Ràfols & Almqvist 2019). Two families of PDFs are considered: Gaussian and Weibull. The Gaussian distribution is chosen due to its ubiquity in nature and engineering applications (Williamson *et al.* 1969; Whitehouse 2023). Examples of Gaussian roughness include turbine blades subject to erosion (Bons 2002), surface finish degradation on gas turbine vanes during service (Bacci *et al.* 2021) and manufactured surfaces such as highly polished steel (Das & Linke 2017). The Weibull distribution is used to represent roughness resulting from tribology and wear (Panda *et al.* 2015), defect populations in materials due to manufacturing processes, environmental factors or operational conditions (Cook & DelRio 2019), as well as geophysical and terrain roughness (Barbosa & Gerke 2022). The inclusion of Weibull roughness in the repository enables the representation of a broader spectrum of asymmetrical and non-Gaussian roughness features.

The Gaussian and Weibull rough surfaces considered are statistically homogeneous along the wall-parallel directions. The Gaussian roughness is generated based on the normal distribution of roughness height by specifying the root-mean-square roughness height in a range from 0.005δ to 0.03δ. The PDF of the Weibull distribution of the random variable $\tilde{k}$ follows

$$PDF_w(\tilde{k}) = \frac{H}{\lambda}\left(\frac{\tilde{k}}{\lambda}\right)^{H-1} e^{-(\frac{\tilde{k}}{\lambda})^H}, \quad \tilde{k} \geqslant 0 \qquad (2.1)$$

where the shape parameter $H > 0$ describes the shape of the probability distribution and is randomly selected within [0.8, 2.3], and $\lambda > 0$ is the scale parameter. The power spectra describing the isotropic self-affine fractal is

$$\begin{aligned} PS(\kappa) &= \kappa^{-2(1+H_f)}, \quad \kappa_0 \leqslant \kappa \leqslant \kappa_1 \\ PS(\kappa) &= \kappa_0^{-2(1+H_f)}, \quad \kappa < \kappa_0 \end{aligned} \qquad (2.2)$$

where $\kappa = \sqrt{\kappa_x^2 + \kappa_z^2}$, and κ_x and κ_z are the non-dimensional wavenumbers in the streamwise (x) and spanwise (z) directions, respectively. The higher bound wavenumber $\kappa_1 = L_x/\lambda_1$ is set by giving the lower bound of the roughness wavelength $\lambda_1 = 0.033\delta$ to ensure that the smallest roughness length scales are resolved by adequate grid points. The power spectra is controlled by two randomized parameters, the roll-off (lower bound) wavenumber κ_0 and the Hurst exponent H_f. The values of κ_0 are selected within the range [3,25], and the values of H_f are varied to obtain the power-law decline rate θ within the range [−4,−3]. The resulting surface generated based on the PDF and power spectra is then scaled from 0 to the root-mean-square height k_{rms} in a range from 0.005δ to 0.03δ. These values are determined to span the range of the roughness parameters for the actual rough surfaces in engineering applications (Bons 2010; Kirschner & Brennan 2012). The roughness repository includes 50 Gaussian rough surfaces and 50 Weibull rough surfaces. Six roughness samples are visualized in figure 1.

The geometric properties of the roughness are characterized by statistical quantities derived from the surface height distribution. The definition of roughness parameters is given in table 1. These include roughness height measures such as mean height k_{avg}, first-order moment of height fluctuations R_a, root-mean-square height k_{rms}, crest height k_c and
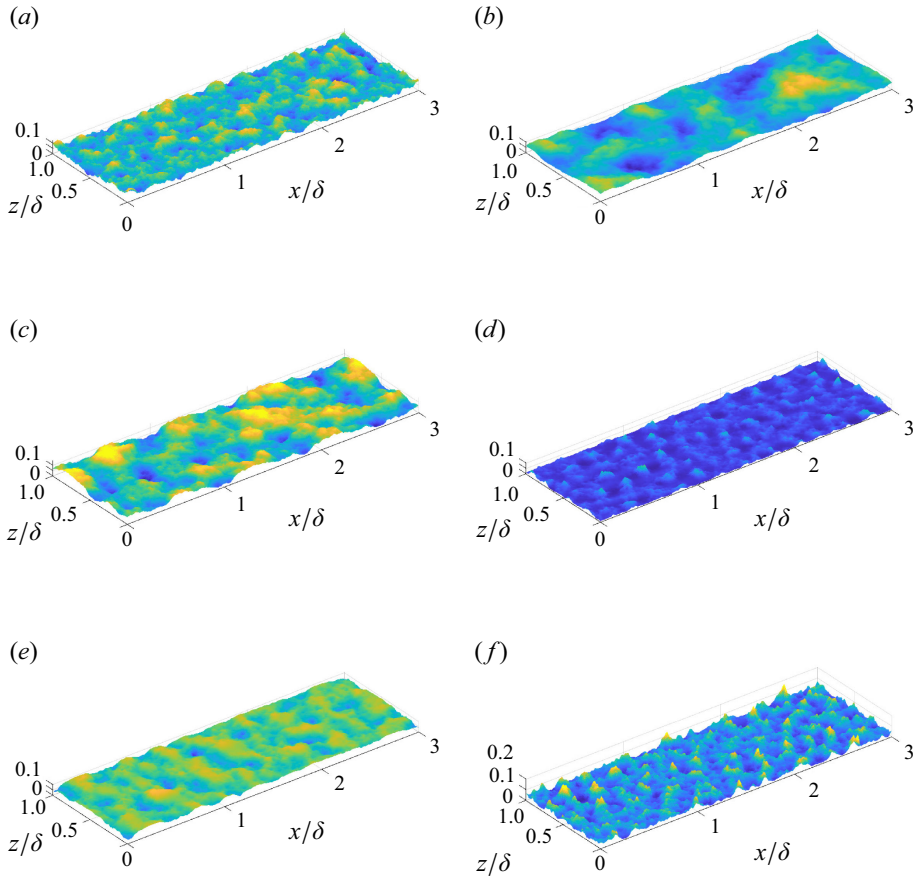
Figure 1. Visualization of roughness height for selected surface samples: (*a–c*) Gaussian roughness; (*d–f*) Weibull roughness. The colours represent the level of k/δ from 0.0 (blue) to 0.12 (yellow), where δ is the channel half-height.

mean peak-to-valley height k_t; high-order moments of height fluctuations such as skewness S_k and kurtosis K_u; height gradients such as the effective slope ES, and inclination angle I; surface porosity P_o; roughness density measures such as frontal solidity λ_f; and the correlation length L_{cor}. The parameters ES, I and L_{cor} remain mainly constant along the wall-parallel direction. However, the generated rough surfaces are not perfectly isotropic. This relaxes the constraint on the applicability of the BFWM-rough to strictly isotropic surfaces and allows the model to generalize to rough surfaces with mild anisotropic features.

2.2. *Active learning*

We employ AL to efficiently build the training repository and minimize the computational expenses associated with DNS. Ideally, DNS simulations of turbulent channel flow using all surfaces from the roughness repository would be conducted to maximize the amount of training data. However, in practice, the number of DNS we can perform is constrained by computational resources. The AL approach iteratively selects the most valuable rough surfaces from the repository for DNS. Active learning focuses on finding the most informative cases to enhance model performance while reducing labelling costs

Mean height	$k_{avg} = \frac{1}{A_t} \int_{x,z} k(x,z) dA$
Crest height	$k_c = max\{k(x,z)\} - min\{k(x,z)\}$
Mean peak-to-valley height	$k_t = mean\{max\|_{\delta \times \delta}\{k(x,z)\} - min\|_{\delta \times \delta}\{k(x,z)\}\}$
Root-mean-square height	$k_{rms} = \sqrt{\frac{1}{A_t} \int_{x,z} (k(x,z) - k_{avg})^2 dA}$
First-order moment of height fluctuations	$R_a = \frac{1}{A_t} \int_{x,z} \|k(x,z) - k_{avg}\| dA$
Skewness	$S_k = \frac{1}{A_t k_{rms}^3} \int_{x,z} (k(x,z) - k_{avg})^3 dA$
Kurtosis	$K_u = \frac{1}{A_t k_{rms}^4} \int_{x,z} (k(x,z) - k_{avg})^4 dA$
Effective slope	$ES = \frac{1}{A_t} \int_{x,z} \|\frac{\partial k(x,z)}{\partial x}\| dA$
Inclination angle	$I = \tan^{-1}\left(\frac{1}{2} S_k\{\frac{\partial k(x,z)}{\partial x}\}\right)$
Surface porosity	$P_o = \frac{1}{A_t k_c} \int_0^{k_c} A_f(y) dy$
Frontal solidity	$\lambda_f = \frac{A_p}{A_t}$
Correlation length	$L_{cor} = min_{\delta x}\{R_h(\delta x, 0) \leqslant 0.2\}$

Table 1. Definitions of roughness geometrical parameters. $k(x,z)$ is the roughness height function, $A_f(y)$ is the fluid area at the y location, A_p is the frontal projected area of the roughness elements, and A_t is the total plan area. The correlation lengths are computed as the horizontal separation at which the roughness height autocorrelation function $R_h(\delta x, \delta z) = \frac{1}{k_{rms}^2} \langle k(x + \delta x, z + \delta z) k(x,z)\rangle_{xz}$ drops below 0.2, where $\langle \cdot \rangle_{xz}$ denotes average over x and z. Given that the rough surfaces considered are isotropic, the parameters ES, I, and L_{cor} are equivalent along any wall-parallel direction. Similar definitions of roughness parameters can be found in Thakkar *et al.* (2017); Ma *et al.* (2021); Jouybari *et al.* (2021) and Chung *et al.* (2021).

(Settles 2009). This strategy ensures effective exploration of the repository by incorporating DNS data that is most useful for training robust, generalizable models. In recent years, the AL approach has been used in fluid applications. For instance, Zhou *et al.* (2022) utilized it for efficient prediction of propeller aerodynamic and acoustic performance, while Yang *et al.* (2023) applied it to predict equivalent sand-grain size. Both studies demonstrate how the AL approach can guide the selection of new experiments or simulations in regions with high discrepancy, significantly improving the trained machine-learning model for various engineering problems.

The AL approach is implemented using a Gaussian process (GP) model to predict the uncertainty from untrained roughness surfaces (Rasmussen & Williams 2006). The GP model is a non-parametric method based on the assumption that the function to be learned is drawn from a GP. This assumption enables the model to make predictions with well-defined uncertainty. The inputs to the GP model include all roughness parameters described in table 1. All roughness parameters are considered in the AL, as it is assumed no *apriori* knowledge of which roughness parameters may be more critical than others during the rough surface selection process. This assumption helps avoid bias in selecting rough surfaces, allowing for the addition of diverse roughness types (e.g. Weibull roughness) to the roughness repository. The output is the non-dimensionalized wall-shear stress $\langle \tau_w \rangle y_1 / (\nu U_1)$, obtained from DNS of turbulent channel flows. Here, y_1 is the wall-normal distance, ν is the kinematic viscosity of the fluid, $\langle \tau_w \rangle$ represents the mean wall shear stress, where the angle brackets denote the average over the homogeneous directions and time, and $U_1 = U(y_1)$ is the mean wall-parallel velocity magnitude at y_1. Detailed computations of $\langle \tau_w \rangle$, $U(y_1)$ and the wall-normal locations considered are presented in § 2.3. The GP model is defined by a mean function and a covariance kernel. A zero mean function is used as the prior mean function, and a squared exponential kernel serves as the prior covariance function. The posterior distribution, given the observed data, is obtained from the prior distribution and is used to predict the uncertainty
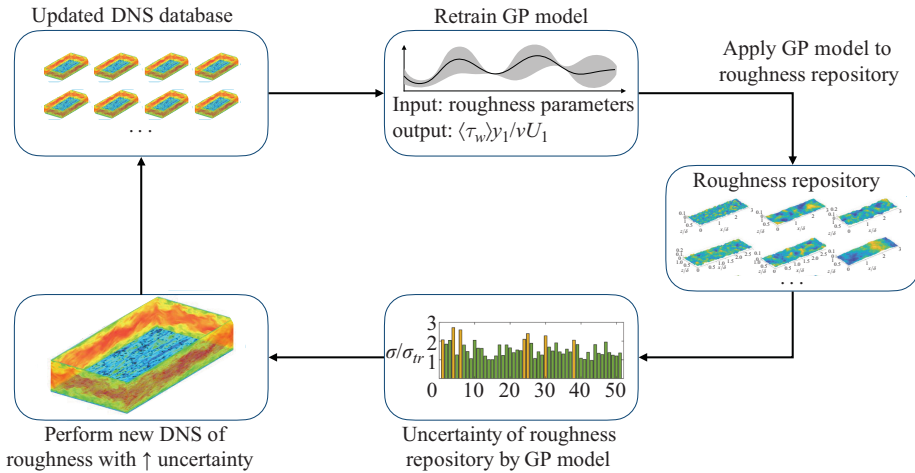
Figure 2. Schematic of the AL to select the rough surfaces for the DNS turbulent channel database.

(namely, predictive variance σ^2) of rough surfaces untrained by the GP model. During the training process, the optimal hyperparameters for the GP model are determined by minimizing the negative logarithm of the marginal likelihood. Readers are referred to Rasmussen & Williams (2006) for additional details about the GP model algorithm.

The steps for AL, summarized in figure 2, are as follows.

 (i) Initialization. A small set of DNS roughness data are used to initialize the process.
 (ii) GP model training. A GP model is trained using the initial labelled DNS data.
(iii) Uncertainty sampling. The GP model is used to predict the uncertainty for all rough surfaces in the repository. Rough surfaces with the highest uncertainty are selected.
(iv) New data generation. The DNS of turbulent channel flows over the selected rough surfaces is performed at various Reynolds numbers.
 (v) Model update. The newly labelled DNS data are added to the training set, and the GP model is retrained.
(vi) Iteration. Steps (iii) to (v) are repeated until a stopping criterion is met.

The rough surfaces selected through the AL framework are labelled as GS# and WB#, where GS and WB denote Gaussian and Weibull roughness, respectively. The value of # is the identification number of the surface used to locate the case in table 2, where the properties of roughness topography are listed. At each iteration, approximately 15 % of the total number of rough surfaces is selected for performing DNS. This value is sufficient to reduce uncertainty of the rough surfaces in the roughness repository at each iteration, and was constrained by our computational resources to conduct new DNS cases in each iteration. A total of four iterations of the GP model are performed after which the value of σ^2/σ_{tr}^2 is less than 2.5 for the whole roughness repository, where σ_{tr}^2 is averaged variance of the last GP model.

As a starting point, six Gaussian rough surfaces not included in the roughness repository, GS01 to GS06, are generated. These six rough surfaces are created using the same method as the other Gaussian rough surfaces in the roughness repository. This initial set of rough surfaces is generated to ensure that the roughness geometries cover a broad range of Gaussian roughness. The choice of this initial set mainly affects the number of iterations

Case	k_{avg}/δ	k_c/δ	k_t/δ	k_{rms}/δ	R_a/δ	S_k	K_u	ES	I	P_o	λ_f	L_{cor}/δ
GS01	0.062	0.125	0.120	0.018	0.014	0.068	2.902	0.540	−0.001	0.503	0.290	0.082
GS02	0.069	0.143	0.125	0.027	0.023	0.130	2.211	0.326	0.004	0.517	0.161	0.486
GS03	0.052	0.107	0.098	0.019	0.015	0.051	2.793	0.197	0.015	0.510	0.102	0.245
GS04	0.049	0.097	0.090	0.015	0.012	−0.001	2.752	0.290	−0.029	0.497	0.153	0.139
GS05	0.053	0.102	0.098	0.014	0.011	−0.095	2.974	0.463	−0.001	0.477	0.240	0.110
GS06	0.069	0.141	0.137	0.023	0.018	0.091	2.778	0.347	0.016	0.510	0.179	0.195
GS07	0.037	0.070	0.068	0.009	0.007	−0.007	3.046	0.272	−0.005	0.468	0.136	0.098
GS08	0.038	0.069	0.064	0.010	0.008	−0.392	3.065	0.124	0.025	0.452	0.056	0.234
GS09	0.035	0.069	0.066	0.012	0.010	0.127	2.543	0.137	0.006	0.498	0.065	0.257
GS10	0.042	0.089	0.084	0.013	0.010	0.045	3.078	0.317	−0.012	0.530	0.158	0.176
GS11	0.045	0.094	0.084	0.012	0.009	0.011	3.352	0.184	0.005	0.525	0.095	0.202
GS12	0.054	0.101	0.099	0.015	0.012	−0.132	2.908	0.327	0.001	0.463	0.160	0.250
GS13	0.036	0.077	0.070	0.011	0.009	0.301	2.733	0.259	−0.009	0.540	0.143	0.164
GS14	0.073	0.131	0.128	0.016	0.013	−0.104	2.967	0.694	−0.003	0.439	0.329	0.084
GS15	0.066	0.149	0.127	0.019	0.015	0.180	3.261	0.397	−0.008	0.556	0.213	0.148
GS16	0.100	0.173	0.170	0.025	0.019	−0.378	3.743	0.259	0.002	0.421	0.134	0.251
GS17	0.088	0.167	0.159	0.022	0.017	−0.167	3.146	0.551	0.009	0.475	0.277	0.119
GS18	0.066	0.128	0.118	0.016	0.012	−0.186	3.180	0.443	0.012	0.488	0.222	0.122
GS19	0.074	0.154	0.151	0.022	0.017	0.001	2.981	0.693	0.004	0.520	0.357	0.082
WB01	0.015	0.091	0.085	0.009	0.007	1.804	7.915	0.235	0.028	0.832	0.110	0.056
WB02	0.023	0.154	0.127	0.010	0.007	2.140	11.970	0.224	0.057	0.851	0.097	0.068
WB03	0.013	0.090	0.070	0.008	0.006	1.752	8.254	0.182	0.069	0.853	0.107	0.060
WB04	0.015	0.094	0.077	0.007	0.005	1.997	9.711	0.279	−0.028	0.843	0.129	0.033
WB05	0.022	0.157	0.139	0.013	0.009	2.055	10.171	0.343	0.072	0.859	0.196	0.053
WB06	0.011	0.084	0.074	0.007	0.005	2.210	10.867	0.152	0.078	0.866	0.064	0.094
WB07	0.025	0.155	0.122	0.012	0.009	1.861	9.430	0.254	−0.062	0.842	0.146	0.066
WB08	0.072	0.104	0.098	0.012	0.010	−0.718	3.655	0.218	0.038	0.304	0.107	0.091
WB09	0.064	0.087	0.087	0.011	0.009	−0.575	3.254	0.288	−0.030	0.265	0.145	0.080
WB10	0.034	0.155	0.148	0.015	0.013	1.355	5.913	0.521	0.016	0.769	0.233	0.050
WB11	0.034	0.123	0.116	0.014	0.011	1.002	4.404	0.577	−0.001	0.728	0.294	0.035
WB12	0.042	0.171	0.161	0.020	0.016	0.980	4.375	0.615	−0.007	0.755	0.336	0.055
WB13	0.039	0.164	0.154	0.017	0.013	1.181	5.146	0.663	−0.032	0.761	0.408	0.037

Table 2. Roughness parameters for rough surfaces in the DNS database.

needed to reach the final dataset; however, the final outcome is generally unaffected. The DNS of turbulent channel flows are performed for each roughness at six different Re_τ values: 180, 360, 540, 720, 900 and 1000. This data are used to initialize the process and train the first GP model (GP model-1). For the Gaussian rough surfaces, two iterations are conducted to improve the initial GP model. Figure 3(*a*) shows that seven new surfaces, GS07 to GS13, with the highest uncertainty, are selected from the roughness repository in the first iteration. Figure 3(*b*) shows that six new surfaces, GS14 to GS19, are selected in the second iteration. The reduced uncertainty in the second iteration, compared with the first, demonstrates that the current strategy effectively explores the repository by adding new data in the most uncertain regions of the parameter space. The GP model-3 is used to test the 50 Weibull rough surfaces from the repository. Figure 3(*c*) shows that seven Weibull rough surfaces, WB01 to WB07, with the highest prediction variance, are selected in the third iteration. The GP model-4 is then trained with the updated DNS data and used to test the Weibull roughness in one more iteration. As shown in figure 3(*d*), six additional Weibull rough surfaces, WB08 to WB13, are selected in the fourth iteration.
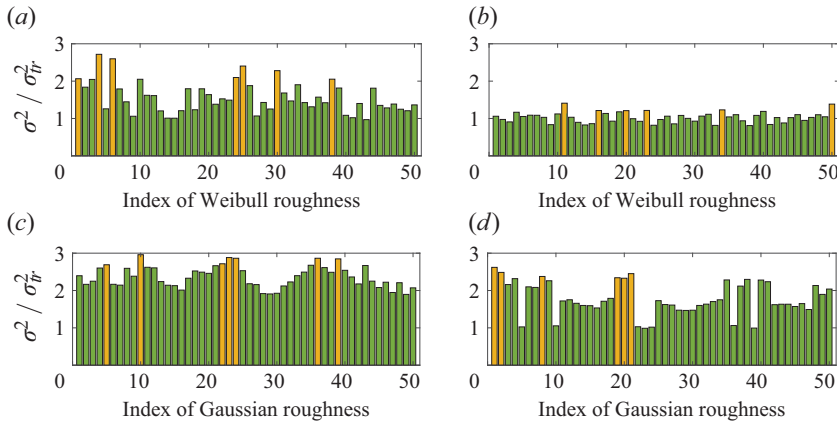
Figure 3. Uncertainty (σ^2) for the rough surfaces in the repository normalized by the mean uncertainty of the most updated GP model (σ_{tr}^2). In the first two iterations of the GP model, only Gaussian roughness is considered. In the third and fourth iterations, Weibull roughness is considered. (*a*) The first iteration using GP model-1; (*b*) the second iteration using GP model-2; (*c*) the third iteration using GP model-3; (*d*) the fourth iteration using GP model-4. The surfaces with the highest prediction variance coloured by yellow are selected for performing DNS.

In summary, the resulting training database contains a total of 19 Gaussian and 13 Weibull rough surfaces, which are used to conduct DNS of turbulent channel flows at six different Re_τ values. As a result, the DNS roughness database includes 192 cases. The statistical parameters of the rough surfaces in the DNS database are summarized in table 2.

Scatter plots of roughness parameters and the PDFs of roughness height are displayed in figures 4 and 5, respectively. The results illustrate the distribution of selected roughness at each iteration, demonstrating how the AL framework assists in exploring uncertain regions within the input roughness feature space. The correlation coefficient between pairs of roughness parameters is shown in figure 4. Strong correlations are observed among k_c/R_a, k_t/R_a and k_{rms}/R_a. This is because all three metrics are sensitive to extreme surface values (i.e. peaks and valleys) in the rough surfaces generated for the current roughness repository. As extreme height variations (measured by k_c and k_t) increase, the overall deviation from the mean (measured by k_{rms}) also tends to increase. Exceptions to this trend could include surfaces with isolated high peaks or deep valleys, which are not represented in the current roughness repository.

Figure 4 also reveals strong correlations among the parameters k_{rms}/R_a and S_k, k_{rms}/R_a and K_u, S_k and K_u, as well as S_k and P_o. For Gaussian roughness, these correlations are less pronounced due to the parameters remaining relatively constant with only small fluctuations. However, for Weibull roughness, the correlations become significant, as these parameters are influenced by the shape parameter H of the Weibull distribution: for smaller H values ($H < 2$), the rough surface is characterized by more prominent peaks, resulting in larger k_{rms}/R_a, higher positive S_k, greater P_o and larger K_u. Notably, ES shows weak correlation with both k_{rms}/R_a and S_k, indicating that ES may be important to characterize the roughness geometry. Additionally, ES has a moderate correlation with L_{cor}/R_a, as high ES corresponds to short-wavelength, steep roughness, while low ES is associated with long-wavelength, shallow roughness (Chung *et al.* 2021). Overall, the high correlations among some of the roughness parameters suggest that only a reduced set may be needed as input variables for the wall model.
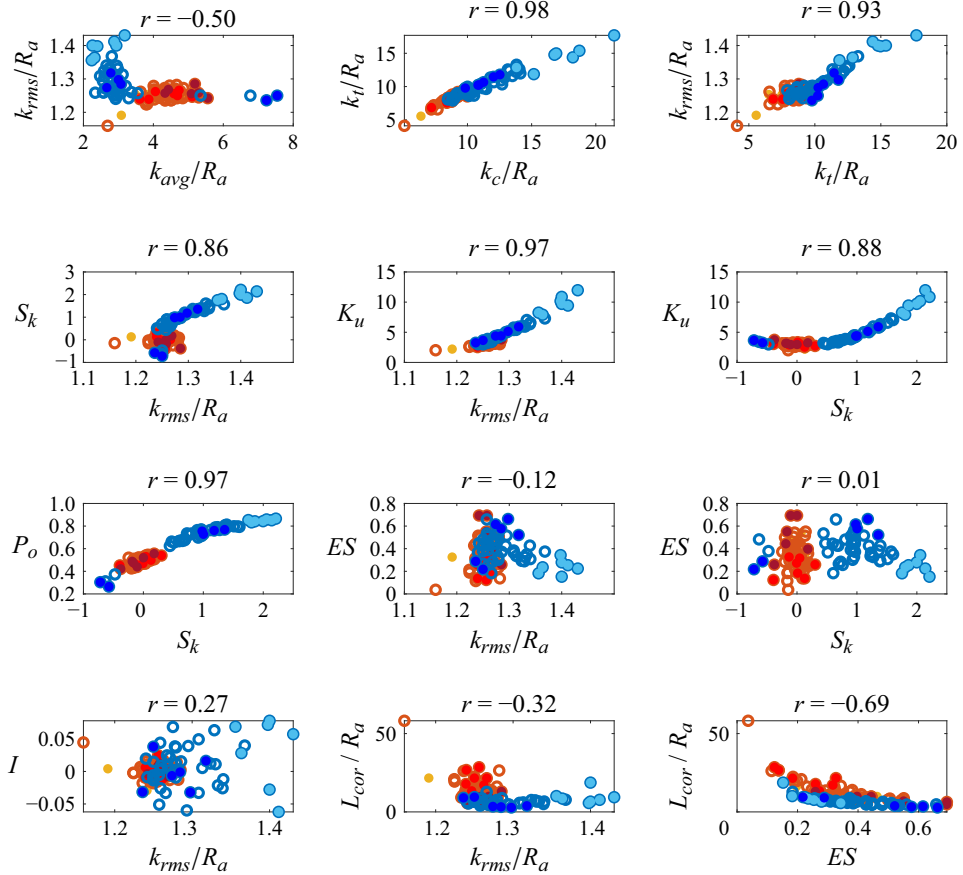
Figure 4. Scatter plots of roughness parameters for the surfaces selected at each iteration in AL. The correlation coefficient r between two parameters is shown on the top of each panel. The roughness repository is circled and the roughness in the training set is filled. Gaussian roughness repository (red); Weibull roughness repository (blue); initial set for AL is GS01 to GS06 (yellow); GS07 to GS13 at the first iteration (light red); GS14 to GS19 at the second iteration (dark red); WB01 to WB07 at the third iteration (light blue); WB08 to WB13 at the fourth iteration (dark blue).
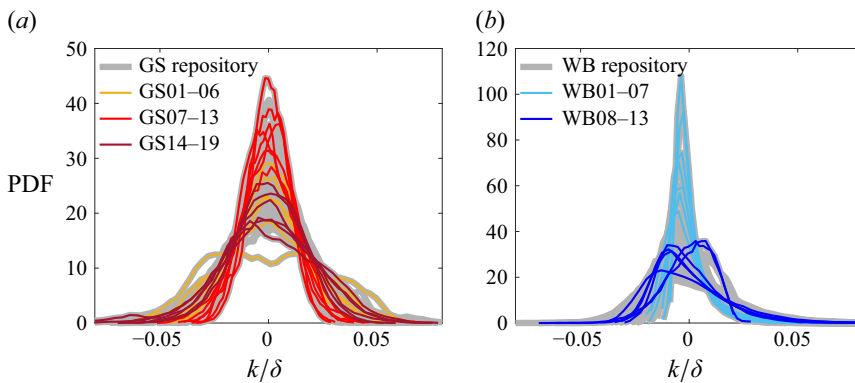

Figure 5. The PDF of roughness height for rough surfaces in the roughness repository and rough surfaces selected at different iterations in AL for (*a*) Gaussian roughness and (*b*) Weibull roughness.

2.3. *The DNS of rough-wall turbulent channel flows*

The DNS of turbulent channel flows with the rough surfaces selected from § 2.2 is performed to generate the training database. The governing equations for momentum and continuity are given by the incompressible Navier–Stokes equations,

$$\frac{\partial u_i}{\partial t} + \frac{\partial u_i u_j}{\partial x_j} = -\frac{1}{\rho}\frac{\partial p}{\partial x_i} + \nu\frac{\partial^2 u_i}{\partial x_j x_j} + F_i, \quad \frac{\partial u_i}{\partial x_i} = 0, \tag{2.3}$$

where u_i is the ith component of the velocity (streamwise: $i = 1$, wall-normal: $i = 2$, spanwise: $i = 3$), p denotes the pressure, ρ is the fluid density and ν is the kinematic viscosity of the fluid. An immersed boundary approach based on the volume-of-fluid method is used, where the no-slip boundary condition on the rough surface is enforced by the body force F_i (Scotti 2006; Yuan & Piomelli 2014*b*). The solver utilizes second-order central finite differences for spatial discretization, second-order Adams–Bashforth semi-implicit time advancement, and is parallelized using a message passing interface method (Keating *et al.* 2004). The code has been extensively validated in previous investigations of rough-wall turbulence (Yuan & Piomelli 2014*b*,c; Yuan & Jouybari 2018; Jouybari *et al.* 2021).

Turbulent open-channel flows are simulated at six different frictional Reynolds numbers: $Re_\tau = 180, 360, 540, 720, 900, 1000$. The Reynolds numbers are determined based on an estimation of the roughness Reynolds number k_s^+ ranging from 0 to 300 for the rough surfaces in the roughness repository. This choice of k_s^+ effectively captures the conditions typically encountered in practical rough-wall flow scenarios, covering both transitionally and fully rough regimes. Note that although the differences in Re_τ may not be highly significant among some cases, the variation in roughness Reynolds number k_s^+ – which quantifies drag and differentiates roughness regimes – can still be substantial among cases with similar Re_τ. A minimal-span channel simulation approach is used to enhance computational efficiency (Jiménez & Moin 1991; Chung *et al.* 2015; MacDonald *et al.* 2017). Chung *et al.* (2015) and MacDonald *et al.* (2017) demonstrated that simulations in a minimal-span domain can accurately capture the near-wall flow dynamics by adhering to the domain constraints. In addition, MacDonald *et al.* (2017) compared DNS of open channel flows over rough surfaces with DNS of standard-height channel flows over rough surfaces in both full-span and minimal-span channels. Their results showed that open channel flows have a negligible effect on the flow characteristics for both full-span and minimal-span channels, with the primary difference occurring in the wake region. The constraints proposed by Chung *et al.* (2015) and MacDonald *et al.* (2017) are

$$L_x \geqslant max(3L_z, 1000\nu/u_\tau, \lambda_{r,x}), \tag{2.4}$$

$$L_y \geqslant k_{ch}/0.15, \tag{2.5}$$

$$L_z \geqslant max(100\nu/u_\tau, k_{ch}/0.4, \lambda_{r,z}), \tag{2.6}$$

where L_x, L_y and L_z are the domain lengths in the streamwise, wall-normal and spanwise directions, respectively; k_{ch} is the characteristic roughness height; and $\lambda_{r,x}$ and $\lambda_{r,z}$ are the streamwise and spanwise length scales of the roughness elements. The use of minimal channels for flow over irregular roughness with a broad range of wavelengths was investigated by Jouybari *et al.* (2021), who compared the minimal-span channel with the full-span channel. Their results confirmed that the minimal-span approach remains valid for multiscale, irregular rough surfaces. The crest roughness height k_c is used as the characteristic roughness height, and surface Taylor microscales $\lambda_{T,x}$ and $\lambda_{T,z}$ are

Re_τ	$N_x \times N_y \times N_z$	$L_x/\delta \times L_y/\delta \times L_z/\delta$	$\triangle x^+$	$\triangle z^+$	$\triangle y_{min}^+$	$\triangle y_{max}^+$
180	$400 \times 300 \times 160$	$3 \times 1 \times 1$	1.35	1.13	0.03	1.66
360	$400 \times 300 \times 160$	$3 \times 1 \times 1$	2.70	2.25	0.05	3.33
540	$400 \times 300 \times 160$	$3 \times 1 \times 1$	4.05	3.38	0.08	4.99
720	$400 \times 300 \times 160$	$3 \times 1 \times 1$	5.40	4.50	0.11	6.65
900	$400 \times 300 \times 160$	$3 \times 1 \times 1$	6.75	5.63	0.14	8.32
1000	$400 \times 300 \times 160$	$3 \times 1 \times 1$	7.50	6.25	0.15	9.24

Table 3. Simulation parameters for DNS of rough-wall channel flows at different Re_τ. N_x, N_y, and N_z are the number of grid points in the streamwise, wall-normal, and spanwise direction, respectively, L_x, L_y, and L_z are the streamwise, wall-normal, and spanwise dimensions of the computational domain, $\triangle x^+$ and $\triangle z^+$ are the streamwise and spanwise grid resolutions, and $\triangle y_{min}^+$ and $\triangle y_{max}^+$ are the minimum and maximum wall-normal grid resolutions. Uniform grids are used in the streamwise and spanwise directions, and non-uniform grids with a hyperbolic tangent function are used in the wall-normal direction. The number of grid points is kept constant across Re_τ to resolve the roughness features and avoid interpolation between cases.

used as the streamwise and spanwise roughness length scales for multiscale random roughness, following Jouybari *et al.* (2021). The surface Taylor microscale represents the length scale of an equivalent roughness element in the context of irregular, multiscale roughness. While other length scales may also characterize the roughness wavelength, the surface Taylor microscale has proven suitable for meeting the constraints required in minimal-span channels, as discussed by Jouybari *et al.* (2021). The simulation domain size is $(L_x, L_y, L_z) = (3\delta, \delta, \delta)$, based on the criteria for small-span channel simulations. Periodic boundary conditions are used in the streamwise and spanwise directions, and no-slip and symmetry boundary conditions are imposed at the bottom surfaces and the top boundary. The grid size is determined to ensure that roughness elements are well resolved by at least four grid points per $\lambda_{T,x}$ and $\lambda_{T,z}$, as recommended by Yuan & Piomelli (2014*a*). Note that although the simulations are conducted in open channels, we still refer to δ as the channel half-height. The simulation details are shown in table 3.

As demonstrated by Chung *et al.* (2015) and MacDonald *et al.* (2017), the minimal-span channel provides reliable results for wall friction and turbulent statistics within the region $y < 0.3\delta$. In the context of wall model development, our primary focus is on the flow close to the wall. Therefore, the minimal-span channel approximation is sufficient to capture the near-wall physics necessary for developing an accurate wall model, provided that the near-wall grid size for WMLES is below 0.3δ. Validation of the minimal-span approach has been done for open channel over roughness and the results compared to a full-span open channel are demonstrated in Appendix A.

In all simulations, mean quantities and statistics are averaged over a time period $T \geqslant 20\delta/u_\tau$ after transients to achieve statistical convergence. The streamwise mean velocity $U(y)$ is calculated as

$$U(y) = \langle u \rangle = \frac{1}{A_f T} \int_T \iint_{A_f} u(x, y, z, t) \mathrm{d}x\mathrm{d}z\mathrm{d}t, \qquad (2.7)$$

where u is the instantaneous streamwise velocity, A_f is the fluid-occupied area at each y location, T is the total time considered, and the angle brackets denote average over homogeneous directions and time. The streamwise mean velocity profiles are shown in figure 6 for four selected rough surfaces. The mean wall shear stress $\langle \tau_w \rangle$ is computed by integrating the time-averaged body force F_1 in the streamwise direction (Yuan & Piomelli 2014*b*,c), and the friction velocity $u_\tau = \sqrt{\langle \tau_w \rangle / \rho}$.
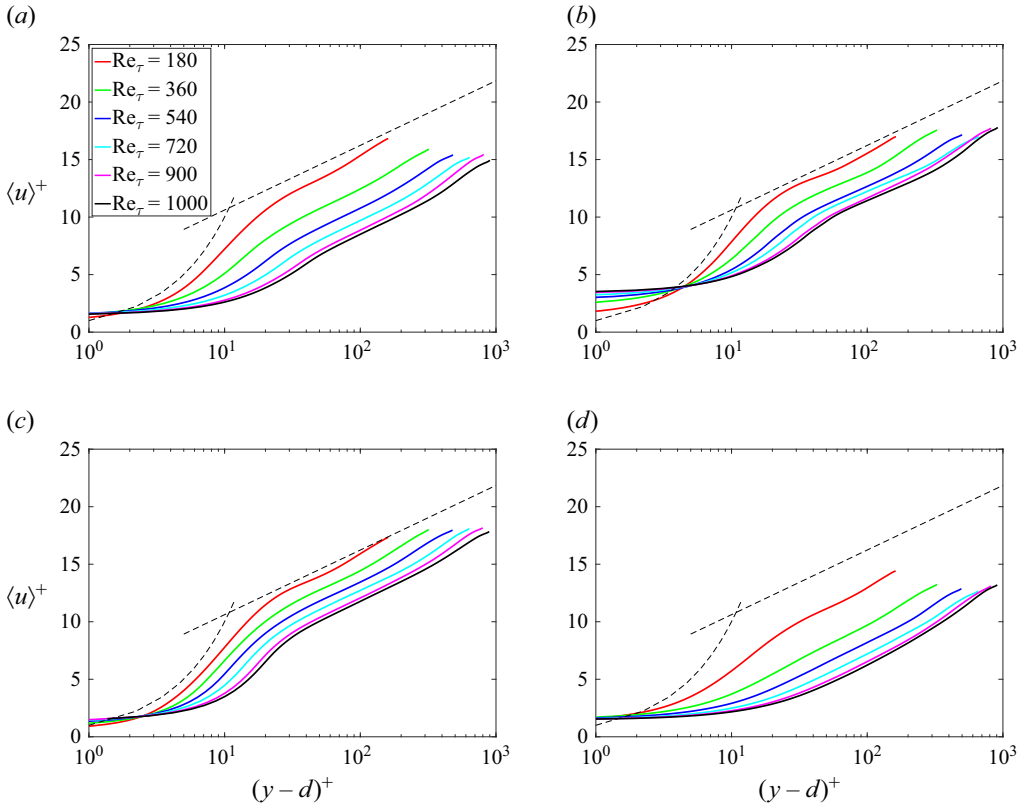
Figure 6. Streamwise mean velocity profiles of selected rough surfaces: (a,b) Gaussian roughness GS01 and GS03; (c,d) Weibull roughness WB08 and WB10. The dashed lines are $U^+ = y^+$ and $U_s^+ = \frac{1}{\kappa} \ln y^+ + 5.0$.

In the logarithmic region, the streamwise mean velocity profile in smooth walls (U_s) can be approximated by

$$U_s^+ \approx \frac{1}{\kappa} \ln y^+ + 5.0, \tag{2.8}$$

where $\kappa \approx 0.41$ is the von Kármán constant. For rough-wall cases, the logarithmic velocity distribution for the mean velocity profile (U_r) also holds in the fully rough regime,

$$U_r^+ \approx \frac{1}{\kappa} \ln \left(\frac{y - d}{\hat{k}_s} \right) + 8.5, \tag{2.9}$$

where $\hat{k}_s^+ = k_s^+$ for fully rough cases (but not for transitionally rough cases), and d is the zero-plane displacement, computed based on the location of the centroid of the wall-normal profile of the averaged drag force (Jackson 1981). For small d, the roughness function $\triangle U^+$ can be obtained by the difference of mean velocities in wall units between smooth and rough walls within the logarithmic layer,

$$\triangle U^+ \approx \frac{1}{\kappa} \ln \hat{k}_s^+ - 3.5. \tag{2.10}$$

To evaluate whether a rough-wall case is in the transitionally or fully rough regime, $\hat{k}_s^+$ can be computed according to (2.10). The values of $\hat{k}_s^+$ for each flow case in the current DNS database are presented in table 4. According to Flack & Schultz (2010), a flow is
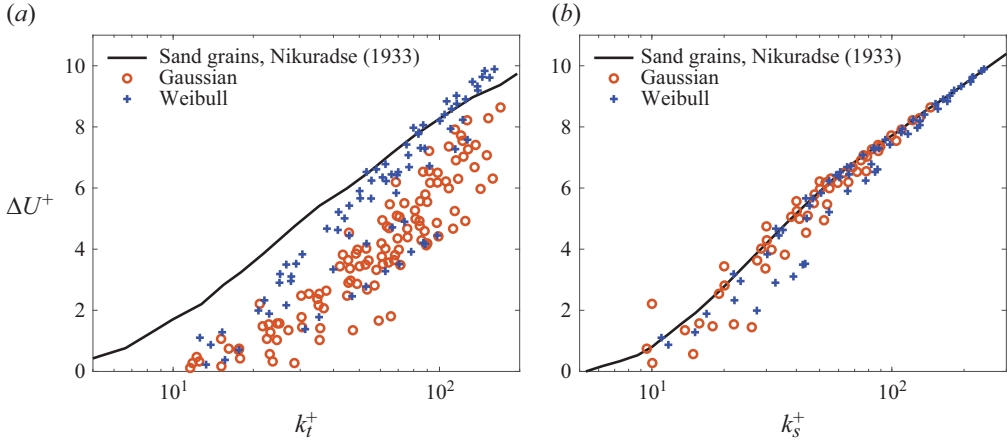
(*a*)

(*b*)

Figure 7. Roughness function $\triangle U^+$ as a function of (*a*) mean peak-to-valley roughness height k_t^+ and (*b*) equivalent sand-grain roughness height k_s^+. A total of 192 rough cases corresponding to table 4 are included in this plot.

considered to be fully rough if $\hat{k}_s^+ \geqslant 70$. Using this reference, it can be seen that the present DNS database spans a wide range of both transitionally and fully rough regimes. The relationship between the roughness function and roughness height is examined in figure 7 for the rough cases presented in table 4. In figure 7(*a*), the mean peak-to-valley roughness height is used as the roughness scale, showing that Gaussian and Weibull rough cases follow distinct trends. The relationship between $\triangle U^+$ and k_t^+ has been summarized by Chung *et al.* (2021), who suggested that this relationship varies for different roughness types, making k_t^+ a less effective scale for collapsing $\triangle U^+$. A better collapse is observed using k_s^+ as the roughness scale in figure 7(*b*), where alignment with the reference sand-grain roughness curve from Nikuradse (1933) indicates the fully rough regime. Note that here, k_s is obtained for each rough surface from the DNS results in the fully rough regime based on (2.10), and k_s^+ in the transitionally rough regime is computed based on $k_s^+ = k_s u_\tau / \nu$. The results suggest that the onset of the fully rough regime occurs at approximately $k_s^+ \approx 60$ for Gaussian roughness and $k_s^+ \approx 90$ for Weibull roughness.

3. Wall model formulation

3.1. *Framework of WMLES*

The wall model is developed within the framework of WMLES, where only the most energetic eddies in the outer layer of the flow are resolved by the computational grid. The effects of the small scales far from the wall are modelled by a SGS model. Close to the wall, the energy-containing eddies are under-resolved, and the wall-shear stress is obtained using a wall model. The flow solver integrates the coarse-grained incompressible Navier–Stokes equations,

$$\frac{\partial \overline{u}_i}{\partial t} + \frac{\partial \overline{u}_i \overline{u}_j}{\partial x_j} = -\frac{1}{\rho} \frac{\partial \overline{p}}{\partial x_i} + \nu \frac{\partial^2 \overline{u}_i}{\partial x_k x_k} + \frac{\partial \tau_{ij}^{SGS}}{\partial x_j}, \quad \frac{\partial \overline{u}_i}{\partial x_i} = 0, \quad (3.1)$$

Case				$\hat{k}_s^+$		
-	$Re_\tau = 180$	$Re_\tau = 360$	$Re_\tau = 540$	$Re_\tau = 720$	$Re_\tau = 900$	$Re_\tau = 1000$
GS01	7.7	20.1	39.5	61.1	85.8	99.5
GS02	6.5	11.6	18.7	23.4	28.5	31.6
GS03	5.6	10.2	17.6	21.5	27.6	30.8
GS04	5.7	11.9	20.4	28.7	36.4	40.8
GS05	5.7	11.9	21.8	33.4	45.1	52.9
GS06	8.0	21.5	40.1	60.3	74.9	87.6
GS07	5.1	6.4	9.8	13.6	17.7	20.0
GS08	4.7	4.8	6.4	7.3	8.3	8.8
GS09	4.5	5.5	7.1	8.1	9.5	10.0
GS10	5.0	7.5	12.6	17.5	22.8	25.7
GS11	4.4	7.1	11.1	14.2	16.4	18.1
GS12	4.5	8.7	13.8	19.6	25.4	29.6
GS13	4.8	8.0	12.4	17.2	21.8	25.0
GS14	5.3	16.7	33.9	52.7	71.3	82.6
GS15	7.3	18.6	32.5	49.0	65.0	76.3
GS16	10.4	17.2	29.4	41.2	53.7	55.6
GS17	7.9	27.0	53.2	80.8	107.7	122.0
GS18	4.7	13.3	24.4	36.1	48.6	55.8
GS19	7.6	31.9	61.6	92.6	125.4	144.9
WB01	7.1	20.2	38.8	56.7	76.4	84.1
WB02	9.1	26.0	46.0	66.0	82.8	93.9
WB03	6.6	15.5	28.5	42.8	54.0	60.9
WB04	6.0	14.1	28.0	42.7	58.4	65.1
WB05	13.8	47.2	88.0	121.1	161.7	181.8
WB06	4.6	10.2	16.5	23.3	29.0	31.3
WB07	10.9	32.7	59.0	83.5	108.8	122.7
WB08	5.6	8.7	13.1	17.7	23.1	26.0
WB09	4.9	7.4	11.5	16.1	20.9	23.6
WB10	17.6	61.0	110.3	157.3	204.4	236.5
WB11	9.5	35.6	67.7	103.6	131.7	152.1
WB12	17.8	63.1	114.3	169.6	218.1	242.1
WB13	15.0	54.3	101.5	142.1	192.3	216.1

Table 4. The roughness parameter $\hat{k}_s^+$ at different Re_τ determined from the DNS results based on the rough-wall logarithmic law.

where the overline denotes coarse-grained quantities and τ_{ij}^{SGS} is the deviatoric part of the SGS stress tensor. At the walls, the non-slip boundary condition is replaced by a shear stress boundary condition, which is obtained from the wall model.

Figure 8 depicts an overview of WMLES and the BFWM-rough. The general steps to build the wall model are as follows.

(i) First, the DNS database of minimal-span turbulent channel flows over rough surfaces, as presented in § 2.3, is used to extract mean velocity profiles and wall shear stress for various rough wall geometries.

(ii) The inputs to the wall model are constructed as non-dimensional numbers based on the mean velocity profile at two wall-normal locations and roughness geometric parameters. The model is trained using mean velocities from multiple wall-normal locations. This approach is motivated by the fact that, in WMLES, the wall-normal location varies with different grid resolutions and flow conditions, requiring the model to learn this variability.
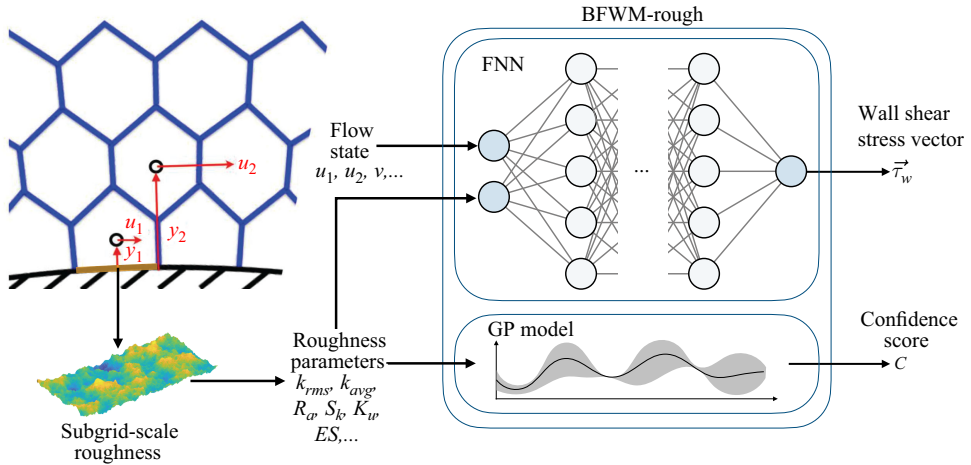
Figure 8. Overview of BFWM-rough for WMLES: for each wall face, the inputs to the wall model include the local flow state and local roughness parameters. The flow state consists of the magnitudes of wall-parallel velocities (u_1 and u_2) at the centre locations of the first and second off-wall control volumes (y_1 and y_2), along with the kinematic viscosity ν. The roughness is assumed to be SGS and known for each control volume attached to the wall. The input roughness parameters are based on the statistical moments of the local roughness height distribution. The output of BFWM-rough is the wall-shear stress vector $\boldsymbol{\tau}_w$ and the confidence score $C \in [0, 1]$. The wall-shear stress vector is predicted with a FNN, while the confidence score is determined using a GP model. The wall-shear stress vector from BFWM-rough is applied as the local boundary condition to the WMLES.

(iii) Wall roughness is assumed to be SGS, meaning that the WMLES grid does not resolve the roughness geometric features. In cases where wall roughness is such that $k_s^+ < 5$, the model includes a sensor to switch back to BFWM-smooth from Lozano-Durán & Bae (2023).
(iv) The output of BFWM-rough is the local wall-shear stress, which serves as the boundary condition for WMLES at the wall.
(v) The relationship between inputs and outputs in BFWM-rough is modelled using feed forward neural network (FNNs).

In the following sections, we discuss the model assumptions, the selection of input and output variables, the architecture of FNNs, and the details of wall model training.

3.2. *Wall model assumptions*

We summarize the main modelling assumptions of the BFWM-rough.

(i) Building block flow assumption. A finite set of simple flows, based on minimal-span turbulent channel flows over rough surfaces, is sufficient to formulate a generalizable rough-wall model.
(ii) Quasi-equilibrium assumption. The near-wall region of complex cases is in a quasi-equilibrium state, i.e. statistically steady flow under mean zero-pressure-gradient effects.
(iii) Velocity/shear-stress alignment assumption. The direction of the wall-shear stress vector is aligned with the relative wall-parallel velocity at the first control volume attached to the wall.
(iv) Statistical roughness description assumption. A collection of statistical parameters of the rough surface, such as the mean and root-mean-square roughness heights,

along with high-order moments of height fluctuations (see table 1), are sufficient to describe the geometrical effects of the surface topology on the wall shear stress.

(v) Space/time locality assumption. The relative wall-parallel velocity from the first two contiguous wall-normal control volumes above the wall, combined with (iv), provides enough information to predict the wall shear stress.

(vi) Subgrid-scale roughness assumption and outer-layer similarity. The roughness effects on the flow are assumed to be SGS and the only impact of roughness on the resolved scales is through the wall shear stress.

(vii) Viscous scaling assumption. The best-performing non-dimensional form of the velocity inputs and the model output is obtained by scaling the variables using the kinematic viscosity and wall-normal distance.

(viii) Mean-flow training data assumption. Inputs and outputs samples based on mean velocity profiles and mean wall-shear stress are suitable for training accurate wall models.

Assumptions (i) and (ii) justify the use of the current DNS training database, which targets equilibrium turbulent flows over rough surfaces. Therefore, the application of the BFWM-rough to complex geometries presupposes that the near-wall flow field maintains quasi-equilibrium conditions. Assumptions (iii), (iv) and (v) are adopted primarily for the sake of model simplicity. Incorporating flow information from farther away from the wall could potentially enhance the accuracy of the model, especially for rough surfaces with larger roughness heights. However, this would also increase the complexity of the model when dealing with unstructured grids in realistic geometries. Similarly, the use of roughness statistics instead of detailed local topography simplifies the model, akin to the use of k_s in other approaches. Assumption (vi) is supported by previous observations from the literature (Raupach *et al.* 1991) and by the Townsend's outer-layer similarity hypothesis (Townsend 1976), which has been confirmed by multiple studies (Flack *et al.* 2005; Flores & Jimenez 2006; Leonardi & Castro 2010; Mizuno & Jiménez 2013; Chung *et al.* 2014; Chan *et al.* 2015; Lozano-Durán & Bae 2019*a*). The viscous scaling from assumption (vii) is appropriate for the flow scenarios addressed in this study but may become inaccurate under conditions with significant pressure gradients or compressibility effects. The rationale behind assumption (viii) is that the fluctuations of the flow are less critical compared with the mean quantities for predicting the mean wall shear stress. Consequently, the training dataset relies on average flow values from DNS, with added Gaussian noise to improve the robustness of the model.

3.3. *Input and output variables*

The goal of this section is to select the non-dimensional input variables for the wall model that are most informative for predicting the wall shear stress across the entire training dataset. The input variables include both flow variables and roughness topography parameters. The variables y_1 and y_2 denote the wall-normal distances to the centres of the first and second control volumes off the wall, respectively (as shown in figure 8), and are related to the WMLES grid resolution by $y_1 \approx \triangle/2$. The magnitudes of the corresponding wall-parallel velocities relative to the wall at y_1 and y_2 are $u_1 = (u_s(y_1)^2 + w_s(y_1)^2)^{1/2}$ and $u_2 = (u_s(y_2)^2 + w_s(y_2)^2)^{1/2}$, where u_s and w_s are samples generated as detailed below. Three different grid resolutions, $\triangle/\delta = 1/20, 1/10, 1/5$, are included in the generation of data for the input and output variables. The grid resolutions correspond to the condition where the first off-wall grid point lies within the buffer layer or the logarithmic layer, which aligns with typical grid resolutions in WMLES for engineering applications. The candidate input variables considered can be organized into five categories:

1) roughness parameters,

$$\frac{k_{avg}}{R_a}, \frac{k_c}{R_a}, \frac{k_t}{R_a}, \frac{k_{rms}}{R_a}, S_k, K_u, ES, P_o, I, \frac{L_{cor}}{R_a}, \lambda_f; \tag{3.2}$$

2) products of roughness parameters,

$$ES^2, S_k^2, ESS_k, ESK_u; \tag{3.3}$$

3) roughness height scaled by y_1,

$$\frac{k_{avg}}{y_1}, \frac{k_c}{y_1}, \frac{k_t}{y_1}, \frac{k_{rms}}{y_1}, \frac{y_1}{R_a}; \tag{3.4}$$

4) products of non-dimensional roughness parameters and roughness height scaled by y_1,

$$\frac{S_k k_{avg}}{y_1}, \frac{K_u k_{avg}}{y_1}, \frac{ES k_{avg}}{y_1}, \frac{P_o k_{avg}}{y_1}, \frac{I k_{avg}}{y_1}, \tag{3.5}$$

$$\frac{S_k k_{rms}}{y_1}, \frac{K_u k_{rms}}{y_1}, \frac{ES k_{rms}}{y_1}, \frac{P_o k_{rms}}{y_1}, \frac{I k_{rms}}{y_1}; \tag{3.5}$$

5) local Reynolds number at y_1 and y_2,

$$\frac{u_1 y_1}{\nu}, \frac{u_2 y_2}{\nu}. \tag{3.6}$$

The samples for $u(y_i)$ and $w(y_i)$ are generated based on a Gaussian distribution $\mathcal{N}[\mu, \sigma^2]$ with mean μ and variance σ^2 as

$$u_s(y_i) \sim \mathcal{N}\left[U(y_i), u_{rms}^2(y_i)\right], \quad w_s(y_i) \sim \mathcal{N}\left[W(y_i), w_{rms}^2(y_i)\right], \tag{3.7}$$

where $U(y_i)$ and $W(y_i)$ are the DNS streamwise and spanwise mean velocities, and $u_{rms}(y_i)$ and $w_{rms}(y_i)$ are the DNS root-mean-squared streamwise and spanwise velocity fluctuations at y_i (with $i = 1, 2$). Products of roughness parameters are also included in categories 2) and 4) as inputs to facilitate model training. Although the FNN can uncover nonlinear relationships between inputs and outputs, it was observed that incorporating quadratic relationships between roughness parameters enabled the training of higher-performing models with less data.

The non-dimensional output of the wall model is

$$\tilde{\tau}_w = \frac{\tau_{w,s} y_1}{\nu u_1}, \tag{3.8}$$

where the scaling factor $\nu u_1/y_1$ represents the naive estimation of the wall shear stress using finite differences. The values of $\tau_{w,s}$ are generated based on a Gaussian distribution $\tau_{w,s} \sim \mathcal{N}[\langle \tau_w \rangle, \tau_{w,rms}^2]$, where $\langle \tau_w \rangle$ is the mean wall-shear stress calculated from DNS, and $\tau_{w,rms}$ is computed based on the correlation $\tau_{w,rms}/\langle \tau_w \rangle = 0.298 + 0.018 \ln Re_\tau$ (Örlü & Schlatter 2011). As discussed in § 3.2, the main goal of the current wall model is to correctly predict mean flow quantities. It was found that this approach facilitates the training of more robust models than the use of actual instantaneous DNS values for velocities and wall shear stress.

We aim to select the best-performing input variables from all the candidates defined above. To that end, we use the 'minimum redundancy maximum relevance' (MRMR) algorithm to aid the finding of the optimal set of the input features (Ding & Peng 2005;

Peng *et al.* 2005). The MRMR algorithm is designed to rank input variables by considering both their relevance to the output and their redundancy with respect to each other. This helps improve the accuracy of the model and reduce overfitting by eliminating unnecessary inputs. Here, we use the MRMR algorithm to determine the order in which input features will be tested when training candidate models in the following section.

The relevance and redundancy are computed using mutual information, which measures the amount of information shared between variables. The mutual information I between a pair of random variables (ϕ, ψ) is defined as

$$I(\phi, \psi) = \iint PDF(\phi, \psi) \log(\frac{PDF(\phi, \psi)}{PDF(\phi)PDF(\psi)}) \mathrm{d}\phi \mathrm{d}\psi, \quad (3.9)$$

where $PDF(\phi, \psi)$ is the joint PDF of ϕ and ψ, and $PDF(\phi)$ and $PDF(\psi)$ are the marginal PDFs of ϕ and ψ, respectively. The MRMR algorithm ranks features by evaluating the importance score (also known as 'mutual information quotient') of an input ϕ_{in},

$$MIQ_{\phi_{in}} = \frac{V_{\phi_{in}}}{W_{\phi_{in}}}, \quad (3.10)$$

where $V_{\phi_{in}}$ is the relevance of the input feature ϕ_{in} with respect to the output variable ϕ_{out},

$$V_{\phi_{in}} = I(\phi_{in}, \phi_{out}) \quad (3.11)$$

and $W_{\phi_{in}}$ is the redundancy of the input feature ϕ_{in} with respect to the rest of input features ψ_{in},

$$W_{\phi_{in}} = \frac{1}{|\mathbf{S}|} \sum_{\psi_{in} \in \mathbf{S}} I(\phi_{in}, \psi_{in}), \quad (3.12)$$

where $|\mathbf{S}|$ is the number of features in the set of the input variables $\mathbf{S}$ maximizing $MIQ_{\phi_{in}}$. Higher relevance indicates a stronger association with the output, while lower redundancy implies less similarity between inputs.

We use MRMR to identify a subset of input variables that collectively maximize relevance to the target while minimizing redundancy among the selected inputs. Table 5 displays the sequence of the candidate input features informed by the MRMR importance scores in descending order, with the top-ranked inputs being the most informative for predicting the output. The results in table 5 suggests that the local Reynolds numbers, $u_1 y_1/\nu$ and $u_2 y_2/\nu$, along with the roughness features k_{rms}/R_a and ES^2, contain the most relevant information for the wall model. It is interesting that the ranking does not prioritize S_k, which has traditionally been a key ingredient in models and correlations for predicting k_s. However, as shown in figure 4, S_k is strongly correlated with k_{rms}/R_a in the present roughness database, where k_{rms}/R_a is the second most important input feature in the ranking. Thus, the effect of S_k is effectively accounted for by including k_{rms}/R_a.

3.4. *Wall model training*

The order of inputs provided by the ranking in table 5 is used in this section to determine the number of model inputs. An FNN is used to parameterize the relationship between inputs and outputs. The layers are connected using hyperbolic tangent sigmoid transfer functions as the activation function except for the last layer, which is connected with rectified linear units. The training algorithm utilized is gradient descent with momentum and adaptive learning rate backpropagation (Yu & Liu 2002). A total of 192 cases of

Sequence	1	2	3	4	5	6	7	8	9	10	11
Input feature	$\frac{u_1 y_1}{\nu}$	$\frac{k_{rms}}{R_a}$	ES^2	$\frac{u_2 y_2}{\nu}$	$\frac{k_c}{y_1}$	P_o	ES	$\frac{K_u k_{rms}}{y_1}$	λ_f	$\frac{k_t}{y_1}$	$\frac{P_o k_{rms}}{y_1}$
Sequence	12	13	14	15	16	17	18	19	20	21	22
Input feature	$\frac{K_u k_{avg}}{y_1}$	$\frac{k_{rms}}{y_1}$	ESS_k	$\frac{P_o k_{avg}}{y_1}$	$\frac{L_{cor}}{R_a}$	$\frac{k_{avg}}{y_1}$	ESK_u	$\frac{y_1}{R_a}$	$\frac{ESk_{avg}}{y_1}$	$\frac{k_{avg}}{R_a}$	$\frac{Ik_{avg}}{y_1}$
Sequence	23	24	25	26	27	28	29	30	31	32	
Input feature	$\frac{k_c}{R_a}$	$\frac{k_t}{R_a}$	K_u	S_k	I	$\frac{Ik_{rms}}{y_1}$	$\frac{S_k k_{rms}}{y_1}$	$\frac{S_k k_{avg}}{y_1}$	S_k^2	$\frac{ESk_{rms}}{y_1}$	

Table 5. Ranking of candidate input variables for the wall model according to MRMR importance score in descending order.

$k_{avg}/\triangle$	$k_c/\triangle$	$k_t/\triangle$	$k_{rms}/\triangle$	$R_a/\triangle$
0.06~2.1	0.38~3.55	0.35~3.49	0.04~0.56	0.03~0.47

Table 6. Range of applicability for BFWM-rough for SGS roughness: ratio of roughness height parameters to the grid size of the training dataset.

turbulent channel flows are randomly divided into training dataset (70 % of the total), validation dataset (15 % of the total) and testing dataset (15 % of the total). The training dataset is used to develop a candidate BFWM-rough model; the validation dataset, while not used for parameter estimation, provides the stopping criterion during training to prevent overfitting and enhance the generalizability of the model. The model architecture – including the number of hidden layers and neurons – and hyperparameters are optimized based on performance metrics from the validation dataset. The testing dataset, not involved in the training process, serves as an independent set of cases to assess the performance of BFWM-rough on untrained scenarios.

To find the best-performing model, different numbers of input features are tested according to the order presented in table 5. Note that this ranking is key to limit the number of possible input combinations. For each number of inputs, 100 random splits of the training, validation and testing datasets are conducted. For each split, the optimal number of hidden layers and neurons per layer is determined through a grid search. The number of hidden layers and neurons per layer considered in the grid search ranges from 3 to 6 and from 5 to 20, respectively. The optimal model for each combination of inputs is then determined based on the minimum L_2-norm error over the entire training, validation and testing datasets from the 100 random splits. A total of 8000 central processing unit hours were taken to obtain the optimal wall model.

The errors for different candidate wall models are plotted against the number of inputs in figure 9. The results show that the wall model requires at least five non-dimensional inputs to achieve errors below 10 %. A less significant reduction in error, from 8 % to 4.6 %, is observed with a higher number of inputs after the fifth ranked input. As an interesting observation, it was not possible to train wall models with errors below 10 % without including roughness parameters non-dimensionalized by the grid size.

According to the Buckingham-π theorem (Buckingham 1914), the number of dimensionless numbers (π terms) required to uniquely determine the wall shear stress in the rough-wall channel simulations equals the total number of parameters needed to set up the cases minus the independent fundamental units. In this study, there are
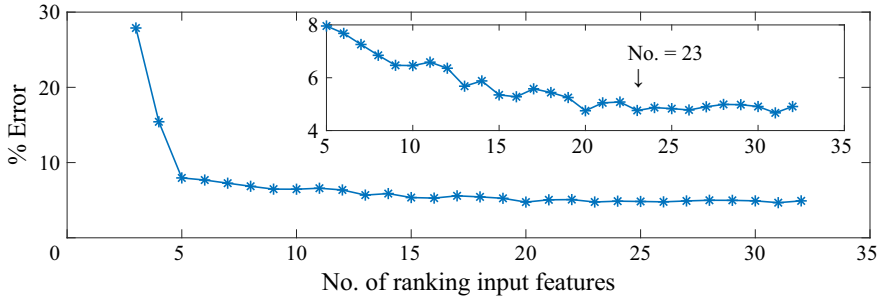
Figure 9. *Apriori* L_2-norm error in the prediction of $\tilde{\tau}_w$ as function of the number of input features as ranked in table 5. The inset shows the errors when the number of inputs ranges from 5 to 32.
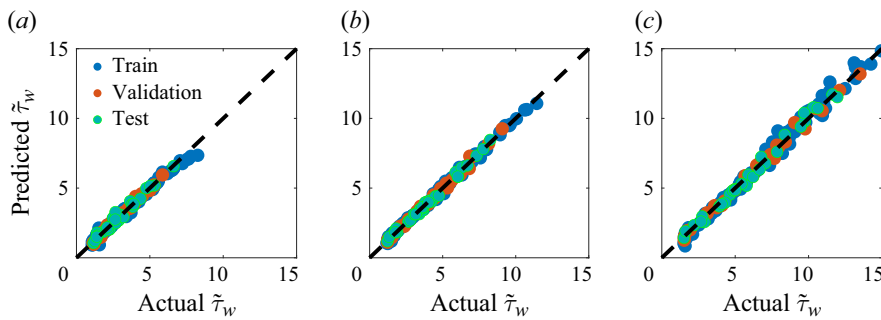

Figure 10. Scatter plot of regression results of actual $\tilde{\tau}_w$ and predicted $\tilde{\tau}_w$ for BFWM-rough. The results are plotted for three grid resolutions: (*a*) $\triangle/\delta = 1/20$; (*b*) $\triangle/\delta = 1/10$; (*c*) $\triangle/\delta = 1/5$.

seven parameters: U_c (mean centreline velocity), δ, ν, k_{rms}, H, H_f and κ_0. The number of fundamental units is two (length and time). Therefore, a total of five dimensionless numbers is required to completely specify the case, and hence the wall shear stress. This implies that any model aiming to accurately predict the wall shear stress is expected to require five non-dimensional inputs. This might explain why the model error drops significantly after five inputs. In practice, the non-dimensional inputs available to the wall model do not include global parameters of the case (such as U_c and δ). Instead, local quantities are used as proxies. For that reason, adding more inputs can still inform the model predictions.

Based on the results shown in figure 9, the model with 23 inputs is selected as the BFWM-rough model. This model was chosen because it has the minimum number of inputs while maintaining an L_2-norm error below 5 %. The corresponding FNN comprises five hidden layers with 15 neurons per layer. The L_2-norm error for BFWM-rough is 4.76 % across all datasets. As an example, regression results for three different grid resolutions are illustrated in figure 10. It is also useful to evaluate the range of applicability of BFWM-rough in terms of typical roughness heights relative to the WMLES grid resolution. The ratio of roughness height parameters in the training database to the training grid size is summarized in table 6. The BFWM-rough covers a peak-to-valley roughness height k_c from $0.38\triangle$ to $3.55\triangle$, and a root-mean-square roughness height k_{rms} from $0.04\triangle$ to $0.56\triangle$. In WMLES scenarios characterized by higher ratios (i.e. finer grid resolutions relative to typical roughness element sizes) it may be more accurate to geometrically resolve the roughness with the WMLES mesh and use a wall model for smooth surfaces. The *aposteriori* performance of BFWM-rough is assessed in actual WMLES in § 4.

3.5. *Confidence score*

A GP model, similar to the one introduced for AL in § 2.2, is used to calculate a confidence score for the roughness topology. The main objective is to identify potential deficiencies in the BFWM-rough when applied to surfaces with roughness features that differ significantly from those in the training dataset. The GP model is trained on all cases considered for BFWM-rough, but it only utilizes the non-dimensional roughness parameters as inputs. The confidence score is defined as $C = min\{\sigma_{tr}^2/\sigma^2, 1\}$, where σ_{tr}^2 is the mean prediction variance of the GP model over the training dataset, and σ^2 is the predictive variance of the case for which the confidence is computed. Note that the confidence score provides an assessment only of the roughness geometry and does not account for the flow conditions. When $\sigma_{tr}^2 \approx \sigma^2$, then $C \approx 1$ and confidence in the roughness geometry is high. This indicates that the geometric properties of the roughness considered are similar to those the BFWM-rough was trained on. Conversely, large uncertainty in the roughness surface properties will increase σ^2, thereby lowering the confidence score $C \ll 1$. In those cases, the prediction from BFWM-rough might be subject to significant errors.

4. Model evaluation

We evaluate the performance of the BFWM-rough across different cases. The testing cases include 556 turbulent channel flows over different rough surfaces. The model is also assessed in a turbulent flow over a HPT blade with two different surface roughnesses.

Two sources of errors can be identified in the evaluation of a wall model (Lozano-Durán *et al.* 2022): errors from the outer LES input data, referred to as external wall-modelling errors, and errors from the wall model physical assumptions, referred to as internal wall-modelling errors. In the former, errors from the SGS model at the matching locations propagate to the value of τ_w predicted by the wall model. These errors can be labelled as external to the wall model inasmuch as they are present even if the wall model provides an exact physical representation of the near-wall region. The second source of errors represents the intrinsic wall-model limitations: even in the presence of exact values for the input data, the prediction might be inaccurate when the physical assumptions the model is rooted in do not hold. In BFWM-rough, internal errors may come from the breakdown of the assumptions discussed in § 3.2 (e.g. lack of quasi-equilibrium conditions, non-local effects, untrained roughness topologies, etc.). The combined external plus internal error is referred to as total error. In the following, we use *apriori* testing to assess the internal errors of BFWM-rough and *aposteriori* testing to evaluate the total errors.

Apriori performance is assessed by the relative error in the BFWM-rough model output when the input is generated from DNS data. The model was implemented in actual WMLES to perform *aposteriori* testing. For turbulent channel flows, WMLES with BFWM-rough was conducted using our in-house code (Bae *et al.* 2019; Lozano-Durán & Bae 2019*a*,b). For HPT blade simulation with wall roughness, the BFWM-rough was implemented into the high-fidelity solver charLES, developed by Cascade Technologies, Inc (Bres *et al.* 2018; Fu *et al.* 2021).

4.1. *Comparison with existing rough-wall models*

The performance of BFWM-rough is compared with existing models in the literature through *apriori* testing. Five models are considered, and their functional forms are summarized in table 7. Note that previous models predict either the roughness function $\triangle U^+$ or the equivalent sand-grain roughness height k_s^+, whereas BFWM-rough predicts

Model form	Reference
$k_s^+ = \begin{cases} 2.73k_{rms}^+(2+S_k)^{-0.45}, \ S_k < 0 \\ 2.11k_{rms}^+, \ S_k = 0 \\ 2.48k_{rms}^+(1+S_k)^{2.24}, \ S_k > 0 \end{cases}$	Flack *et al.* (2020)
$k_s^+ = \begin{cases} 4.0k_{rms}^+(1+0.17S_k)^4, \ S_k \leqslant 0 \\ 2.48k_{rms}^+(1+S_k)^{2.24}, \ S_k > 0 \end{cases}$	Kuwata & Kawaguchi (2019)
$\triangle U^+ = \frac{1}{\kappa}\ln(R_a^+) + 1.12ES + 1.47$	Chan *et al.* (2015)
$\triangle U^+ = \frac{1}{\kappa}\ln(ESk_{rms}^+) + 3.5$	De Marchis *et al.* (2020)
$\triangle U^+ = c_1\ln(c_2k_{rms}^+ES)[(c_3-1)\tanh(c_4S_k)+1]e^{-c_5ES^{c_6}}$,	Bornhoft *et al.* (2024)
$c_1 = 3.026, c_2 = 3.444, c_3 = 28.56, c_4 = 0.0031, c_5 = 0.353, c_6 = 0.894.$	

Table 7. Models considered for comparison with the BFWM-rough.

the wall shear stress. We use $\triangle U^+$ to enable comparisons across different models. For models based on correlations for k_s^+, the predicted roughness function $\triangle U_{pred}^+$ is computed based on (2.10). For BFWM-rough, $\triangle U_{pred}^+$ is computed according to the equation $\triangle U^+ = U_s^+ - U_r^+$ in the logarithmic layer (Flack & Schultz 2010). This equation can be rewritten in the following formulation to relate $\triangle U^+$ with the predicted wall-shear stress from BFWM-rough:

$$\triangle U_{pred}^+ = U_s^+ \left(1 - \sqrt{\frac{\tau_{w,s}}{\tau_{w,r}}}\frac{U_r}{U_s}\right), \tag{4.1}$$

where $\tau_{w,r}$ is the predicted wall-shear stress from BFWM-rough, $\tau_{w,s}$ is the wall-shear stress from the DNS of smooth-wall channel flow, and U_s and U_r are obtained from the mean velocity profiles of DNS for smooth- and rough-wall channel flows, respectively. The equation assumes that U_s^+, $\tau_{w,s}$ and U_r are the true values obtained from DNS, while $\tau_{w,r}$ is derived from the model prediction. This establishes an *apriori* relationship between $\tau_{w,r}$ and $\triangle U^+$.

The performance of different models for Gaussian and Weibull roughness is shown in figure 11. The correlation proposed by Flack *et al.* (2020) provides good predictions for Gaussian roughness with high ES and Weibull roughness with low S_k, as shown in figure 11(*a*). This aligns with the assertion from Flack *et al.* (2020) that this correlation does not adequately capture wavy surfaces with low ES or surfaces with high positive skewness. A similar correlation proposed by Kuwata & Kawaguchi (2019) improves predictions for Gaussian roughness with negative S_k (figure 11*b*), although it still overestimates most Weibull roughness with high S_k. This may be due to the fact that the correlation was developed for rough surfaces with moderate skewness ($-1 < S_k < 1$). The correlation proposed by Chan *et al.* (2015) (figure 11*c*) provides accurate predictions for Weibull roughness with positive S_k but overpredicts drag for both Gaussian roughness and Weibull roughness with negative S_k. De Marchis *et al.* (2020) found a linear-log correlation between $\triangle U^+$ and $ES \cdot k_{rms}^+$, which overestimates Gaussian roughness and Weibull roughness with negative S_k while underpredicting Weibull roughness with high positive S_k, as shown in figure 11(*d*). Figure 11(*e*) displays results from a correlation with a more comprehensive treatment of roughness parameters, proposed by Bornhoft *et al.* (2024). This model offers improved predictions compared with the previously discussed models; however, it still overpredicts Gaussian roughness and shows significant errors
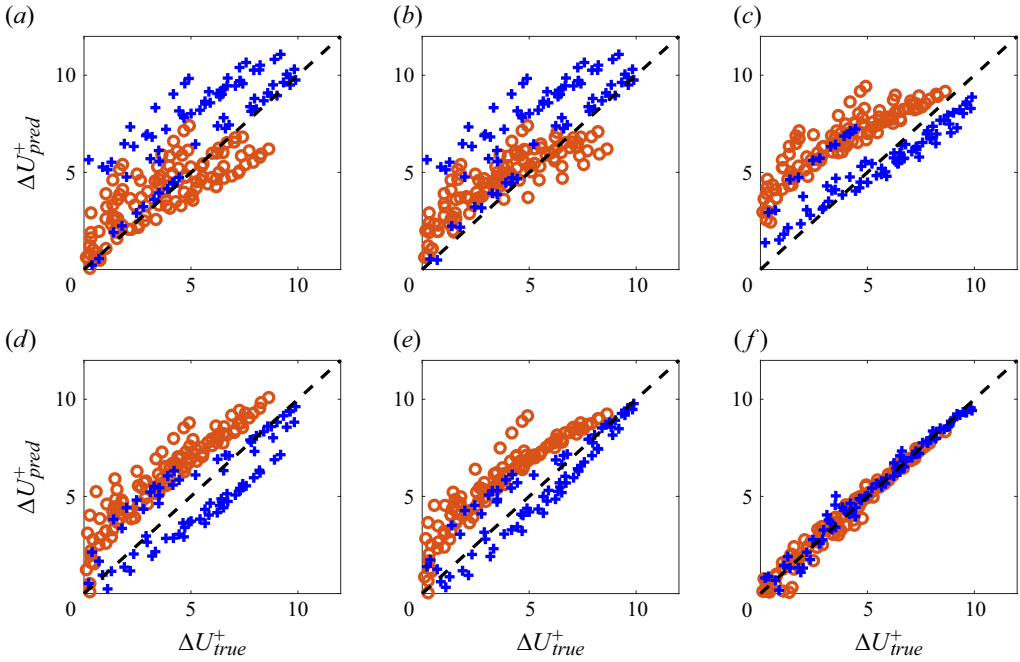
Figure 11. Comparison with rough-wall models from table 7 on the overall 192 rough cases in § 3.4: (*a*) model for k_s^+ from Flack *et al.* (2020); (*b*) model for k_s^+ from Kuwata & Kawaguchi (2019); (*c*) model for $\triangle U^+$ from Chan *et al.* (2015); (*d*) model for $\triangle U^+$ from De Marchis *et al.* (2020); (*e*) model for $\triangle U^+$ from Bornhoft *et al.* (2024); (*f*) results from the BFWM-rough. Here red circle, Gaussian roughness; blue cross, Weibull roughness.

for certain Weibull roughness profiles. Finally, results for BFWM-rough are shown in figure 11(*f*), where it provides the most accurate predictions for both training and testing rough cases. For fairness, it should be noted that our model achieves the highest accuracy partly because it was specifically trained on Gaussian and Weibull roughness. However, it is also worth emphasizing that BFWM-rough can be retrained with an expanded roughness database using the workflow outlined in this study.

4.2. *Rough-wall turbulent channel flow*

The performance of BFWM-rough is evaluated first in WMLES of turbulent channel flows. The WMLES equations are solved using staggered second-order finite differences and a fractional-step method with a third-order Runge–Kutta time advancement scheme (Bae *et al.* 2019; Lozano-Durán & Bae 2019*a*,b). The dynamic Smagorinsky model (Germano *et al.* 1991; Lilly 1992) is used as the SGS model. The simulations are conducted by fixing the Reynolds number based on the mean centreline velocity, $Re_c = U_c \delta / \nu$. The streamwise, wall-normal and spanwise lengths of the computational domain are $2\pi\delta$, 2δ and $\pi\delta$, respectively. Six grid resolutions are considered, all of them with equal grid size in each spatial direction: $\triangle/\delta = 1/5, 1/8, 1/10, 1/15, 1/20$ and $1/30$. The simulations were carried out for 30 eddy turnover times after transients.

Hereafter, we use the term 'untrained' (or testing) to indicate that the model was never trained for that particular rough surface, rough Reynolds number (k_s^+) and/or grid resolution. The term 'trained' is used when the model was explicitly trained for that condition. The four categories of cases examined are as follows:

Case	Re_τ	$\hat{k}_s^+$	A-priori error of τ_w						A-posteriori error of τ_w					
			Training grid $\triangle/\delta$			Testing grid $\triangle/\delta$			Training grid $\triangle/\delta$			Testing grid $\triangle/\delta$		
			1/20	1/10	1/5	1/15	1/8	1/30	1/20	1/10	1/5	1/15	1/8	1/30
WB13	720	142.1	−2.42	5.48	4.52	1.50	2.41	34.41	−17.05	−7.38	−7.38	−12.5	−6.76	−23.56
WB06	360	10.2	−0.51	−2.99	−11.76	−9.53	−0.03	−26.45	2.55	7.64	17.10	1.41	11.47	−10.06
GS03	720	21.5	0.75	4.56	1.17	0.06	4.51	6.84	−4.14	4.68	9.77	−1.02	6.72	−11.00
WB05	360	32.7	1.60	−2.40	−5.92	−3.41	−3.82	13.49	−8.33	2.64	11.99	−4.67	8.33	−16.46
GS14	540	33.9	1.97	−1.85	−2.31	8.27	5.53	−5.65	−0.77	10.95	16.11	5.16	13.84	−9.09
GS18	1000	55.8	4.95	2.61	7.23	−0.11	−6.34	10.42	2.34	13.04	12.42	7.04	15.11	−5.59

Table 8. Relative errors (in %) of τ_w for the trained Gaussian/Weibull rough surfaces at trained/untrained grid resolutions and untrained Reynolds numbers. Note that these cases are from the testing dataset in § 3.4. The relative error is computed based on the predicted value from the BFWM-rough and the actual value from DNS in turbulent channel flows. The table shows case name, Re_τ, and $\hat{k}_s^+$.

(i) trained Gaussian/Weibull rough surfaces;
(ii) untrained Gaussian/Weibull rough surfaces;
(iii) untrained bimodal Gaussian rough surfaces;
(iv) untrained ellipsoidal, sinusoidal, Fourier-mode and sand-grain rough surfaces from Jouybari *et al.* (2021).

Cases in (i), (ii) and (iii) are evaluated at untrained grid resolutions and untrained k_s^+, whereas cases in (iv) are only conducted for untrained k_s^+.

4.2.1. *Trained Gaussian/Weibull rough surfaces*

Six cases from the testing datasets in § 3.4, spanning both fully and transitionally rough regimes, are selected for evaluation: WB13-$Re_\tau = 720$, WB06-$Re_\tau = 360$, GS03-$Re_\tau = 720$, WB05-$Re_\tau = 360$, GS14-$Re_\tau = 540$ and GS18-$Re_\tau = 1000$. Note that the six rough surfaces are seen during the training process, however, they are evaluated at untrained rough Reynolds numbers. The cases are also examined for three training grid sizes ($\triangle/\delta = 1/20, 1/10, 1/5$) and three testing grid sizes ($\triangle/\delta = 1/30, 1/15, 1/8$). The grid sizes $\triangle/\delta = 1/15$ and $\triangle/\delta = 1/8$ fall within the range of the training grids, while $\triangle/\delta = 1/30$ is outside this range.

The *apriori* and *aposteriori* errors for the predicted wall shear stress are listed in table 8. The *apriori* mean error is 6 % with a standard deviation of 7 %. The errors range from 0 % to 30 %, with the latter occurring for rough surfaces with Weibull distributions at the finest grid resolution considered ($\triangle/\delta = 1/30$). This discrepancy could be attributed to two factors. First, the characterization of Weibull surfaces requires more parameters than Gaussian roughness, complicating predictions due to the increased dimensionality of the input space. Second, the finest grid resolution falls outside the bounds of the training set. These factors combined make Weibull surfaces at fine grid resolutions more susceptible to inaccurate results due to model extrapolation. The mean *aposteriori* error is 9 %, with a standard deviation of 5 %. In some instances, *aposteriori* errors are lower than *apriori* errors. However, this apparent improvement is due to error cancellation arising from external errors in the SGS model.

The streamwise mean velocity profiles from WMLES are plotted and compared with DNS results in figure 12. To facilitate the visualization of near-wall errors from WMLES, the mean velocity U is normalized by U_c instead of u_τ. The wall-normal distance is non-dimensionalized by $(y - d)/\delta$. Three grid resolutions are shown: $\triangle/\delta = 1/20$, $\triangle/\delta = 1/15$ and $\triangle/\delta = 1/30$. The agreement between DNS and WMLES is within 5 %,
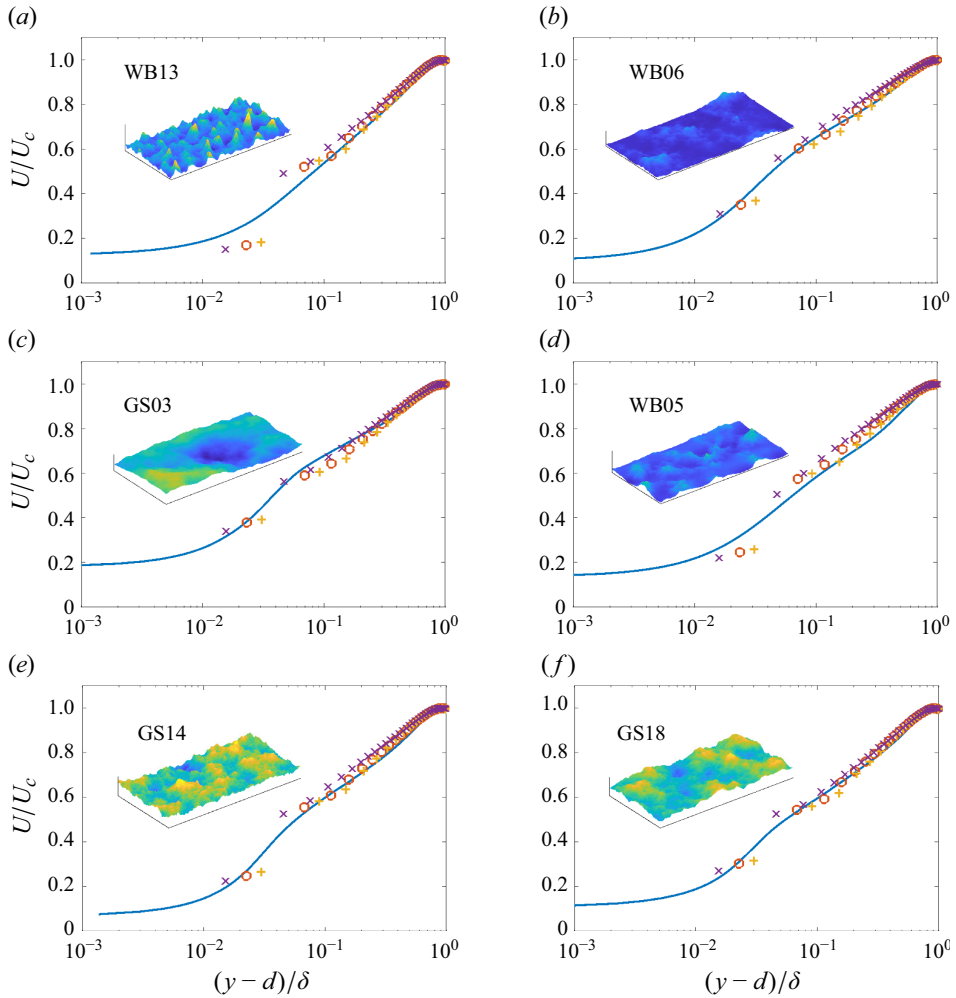
Figure 12. Mean velocity profiles for DNS (line) and WMLES (symbols) of turbulent channel flows for selected test cases: (*a*) WB13 at $Re_\tau = 720$; (*b*) WB06 at $Re_\tau = 360$; (*c*) GS03 at $Re_\tau = 720$; (*d*) WB05 at $Re_\tau = 360$; (*e*) GS14 at $Re_\tau = 540$; (*f*) GS18 at $Re_\tau = 1000$. The roughness geometries are visualized with a section of δ in x and 0.5δ in z. Three grid resolutions are visualized: $\triangle/\delta = 1/20$ (○); $\triangle/\delta = 1/15$ (+); $\triangle/\delta = 1/30$ (×).

demonstrating the capability of BFWM-rough in predicting the mean velocity profiles for both fully and transitionally rough cases. It is important to note that the SGS model plays a crucial role in predicting the mean velocity profile. Therefore, most of the errors observed in figure 12 are likely dominated by deficiencies in the SGS model rather than internal errors from BFWM-rough.

4.2.2. *Untrained Gaussian/Weibull rough surfaces*

We evaluate BFWM-rough on surfaces from the roughness repository whose geometrical features are not included in the training process. Additionally, these cases contain untrained k_s and are assessed in untrained grid resolutions. To identify the most challenging cases, we use the confidence score C introduced in § 3.5 to select rough surfaces with the lowest confidence for evaluation. Figure 13 displays the confidence
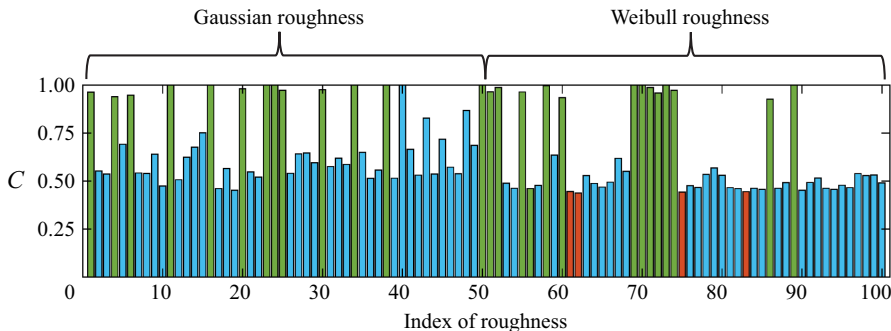
Figure 13. Confidence score for each surface in the roughness repository. Surface indices 1–50 are Gaussian roughness, and surface indices 51–100 are Weibull roughness. The training cases are coloured in green. The four cases with the smallest C (coloured in red) are selected for evaluation of BFWM-rough.

Case	k_{avg}	k_c	k_t	k_{rms}	R_a	S_k	K_u	ES	I	P_o	λ_f	L_{cor}
WB14	0.026	0.128	0.122	0.013	0.010	1.510	6.593	0.308	−0.012	0.798	0.172	0.064
WB15	0.031	0.140	0.119	0.013	0.010	1.227	5.402	0.336	0.005	0.779	0.144	0.060
WB16	0.040	0.184	0.162	0.019	0.014	1.384	6.008	0.289	0.063	0.786	0.201	0.105
WB17	0.028	0.109	0.089	0.011	0.009	0.952	4.495	0.327	0.012	0.748	0.177	0.061
BM01	0.047	0.151	0.139	0.022	0.018	0.597	3.007	0.574	−0.001	0.693	0.264	0.064
BM02	0.046	0.136	0.132	0.020	0.016	0.468	2.804	0.547	0.008	0.666	0.264	0.067
BM03	0.039	0.109	0.105	0.017	0.014	0.516	2.810	0.617	−0.014	0.646	0.298	0.058
BM04	0.084	0.123	0.122	0.018	0.014	−0.530	2.814	0.635	−0.028	0.315	0.276	0.055

Table 9. Roughness parameters of Weibull and Bimodal rough surfaces for model evaluation.

scores for each surface in the roughness repository. As expected, the confidence levels for all training cases (coloured in green) are around 100 %. The confidence scores for the untrained rough surfaces range from 45 % to 90 %. The four cases with the lowest C scores (WB14, WB15, WB16 and WB17, coloured in red) are selected for evaluation of BFWM-rough. The geometrical parameters of these four surfaces are summarized in table 9.

The errors in τ_w are shown in figures 14 and 15 for both *apriori* and *aposteriori* testing across different training and testing grid resolutions. The streamwise mean velocity profiles from WMLES are plotted in figure 16 and compared with those from DNS for some selected cases. We should note that these selected untrained rough surfaces are the most challenging test cases with the least confidence scores. Generally, in *apriori* testing, BFWM-rough tends to overpredict drag in the transitionally rough regime and underpredict drag in the fully rough regime. However, this pattern is not always seen in *aposteriori* testing due to possible error cancellation. Both *apriori* and *aposteriori* errors decrease with increasing k_s^+ in the transitionally rough regime and stabilize in the fully rough regime. The BFWM-rough exhibits *apriori* errors ranging from −20 % to 40 % across the training grids, and errors within 25 % in *aposteriori* testing due to error cancellation. As summarized in table 10, the mean and standard deviation of drag prediction errors on Weibull roughness are within 8 % and 10 %, respectively. In the transitionally rough regime, the mean prediction error is smaller than in the fully rough regime, but the standard deviation is larger. The error variation in *aposteriori* testing shows a similar trend with increasing k_s^+, with values generally lower than in *apriori* testing. The
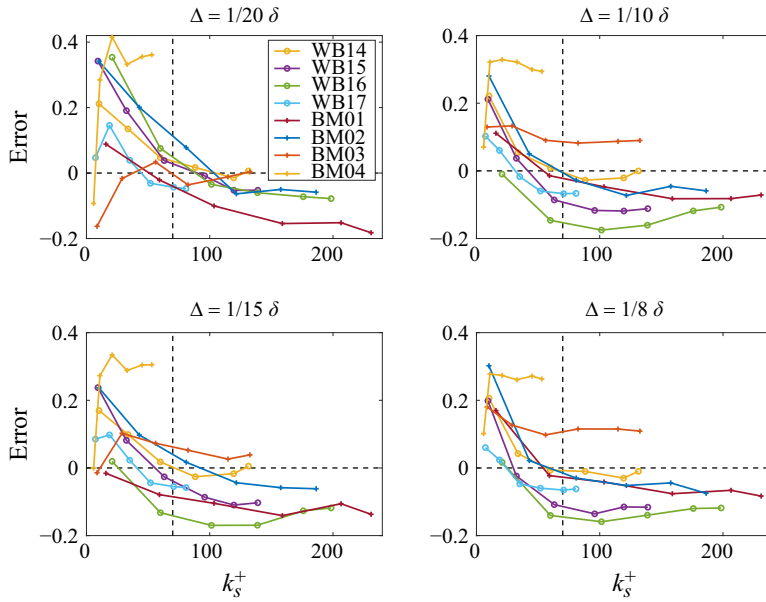
Figure 14. *Apriori* error variation of wall-shear stress against k_s^+. Untrained Weibull and Bimodal rough surfaces are tested for $Re_\tau = 180, 360, 540, 720, 900$ and 1000 with trained and untrained grid resolutions, as shown by the symbols with increasing k_s^+. The error is given as a fraction. Note that these selected untrained rough surfaces are the most challenging test cases with the least confidence scores.
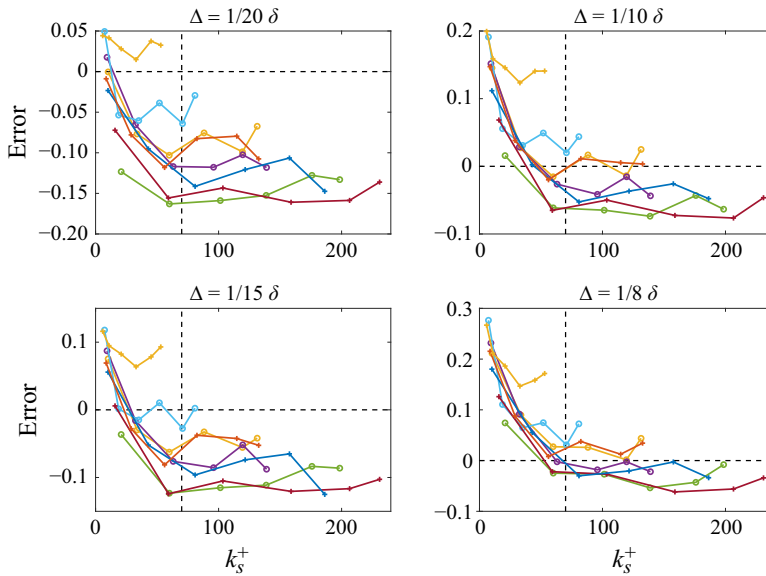

Figure 15. *Aposteriori* error variation of wall-shear stress against k_s^+. Untrained Weibull and Bimodal rough surfaces are tested in WMLES of turbulent channel flow at $Re_\tau = 180, 360, 540, 720, 900$ and 1000 with trained and untrained grid resolutions, as shown by the symbols with increasing k_s^+. The error is given as a fraction, and the legend is the same as figure 14.
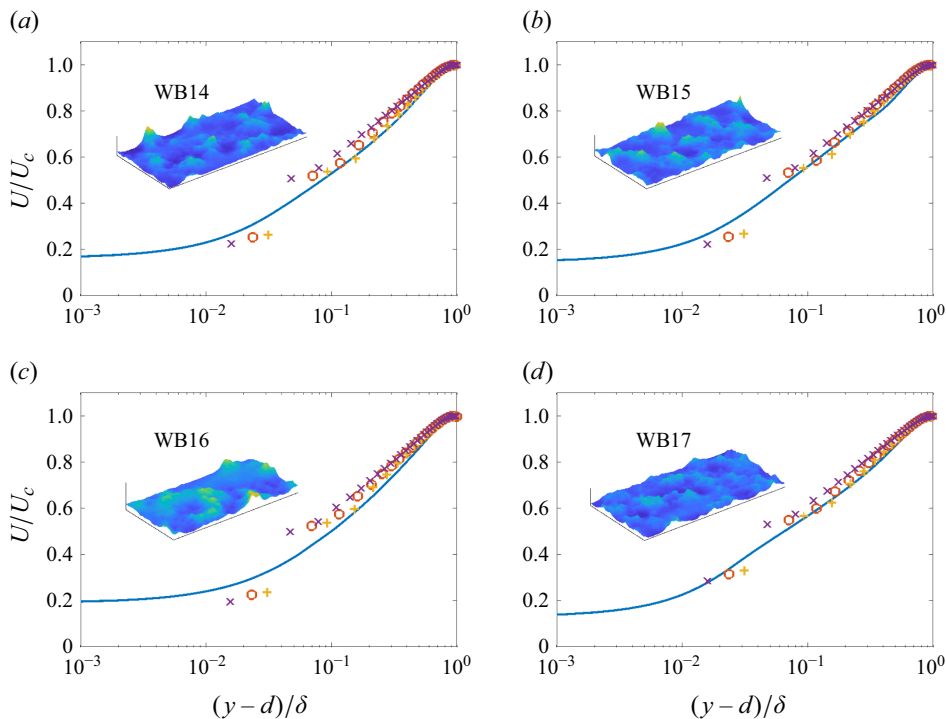
Figure 16. Mean velocity profiles for DNS (line) and WMLES (symbols) of turbulent channel flows at $Re_\tau = 1000$ with test Weibull surfaces: (*a*) WB14; (*b*) WB15; (*c*) WB16; (*d*) WB17. The roughness geometries are visualized with a section of δ in x and 0.5δ in z. Three grid resolutions are visualized: $\triangle/\delta = 1/20$ ($\circ$); $\triangle/\delta = 1/15$ (+); $\triangle/\delta = 1/30$ ($\times$).

Test untrained roughness	*A priori* error of τ_w				*A posteriori* error of τ_w			
	Transitionally		Fully		Transitionally		Fully	
	μ_{err}	σ_{err}	μ_{err}	σ_{err}	μ_{err}	σ_{err}	μ_{err}	σ_{err}
Weibull	4.4	10.4	−7.7	5.0	1.2	7.8	−6.5	3.9
Bimodal	17.1	13.8	−2.6	7.9	4.5	8.0	−8.5	3.1

Table 10. Model evaluation in the transitionally and fully rough regimes, demonstrated by the mean and standard deviation (in %) of the errors in the testing untrained Weibull and Bimodal rough cases (crossing different Re_τ with training and testing grids).

best predictions are achieved for WB17, with errors mostly under 10 % in both *apriori* and *aposteriori* testing across all grid resolutions. In contrast, WB16 shows the largest errors, likely due to its high k_{rms} and large correlation wavelength (small ES), which fall outside the characteristics of Weibull rough cases in the training database.

4.2.3. *Untrained bimodal Gaussian rough surfaces*

In this section, BFWM-rough is evaluated on rough surfaces with a bimodal Gaussian distribution. The goal is to assess the performance of BFWM-rough on a new roughness type that shares some similarities with the surfaces included in the roughness repository
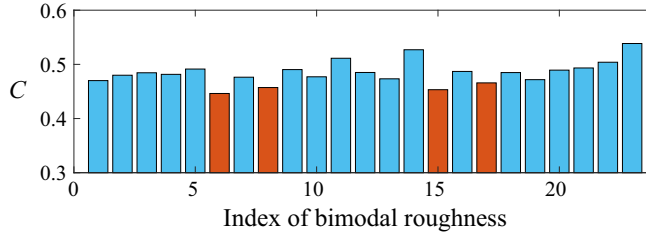
Figure 17. Confidence scores for 23 bimodal rough surfaces not included in the roughness repository. The four cases with the smallest C (coloured in red) are selected for evaluation of BFWM-rough.

but does not follow the same generation process. The selection of bimodal Gaussian roughness is motivated by the presence of engineering surfaces, which are frequently generated by successive processes or multiple factors. This introduces two or more different modal roughness distributions into the final surface (Peng & Bhushan 2000). One example of a bimodal roughness distribution is observed in ice accretion on airfoils (Bornhoft *et al.* 2022). The bimodal Gaussian distributions are constructed by combining two Gaussian distributions according to (Peng & Bhushan 2000)

$$PDF_B(\tilde{k}) = PDF_G(\tilde{k}; 0, 1) + PDF_G(\tilde{k}; \mu_G, \sigma_G^2) - PDF_G(\tilde{k}; 0, 1) PDF_G(\tilde{k}; \mu_G, \sigma_G^2), \tag{4.2}$$

where $PDF_B(\tilde{k})$ is the PDF of the bimodal distribution, $PDF_G(\tilde{k}; 0, 1)$ is a normal distribution and $PDF_G(\tilde{k}; \mu_G, \sigma_G^2)$ is the Gaussian distribution with randomized mean $0 < \mu_G < 0.5$ and randomized variance $0 < \sigma_G^2 < 0.5$. The power spectra is specified in the same way as described in § 2.1. The resulting surface map is then scaled from 0 to the root mean square height normalized by the channel half-height k_{rms}/δ, whose values are randomly chosen in the range of $0.005 < k_{rms}/\delta < 0.030$. A total of 23 bimodal rough surfaces are generated. Their confidence scores, shown in figure 17, are approximately 50 %. The most challenging surfaces for evaluation are identified as those with the lowest confidence score: BM01, BM02, BM03 and BM04 (where BM refers to bimodal roughness). Their roughness parameters are listed in table 9.

The performance of BFWM-rough for BM01, BM02, BM03 and BM04 is shown in figures 14 and 15. Similar error variations as Weibull roughness are observed with increasing k_s^+. Table 10 shows that the mean and standard deviation of errors for Bimodal roughness are larger than Weibull in the transitionally rough regime, but comparable to Weibull in the fully rough regime. It is also worth noting that for both Weibull and Bimodal roughness, the errors are generally smaller in *aposteriori* testing than *apriori* testing due to the internal and external error cancellation. Despite the mean errors being larger than in the previous sections, the results still demonstrate that BFWM-rough can offer reasonable predictions for new roughness types as long as these follow similar distributions to those it was trained for. The largest errors are obtained for BM04. The primary distinction between BM04 and the other three surfaces, as well as surfaces in the training database, is the combination of large negative S_k and large effective slopes for BM04. These roughness features can lead to less drag; however, since the training database lacks this information, the wall model results in an overestimation of the wall shear stress for BM04. The mean velocity profiles of WMLES compared with DNS are shown in figure 18. The improved agreement with the DNS results for BM04 at the grid resolutions of $\triangle/\delta = 1/20$ and $\triangle/\delta = 1/30$ is due to error cancellation in *aposteriori* evaluation.
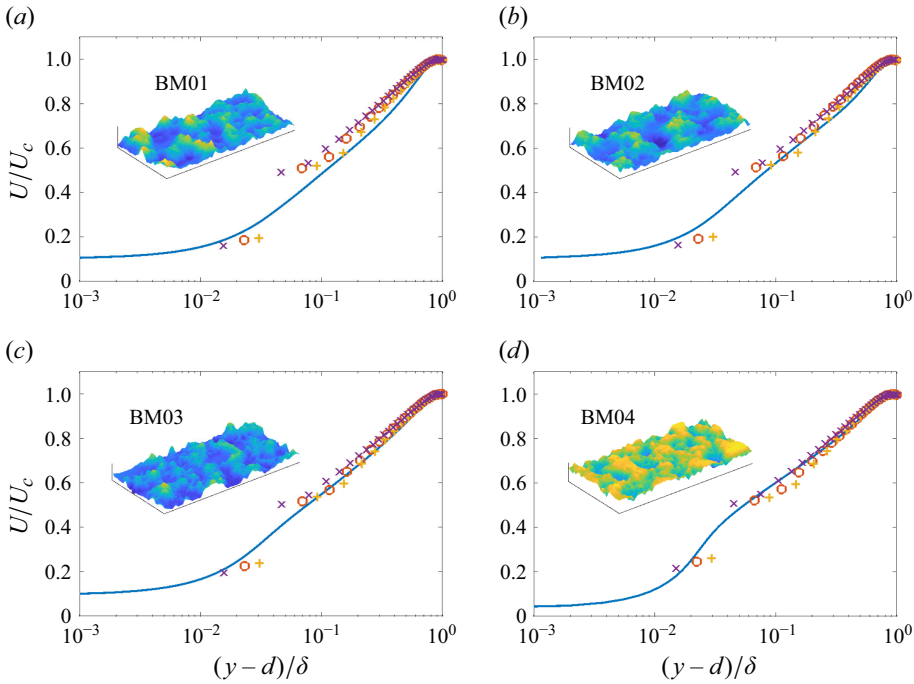
Figure 18. Mean velocity profiles for DNS (line) and WMLES (symbols) of turbulent channel flows at $Re_\tau = 1000$ with bimodal surfaces: (*a*) BM01; (*b*) BM02; (*c*) BM03; (*d*) BM04. The roughness geometries are visualized with a section of δ in x and 0.5δ in z. Three grid resolutions are visualized: $\triangle/\delta = 1/20$ (○); $\triangle/\delta = 1/15$ (+); $\triangle/\delta = 1/30$ (x).

4.2.4. *Untrained rough surfaces from Jouybari et al. (2021)*

The final cases analysed are the roughness types from Jouybari *et al.* (2021). The goal is to assess the performance of BFWM-rough on surfaces with topologies significantly different from the roughness repository used for training. A total of 42 rough surfaces are evaluated (labelled from C01 to C42), including 27 fully rough and 15 transitionally rough cases. The cases comprise ellipsoidal, sinusoidal, Fourier-mode and sand-grain roughnesses. The topology of some of the surfaces is shown in figure 19. The exact $\tilde{\tau}_w$ for all 42 rough cases is obtained from the DNS of turbulent channel flows at $Re_\tau = 1000$ as reported by Jouybari *et al.* (2021).

The confidence score for each surface is presented in figure 19(*a*). Overall, the results demonstrate the capability of the confidence score to identify untrained surfaces. The confidence levels range from 0.35 to 0.45, which are lower than those for the previous test cases involving untrained Weibull and Bimodal roughnesses. The highest confidence scores are found for cases C31 to C42, which correspond to sand-grain roughness and surfaces generated by low-order Fourier modes. These cases share more similarities with the isotropic roughness in the training dataset, explaining their higher confidence scores. The lowest confidence is observed for ellipsoidal roughness (C01 to C24) and roughness with streamwise sinusoidal waves (C25 to C30). This is expected, as the ellipsoidal elements are generated with varying orientations and semiaxis lengths, and the sinusoidal rough surfaces feature only streamwise waves. Both of these roughness types represent strongly anisotropic roughness, differing significantly from the surfaces in the training database.
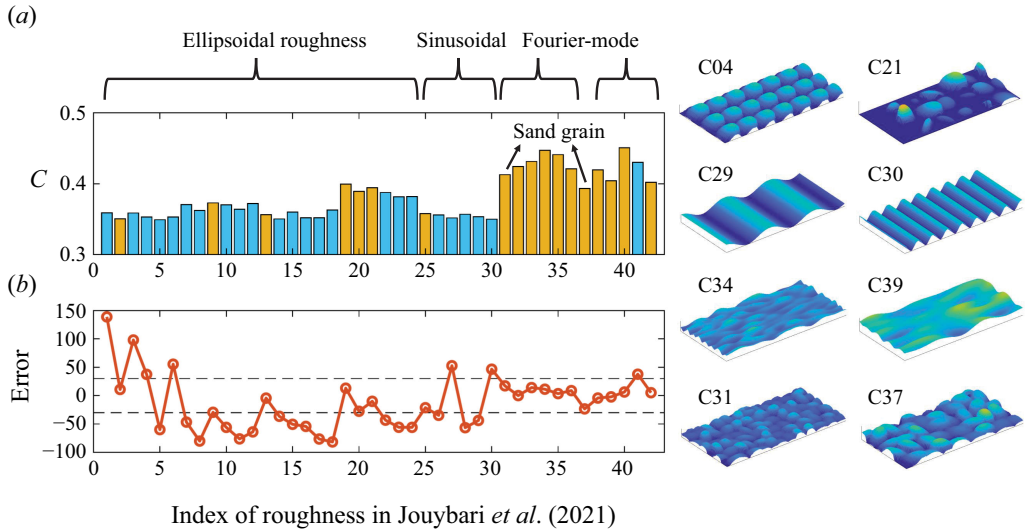
(*a*)

(*b*)

Index of roughness in Jouybari *et al.* (2021)

Figure 19. (*a*) Confidence score for each rough surface and (*b*) *apriori* relative error (in %) of predicted $\tilde{\tau}_w$ using the BFWM-rough and actual $\tilde{\tau}_w$ from DNS of Jouybari *et al.* (2021). The cases with the model error within $\pm 30\,\%$ are highlighted in yellow in (*a*). The dashed lines in (*b*) denote the error values at $\pm 30\,\%$. Roughness index 1–42 corresponds to Cases C01–C42 in Jouybari *et al.* (2021). The topology of some of the surfaces is visualized on the right-hand side: ellipsoidal roughness, C04 and C21; surfaces with sinusoidal waves, C29 and C30; roughness generated by Fourier modes, C34 and C39; sandgrain roughness, C31 and C37.

Apriori relative errors for $\tilde{\tau}_w$ are plotted in figure 19(*b*). The grid resolution considered is $\triangle/\delta = 1/20$. Consistent with the evaluation by the confidence score, BFWM-rough tends to perform best for cases C31 to C42. However, the predictions can exhibit errors of up to 30 %, which is significantly higher than the errors reported for previous test cases. The majority of cases from C01 to C30 show errors above 30 %, correlating with their low confidence scores. There are instances where the confidence score is low, yet the errors remain below 30 %. This might be coincidental, and caution should always be exercised for predictions accompanied by low values of the confidence score.

4.3. *High-pressure turbine blade with roughness*

We assess the performance of BFWM-rough in a HPT with wall roughness. Surface roughness on HPT blades can result from the manufacturing process or in-service degradation, significantly affecting the aerothermal performance of the blade (Nardini *et al.* 2023). The case selected for evaluation is the VKI LS-89 HPT blade with surface roughness, and the results are compared with previous numerical studies by Jelly *et al.* (2023) and Nardini *et al.* (2023). This case involves laminar–turbulent transition, strong pressure gradient effects, shock waves and vortical wakes. Therefore, it presents a challenging scenario to evaluate the predictive capabilities of WMLES in practical flow conditions involving complex geometries and flow physics.

Our WMLES follows the WRLES set-up from Jelly *et al.* (2023). The exit Reynolds number is $Re_{ex} = \rho_{ex} U_{ex} C_{ax}/\mu_\infty = 590\,000$, and the exit Mach number is $Ma_{ex} = U_{ex}/c_\infty = 0.9$, where ρ_{ex} and U_{ex} are the mean exit density and velocity, respectively. Here C_{ax} is the axial chord length, μ_∞ is the dynamic viscosity of the reference state and c_∞ is the acoustic velocity for the reference state. The inflow freestream turbulence is generated by a spanwise array of parallel bars. The turbulence intensity
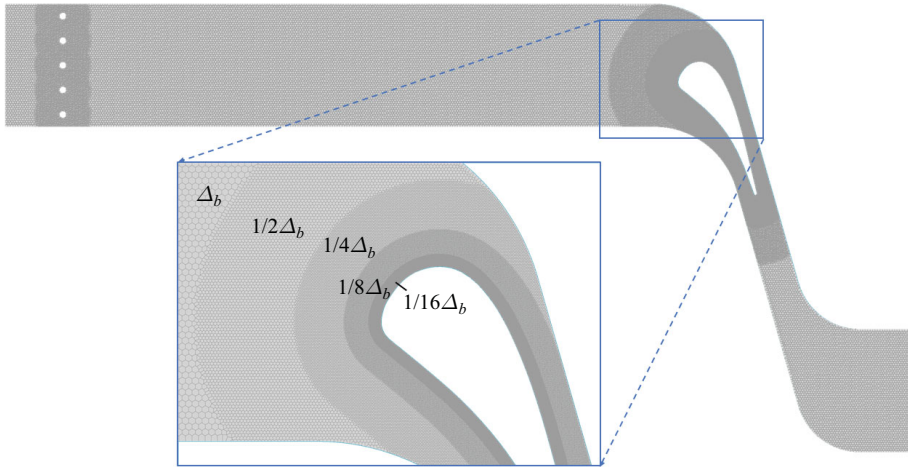
Figure 20. Visualization of Voronoi control volumes for WMLES of the HPT blade: the whole computational domain with a zoom-in view near the leading edge of the blade.

(TI) is defined as $TI = 1/3({u'_{rms}}^2 + {v'_{rms}}^2 + {w'_{rms}}^2)$. The value of TI is set to 8 % of the axial mean inlet velocity U_{in} by adjusting the distance of the bars upstream of the blade. The total pressure and temperature are specified at the inlet with the characteristic boundary conditions. At the exit, the non-reflective Navier–Stokes characteristic boundary conditions are enforced. The static pressure is prescribed at the exit based on the isentropic Mach number relationships. Periodic boundary conditions are prescribed at the upper, lower and spanwise boundaries. The spanwise extent of the domain is set to $0.4C_{ax}$, as suggested by Pichler *et al.* (2017), to ensure the correct development of the largest inflow turbulence structures.

The mesh generation is based on the Voronoi tessellation of a collection of points. The Voronoi control volumes are visualized in figure 20. The size of the background grid is $\triangle_b = 0.0311C_{ax}$. The grid size near the upstream bars is refined to $(1/2)\triangle_b$ with 20 layers, and the flow field near the bars is wall-resolved. As shown in the zoom-in view of figure 20, the grid size near the blade is refined by four levels, and for each level, the grid size is reduced by 50 % with 30 layers. As a result, the minimum grid size near the blade surface is $(1/16)\triangle_b = 0.00195C_{ax}$, and the number of control volumes per boundary layer thickness ranges from 0 (at the blade leading edge) to 30 (at the blade trailing edge). The total number of control volumes is 26×10^6.

Two different surface roughness profiles are considered for the blade. The cases, denoted as BS01 and BS02, feature three-dimensional, irregular Gaussian roughness. The key geometrical parameters of the roughness are summarized in table 11, which also includes the confidence scores for both rough surfaces. The confidence scores are $C = 0.73$ for BS01 and $C = 0.67$ for BS02, indicating that the roughness topologies can potentially be well-predicted by BFWM-rough. The WMLES with BFWM-rough is conducted for both blade surfaces, using the Vreman model (Vreman 2004) as the SGS model. The instantaneous axial velocity field from WMLES with BFWM-rough for roughness BS01 is visualized in figure 21, with the boundary layer transition observable in the zoom-in view. The results are compared with the WRLES results of Jelly *et al.* (2023) and DNS data of Nardini *et al.* (2023) for the same rough surfaces.

	k_{rms}/C_{ax}	S_k	K_u	ES	L_{cor}/C_{ax}	C
BS01	0.4×10^{-3}	0.0	3.0	0.16	3.6×10^{-3}	0.73
BS02	0.6×10^{-3}	0.0	3.0	0.18	9.5×10^{-3}	0.67

Table 11. The key geometrical roughness parameters of the blade surface roughness BS01 and BS02 from Jelly *et al.* (2023). The last column contains the confidence score of BFWM-rough for BS01 and BS02.
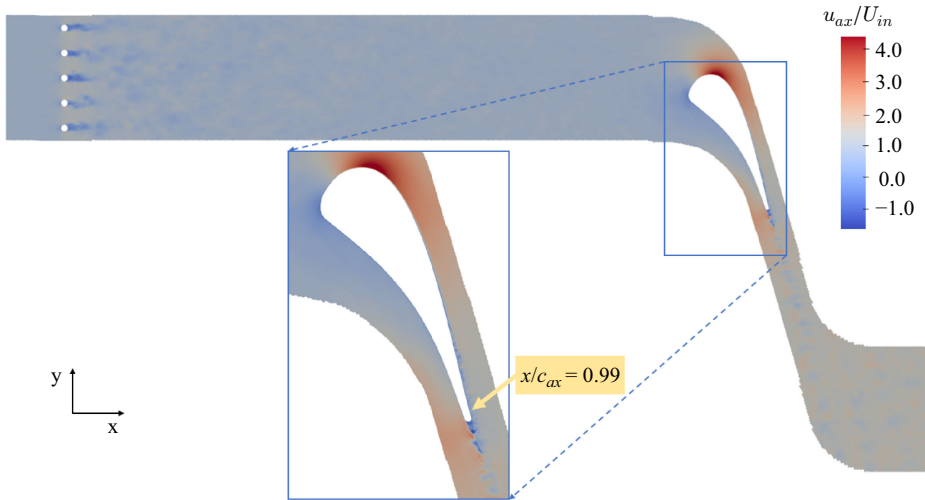
Figure 21. Visualization of the instantaneous axial velocity field normalized by the axial mean inlet velocity u_{ax}/U_{in} for WMLES of the HPT blade. The arrow indicates the location at 99 % of the axial chord length for probing the mean tangential velocity.

Figure 22 shows the time- and spanwise-averaged skin-friction coefficient C_f for WMLES and DNS. The C_f for a smooth blade obtained from DNS is included as a reference to demonstrate the effects of roughness. The WMLES using an equilibrium wall model with prescribed k_s (denoted by EQWM-k_s) is also shown in figure 22 for a comparison with BFWM-rough. The value of k_s is taken to be $k_s = \alpha k_{rms}$, with α set to a constant value. Hama (1954) suggested a scaling factor of $\alpha = 5$ for Gaussian roughness. However, this is not always accurate as the other roughness parameters such as ES and L_{cor}/R_a may also affect the value of k_s. For example, we found that α ranges from 3.5 to 6.5 for Gaussian roughness in our DNS database. For that reason, the EQWM-k_s is tested within the range $\alpha = 3 \sim 7$. The shaded area in figure 22 shows the variation of the results by EQWM-k_s for the range of values considered for k_s.

On the pressure side (PS) of the blade, the increase in C_f is due to the boundary layer transition, which occurs in the trailing-edge region. The increase in roughness on the PS of the blade is accurately captured by BFWM-rough and EQWM-k_s for the two rough surfaces. The results also indicate that the skin friction on the PS of the blade is only mildly sensitive to the value of k_s. The roughness effects are more significant on the suction side (SS) of the blade. The sharp increase in skin friction corresponds to the laminar-to-turbulent transition of the boundary layer, with larger roughness elements (BS02) triggering an earlier transition compared with the smaller roughness (BS01). Both WMLES with BFWM-rough and EQWM-k_s capture the key trends of the rough wall blade that are absent in the smooth wall case, such as the faster and larger increase in C_f and the
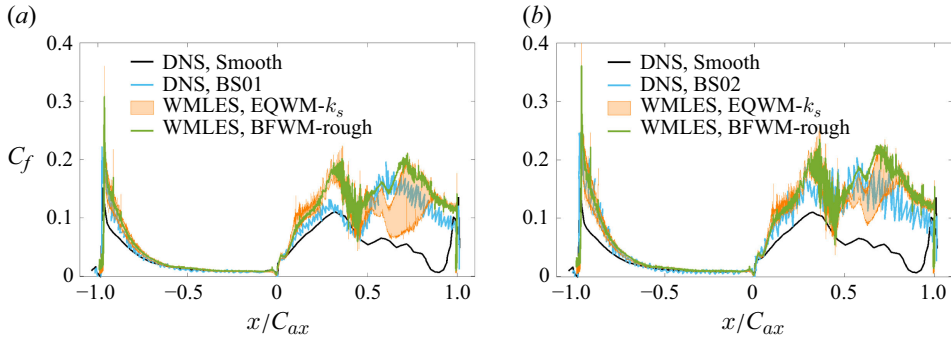
1007 A17-36

Figure 22. Time and spanwise-averaged distribution of skin friction coefficient C_f along the axial position of the blade normalized by axial chord length x/C_{ax}. (*a*) Roughness BS01 and (*b*) roughness BS02. The DNS results of smooth and rough surfaces are from Nardini *et al.* (2023). The blade roughness BS01 and BS02 correspond to cases k_2^s and k_3^s in Nardini *et al.* (2023). The shaded area denotes the WMLES with EQWM-k_s, where $k_s = \alpha k_{rms}$, and $\alpha = 3 \sim 7$. Note that $x/C_{ax} > 0$ and $x/C_{ax} < 0$ correspond to the SS and PS, respectively, with $x/C_{ax} = 0$ locate at the leading edge of the blade.
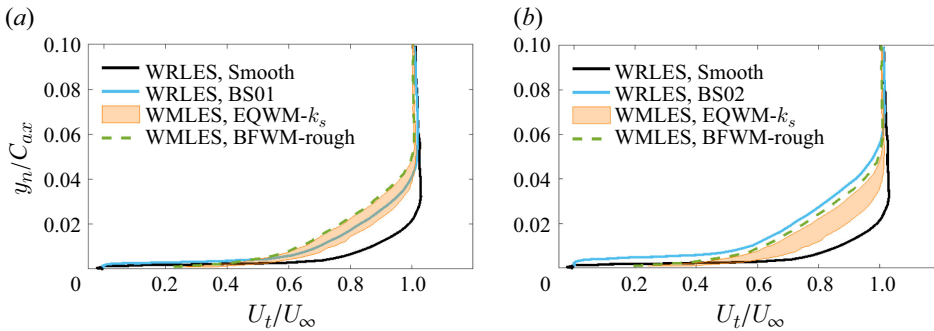
Figure 23. Mean tangential velocity at $x/C_{ax} = 0.99$ normalized by the local freestream velocity U_t/U_∞ along the blade-normal direction normalized by axial chord length y_n/C_{ax}. (*a*) Roughness BS01 and (*b*) roughness BS02. The WRLES results of smooth and rough surfaces are from Jelly *et al.* (2023). The shaded area denotes the WMLES with EQWM-k_s, where $k_s = \alpha k_{rms}$, and $\alpha = 3 \sim 7$.

second peak after $x/C_{ax} > 0.5$. It is also observed that the BFWM-rough outperforms the EQWM-k_s in the region $0.5 < x/C_{ax} < 0.8$ where the fully developed turbulence occurs in the two rough cases, while the prediction based on EQWM-k_s shows sensitivity to the variation of k_s in this region. In addition, for both BFWM-rough and EQWM-k_s, the value of C_f is overpredicted from $x/C_{ax} = 0.7$ to the trailing edge for the two rough cases. This might be related to the interaction of the TBL over the blade with the shock waves within the region $0.7 < x/C_{ax} < 0.9$. This effect was described by Nardini *et al.* (2023), who noted that a normal shock wave is induced by the roughness, with larger roughness magnitude increasing the intensity of normal shock patterns. The presence of shock waves may lead to the formation of shock-induced vortices, and these vortices can interact with the boundary layer, influencing its stability and altering the skin friction pattern. This intricate effect might not be correctly captured by WMLES and multiple factors beyond the wall model (e.g., SGS models and grid resolution) are probably at play.

Figure 23 shows the mean tangential velocity profiles close to the trailing edge at the axial location $x/C_{ax} = 0.99$. The results from WMLES are compared with WRLES from

Jelly *et al.* (2023). Although both the smooth- and rough-blade boundary layers have transitioned to turbulence in the trailing-edge region of the SS surface, the flow has yet to reach fully rough conditions (Jelly *et al.* 2023). Figure 23 illustrates that the velocity deficit is generally well captured by the WMLES with BFWM-rough for both rough surfaces considered. For BS01, BFWM-rough slightly overpredicts the velocity deficit, while for BS02, the predictions exhibit high accuracy. These results indicate that BFWM-rough performs well in modelling the integrated momentum deficit along the blade, which is ultimately responsible for the lower-speed mean velocity profiles at the trailing edge. The WMLES with EQWM-k_s also demonstrates good performance for case BS01; however, it clearly underpredicts the velocity deficit for case BS02.

5. Conclusions

We introduce a wall model for WMLES applicable to rough surfaces with Gaussian and Weibull distributions for both the transitionally and fully rough regimes. The model can be applied to arbitrary complex geometries where the surface roughnesses are assumed to be underresolved. The wall model is implemented using a multi-hidden-layer FNN, with the statistical geometric parameters of the surface roughness and near-wall flow quantities serving as input. The wall model, referred to as BFWM-rough, extends the BFWM introduced by Lozano-Durán & Bae (2023) to rough walls.

A roughness repository containing 100 random rough surfaces was created, encompassing a wide range of Gaussian and Weibull roughness features. Active learning was then employed to optimally construct the DNS training database by selecting rough surfaces from the repository with maximum uncertainty. This approach effectively improved the wall model prediction capabilities while minimizing the number of DNS cases required for training. A total of 19 Gaussian and 13 Weibull rough surfaces were selected through AL and used to conduct the DNS of turbulent channel flows. The DNS cases were performed in a minimal-span channel flow at six different Re_τ ranging from 180 to 1000. The final DNS database comprises approximately 200 cases.

The optimal set of non-dimensional inputs to the model was selected using information theory. The approach identifies a collection of non-dimensional inputs with minimum redundant information among them and maximum information about the output. Over 30 input candidates were ranked in order of importance to predict the wall shear stress. The most informative inputs were found to be based on flow state variables such as local Reynolds numbers at the first and second points off the wall ($u_1 y_1/\nu$ and $u_2 y_2/\nu$) and mean roughness features related to roughness height fluctuations and effective slope (k_{rms}/R_a and ES^2).

The wall model also incorporates a confidence score to detect potential low performance in the presence of untrained rough surfaces. The score is computed using GP model to evaluate the uncertainty of the roughness topology compared with the roughness repository used for training. The confidence score was calculated for different roughness types, such as Gaussian, Weibull, Bimodal, sandgrain and Fourier-mode roughness. The results demonstrated the ability of the confidence score to highlight potential model deficiencies for rough surfaces with strong anisotropic characteristics. In such cases, the confidence score can be used for *apriori* detection of low performance scenarios for BFWM-rough. This information can also be leveraged to inform future extensions of the model by incorporating roughness types with low confidence scores into the training dataset.

The BFWM-rough model has been tested *apriori* and *aposteriori* in more than 550 turbulent channel flows across various rough surfaces and flow conditions. These include

cases with untrained rough Reynolds numbers and grid resolutions over untrained surface roughness with Gaussian and Weibull distributions both at transitionally and fully rough flow conditions. The model was compared with five different models from previous studies and demonstrated superior predictive accuracy for both Gaussian and Weibull roughness. The model was also evaluated in over 40 rough surfaces whose geometrical features were not incorporated in the training process, such as bimodal rough surfaces and rough surfaces from Jouybari *et al.* (2021) that contain ellipsoidal, sinusoidal, sand-grain roughness and rough surfaces generated with low-order Fourier modes. The results show that the rough-wall model typically predicts the wall shear stress within a 1%–15 % accuracy range for roughness types resembling Gaussian and Weibull distributions. While the performance of BFWM-rough degrades for bimodal distributions, the accuracy remains comparable. Model errors increase to around 30%–50 % for roughness types exhibiting strong anisotropy, as these geometries differ considerably from the training set. However, the low performance in these cases can be anticipated by the low confidence score.

The BFWM-rough has also been evaluated in a complex flow involving a HPT blade with two different rough surfaces. This case includes laminar–turbulent transition, strong pressure gradient effects, shock waves and vortical wakes, making it a challenging scenario for assessing the predictive capabilities of WMLES in practical flow conditions. The results show that BFWM-rough outperforms the equilibrium wall model with the prescribed k_s, and captures key trends of the rough wall blade that are absent in the smooth wall case, such as the faster and larger increase in friction coefficient, with errors typically ranging between 1 % and 10 %. The BFWM-rough also accurately predicts the integrated momentum deficit along the blade, which is ultimately responsible for the slower mean velocity profiles at the trailing edge.

The current version of BFWM-rough is designed for equilibrium turbulence over isotropic rough surfaces. Future developments aim to extend its applicability to more general flow conditions, including mean pressure gradient effects and separation, following the approach from Lozano-Durán & Bae (2023). Additionally, the wall model is expected to be expanded to cover a broader range of roughness types, including those with anisotropic geometries.

Appendix A

We conduct a validation case to confirm that the minimal-span domain size is statistically representative of the roughness distribution. The DNS of a minimal-span open channel and a full-span open channel over the same rough surface are carried out at $Re_\tau = 1000$ for a comparison. The selected roughness is from GS02, which has the largest wavelength among the rough surfaces in the DNS database (i.e. the least sampling of the geometrical elements with this wavelength), making it the most constraining case. The minimal-span channel corresponds to a domain size $L_x = 3\delta$ and $L_z = \delta$. For the full-span channel, the rough surface GS02 is duplicated in the streamwise and spanwise directions, resulting in
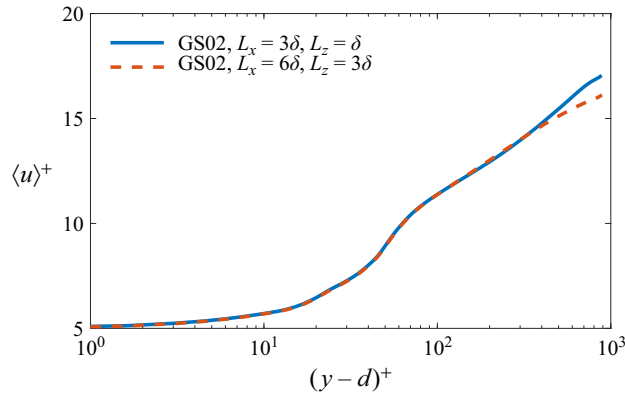
Figure 24. Comparison of streamwise mean velocity profiles for DNS of turbulent channel flow with a minimal-span domain and a larger domain over the same rough surface GS02.

a domain size $L_x = 6\delta$ and $L_z = 3\delta$. The results in figure 24 show that $\langle u \rangle^+$ is identical below $y/\delta = 0.3$ between the minimal-span and full-span channels. This result confirms that the chosen domain size is adequate for the rough surfaces considered in this work. Since only the mean velocity below $y/\delta = 0.2$ from DNS is taken to construct the input of the BFWM-rough, truncating larger roughness wavelengths does not impact the accuracy of the training data used for the wall model.

REFERENCES

ANDERSON, W. & MENEVEAU, C. 2011 Dynamic roughness model for large-eddy simulation of turbulent flow over multiscale, fractal-like rough surfaces. *J. Fluid Mech.* **679**, 288–314.

AUPOIX, B. 2016 Revisiting the discrete element method for predictions of flows over rough surfaces. *J. Fluids Engng* **138** (3), 031205.

AUPOIX, B. & SPALART, P.R. 2003 Extensions of the Spalart–Allmaras turbulence model to account for wall roughness. *Intl J. Heat Fluid Flow* **24** (4), 454–462.

BACCI, T., PICCHI, A., LENZI, T., FACCHINI, B. & INNOCENTI, L. 2021 Effect of surface roughness and inlet turbulence intensity on a turbine nozzle guide vane external heat transfer: experimental investigation on a literature test case. *J. Turbomach.* **143** (4), 041006.

BAE, H.J., LOZANO-DURÁN, A., BOSE, S.T. & MOIN, P. 2019 Dynamic slip wall model for large-eddy simulation. *J. Fluid Mech.* **859**, 400–432.

BARBOSA, L.A.P. & GERKE, H.H. 2022 Structural heterogeneity of soil clods: correlating Weibull parameters to fracture surface topography. *Geoderma* **428**, 116161.

BONS, J.P. 2002 St and C_f augmentation for real turbine roughness with elevated freestream turbulence. In *Turbo Expo: Power for Land, Sea, and Air*, vol. **36088**, pp. 349–363. The American Society of Mechanical Engineers (ASME).

BONS, J.P. 2010 A review of surface roughness effects in gas turbines. *J. Turbomach.* **132**, 021004-1–16.

BONS, J.P., TAYLOR, R.P., MCCLAIN, S.T. & RIVIR, R.B. 2001 The many faces of turbine surface roughness. *J. Turbomach.* **123** (4), 739–748.

BORNHOFT, B., JAIN, S.S., GOC, K., BOSE, S.T. & MOIN, P. 2022 Wall-modeled LES of laser-scanned rime, glaze, and horn ice shapes. *Annu. Res. Briefs Center Turbul. Res.* pp. 71–85.

BORNHOFT, B.J., JAIN, S.S., BOSE, S. & MOIN, P. 2024 Roughness modeling investigation in large-eddy simulations of a NACA23012 airfoil under rime ice conditions. *AIAA Paper* 2024-4001.

BRERETON, G.J. & YUAN, J. 2018 Wall-roughness eddy viscosity for Reynolds-averaged closures. *Intl J. Heat Fluid Flow* **73**, 74–81.

BRES, G.A., BOSE, S.T., EMORY, M., HAM, F.E., SCHMIDT, O.T., RIGAS, G. & COLONIUS, T. 2018 Large-eddy simulations of co-annular turbulent jet using a voronoi-based mesh generation framework. In 2018 AIAA/CEAS Aeroacoustics Conference, pp. 3302. The American Institute of Aeronautics and Astronautics (AIAA).

BUCKINGHAM, E. 1914 On physically similar systems; illustrations of the use of dimensional equations. *Phys. Rev* **4** (4), 345–376.

BUSSE, A. & SANDHAM, N.D. 2012 Parametric forcing approach to rough-wall turbulent channel flow. *J. Fluid Mech.* **712**, 169–202.

CEBECI, T. & CHANG, K.C. 1978 Calculation of incompressible rough-wall boundary-layer flows. *AIAA J.* **16** (7), 730–735.

CHAN, L., MACDONALD, M., CHUNG, D., HUTCHINS, N. & OOI, A. 2015 A systematic investigation of roughness height and wavelength in turbulent pipe flow in the transitionally rough regime. *J. Fluid Mech.* **771**, 743–777.

CHEDEVERGNE, F. & FOROOGHI, P. 2020 On the importance of the drag coefficient modelling in the double averaged Navier–Stokes equations for prediction of the roughness effects. *J. Turbul.* **21** (8), 463–482.

CHUNG, D., CHAN, L., MACDONALD, M., HUTCHINS, N. & OOI, A. 2015 A fast direct numerical simulation method for characterising hydraulic roughness. *J. Fluid Mech.* **773**, 418–431.

CHUNG, D., HUTCHINS, N., SCHULTZ, M.P. & FLACK, K.A. 2021 Predicting the drag of rough surfaces. *Annu. Rev. Fluid Mech.* **53** (1), 439–471.

CHUNG, D., MONTY, J.P. & OOI, A. 2014 An idealised assessment of Townsend's outer-layer similarity hypothesis for wall turbulence. *J. Fluid Mech.* **742**, R3.

COOK, R.F. & DELRIO, F.W. 2019 Material flaw populations and component strength distributions in the context of the Weibull function. *Exp. Mech.* **59** (3), 279–293.

DAS, J. & LINKE, B. 2017 Evaluation and systematic selection of significant multi-scale surface roughness parameters (SRPs) as process monitoring index. *J. Mater. Process. Technol.* **244**, 157–165.

DE MARCHIS, M., SACCONE, D., MILICI, B. & NAPOLI, E. 2020 Large eddy simulations of rough turbulent channel flows bounded by irregular roughness: advances toward a universal roughness correlation. *Flow Turbul. Combust.* **105** (2), 627–648.

DING, C. & PENG, H. 2005 Minimum redundancy feature selection from microarray gene expression data. *J. Bioinform. Comput. Biol.* **3** (02), 185–205.

DURBIN, P.A. 2023 Reflections on roughness modelling in turbulent flow. *J. Turbul.* **24** (1-2), 3–13.

DURBIN, P.A., MEDIC, G., SEO, J., EATON, J.K. & SONG, S. 2001 Rough wall modification of two-layer k-ε. *J. Fluids. Eng.* **123**, 16–21.

FEIEREISEN, W.J. & ACHARYA, M. 1986 Modeling of transition and surface roughness effects in boundary-layer flows. *AIAA J.* **24** (10), 1642–1649.

FLACK, K.A. & SCHULTZ, M.P. 2010 Review of hydraulic roughness scales in the fully rough regime. *J. Fluids. Eng.* **132**, 041203.

FLACK, K.A., SCHULTZ, M.P. & BARROS, J.M. 2020 Skin friction measurements of systematically-varied roughness: probing the role of roughness amplitude and skewness. *Flow Turbul. Combust.* **104** (2-3), 317–329.

FLACK, K.A., SCHULTZ, M.P. & SHAPIRO, T.A. 2005 Experimental support for Townsend's Reynolds number similarity hypothesis on rough walls. *Phys. Fluids* **17** (3), 035102.

FLORES, O. & JIMENEZ, J. 2006 Effect of wall-boundary disturbances on turbulent channel flows. *J. Fluid Mech.* **566**, 357–376.

FOROOGHI, P., STROH, A., MAGAGNATO, F., JAKIRLIĆ, S. & FROHNAPFEL, B. 2017 Toward a universal roughness correlation. *J. Fluids Engng* **139** (12), 121201.

FU, L., KARP, M., BOSE, S.T., MOIN, P. & URZAY, J. 2021 Shock-induced heating and transition to turbulence in a hypersonic boundary layer. *J. Fluid Mech.* **909**, A8-1-49.

GERMANO, M., PIOMELLI, U., MOIN, P. & CABOT, W. 1991 A dynamic subgrid-scale eddy viscosity model. *Phys. Fluids* **3** (7), 1760–1765.

GRIMMOND, C.S.B. & OKE, T.R. 1999 Aerodynamic properties of urban areas derived from analysis of surface form. *J. Appl. Meteorol. Climatol.* **38** (9), 1262–1292.

HAMA, F.R. 1954 Boundary layer characteristics for smooth and rough surfaces. *Trans. Soc. Nav. Archit. Mar. Engrs* **62**, 333–358.

HANNA, S.R. & BRITTER, R.E. 2010 *Wind Flow and Vapor Cloud Dispersion at Industrial and Urban Sites*. John Wiley & Sons.

JACKSON, P.S. 1981 On the displacement height in the logarithmic velocity profile. *J. Fluid Mech.* **111** (1), 15–25.

JELLY, T.O., NARDINI, M., ROSENZWEIG, M., LEGGETT, J., MARUSIC, I. & SANDBERG, R.D. 2023 High-fidelity computational study of roughness effects on high pressure turbine performance and heat transfer. *Intl J. Heat Fluid Flow* **101**, 109134.

JIMÉNEZ, J. 2004 Turbulent flows over rough walls. *Annu. Rev. Fluid Mech.* **36** (1), 173–196.

JIMÉNEZ, J. & MOIN, P. 1991 The minimal flow unit in near-wall turbulence. *J. Fluid Mech.* **225**, 213–240.

JOUYBARI, M.A., YUAN, J., BRERETON, G.J. & MURILLO, M.S. 2021 Data-driven prediction of the equivalent sand-grain height in rough-wall turbulent flows. *J. Fluid Mech.* **912**, A8.

KEATING, A., PIOMELLI, U., BREMHORST, K. & NEŠIĆ, S. 2004 Large-eddy simulation of heat transfer downstream of a backward-facing step. *J. Turbul.* **5** (1), 020.

KIRSCHNER, C.M. & BRENNAN, A.B. 2012 Bio-inspired antifouling strategies. *Annu. Rev. Mater. Res.* **42** (1), 211–229.

KNOPP, T., EISFELD, B. & CALVO, J.B. 2009 A new extension for k–ω turbulence models to account for wall roughness. *Intl J. Heat Fluid Flow* **30** (1), 54–65.

KUWATA, Y. & KAWAGUCHI, Y. 2019 Direct numerical simulation of turbulence over systematically varied irregular rough surfaces. *J. Fluid Mech.* **862**, 781–815.

LEONARDI, S. & CASTRO, I.P. 2010 Channel flow over large cube roughness: a direct numerical simulation study. *J. Fluid Mech.* **651**, 519–539.

LI, S., YANG, X. & LV, Y. 2022 Predictive capability of the logarithmic law for roughness-modeled large-eddy simulation of turbulent channel flows with rough walls. *Phys. Fluids* **34** (8), 085112.

LI, Z. & YANG, X. 2021 Large-eddy simulation on the similarity between wakes of wind turbines with different yaw angles. *J. Fluid Mech.* **921**, A11.

LILLY, D.K. 1992 A proposed modification of the Germano subgrid-scale closure method. *Phys. Fluids* **4** (3), 633–635.

LOZANO-DURÁN, A. & BAE, H.J. 2019*a* Characteristic scales of townsend's wall-attached eddies. *J. Fluid Mech.* **868**, 698–725.

LOZANO-DURÁN, A. & BAE, H.J. 2019*b* Error scaling of large-eddy simulation in the outer region of wall-bounded turbulence. *J. Comput. Phys.* **392**, 532–555.

LOZANO-DURÁN, A. & BAE, H.J. 2023 Machine learning building-block-flow wall model for large-eddy simulation. *J. Fluid Mech.* **963**, A35.

LOZANO-DURÁN, A., BOSE, S.T. & MOIN, P. 2022 Performance of wall-modeled LES with boundary-layer-conforming grids for external aerodynamics. *AIAA J.* **60** (2), 747–766.

MA, H., LI, Y., YANG, X. & YE, L. 2023 Data-driven prediction of the equivalent sand-grain roughness. *Sci. Rep.* **13** (1), 19108.

MA, R., ALAMÉ, K. & MAHESH, K. 2021 Direct numerical simulation of turbulent channel flow over random rough surfaces. *J. Fluid Mech.* **908**, A40.

MA, R., YUAN, Y., ARRANZ, G., LING, Y., AGRAWAL, R., WHITMORE, M., ELNAHHAS, A., PEDERSON, C. & LOZANO-DURÁN, A. 2024 Building-block-flow model for LES of high-speed flows. In Proceedings of the CTR Summer Program, pp. A40-1-33. Stanford University.

MACDONALD, M., CHUNG, D., HUTCHINS, N., CHAN, L., OOI, A. & GARCÍA-MAYORAL, R. 2017 The minimal-span channel for rough-wall turbulent flows. *J. Fluid Mech.* **816**, 5–42.

MACDONALD, R.W., GRIFFITHS, R.F. & HALL, D.J. 1998 An improved method for the estimation of surface roughness of obstacle arrays. *Atmos. Environ.* **32** (11), 1857–1864.

MIZUNO, Y. & JIMÉNEZ, J. 2013 Wall turbulence without walls. *J. Fluid Mech.* **723**, 429–455.

MUNK, T., KANE, D. & YEBRA, D.M. 2009 The effects of corrosion and fouling on the performance of ocean-going vessels: a naval architectural perspective. In *Advances in Marine Antifouling Coatings and Technologies*, pp. 148–176. Elsevier.

NARDINI, M., JELLY, T.O., KOZUL, M., SANDBERG, R.D., VITT, P. & SLUYTER, G. 2023 Direct numerical simulation of transitional and turbulent flows over multi-scale surface roughness—Part II: The effect of roughness on the performance of a high-pressure turbine blade. *J. Turbomach.* **146** (3), 031009.

NIKURADSE, J. 1933 Laws of flow in rough pipes. Technical Memorandum 1292. NACA.

ÖRLÜ, R. & SCHLATTER, P. 2011 On the fluctuating wall-shear stress in zero pressure-gradient turbulent boundary layer flows. *Phys. Fluids* **23** (2), 021704.

PANDA, S., CHOWDHURY, S.K.R. & SARANGI, M. 2015 Effects of non-gaussian counter-surface roughness parameters on wear of engineering polymers. *Wear* **332**, 827–835.

PENG, H., LONG, F. & DING, C. 2005 Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **27** (8), 1226–1238.

PENG, W. & BHUSHAN, B. 2000 Modelling of surfaces with a bimodal roughness distribution. *Proc. Inst. Mech. Engrs J: J. Engng Tribol.* **214** (5), 459–470.

PÉREZ-RÀFOLS, F. & ALMQVIST, A. 2019 Generating randomly rough surfaces with given height probability distribution and power spectrum. *Tribol. Intl* **131**, 591–604.

PICHLER, R., SANDBERG, R.D., LASKOWSKI, G. & MICHELASSI, V. 2017 High-fidelity simulations of a linear HPT vane cascade subject to varying inlet turbulence. In *Turbo Expo: Power for Land, Sea, and Air*, vol. **50787**, pp. V02AT40A001. American Society of Mechanical Engineers.

RASMUSSEN, C.E. & WILLIAMS, C.K.I. 2006 *Gaussian Processes for Machine Learning*. Vol. **1**. Springer.

RAUPACH, M.R., ANTONIA, R.A. & RAJAGOPALAN, S. 1991 Rough-wall turbulent boundary layers. *Appl. Mech. Rev.* **44**, 1–25.

SCOTTI, A. 2006 Direct numerical simulation of turbulent channel flows with boundary roughened with virtual sandpaper. *Phys. Fluids* **18** (3), 031701.

SETTLES, B. 2009 Active learning literature survey. *Tech Rep*.

SHAW, R.H. & SCHUMANN, U. 1992 Large-eddy simulation of turbulent flow above and within a forest. *Boundary-Layer Meteorol.* **61** (1), 47–64.

STRIPF, M., SCHULZ, A., BAUER, H. & WITTIG, S. 2009 Extended models for transitional rough wall boundary layers with heat transfer—part i: model formulations. *J. Turbomach.* **131**, 031016.

SUGA, K., CRAFT, T.J. & IACOVIDES, H. 2006 An analytical wall-function for turbulent flows and heat transfer over rough walls. *Intl J. Heat Fluid Flow* **27** (5), 852–866.

THAKKAR, M., BUSSE, A. & SANDHAM, N. 2017 Surface correlations of hydrodynamic drag for transitionally rough engineering surfaces. *J. Turbul.* **18** (2), 138–169.

THEURER, W. 1993 Dispersion of ground-level emissions in complex built-up areas. *J. Wind Engng Ind. Aerodyn.* **44**, 2721–2732.

TOWNSEND, A.A.R. 1976 *The Structure of Turbulent Shear Flow*. Cambridge University Press.

VREMAN, A.W. 2004 An eddy-viscosity subgrid-scale model for turbulent shear flow: algebraic theory and applications. *Phys. Fluids* **16** (10), 3670–3681.

WHITEHOUSE, D.J. 2023 *Handbook of Surface Metrology*. Routledge.

WILCOX, D.C. 1998 *Turbulence Modeling for CFD*, vol. **2**. DCW Industries La Canada.

WILLIAMSON, J.B.P., PULLEN, J., HUNT, R.T. & LEONARD, D. 1969 The shape of solid surfaces. In *Surface Mechanics*, pp. 24–35. ASME.

YANG, J., STROH, A., LEE, S., BAGHERI, S., FROHNAPFEL, B. & FOROOGHI, P. 2023 Prediction of equivalent sand-grain size and identification of drag-relevant scales of roughness – a data-driven approach. *J. Fluid Mech.* **975**, A34.

YANG, X., KHOSRONEJAD, A. & SOTIROPOULOS, F. 2017 Large-eddy simulation of a hydrokinetic turbine mounted on an erodible bed. *Renew. Energy* **113**, 1419–1433.

YANG, X.I.A., SADIQUE, J., MITTAL, R. & MENEVEAU, C. 2016 Exponential roughness layer and analytical model for turbulent boundary layer flow over rectangular-prism roughness elements. *J. Fluid Mech.* **789**, 127–165.

YU, C. & LIU, B. 2002 A backpropagation algorithm with adaptive learning rate and momentum coefficient. In *Proceedings of the 2002 International Joint Conference On Neural Networks*, 02CH37290), Vol. 2, pp. 1218–1223. IEEE.

YUAN, J. & JOUYBARI, M.A. 2018 Topographical effects of roughness on turbulence statistics in roughness sublayer. *Phys. Rev. Fluids* **3** (11), 114603.

YUAN, J. & PIOMELLI, U. 2014*a* Estimation and prediction of the roughness function on realistic surfaces. *J. Turbul.* **15** (6), 350–365.

YUAN, J. & PIOMELLI, U. 2014*b* Numerical simulations of sink-flow boundary layers over rough surfaces. *Phys. Fluids* **26** (1), 015113.

YUAN, J. & PIOMELLI, U. 2014*c* Roughness effects on the Reynolds stress budgets in near-wall turbulence. *J. Fluid Mech.* **760**, R1.

ZHOU, B.Y., HANSON, L.P., PULLIN, S.F., ZANG, B., HAUTH, J. & HUAN, X. 2022 A data-driven approach for enhancement of propeller performance prediction, 28th AIAA/CEAS Aeroacoustics 2022 Conference, pp. 3106. The American Institute of Aeronautics and Astronautics (AIAA).