Data-Driven Approach for Predicting and Explaining Drag of Irregular Roughness in Turbulent Flow

Zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften (Dr.-Ing.)

von der KIT-Fakultät für Maschinenbau des
Karlsruher Instituts für Technologie (KIT)
angenommene

Dissertation

von

M.Sc. Jiasheng Yang

Tag der mündlichen Prüfung: 18.10.2024

Hauptreferentin:	Prof. Dr.-Ing. Bettina Frohnapfel
Korreferent:	Prof. Dr. Neil Sandham
	Prof. Dr. Pourya Forooghi

Acknowledgements

I would like to express my deepest gratitude to my doctoral advisors, Prof. Dr. Pourya Forooghi, Prof. Dr.-Ing. Bettina Frohnapfel and Dr.-Ing. Alexander Stroh, for their unwavering trust and for granting me the opportunity to conduct research under their supervision. The free and creative research environment they created has been a wellspring of inspiration, allowing me to develop new ideas and persevere through any research challenges I encountered during my four-year PhD journey. None of these achievements would have been possible without their immense, unconditional support and invaluable input from every perspective. A special thank you extends to Prof. Dr. Willy Dörfler at the Institute of Applied and Numerical Mathematics at KIT for his interest in my research work and the related support. I gratefully acknowledge the financial support of the Friedrich und Elisabeth Boysen-Stiftung (Project number BOY-151). This project would not have been possible without their generosity and commitment to advancing research. Furthermore, I would like to thank Prof. Dr. Neil Sandham for the review of this thesis. A research stay at KTH Stockholm, Sweden was enabled through a KHYS travel grant. At KTH Stockholm I would like to thank Prof. Shervin Bagheri, Prof. Dr. Ramis Örlü, Dr. Patricia Sujar Garrido, and the FSG group members – Dr. Willy Bonneuil, Dr. Si Suo, Dr. Zhaoyu Shi, Dr. Romain Vallon, Dr. Wenhai Lei, Dr. Seyed Morteza Habibi Khorasani, Sofia Saoncella, Cornelius Wittig, and Michele Pellegrino at KTH for their warm hospitality in Stockholm and for their creative ideas and inspirations that greatly enriched my research. Special thanks go to Prof. Dr. Sangseung Lee and Heesoo Shih for their significant contributions to the machine-learning aspects of my work. Moreover, I would like to thank all my external collaborators on various publications throughout my PhD: Apl. Prof. Dr.-Ing. habil. Suad Jakirlic, Prof. Dr. Daniel Chung, Dr. François Chedevergne, Dr. Stephan Bansmer, Dr. Juan Velandia and Florian Theobald. Their support has been invaluable. Lastly, but most importantly, I want to thank my colleagues: Dr. Davide Gatti, Dr. Jochen Kriegseis, Dr. Franco Magagnato, Dr. Arjun John Kaithakkal, Dr. Leister Robin, Dr. Kay Schäfer, Dr. Lars von Deyn, Dr. Naseem Ali, Dr. Erik Arne Hansen, Dr. Nima Samkhaniani, Dr. Jacopo Serpieri, Andrea Andreolli, Sebastian Blessing, Stefano Cipelli, Simon Benedikt Dalpke, Maximilian Dreisbach, Georg Fahland, Annika Frede, Tobias Karl, David Müller, Jonathan Neuhauser, Saskia Pasch, Dominic Pöppe, David Rautenberg, Christian Sax, Carola Sophie Constanze Schmidt, Wiebke Schrader, Francesco Secchi and Gabriel Thäter for creating a fun, relaxed, and optimistic work environment, which has been crucial for me.

I extend my heartfelt thanks to my parents for their unwavering love and support. I am profoundly grateful to everyone I have met throughout my journey in Karlsruhe and Stockholm, each of whom has contributed to my growth in meaningful ways. Most importantly, I wish to express my deepest admiration and gratitude to my beloved wife, Yujia Fu. From our early days as undergraduates in Dalian, Yujia has been my steadfast companion, offering unwavering support and care. Her loving spirit has truly illuminated my life. As I document our journey, I hope our love continues to flourish eternally. To all who read this acknowledgement, may you bear witness to the enduring bond we share.

Abstract

Accurate prediction of skin friction of arbitrary rough surface in turbulent flow has been a longstanding focus in the field of roughness research. Motivated by the recent advancements in machine learning (ML), this study introduces a cost-effective technical framework for developing a universally applicable ML model for this task. To achieve this, a roughness database is created by compiling artificially generated roughness topographies and their corresponding equivalent sand-grain size k_s, calculated through direct numerical simulations (DNS) in minimal channels. A pseudo-random algorithm is utilized to generate artificial roughness topographies. This algorithm allows manipulation of the power spectrum (PS) and probability density function (p.d.f.) of surface height while maintaining stochastic nature of the generated roughness. Initially, 12 types of roughness with systematically varied PS and p.d.f. are investigated at both transitionally and fully rough regimes with the roughness height k^+ ranging in the range of 25-100. It is demonstrated that the roughness function ΔU^+ and temperature roughness function $\Delta\Theta^+$ can be recovered by the minimal-channel DNS compared to its full-span DNS counterpart with an accuracy within $\pm 5\%$. Same level of agreement is achieved by a series of individually generated roughness instances that share identical PS and p.d.f. but with deterministically different surface topographies. With the same set of roughness data, the impact of different roughness properties on the resulting skin friction as well as the distribution of time-averaged surface force exerted by roughness structures onto the fluid are investigated. Subsequently, the PS and p.d.f. extracted from realistic rough surface scans are employed to generate artificial roughness surrogates. The turbulent flows over the artificial roughness surrogates are investigated through DNS at $\mathrm{Re}_\tau = 500 - 2000$. Excellent agreement in terms of the skin-friction coefficient C_f and the Stanton number St, mean velocity and temperature profiles, Reynolds stresses, and the equivalent sand-grain size k_s values are achieved compared to their corresponding original realistic surface scans with exception of a strongly anisotropic sample. Moreover, the experimental measurement of C_f and the mean velocity profile over P60 sandpaper at $\mathrm{Re}_b = 18000$ are successfully reproduced via DNS on the numerical artificial surrogate. This surrogate is generated based on the PS and p.d.f. extracted from perthometer measurement and 3-D photogrammetry, respectively. Finally, the ML model is constructed in the fashion of ensemble neural network (ENN) consisting of 50 individual multi-layer perceptron (MLP) models. Wherein, each MLP model incorporates roughness PS and p.d.f. as inputs. The training database is obtained through the proposed data-generation process. The efficiency of the database construction is further enhanced through active learning (AL), which offers the model liberty to navigate the expansion of training database by prioritizing the simulation of the most informative roughness topographies, thereby maximizing the generalizability of the prediction with a certain amount of data. Following the AL framework, while 4200 roughness topographies are generated through the roughness generation algorithm to form the repository of roughness topographies, only 85 most informative roughness are selected to compile the training database. Their k_s values are calculated in the fully rough regime at friction Reynolds number $\mathrm{Re}_\tau = 800$. The final model performance is examined using three different testing data sets with different types of roughness, including 21 surfaces from the literature. The model yields an overall mean error of 5% to 10% on different testing data sets. Based on the trained model, a data interpretation technique, namely layer-wise relevance propagation (LRP), is applied to explain the model prediction. High-pass filtering is applied to the roughness PS to exclude the wavenumbers identified as drag-irrelevant. The DNS investigations on the filtered rough surfaces demonstrate the successful preservation of the skin-friction coefficient of the original roughness. In the end, two additional ENN models are developed, each employing different sets of frequently studied single-valued roughness parameters as inputs, yet being furnished with the same training database. The significant decline in performance of parameter-based models on external testing data suggests that the characterization method using PS and p.d.f. offers a universal and precise representation of roughness properties.

Kurzfassung

Eine genaue Vorhersage der Hautreibung beliebiger Oberflächen in turbulenter Strömung ist seit langem ein Schwerpunkt im Forschungsbereich der Rauigkeit. Motiviert durch die rasante Entwicklung des maschinellen Lernens, stellt diese Studie einen kosteneffizienten Rahmen für die Entwicklung eines universell anwendbaren ML-Modells vor. Eine Datenbank von Rauheiten wird durch die Sammlung von generierten Rauhigkeitstopographien und deren entsprechenden äquivalenten Sandkorngrößen k_s, berechnet durch direkte numerische Simulationen (DNS), erzeugt. Ein pseudozufälliger Algorithmus erzeugt künstliche Rauheitstopographien, manipuliert das Leistungsspektrum (PS) und die Wahrscheinlichkeitsdichtefunktion (p.d.f.) der Oberflächenhöhe, während der stochastische Charakter beibehalten wird. Zunächst werden 12 Rauheiten unterschiedlicher Typen mit systematisch variierten PS und p.d.f. sowohl im Übergangs- als auch im Vollrauhigkeitsregime untersucht, wobei die Rauheitshöhe k^+ im Bereich von 25-100 variiert. Es wird gezeigt, dass die Rauheitsfunktion ΔU^+ und die Temperaturrauheitsfunktion $\Delta\Theta^+$ durch die Minimal-Channel-DNS im Vergleich zur Full-Span-DNS mit einer Fehlerrate von $\pm 5\%$ wiederhergestellt werden können. Das gleiche Maß an Übereinstimmung wird durch eine Reihe individuell erzeugter Rauheiten mit identischem PS und p.d.f. -– jedoch unterschiedlichen Oberflächenstrukturen -– erreicht. Mit demselbem Satz von Rauheitsdaten werden die Auswirkungen verschiedener Rauheitseigenschaften auf die resultierende Hautreibung sowie die Verteilung der zeitlich gemittelten Oberflächenkraft, die von den Rauheitsstrukturen auf die Strömung ausgeübt wird, untersucht. Anschließend werden die PS und p.d.f., die aus realistischen Oberflächenscans extrahiert wurden, zur Erzeugung künstlicher Rauheitssurrogate verwendet. Die turbulente Strömung über den künstlichen Rauheitssurrogaten wird mittels DNS im Bereich von Re_τ von 500 bis 2000 untersucht. Es wird bemerkenswerte Übereinstimmungen zwischen dem Hautreibungskoeffizienten C_f, der Stanton-Zahl St, der mittleren Geschwindigkeits- und Temperaturprofile, den Reynoldsspannungen und der äquivalenten Sandkorngröße k_s im Vergleich zu den entsprechenden Originalscans der realistischen Oberflächen beobachtet, abgesehen von einer stark anisotropen Rauheit. Darüber hinaus werden die experimentelle Messung von C_f und das Geschwindigkeitsprofil über P60-Schleifpapier bei $\mathrm{Re}_b = 18000$ erfolgreich mittels DNS auf dem numerischen künstlichen Surrogat reproduziert. Dieses Surrogat wird basierend auf PS und p.d.f. aus der Perthometermessung bzw. der 3D-Photogrammetrie erzeugt. Schließlich wird das ML-Modell in Form eines neuronalen Netzes (ENN) aufgebaut, das aus 50 individuellen Multi-Layer Perceptron Modellen (MLP) besteht. Jedes MLP-Modell enthält das PS und die p.d.f. als Eingabedaten. Die Trainingsdatenbank wird durch den oben beschriebenen Datenerzeugungsprozess erstellt. Die Effizienz des Datenbankaufbaus wird durch aktives Lernen (AL) weiter verbessert, das dem Modell die Freiheit gibt, die Erweiterung der Trainingsdatenbank zu steuern, indem es der Simulation der informativsten Rauheitstopographien Vorrang einräumt und so die Verallgemeinerbarkeit der Vorhersage mit einer begrenzten Datenmenge maximiert. Nach dem AL-Rahmen werden 4200 Rauheitstopografien durch den Rauheitsgenerierungsalgorithmus erzeugt, um das Repository an Rauheitstopografien zu bilden, aber nur die 85 informativsten Rauheiten werden für den Aufbau der Trainingsdatenbank ausgewählt. Ihre k_s-Werte werden im Vollrauhigkeitsregime bei der Reibungs-Reynoldszahl $\mathrm{Re}_\tau = 800$ berechnet. Das Modell liefert einen mittleren Gesamtfehlerraten von 5% bis 10% auf verschiedenen Testdatensätzen. Auf der Grundlage des trainierten Modells wird eine Dateninterpretationstechnik, nämlich die schichtweise Relevanzausbreitung (LRP), angewandt, um die Modellvorhersage zu erklären. Die Hochpassfilterung wird auf die Rauheit PS angewendet, um die als widerstandsunrelevant identifizierten Wellenzahlen auszuschließen. Die DNS-Untersuchungen an den gefilterten rauen Oberflächen zeigen, dass der Hautreibungskoeffizient der ursprünglichen Rauheit erfolgreich beibehalten wird. Schließlich werden zwei weitere ENN-Modelle entwickelt, die jeweils verschiedene Sets häufig untersuchter Rauheitsparameter als Eingaben verwenden, aber mit derselben Trainingsdatenbank ausgestattet sind. Die deutliche Leistungsabfall der parameterbasierten Modelle bei externen Testdaten deutet darauf hin, dass die Charakterisierungsmethode durch PS und p.d.f. eine universelle und genaue Darstellung der Rauheitseigenschaften bietet.

Contents

Nomenclature

Latin letters

upper case

Symbol	Description
A	log-law intercept
A_f	total projected frontal area of roughness
A_h	temperature log-law intercept
A_t	wall-projected area of roughness patch
B	log-law intercept over roughness
$B^{(\mathcal{J})}$	neural bias in layer $\mathcal{J}$
B_s	constant in asymptotic fully-rough roughness function
C_0	constant in power-law PS
C_f	skin-friction coefficient
D	pipe diameter
E_k	function of power spectrum for roughness height map
E_{kf}	function of cross power spectrum for roughness height map and streamwise surface force map
$\mathbf{F}$	set of features in database
$\mathcal{F}_i$	i-th model output
H	channel half height
$\mathbf{H}$	set of hyperparameters
H_{eff}	effective channel half height
H_f	Hurst exponent
$\mathcal{I}$	input layer
K_W	shape parameter for the Weibull distribution
Ku	kurtosis
$\mathcal{L}$	labeled training roughness data set
$\mathbf{L}$	likelihood
$\mathscr{L}$	loss function
L_f	integral length scale of streamwise surface force
L_k	integral length scale of roughness height distribution
L_x	streamwise length of computational domain
L_y	wall-normal height of computational domain
L_z	spanwise width of computational domain
L^{Corr}	correlation length

M	moment of the drag acting on the roughness
N_x	number of simulation grid points in streamwise direction
N_y	number of simulation grid points in wall-normal direction
N_z	number of simulation grid points in spanwise direction
$\mathcal{O}$	output layer
P	probability distribution function
Por	porosity of roughness
Pr	Prandtl number
Pr_t	turbulent Prandtl number
Q	heat source
Q_Θ	heat source term corresponding to the mixed-type thermal boundary condition
R_a	mean height deviation
R_i	contribution score of i-neuron
R_{ff}	streamwise auto-correlation function of the streamwise surface force
R_{kf}	streamwise correlation function of the streamwise surface force and roughness height distribution
Re	Reynolds number
Re_b	bulk Reynolds number
$\mathrm{Re}_{\mathrm{cl}}$	center-line Reynolds number
Re_τ	friction Reynolds number
$\mathcal{S}_i$	neural network global sensitivity score of i-input neuron
Sk	skewness
St	Stanton number
$\mathcal{T}_{\mathrm{ext}}$	external testing data set
$\mathcal{T}_{\mathrm{inter}}$	internal testing data set
U	double-averaged streamwise velocity
$\mathcal{U}$	unlabeled roughness topography repository
U_b	bulk velocity
V	double-averaged wall-normal velocity
W	double-averaged spanmwise velocity
$\mathbf{W}$	neural network weight matrix
$W_{ij}^{(\mathcal{J})}$	neural weight for j-neuron in layer $\mathcal{J}$ and i-neuron in layer $\mathcal{J}-1$
$\mathbf{X}$	data set for conducting hyperparameter Bayesian optimization

lower case

Symbol	Description
$a_i^{(\mathcal{J})}$	neural activation of i-neuron at layer $\mathcal{J}$
c_p	specific heat
d	zero-plane displacement
$\mathbf{f}_{\mathrm{in}}$	neural network input vector
f_i	i-component input features in the input vector $\mathbf{f}_{\mathrm{in}}$
f_{B}	bimodal distribution
f_{BO}	objective function for Bayesian optimization

f_{G}	Gaussian distribution
$f_{\mathrm{IBM},i}$	volume force term for immersed boundary method in i-direction
f_i	wall-normal integrated time-averaged local IBM forcing in i-direction
f_{W}	Weibull distribution
f_θ	heat source terms for immersed boundary method
g_i	volume body force term in i-direction
h_{ref}	reference plane height due to roughness
k	roughness height map
k_{99}	99% confidence interval of roughness height p.d.f.
k_{rms}	standard deviation or roughness height
k_h	roughness height map exhibiting prescribed p.d.f.
k_{md}	melt-down height
k_{ps}	roughness height map exhibiting prescribed PS
k_r	dimensionless equivalent sand-grain roughness
k_s	equivalent sand grain size
k_t	sub-sample-average peak-to-trough height
k_z	peak-to-trough height
n_h	frequency corresponding to the histogram of the target p.d.f.
n_{ps}	frequency corresponding to the histogram of the p.d.f. of k_{ps}
p	pressure
$p_i^{(\mathcal{J})}$	neural pre-activation of i-neuron at layer $\mathcal{J}$
q	wavenumber
q_0	minimum in-plane wavenumber
q_1	maximum in-plane wavenumber
q_r	roll-off wavenumber
q_w	wall heat flux
r	sand grain size
t	time
u	streamwise velocity component
u_i	instantaneous velocity in i-direction
u_τ	friction velocity
v	wall-normal velocity component
w	spanwise velocity component
x	streamwise coordinate
x_i	basis coordinates
y	wall-normal coordinate
y_c	critical height of minimal channel
y_t	target value of model prediction
$\mathbf{y}_t$	set of target values in database
z	spanwise coordinate

Greek letters

upper case

Symbol	Description
ΔU^+	roughness function
$\Delta\Theta^+$	temperature roughness function
Δ_x	streamwise simulation grid size
Δ_x	wall-normal simulation grid size
Δ_z	spanwise simulation grid size
Δ^L	separation distance in correlation function
$\Delta D_{\overline{u}^+=5}$	blanketing layer depth
Θ	double-averaged temperature
Λ_s	roughness density parameter
Π	streamwise constant pressure gradient

lower case

Symbol	Description
β_W	scale parameter of the Weibull distribution
γ^2_{kf}	coherence function of roughness height map and streamwise surface force map
δ_ν	viscous length scale
δ_{ij}	Kronecker delta
ϵ	noise in machine learning model prediction
θ	temperature difference compared to the wall temperature
θ_m	mixed mean temperature
θ_τ	friction temperature
θ_{ps}	power-law PS slope
κ	*von Kármán* constant
κ_h	temperature log-law slope coefficient
λ	wavelength
λ_f	friction factor
λ_s	frontal solidity
λ_0	maximum in-plane wavelength
λ_1	minimum in-plane wavelength
λ_{Coh}	streamwise coherence dropping wavelength
λ_f	streamwise force peak separation length
λ_{L2}	L2-regularization parameter
μ	dynamic viscosity
μ_{G}	mean value assigned to Gaussian distribution
μ_{k_r}	mean predicted k_r value across the ensemble neural network members
ν	kinetic viscosity
ρ	flow density
σ_{G}	standard deviation assigned to Gaussian distribution
σ_{k_r}	standard deviation of the predicted k_r values across the ensemble neural network members
σ_ϵ	standard deviation of the noise in machine learning model prediction

τ_{ij}	surface stress
τ_{tot}	total shear stress
τ_w	wall shear stress
φ	activation function

Abbreviations

Symbol	Description
1-D	one-dimensional
2-D	two-dimensional
3-D	three-dimensional
Adam	adaptive moment estimation
AI	artificial intelligence
AL	active learning
ANN	artificial neural network
BO	Bayesian optimization
CFR	constant flow rate
CNN	convolutional neural network
CPG	constant pressure gradient
CV	computer vision
DA	double averaging
DANS	double averaged Navier-Stokes equation
DFFT	discrete fast Fourier transform
DNS	direct numerical simulation
ENN	ensemble neural network
ES	effective slope
GPR	Gaussian process regression
HPC	high-performance computing
IBM	immersed boundary method
LIME	local interpretable model-agnostic explanations
LLM	large language model
LRP	layer-wise relevance propagation
MAE	mean absolute error
MAPE	mean absolute percentage error
MT	mixed-type thermal boundary condition
ML	machine learning
MLP	multi-layer perceptron
NLP	natural language processing
NN	neural network
p.d.f.	probability density function
PINN	physics informed neural network
PS	power spectrum
RANS	Reynolds-averaged Navier-Stokes equation
ReLU	rectified linear unit

QBC	query by committee
RMSE	root mean squared error
SA	sensitivity analysis
SGD	mini-batch stochastic gradient descent
XAI	explainable artificial intelligence

1 Introduction

1.1 Motivation

Engineering applications associated with surface-bounded turbulent flow permeate various aspects of daily life, spanning crucial domains including pipeline systems [1], marine vessels [2], turbine machinery [3], and combustion engines [4]. In these applications, surface skin friction is one of the predominant factors in determining the equipment's performance. In order to achieve the best energy efficiency, meticulous design on flow-related equipment is required. However, it is well understood that surfaces may undergo various forms of degradation, such as bio-fouling [5] and ice accretion [6]. The presence of roughness caused by surface degradation can significantly alter flow behaviour, leading to performance deterioration in delicately designed equipment.

Due to the critical economic and environmental implications it holds for the industry, a substantial body of literature is devoted to the increased drag caused by the enhanced momentum transfer over rough surfaces. For instance, addressing bio-fouling on ship hulls through anti-fouling coating application, removal, and hull cleaning costs several tens of millions of dollars [7]. This expenditure is, however, considered cost-effective in comparison to the resulting fuel consumption due to roughness. Nevertheless, an economically informed maintenance of ship hulls relies on accurate drag estimations, and the prediction uncertainty of the current state-of-the-art model results in approximately $67M uncertainty in annual global costs [8]. Furthermore, any form of profile variation on wind turbine blades, for instance, caused by ice accretion or erosion, can lead to a significant reduction in energy harvesting efficiency and potentially result in turbine stoppage [9]. Therefore, improving the roughness prediction accuracy is critical to optimizing long-term equipment performance and lifespan. This necessitates an in-depth understanding of the physical mechanisms through which roughness structures affect surface skin friction.

To unravel this intricate influence of wall roughness on the turbulent flow, extensive research efforts have been dedicated to investigating various types of roughness. These studies aim to scale the degree of roughness based on various surface properties and subsequently propose predictive models for the drag induced by roughness. Despite numerous attempts, a universally applicable prediction method remains elusive. In continuation of the preceding endeavors, this work focuses on developing the methodologies for predicting the drag of roughness in turbulent flows by leveraging data-driven approaches. Throughout this thesis, novel insights into the underlying physics of roughness-induced drag are presented. The ultimate objective is to facilitate a deeper comprehension of the impact of arbitrary rough surfaces on the turbulent flow, thereby paving the way for developing more efficient engineering solutions in the prediction and management of surface skin friction.

1.2 Background

Research on the impact of roughness on the flow dates back to the pioneering work carried out by J. Nikuradse [10]. In this work, laboratory experiments are conducted on circular pipes roughened with uniform sand grains. A foundational dataset is established, featuring the relationship between the friction factor (λ) and sand-grain size (r) across a wide range of Reynolds number (Re). It is worth noting that the skin-friction coefficient (C_f) is proportional to the friction factor with $\lambda = 4C_f$. These experiments lead to the identification of three distinct flow regimes, each characterized by the impact of roughness on turbulent flow. Firstly, while Re is situated at a low range such that the viscous sublayer envelops the entire roughness topography, the perturbation of the wall structures is effectively damped out, thus inducing no effect on the bulk flow. This regime is referred to as the *hydraulically smooth* regime, wherein the behaviour of C_f is identical to that of a smooth wall [10]. Subsequently, the size of inner-scaled roughness topography grows with increasing Re, its perturbations to the flow can no longer be completely suppressed within the viscous sublayer. The remaining roughness effect leads to an augmented drag penalty on the flow. As a consequence, the C_f in this *transitionally rough* regime, is a function of both roughness topographical features and Re. Eventually, with a further increase in Re, the inner-scaled roughness topography completely penetrates the viscous sublayer, and the influence of form drag becomes predominant. This dominance of form drag leads to C_f reaching an asymptotic constant value. Therefore, in this termed *fully rough* regime, C_f is no longer a function of Re but solely depends on the characteristics of the roughness topography. Inspired by this work, the *Nikuradse-Colebrook-Moody-Diagram* (or *Moody diagram*) is produced by L. Moody [11] which is frequently utilized in engineering applications as a standard approach. As can be seen from figure 1.1, the Moody diagram relates λ with the Reynolds number for various values of relative roughness k_s/D. Here D represents pipe diameter. The *equivalent sand-grain size*, denoted as k_s, is defined as the uniform sand-grain size [10] that yields an identical value of C_f in the fully rough regime as the investigated roughness. Therefore, the k_s value is of central importance for estimating the value of C_f and can only be determined *a posteriori* through either resource-intensive laboratory experiments or numerical simulations.

In order to avoid repetition of the aforementioned investigations for any new roughness topography, a long-existing interest in this field aims at predicting roughness k_s values based solely on roughness topographies. This is achieved through the utilization of cost-effective mathematical correlations or data-driven models. These methods typically incorporate statistical roughness parameters, such as effective slope (ES) or skewness (Sk). In order to derive such correlations, surfaces roughened by regular elements, typically in the form of spheres [12], 2-D sinusoidal waves [13], cones [14] or pyramids [15], are frequently employed for systematic research involving specific roughness parameters. The utilization of regular roughness offers favorable flexibility for conducting parametric studies on roughness properties, thanks to the relatively simplified geometry. Irregular roughness, on the other hand, is often studied through realistic surface samples e.g. [4, 16], which are understandably rare and challenging to produce due to the stringent experimental conditions required for their creation. Alternatively, researchers generate artificial irregular roughness through mathematical methods such as the Fourier transform [17] or the moving average technique [18]. These roughness generation methodologies enhance the accessibility of irregular roughness for research purposes.

Through a substantial body of preceding research on the artificial and realistic roughness samples, numerous empirical correlations are proposed, see, e.g. [14, 19, 20]. Despite these endeavors, it must be acknowledged that these correlations have limited applicability to specific parameter spaces for which they were derived. Establishing a universally applicable empirical correlation for a broad spectrum of naturally occurring roughness remains challenging [8]. The challenges impeding researchers from establishing a universally applicable correlation primarily stem from the complex and multifaceted nature of realistic roughness, which poses a multivariate regression problem. Confronted with the challenges of correlating numerous roughness topographical variables to drag, researchers have shifted their focus towards machine learning (ML) approaches. Jouybari *et al.* [21] employed a *multi-layer perceptron* (MLP) to predict k_s values using a set of 17 statistical roughness parameters, including their dot products. Lee *et al.* [22] applied transfer learning to augment available training data by leveraging the "coarse" physical knowledge from empirical equations to pre-train the MLP model. Subsequently, the model was fine-tuned using a set of Direct Numerical Simulations (DNS) on irregular roughness, serving as high-fidelity data. Nevertheless, the

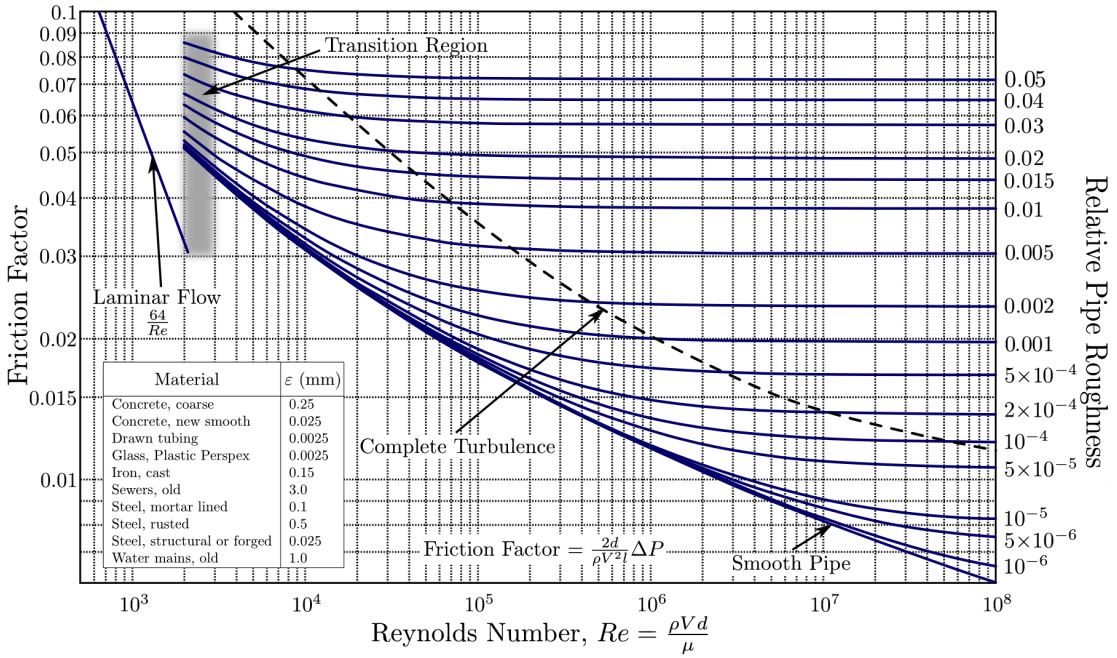


Figure 1.1: Moody chart [11].

training of data-driven models encounters a substantial challenge attributed to the constrained availability of naturally occurring roughness data. Moreover, as will be elucidated in the present dissertation, conventional single-valued roughness parameters manifest limitations in capturing the multi-scale nature of realistic roughness, potentially leading to inaccuracies in predictions. Faced with these challenges, the imperative to advance data-driven prediction methodologies becomes evident and underscores the critical necessity for innovative approaches.

1.3 Objectives and Procedure

The present dissertation is motivated by the aforementioned challenges as well as the already established knowledge in the field of roughness research. The primary objective is to apply innovative data-driven approaches for the prediction of k_s values of irregular roughness. Simultaneously, an in-depth exploration of the physical insights into the roughness effect on turbulent flow is planned. This investigation leverages the roughness database as well as the interpretation techniques for machine learning models. The aim is to provide novel insights into the behaviour of turbulent flows over irregular rough surfaces and to identify the relevant roughness structures for the resulting drag.
In order to achieve the above objectives, the initial step in the present research involves establishing and validating a training data acquisition framework. Therein, a mathematical roughness generation algorithm proposed by Pérez-Ràfols and Almqvist [17] is employed to generate a massive repository of irregular roughness topography. The roughness samples in the repository are generated based on the prescribed height *probability density function* (p.d.f.) and *power spectrum* (PS). As the main means of this study, the DNS of turbulent flow over roughness is performed using the pseudo-spectral code SIMSON [23]. The roughness structures are represented by *immersed boundary method* (IBM) following Goldstein's method [24]. The simulation code as well as the implemented IBM are thoroughly examined in preceding research, see, e.g. [14, 25]. In order to achieve the highest possible computational efficiency, the concept of minimal-channel DNS [13, 26] is employed for the calculation of the k_s values. While some researchers have already successfully applied the concept, it is essential to note that the applicability of minimal channel has only been validated in the context of regular 2-D sinusoidal wall structures. Recently, the documented failure of the minimal-channel simulation [27] underscores the necessity of a systematic verification of the minimal channel approach in the context of broader types of roughness. Bearing this in mind, the suitability of the minimal channel for irregular roughness is first assessed using systematically varied artificial irregular roughness generated by the chosen algorithm across transitionally and fully rough regimes. These roughnesses are generated following systematically varied combinations of roughness p.d.f. and PS. The uncertainty of the results obtained through minimal-channel DNS introduced by the randomness of irregular roughness is investigated. The correlation and coherence function between surface force and roughness topography are computed to provide deeper insights into the mechanisms of drag induced by roughness. Utilizing the current roughness datasets, the impacts of several roughness statistical parameters on their resulting drag are investigated. Three widely cited roughness empirical correlations are assessed based on the same datasets.
The efficacy of the chosen pseudo-random roughness generation algorithm in representing the properties of naturally occurring roughness is examined through reproducing several realistic surface scans. The impact of the artificial roughness surrogates on the turbulent flow is compared with their corresponding original realistic surface scans through DNS. Furthermore, the current DNS investigation routine is used to replicate the laboratory experiment on P60 grit sandpaper. This comparative analysis seeks to examine the alignment between the numerical outcomes and the real-world laboratory observations, thereby confirming the validity of current roughness data acquisition tools in accurately capturing the intricate physical phenomena associated with the alteration of skin friction due to surface roughness.
With the data acquisition framework validated, a massive training dataset for a data-driven predictive model can then be established. However, despite the significantly enhanced computational efficiency provided by minimal channels, constructing a comprehensive training database remains prohibitively expensive. In response to this challenge, an *active learning* (AL)-based database construction approach [28, 29] is implemented, entailing iteratively expanding the database to maximize information gain while minimizing the number of necessary DNS investigations. Eventually, the ultimate ML model is trained based on this AL-constructed database. The model is developed in the fashion of ensemble prediction technique, comprising 50 independent *multi-layer perceptron* (MLP) members. These MLP members incorporate roughness p.d.f. and PS as input features. This ensemble model is referred to as *ensemble neural network* (ENN). The model performance is evaluated on a number of testing roughness samples, stemming from both internal and external roughness databases. Finally, a model interpretation technique, namely *layer-wise relevance propagation* (LRP) [30], is applied to extract human-readable physical knowledge from the trained model. The drag-relevant roughness structures are successfully identified through the interpretation. The effects of these drag-relevant roughness structures are demonstrated by means of the analysis of

the time-averaged flow field and turbulent statistics. Recognizing the irreplaceable practical significance of the predictions derived from conventional single-valued roughness statistical parameters, additional ENN models are developed based on these parameters using the same AL-constructed database. The architectures of these ENN models in the comparison are optimized through Bayesian optimization to achieve their respective best performance. The comparative analysis examines how different roughness characterization methods, including those based on p.d.f. and PS, or single-valued roughness parameters, impact the model when used as inputs. The *sensitivity analysis* (SA) is carried out on these parameter-based ENN models. This analysis derives data-driven insights into the sensitivity of the roughness-induced drag to each roughness parameter. By comparing the mentioned ML models, the distinct advantages of all characterization methods are revealed. This insight is expected to provide valuable guidance for conducting an informed selection of roughness predictive models for future applications.

In summary, the present dissertation introduces an innovative data acquisition and ML framework for predicting k_s across a wide range of roughness types. Furthermore, to shed further light on the underlying physical mechanism of the roughness impact on the turbulent flow, comprehensive analyses on various types of roughness as well as on the ML models are carried out.

This dissertation is based on the research documented in several articles authored or co-authored by Jiasheng Yang. These articles are cited using the author-year citation style to distinguish them from other literature, simplifying identification for the readers. Specifically, chapter 3 is based on the journal publication [Yang et al. 2022] and a conference contribution to the 13th International ERCOFTAC Symposium (ETMM13) [Yang et al. 2021]. The DNS-based comparison of artificial roughness surrogates with realistic roughness, detailed in chapter 4, draws on the results presented in the journal publications [Yang et al. 2023b] and [Frohnapfel et al. 2024]. Finally, chapter 5 illustrates the implementation and evaluation of data-driven predictive models based on the journal publications [Yang et al. 2023a] and [Yang et al. 2024]. The contents directly adopted from these publications are marked using quotation marks along with the citations to the corresponding publications. Please note that the numbering of sections, figures, citations, tables, and mathematical symbols, have been adapted to align with the format of this thesis without specific notification.

2 Fundamentals

This chapter presents fundamental concepts and methodologies that are relevant to the research objectives of this dissertation. Section 2.1 offers a concise overview of the characterization parameters used for describing roughness properties and the numerical methodologies for generating irregular roughness topographies. A literature review about the observations and formulations regarding turbulent channel flows subjected to both smooth and rough surfaces is presented in section 2.2. In section 2.3, detailed numerical procedures for conducting direct numerical simulations of turbulent channel flows over rough surfaces are introduced. Lastly, the foundational operational principles of artificial neural networks and the model interpretation techniques utilized in this work are elucidated in section 2.4.

2.1 Irregular Roughness Topography

This section initiates with an overview of the traditional methodologies used to describe roughness topography, covering fundamental formulations and diverse statistical characterization methods. Subsequently, several methodologies for generating irregular roughness are introduced.

2.1.1 Characterization of Roughness Topography

The roughness topographies are typically represented by 2-D discrete height elevation maps denoted as $k(x, z)$, where x and z denote the wall-parallel coordinates. However, characterizing roughness based on the shapes and the distributions of roughness elements is exceptionally challenging due to the stochastic nature of irregular surface topographies, which renders each rough surface unique. To address this challenge, statistical measurements of roughness topographies are frequently employed. These statistics offer valuable characterization of roughness from different perspectives. Typically, these statistical parameters can be categorized into two groups based on the aspects of roughness properties they describe: those related to the height distribution of roughness and those related to its spatial (wall-parallel) distribution [17].

Height Distribution of Roughness

The height (vertical) distribution properties of roughness structures is responsible for describing the variation of roughness topography, e.g., the volume occupation, along the wall-normal direction. These properties can be manifested through the roughness height p.d.f.. However, in practical applications, a set of single-valued statistical parameters are suggested by the researchers to characterize this aspect of roughness properties.

Originating from the earliest studies, the roughness peak-to-trough height $k_z = max\{k\} - min\{k\}$ was initially employed as a metric of roughness size [10]. However, in the context of irregular roughness, extreme peaks or valleys may occur. The presence of these rare but extreme roughness heights can potentially compromise the representativeness of k_z. To address this challenge, the concept of subsample-average peak-to-trough height $k_t = \sum_{i=1}^{N} k_{z,i}/N$ across N sub-areas of the surface is introduced [37]. Alternatively, the extent of the 99% confidence interval of the roughness height p.d.f. – k_{99} – is utilized in the present work.

In addition to the roughness height measurements, the roughness parameter space is complemented by various statistical moments of the p.d.f. The *melt-down height*, also denoted as averaged roughness height, $k_{\mathrm{md}} = \int k \mathrm{d}A/A_t$ is frequently used for determining the effective roughness blockage within the channel. Here A_t denotes the wall-projected area of the roughness patch. As implied by the mathematical formulation, k_{md} represents the height that would result through melting down all the roughness elements into a smooth surface. The *porosity* of a roughness within the roughness region is defined as $Por = k_{\mathrm{md}}/k_z$. Moreover, the

mean height deviation $R_a = \int |k - k_{\mathrm{md}}| \mathrm{d}A/A_t$ and the *standard deviation* $k_{\mathrm{rms}} = \sqrt{(1/A_t)\int (k - k_{\mathrm{md}})^2 \mathrm{d}A}$ are utilized in the literature to describe the undulation of roughness profiles. [37–40]. The *Skewness* is delineated as the third central moment of roughness height p.d.f.:

$$Sk = \frac{1}{A_t k_{\mathrm{rms}}^3} \int (k - k_{\mathrm{md}})^3 \mathrm{d}A \ . \tag{2.1}$$

According to the mathematical definition, Sk effectively represents the asymmetry of the p.d.f. [41]. A roughness exhibiting $Sk > 0$ is regarded as peak-dominant, while $Sk < 0$ indicates a pitted surface. Numerous studies [42–44] suggest that Sk is among the key parameters for correlating topography to roughness drag. Lastly, the *Kurtosis* is a parameter used in various studies on roughness characterization [41, 44]. It measures the tailedness of the p.d.f., representing how the distribution of roughness heights deviates from a normal distribution. Mathematically, it is defined as the fourth central moment of the p.d.f. following Pearson's definition [45]:

$$Ku = \frac{1}{A_t k_{\mathrm{rms}}^4} \int (k - k_{\mathrm{md}})^4 \mathrm{d}A \ . \tag{2.2}$$

In practical terms, the Gaussian distribution exhibits $Ku = 3$. A $Ku > 3$ suggests that the distribution has heavier tails compared to the Gaussian distribution. Conversely, a $Ku < 3$ indicates lighter tails, implying a flatter distribution of roughness heights.
It can be understood that the aforementioned roughness parameters effectively capture specific topographical features embedded in the roughness p.d.f. Hence, a direct representation of the roughness p.d.f. can potentially provide a more comprehensive characterization of the height distribution properties of the roughness. Nevertheless, it is worth noting that realistic rough surfaces often exhibit unique p.d.f.s. These distinctive p.d.f.s can mostly be modeled using various analytical distribution functions governed by a finite set of shape parameters. For instance, the *Gaussian distribution* is expressed as:

$$f_{\mathrm{G}}(k|\mu_{\mathrm{G}}, \sigma_{\mathrm{G}}^2) = \frac{1}{\sigma_{\mathrm{G}}\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{k-\mu_{\mathrm{G}}}{\sigma_{\mathrm{G}}})^2} \ , \tag{2.3}$$

where μ_{G} and σ_{G} denotes the mean value and standard deviation of the Gaussian distribution, respectively. Another frequently used distribution function is the *Weibull distribution*, given by:

$$f_{\mathrm{W}}(k|K_W, \beta_W) = K_W \beta_W^{K_W} k^{(K_W - 1)} e^{-(\beta_W k)^{K_W}} \ , \tag{2.4}$$

where K_W and β_W are the shape parameter and scale parameter of the Weibull distribution, respectively. Additionally, the *bimodal distribution* is constructed through combining two Gaussian distributions, as outlined in [46]:

$$f_{\mathrm{B}}(k|\mu_{\mathrm{G}}, \sigma_{\mathrm{G}}^2) = f_{\mathrm{G}}(k|0,1) + f_{\mathrm{G}}(k|\mu_{\mathrm{G}}, \sigma_{\mathrm{G}}^2) - f_{\mathrm{G}}(k|0,1) f_{\mathrm{G}}(k|\mu_{\mathrm{G}}, \sigma_{\mathrm{G}}^2) \ . \tag{2.5}$$

Spatial Distribution of Roughness

The spatial distribution of roughness is widely identified as one of the significant factors determining the roughness drag [14, 47–49]. In the pioneering work by Schlichting [50], the *frontal solidity* $\lambda_s = A_f/A_t$ is defined as one of the roughness parameters for correlating the roughness drag, where A_f is the *total projected frontal area*. This parameter is demonstrated to correlate well with the surface skin friction exerted by regular roughness. To account for the various shapes of the roughness elements, the *roughness density parameter* is delineated as [51, 52]

$$\Lambda_s = \frac{A_t}{A_f} \frac{A_f}{A_s}^{-1.6} \ , \tag{2.6}$$

where A_s is the *total windward wetted surface area*. However, the definitions mentioned above may introduce ambiguities in their calculation when dealing with the surface covered by irregular roughness. To address this challenge, the *effective slope* is defined as the mean absolute gradient of the roughness [53]:

$$\mathrm{ES} = \frac{1}{A_t} \int \left| \frac{\partial k}{\partial x} \right| \mathrm{d}A \ . \tag{2.7}$$

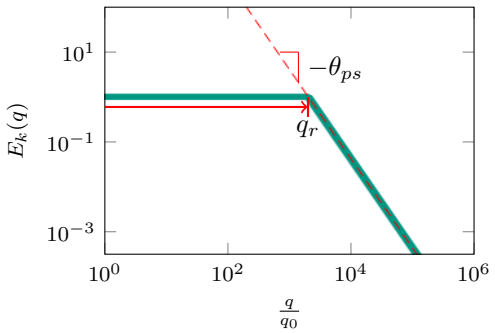


Figure 2.1: Illustration of the modeled PS incorporating the roll-off wavenumber q_r and PS slope $-\theta_{ps}$.

Furthermore, the significance of the streamwise *correlation length* in influencing the roughness drag is reported by Thakkar *et al.* [37]. This parameter – denoted as L_x^{Corr} – signifies the mean streamwise separation length at which the auto-correlation function of the roughness topography drops to a certain threshold – typically 0.2 [54].
Nevertheless, it is essential to acknowledge that the roughness element at different length scales may yield different impact on the turbulent flow. This multi-scale nature is substantial in determining the roughness drag [55–57]. However, despite the introduced variety of roughness parameters for characterizing the spatial distribution of surface roughness, these single-valued parameters are arguably inadequate for capturing the said multi-scale nature of irregular roughness. Acknowledging this limitation, this dissertation utilizes the roughness height power spectrum (PS) to capture roughness undulations across different wavenumbers. Although the utilization of PS for predicting the surface skin friction has received relatively less attention, it is extensively applied in fields such as tribology and adhesion studies [58]. In these investigations, the roughness is usually simplified as *self-affine fractal* which is featured by *power-law* descending PS. The power-law PS is expressed as:

$$E_k(q) = C_0 \left(\frac{q}{q_0} \right)^{-2(1-H_f)} , \tag{2.8}$$

where q_0 is the reference wavenumber, C_0 is the constant for scaling the roughness height, H_f represents the *Hurst exponent*. The self-affine power spectrum can be observed in various real-world scenarios, such as on fracture surfaces. However, as schematically displayed in figure 2.1, the power spectra for most naturally occurring roughness can be simplified into two segments. With this simplification, the PS of realistic roughness can be modeled using two parameters: the roll-off wavenumber q_r [58] and the power-law slope $-\theta_{ps} = -2(1 - H_f)$ [59]. Wherein, the parameter q_r denotes the wavenumber below which the PS remains constant, while beyond q_r, it decreases with an exponent of $-\theta_{ps}$.

2.1.2 Pseudo-Random Roughness Generation

Roughness encountered in realistic applications typically exhibits stochastic morphology. In pursuit of a comprehensive and systematic study of rough surfaces that closely resemble real-world roughness effects, a random roughness generation process is sought while retaining control over their statistical properties, thus it is termed "pseudo-random". Napoli *et al.* [53] generated irregular 1-D roughness profiles through the superposition of multiple sinusoidal waves with different wavelengths and amplitudes. Busse *et al.* [60] employed a moving average technique to generate 2-D rough surfaces through linear combinations of uncorrelated Gaussian random number matrices, attaining the desired auto-correlation function. Forooghi *et al.* [14, 61] placed circular cones of random sizes and shapes on the smooth surface to generate irregular roughness. The statistical properties of the generated roughness, e.g., ES and Sk, were achieved through iterative adjustment of the shapes of these roughness elements.
In addition to the previously mentioned roughness generation methods, the present work employs a mathematical algorithm outlined by Pérez-Ràfols and Almqvist [17]. This method is deemed advantageous in producing artificial roughness since it provides flexibility to prescribe any desired combination of p.d.f.

and PS. This algorithm contains iterative adjustment of roughness p.d.f. and PS in physical and spectral spaces, respectively. The target PS is defined within a specified wavenumber range, denoted by $q_0 \leq q \leq q_1$, where q_0 and q_1 represent the lower and upper wavenumber limits of the target PS, respectively. The PS outside this wavenumber range is automatically prescribed to zero. The transform between physical space and spectral space is accomplished by *discrete fast Fourier transform* (DFFT). It is worth mentioning that due to the use of DFFT, a periodic boundary condition of the roughness topography is naturally fulfilled. Initially, two initial roughness maps, denoted as k_h^0 and k_{ps}^0, are independently generated based on the prescribed p.d.f. and PS, respectively. The subscripts h and ps indicate that the roughness maps contain the target p.d.f. and PS, respectively. Subsequently, these initial maps are iteratively corrected to approach the desired combination of p.d.f. and PS. This iteration scheme consists of the following procedures:

1. Correcting PS of k_h^i to get k_{ps}^{i+1} through:

$$\hat{k}_{ps}^{i+1} = \hat{k}_h^i \frac{|\hat{k}_{ps}^i|}{|\hat{k}_h^i|} \ , \tag{2.9}$$

   where $\hat{k}$ represents the roughness map after Fourier transform

2. Adjusting p.d.f. of k_{ps}^{i+1} through height rank ordering in physical space. The updated surface is denoted as k_h^{i+1}.

3. Sending updated maps k_{ps}^{i+1} and k_h^{i+1} to step 1 until both surfaces converge to stationary state.

In order to quantify the departure of the generated p.d.f. and PS from their target counterparts, three error metrics are defined as follows:

- Error in p.d.f.

$$err_h = \sum_{i=1}^{N_b} |n_{h,i} - n_{ps,i}|. \tag{2.10}$$

- Error of PS within the prescribed wavenumber range

$$err_{ps} = \sqrt{\frac{1}{N_{ps}} \sum_{q_0 \leq q_i \leq q_1} \left(\frac{|\hat{k}_h(q_i)| - |\hat{k}_{ps}(q_i)|}{|\hat{k}_{ps}(q_i)|}\right)^2}. \tag{2.11}$$

- Error of PS outside the prescribed wavenumber range

$$err_{ps0} = \frac{\sqrt{1/N_{ps0} \sum_{q_i < q_0 | q_1 < q_i} (|\hat{k}_h(q_i)|)^2}}{1/N_{ps} \sum_{q_0 \leq q_i \leq q_1} (|\hat{k}_{ps}(q_i)|)}. \tag{2.12}$$

Here, n_h is the frequency of roughness height within the i-th height interval (denoted as bin) of the target p.d.f., and n_{ps} is that of k_{ps}, N_b is the total number of bins of the p.d.f. N_{ps} is the number of discrete wavenumbers that are located within the range of $q_0 \leq q \leq q_1$ while N_{ps0} denotes the number of discrete wavenumbers located outside this wavenumber range, i.e., at $q < q_0$ or $q > q_1$.

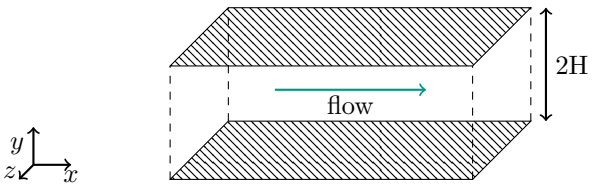


Figure 2.2: Schematics of canonical channel flow of infinite length and width, hatched planes represent walls.

2.2 Turbulent Channel Flow

The canonical channel flow entails the confinement of a fluid between two parallel plates of infinite extent, subject to a pressure gradient that drives the flow. The geometry of the idealized channel flow is depicted in figure 2.2. The channel comprises two parallel planes separated by a distance of 2H along the vertical coordinate y, with the primary flow aligned with the x coordinate, and z representing the spanwise coordinate. Here, H denotes the channel half height. The instantaneous velocity components along the x, y and z coordinates are denoted by u, v, w, respectively. Alternatively, the coordinates and velocity components can be expressed using index notation, i.e., x_i and u_i, respectively. The simplicity of the geometry and boundary conditions has facilitated both experimental and theoretical investigations on wall-bounded turbulence by means of channel flow [62]. This section provides a concise overview of the fundamental theories and formulations governing turbulent channel flows subject to smooth and rough channel walls.

2.2.1 Governing Equations

The governing equations for the fluid motion are derived based on the reference [63]. The mass conservation equation in differential form reads:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i}(\rho u_i) = \underbrace{\frac{\partial \rho}{\partial t} + u_i \frac{\partial \rho}{\partial x_i}}_{\text{incompressible}=0} + \rho \frac{\partial u_i}{\partial x_i} = 0 \ , \tag{2.13}$$

where u_i denotes the velocity component in the i-direction. Throughout this dissertation, the fluid is assumed to be incompressible. Under this assumption, the material derivative of the density is zero, resulting in the divergence-free condition of the flow field:

$$\frac{\partial u_i}{\partial x_i} = 0 \ . \tag{2.14}$$

The momentum equation of the flow is derived from Newton's second law, describing the acceleration of the fluid motion due to surface and body forces. Under the same assumption of incrompressible flow, the momentum equation is formulated as

$$\frac{\partial u_j}{\partial t} + u_i \frac{\partial u_j}{\partial x_i} = \frac{1}{\rho} \frac{\partial \tau_{ij}}{\partial x_i} - g_j \ , \tag{2.15}$$

where τ_{ij} represents the surface stress tensor, and g_j represents the body force, e.g., gravitational force. Moreover, the surface stress τ_{ij} can be decomposed into pressure and shear stress components. In the case of a *Newtonian fluid*, the shear stress exhibits a linear relation to the flow shear rate, thus the decomposition of τ_{ij} is given as:

$$\tau_{ij} = -p\delta_{ij} + \mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \ , \tag{2.16}$$

where μ represents dynamic viscosity, δ_{ij} is the *Kronecker-delta* function. Finally, inserting equation 2.16 into equation 2.15 yields the *Navier-Stokes equations*:

$$\frac{\partial u_j}{\partial t} + u_i \frac{\partial u_j}{\partial x_i} = -\frac{1}{\rho} \frac{\partial p}{x_j} + \nu \frac{\partial^2 u_j}{\partial x_i \partial x_i} - g_j \ , \tag{2.17}$$

where $\nu = \mu/\rho$ represents the *kinematic viscosity*.
Furthermore, the applicability of the current numerical framework for studying heat transfer across roughness is examined in this dissertation as well. The temperature field θ is considered a passive scalar. The transport equation of the passive temperature scalar is derived based on the energy conservation:

$$\frac{\partial \theta}{\partial t} + u_i \frac{\partial \theta}{\partial x_i} = \alpha \frac{\partial^2 \theta}{\partial x_i \partial x_i} + Q \ , \tag{2.18}$$

where α is the temperature diffusivity, and Q represents the heat source term, such as radiation in the fluid.
It is well known that the turbulent flows are characterized by unsteady and chaotic fluid motion. As the common practice of turbulence research, turbulent flows are frequently described using statistical methods. To this end, time-averaged flow properties are extracted by applying Reynolds decomposition, which yields

$$\varphi(x_j, t) = \overline{\varphi}(x_j) + \varphi'(x_j, t) \ , \tag{2.19}$$

where φ represents any interested flow property, such as u_i or θ. The overbar $\overline{\cdot}$ is the time averaging operation and $'$ denotes the temporal fluctuation. Introducing the Reynolds decomposition to the mass conservation equation (Eqn. 2.14), Navier-Stokes equations (Eqn. 2.17), and energy transport equation(Eqn. 2.18), followed by time averaging, leads to the Reynolds-averaged Navier-Stokes equation (RANS):

$$\frac{\partial \overline{u}_i}{\partial x_i} = 0, \tag{2.20a}$$

$$\frac{\partial \overline{u}_j}{\partial t} + \overline{u}_i \frac{\partial \overline{u}_j}{\partial x_i} = -\frac{1}{\rho} \frac{\partial \overline{p}}{x_j} + \nu \frac{\partial^2 \overline{u}_j}{\partial x_i \partial x_i} - \frac{\partial}{\partial x_i} \overline{u_i' u_j'} - g_j, \tag{2.20b}$$

$$\frac{\partial \overline{\theta}}{\partial t} + \overline{u}_i \frac{\partial \overline{\theta}}{\partial x_i} = \alpha \frac{\partial^2 \overline{\theta}}{\partial x_i \partial x_i} + \overline{Q} - \frac{\partial}{\partial x_i} \overline{u_i' \theta'}, \tag{2.20c}$$

here $\overline{u_i u_j}$ is referred to as the *Reynolds stress* and $\overline{u_i \theta'}$ denotes *turbulent heat flux*. The time-averaged Navier-Stokes equation for fully developed turbulent channel flow in streamwise direction is expressed as

$$0 = -\frac{\partial \overline{p}}{\partial x} + \frac{\partial}{\partial y} \underbrace{\left(\mu \frac{\partial \overline{u}}{\partial y} - \rho \overline{u' v'} \right)}_{\tau_{tot}} \ , \tag{2.21}$$

where the summation of viscous and Reynolds shear stresses is denoted as the total shear stress τ_{tot}.
Throughout the present dissertation, *no-slip* boundary condition is employed at the upper and lower wall. The flow is driven by a *constant pressure gradient* (CPG), denoted by $\Pi = \partial \overline{p}/\partial x = const$. A specific thermal boundary condition is employed to mimic realistic conditions [64]. This involves an isothermal boundary condition with zero wall temperature, while the averaged temperature increases linearly in the streamwise direction, similar to isoflux boundary conditions. This thermal boundary condition, which simultaneously satisfies Dirichlet and Neumann boundary conditions, is referred to as a *mixed-type* (MT) boundary condition in the following text. However, it is important to note that the current MT boundary condition differs from the Robin boundary condition [65] or mixed boundary condition on disjoint boundary parts [66]. The constant temperature gradient in the streamwise direction is enforced by assigning the specific heat source Q_Θ in the heat transport equation 2.18. Details on the implementation of the MT boundary condition in the numerical simulations are provided in section 2.3.1.

2.2.2 Smooth Wall Turbulent Channel Flow

Nikuradse [10] and Reichardt [67] were among the first few researchers to investigate fully developed turbulent channel flows. The former investigated mean flow properties, while the latter's experiments provided insight into turbulent fluctuation in the flow. Following these pioneering investigations, numerous novel investigations (e.g., [68, 69]) utilizing advancing instruments were carried out. These studies provide detailed observation of turbulent flow behaviour in the vicinity of the wall at a broad range of Reynolds

numbers. In the context of channel flow, the *bulk Reynolds number* – denoted as Re_b – is frequently utilized to characterize the bulk flow. The bulk Reynolds number is delineated as

$$Re_b = \frac{2\mathrm{H}U_b}{\nu} \ . \tag{2.22}$$

where $U_b = (1/2\mathrm{H}) \int_{y=0}^{2\mathrm{H}} \overline{u} \mathrm{d}y$ is the *bulk velocity*.

Momentum and Heat Flux at Smooth Wall

Considering turbulent channel flow driven by CPG, the solution for τ_{tot} can be derived as a linear function in wall-normal direction within the smooth channel $y \in [0, 2\mathrm{H}]$:

$$\tau_{tot} = \tau_w - \Pi \frac{y}{\mathrm{H}} \ , \tag{2.23}$$

where $\tau_w = \tau_{tot}(y = 0)$ denotes the *wall shear stress*, which represents the momentum flux at the wall. Hence, the wall shear stress can be calculated through:

$$\tau_w = \mu \frac{\mathrm{d}\overline{u}}{\mathrm{d}y}\bigg|_{y=0} \ . \tag{2.24}$$

Alternatively, τ_w can be determined in terms of Π by solving the force equilibrium within the channel using the equation:

$$\tau_w = \Pi \, \mathrm{H} \ . \tag{2.25}$$

Consequently, the linear function of τ_{tot} in equation 2.23 can be reformulated as:

$$\tau_{tot} = \Pi \, \mathrm{H}(1 - \frac{y}{\mathrm{H}}) \ . \tag{2.26}$$

The dimensionless form of τ_w, referred to as the *skin-friction coefficient* C_f, is delineated as:

$$C_f = 2\frac{\tau_w}{\rho U_b^2} \ . \tag{2.27}$$

On the other hand, the *wall heat flux* over a smooth wall is defined as

$$q_w = -\alpha \frac{\partial \overline{\theta}}{\partial y}\bigg|_{y=0} \ . \tag{2.28}$$

Under the current MT thermal boundary condition, q_w can be derived through energy conservation:

$$q_w = \int_{y=0}^{\mathrm{H}} Q_\Theta \mathrm{d}y \tag{2.29}$$

The dimensionless number that characterizes the heat transfer across a surface is referred to as the *Stanton number* St, which account for the influence by the flow rate. It is defined as:

$$St = \frac{q_w}{\rho c_p U_b \theta_m} \ , \tag{2.30}$$

where $\theta_m = \int_{y=0}^{2\mathrm{H}} \overline{u}\overline{\theta} \mathrm{d}y/(2U_b\mathrm{H})$ is the *mixed mean temperature* [70], and c_p represents *specific heat*.

Viscous Units and the Law of the Wall

The important physical parameters of the flow in the vicinity of the wall are the viscosity and the wall shear stress [63]. With these quantities, a set of viscous units can be defined. Wherein, the friction velocity u_τ and friction temperature θ_τ are defined as

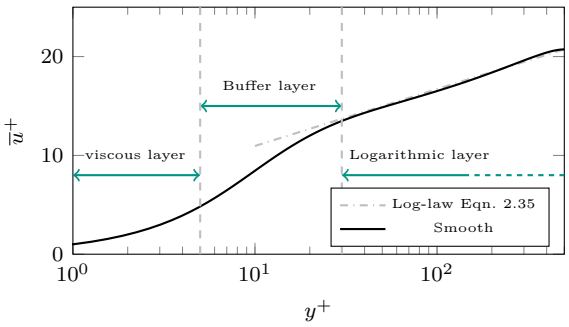


Figure 2.3: Inner-scaled velocity profile over smooth wall.

$$u_\tau = \sqrt{\frac{\tau_w}{\rho}} \ , \tag{2.31a}$$

$$\theta_\tau = \frac{q_w}{\rho c_p u_\tau} \ . \tag{2.31b}$$

The viscous length scale is delineated as

$$\delta_\nu = \frac{\nu}{u_\tau} \ . \tag{2.32}$$

The velocity and temperature in the viscous units are calculated through $u_i^+ = u_i/u_\tau$ and $\theta^+ = \theta/\theta_\tau$. Analogously, the distance in viscous unit is $l^+ = l/\delta_\nu$. Here the superscript + represents the quantities in the viscous unit. The dimensionless parameter that characterizes the wall-bounded turbulence is the *friction Reynolds number*

$$\mathrm{Re}_\tau = \frac{u_\tau \mathrm{H}}{\nu} \ . \tag{2.33}$$

In the vicinity of the wall, the *inner layer* is delineated where the mean velocity profile is predominantly influenced by the viscous scales, remaining mostly independent of the outer scales, such as H and U_b. An exemplary inner-scaled velocity profile over smooth wall is depicted in figure 2.3. It is supposed that the influence of viscosity is dominant in the region $y^+ \lessapprox 5$. This region is thus referred to as the *viscous layer*. The inner-scaled velocity profile in this layer exhibits linearity with respect to the distance from the wall:

$$\overline{u}^+ = y^+ \ . \tag{2.34}$$

Further away from the wall ($y/\delta \ll 1$ and $y^+ \gtrapprox 30$), it is supposed that the viscosity has little effect, the inner-scaled velocity profile in this region exhibits logarithmic-linear form – thus is referred to as *logarithmic layer*

$$\overline{u}^+ = \frac{1}{\kappa} \ln y^+ + A \ , \tag{2.35}$$

where $A \approx 5.2$ is the log-law intercept, and $\kappa \approx 0.4$ is the *von Kármán* constant. Additionally, the transitional zone is located between the viscous sublayer and the logarithmic layer, known as the *buffer layer*. It delineates the transition from the viscosity-dominated to turbulence-dominated regions [63].

2.2.3 Rough Wall Turbulent Channel Flow

In general, surfaces are deemed hydraulically rough when their roughness structures are sufficiently large to perturb the smallest eddies near the wall, exerting additional drag in comparison to a smooth wall. The presence of a rough wall introduces complexity to the flow dynamics, thereby complicating the associated investigation. Hence, this section provides an introductory overview of fundamental concepts and methodologies essential for studying turbulent flows bounded by rough walls. In pursuit of a close correspondence to investigations on smooth surfaces, the no-slip boundary condition is applied to the rough surface, with the flow driven by CPG. Additionally, the MT thermal boundary condition is employed, assigning a constant zero temperature to the roughness elements.

Double Average Methodology

Turbulent flow over rough surfaces can exhibit substantial spatial heterogeneity, such as the spatial variations in mean velocity and the generation of wake turbulence.The presence of such flow heterogeneity declines the application of the above mentioned turbulent description approaches through Reynolds averaging for determining global flow quantities over rough surfaces. Wilson and Shaw [71] pointed out that the inadequacy of the one-dimensional representation of wake turbulence by including an external drag term thereby short-circuits a portion of inertial eddy cascade process. As a general investigation tool for turbulent flow over roughness, these authors suggested spatial averaging of time-averaged velocity field. This spatial and time averaging process is usually referred to as the *double averaging* (DA). Following their suggestions, Raupach *et al.* [72] first evaluated the DA approach to momentum conservation in canopy flow and found it capable of clarifying the nature of wake influence on canopy flow. The mathematical procedure of DA is expressed as:

$$\langle\overline{\varphi}\rangle = \frac{1}{S^*}\iint_S \overline{\varphi}(x,y,z)\mathrm{d}x\mathrm{d}z \ , \tag{2.36}$$

where S^* represents the projected area over which spatial averaging is performed, $\langle\cdot\rangle$ represents the spatial-averaging operation, and $\overline{\varphi}$ denotes the local time-averaging operation. Typically, DA is classified as *intrinsic* or *extrinsic* based on its treatment of the solid volume within the roughness structures. In intrinsic averaging, intersections of the wall-parallel spatial averaging plane with roughness elements are addressed by assigning zero values inside the roughness. Conversely, extrinsic averaging only considers areas within the fluid domain. In the following content, intrinsic averaging is employed by default due to its simplicity of implementation and its close relevance to roughness modeling [73, 74], unless otherwise specified. For the sake of brevity, DA velocity ($\langle\overline{u}\rangle, \langle\overline{v}\rangle, \langle\overline{w}\rangle$) and temperature $\langle\overline{\theta}\rangle$ are denoted as (U, V, W) and Θ, respectively..
Following the DA approach, the *triple decomposition* of simultaneous flow field can be performed through:

$$\varphi(x,y,z,t) = \langle\overline{\varphi}\rangle(y) + \tilde{\varphi}(x,y,z) + \varphi'(x,y,z,t) \ . \tag{2.37}$$

By applying the triple decomposition to Equation 2.17 and employing DA, the *double-averaged Navier-Stokes equation* (DANS) is obtained:

$$\begin{aligned}\frac{\partial<\overline{u}_i>}{\partial t} + <\overline{u}_j>\frac{\partial<\overline{u}_i>}{\partial x_j} + \frac{\partial}{\partial x_j}<\tilde{u}_i\tilde{u}_j> + \frac{\partial}{\partial x_j}<\overline{u_i'u_j'}> = \\ -\frac{1}{\rho}\frac{\partial<\overline{p}>}{\partial x_i} - \frac{1}{\rho}<\frac{\partial\tilde{p}}{\partial x_i}> + \nu\nabla^2<\overline{u}_i> + \nu<\nabla^2\tilde{u}> - g_j \ .\end{aligned} \tag{2.38}$$

The term $\langle\tilde{u}_i\tilde{u}_j\rangle$ is referred to as the *dispersive stress*, arising from the spatial fluctuation of the time-averaged flow. Studies have demonstrated that both Reynolds stresses and dispersive stresses are crucial in determining the surface skin friction [43].

Outer-Layer Similarity

Based on the extensive observations of turbulent flows over roughness, it is noted that the influence of roughness is most pronounced in the vicinity of the wall. This roughness impact is often reflected by the augmentation of dispersive stresses. Hence, the near-wall region, where dispersive stresses display significant magnitudes, delineates the wall-normal extent of the roughness effect, and is referred to as the *roughness sublayer*. It is reported from numerous investigations on different types of roughness, that despite distinct flow responses within the roughness sublayer, turbulence behaviour becomes similar beyond this region. The general existence of this outer-layer similarity is hypothesised by Townsend [62], and articulated by Raupach *et al.* [47]. Based on the outer-layer similarity hypothesis, rough channels are anticipated to exhibit parallel mean velocity profiles and agreement in the Reynolds stresses in the outer layer, compared to their smooth channel counterparts. Consequently, the impact of the presence of roughness on the bulk flow – hence the surface skin friction – is characterized by the wall-normal offset of the mean velocity profile. This simplification has widely facilitated development of roughness predictive models to map a set of roughness topographical features to this single-valued prediction.

It is widely accepted, that the existence of outer-layer similarity relies on the sufficiently large scale separation of roughness height k and viscous length scale δ_ν in contrast to the outer layer thickness, i.e., H. These conditions can be expressed as: $k/\delta \ll 1$ and $\mathrm{Re}_\tau \gg 1$, respectively. However, despite extensive investigations on the outer-layer similarity [75, 76], a clear quantitative requirement to achieve outer-layer similarity is still missing. Jimémez [77] derived theoretical scale separation of $\mathrm{H}/k > 40$ to achieve outer-layer similarity. This criterion is determined by suppressing the roughness influence to the under half of the logarithmic layer. Nonetheless, the validity of the outer-layer similarity hypothesis has been verified by numerous laboratory or numerical experiments at lower scale separations. Macdonald *et al.* [13] reported outer-layer similarity over 2-D sinusoidal wave surfaces with the smallest scale separation of $\mathrm{H}/k = 18$. The collapse of outer-layer Reynolds stresses as well as mean velocity deficit profiles over irregular roughness compared to the smooth surface is reported with the scale separation of $\mathrm{H}/k \approx 7$ by Forooghi *et al.* [14]. Moreover, Busse *et al.* [54] investigated realistic rough surface scans with $\mathrm{H}/k = 6$ and observed outer-layer similarity. The observation of outer-layer similarity at relatively low scale separations carries significant implications for numerical simulations, given the demanding computational requirements necessary to resolve the outer-layer length scale H down to the roughness length scale k. While the present thesis heavily relies on the outer-layer similarity hypothesis, it's important to note that several experimental and numerical investigations, even at sufficiently large scale separation, have reported absence of outer-layer similarity due to extended roughness impact on the outer-layer flow, e.g., in [49, 78, 79]. In this respect, turbulent secondary motion induced by roughness is considered to be one of the significant factors leading to such deviation [25, 80]. Therefore, a general guarantee of the outer-layer similarity is yet still ambiguous and might necessitate case-by-case confirmation.

Zero-Plane Displacement

In contrast to the well-defined wall shear stress τ_w for smooth surfaces, measuring wall shear stress for rough surfaces can be subject to considerable uncertainty due to the undulating topography of the structured wall [12]. Moreover, the lack of a clear definition of the origin of wall-normal coordinate could lead to misalignment of mean velocity profiles, potentially affecting the assessment of the outer-layer similarity. In particular, the misalignment of the logarithmic region, as will be elucidated in the subsequent sections, may exert considerable impact on the investigation of roughness effects in the present study. For these reasons, a virtual origin of the logarithmic law for the mean velocity profile over roughness is to be defined. The position of this virtual plane is indicated by its displacement from the bottom wall of the roughness, known as the *zero-plane displacement* d, or denoted as the *wall-offset*. One of the common *a priori* definitions of the zero-plane displacement is $d = k_{\mathrm{md}}$ [14, 20, 54]. Alternatively, Raupach *et al.* [47] suggest defining this displacement *a posterior* by aligning logarithmic velocity profile with that over a smooth wall at the corresponding Reynolds number.
Jackson [81] proposed the centroid of the distributed drag on the roughness surface as the position for the zero-plane. This definition has been widely adopted in numerous studies, e.g., in [21, 82, 83]. "The moment of the drag can be calculated by projecting the drag forces on a y-z plane. However, it is demonstrated that d can be calculated in terms of mean flow properties, i.e., total shear stress τ_{tot}. The time averaged Navier-Stokes equation of the flow in streamwise direction writes:

$$\rho\frac{\partial(\bar{u}\bar{u})}{\partial x} + \rho\frac{\partial(\bar{u}\bar{v})}{\partial y} + \rho\frac{\partial(\bar{u}\bar{w})}{\partial z} = -\frac{\partial\bar{p}}{\partial x} + \frac{\partial T_{11}}{\partial x} + \frac{\partial T_{12}}{\partial y} + \frac{\partial T_{13}}{\partial z} - \Pi + \bar{f}_{\mathrm{IBM},x} \ , \tag{2.39}$$

where $\bar{u}$, $\bar{v}$ and $\bar{w}$ are the mean velocity components. $\bar{p}$ is the mean pressure and T_{11}, T_{12} and T_{13} are the stresses including Reynolds stresses. $\bar{f}_{\mathrm{IBM},x}$ is the streamwise component of the mean IBM force. Π is the constant pressure gradient added to the flow. With the idealized geometry proposed by Jackson [81], if (2.39) is integrated over wall-parallel directions, then multiplied by y and integrated over y from the

bottom $y = 0$ to the tip of the roughness $y = k_z$ we obtain:

$$\begin{aligned}
&\underbrace{\int_{L_z}\int_0^{k_z} y\rho[\bar{u}\bar{u}]_0^{L_x} dydz}_{=0} + \int_{L_z}\int_{L_x}[y\rho\bar{u}\bar{v}]_0^{k_z} dxdz - \int_{L_z}\int_{L_x}\int_0^{k_z}\rho\bar{u}\bar{v}dydxdz \\
&+ \underbrace{\int_{L_x}\int_0^{k_z} y\rho[\bar{u}\bar{w}]_0^{L_z} dydx}_{=0} = \underbrace{\int_{L_z}\int_0^{k_z} -y[\bar{p}]_0^{L_x} dydz}_{=0} + \underbrace{\int_{L_z}\int_0^{k_z} y[T_{11}]_0^{L_x} dydz}_{=0} \\
&+ \int_{L_z}\int_{L_x}[yT_{12}]_0^{k_z} dxdz - \int_{L_z}\int_{L_x}\int_0^{k_z} T_{12}dydxdz + \underbrace{\int_{L_x}\int_0^{k_z} y[T_{13}]_0^{L_z} dydx}_{=0} \\
&- \int_{L_z}\int_{L_x}\int_0^{k_z} y\Pi dydxdz + \int_{L_z}\int_{L_x}\int_0^{k_z} y\bar{f}_{\mathrm{IBM},x} dydxdz \ .
\end{aligned} \tag{2.40}$$

As marked in the equation, some of the terms vanish due to the periodic boundary condition in wall parallel directions. Thus, the moment of the drag acting on the roughness is formulated as:

$$\begin{aligned}
M = -\int_{L_z}\int_{L_x}\int_0^{k_z} y\bar{f}_{\mathrm{IBM},x} dydxdz &= \int_{L_x}\int_{L_z}[yT_{12} - y\rho\bar{u}\bar{v}]_{y=k_z} dzdx \\
&- \int_{L_x}\int_{L_z}\int_0^{k_z}[T_{12} - \rho\bar{u}\bar{v}]dydzdx - \int_{L_x}\int_{L_z}\int_0^{k_z} y\Pi dydzdx \ .
\end{aligned} \tag{2.41}$$

Here M is the moment on the surface. Following which, zero-plane displacement $d = k_z - \frac{M}{\tau_w L_x L_z}$ is calculated. With the operation $(L_x L_z)^{-1}\int_{L_x}\int_{L_z}[T_{12} - \rho\bar{u}\bar{v}]dzdx$ dispersive stress is included in the total shear stress τ_{tot} [81]. Kameda *et al.* [84] calculated the displacement d by setting the wall coordinate origin at k_z, thus the equation for the zero-plane displacement d writes:"[Yang et al. 2022]

$$d = k_z - \frac{\int_0^{k_z}(\tau_{tot} + y\Pi)dy}{\tau_w} \ . \tag{2.42}$$

Throughout the present thesis, the definition of d according to Eqn. 2.42 is employed.

Skin friction and Heat Transfer of Rough Surface

As discussed in the preceding section, in the context of roughness, a direct measurement of wall shear stress τ_w through Eqn. 2.24 is not feasible due to the presence of wall structure. For the purpose of measuring the skin friction, direct methods such as through floating-element force balance [85] in experiments or integrating roughness surface force [12, 34] in simulations are commonly utilized.
Alternatively, an estimation of τ_w can be obtained through indirect methods through investigating the mean flow quantities. As a common approach, τ_w can be obtained through the force balance between the applied constant pressure gradient Π and τ_w [34, 86]

$$\tau_w = \Pi\,\mathrm{H}_{\mathrm{eff}} \ , \tag{2.43}$$

where $\mathrm{H}_{\mathrm{eff}} = \mathrm{H} - h_{\mathrm{ref}}$ is the effective channel half height, h_{ref} represents the reference plane height due to the blockage of the roughness in the channel. In the present thesis $h_{\mathrm{ref}} = k_{\mathrm{md}}$ is chosen, which is arbitrary but commonly used in the literature, for instance in [13, 86, 87]. Further selections of h_{ref}, along with their consequential influence on the calculation of skin-friction coefficient, are discussed in the literature [12, 34]. An alternative approach involves calculating τ_w by extrapolating the linear part of τ_{tot} in the outer layer down to h_{ref} [14]. It is demonstrated that both indirect methods closely resemble the values for τ_w [34]. The calculation of C_f is identical to Eqn. 2.27, with the distinction that the bulk velocity is computed based on the effective half height of the channel, i.e., $U_b = \int_{y=0}^{2\mathrm{H}} U\mathrm{d}y/(2\mathrm{H}_{\mathrm{eff}})$. Consequently, Re_τ for rough surface is defined with $\mathrm{H}_{\mathrm{eff}}$:

$$\mathrm{Re}_\tau = \frac{u_\tau \mathrm{H}_{\mathrm{eff}}}{\nu} \ . \tag{2.44}$$

With the current MT thermal boundary condition, the heat flux q_w across the roughness, on the other hand, can be calculated through Eqn. 2.29. Finally, the calculation of St follows Eqn. 2.30 with $\theta_m = \int_{y=0}^{2\mathrm{H}} U\Theta \mathrm{d}y/(2U_b\mathrm{H}_{\mathrm{eff}})$.

Roughness Function

Analogous to the derivation of log-law over smooth surfaces, where wall-normal position y is required to be sufficiently larger than the viscous sublayer scaled by δ_ν, i.e., $y \gg \delta_\nu$. A further significant parameter – namely the topography of roughness k– is introduced into the system when the viscous-scaled roughness is sufficiently large $k \gg \delta_\nu$. Under the outer-layer similarity hypothesis, where the flow in the logarithmic layer remains unaffected by the roughness, it is essential that y is significantly larger than the characteristic roughness length scale, denoted as $y^+ \gg k^+$. To achieve the same slope of $y^+\mathrm{d}U/\mathrm{d}y = 1/\kappa$, the log-law over rough surface in the y^+/k^+ coordinate, incorporating zero-plane displacement d, can be formulated as:

$$U_R^+ = \frac{1}{\kappa}\ln\frac{(y-d)^+}{k^+} + B(k^+) \ , \tag{2.45}$$

where $B(k^+)$ is a function of both roughness topography and roughness Reynolds number. As can be understood, the log-law intercept $B(k^+)$ indicates the mean velocity offset in the logarithmic layer, Hama[88] and Clauser [89] independently defined the *roughness function* ΔU^+ to describe this log-law intercept associated to the roughness effect. The mathematical formulation of log-law over roughness is thus expressed as

$$U_R^+ = \frac{1}{\kappa}\ln(y-d)^+ + A - \Delta U^+ \ . \tag{2.46}$$

Therefore, the impact of a given rough surface boils down to determining ΔU^+. A positive value of the roughness function ΔU^+ can quantify drag penalty while negative value of ΔU^+ indicates drag reduction [90]. With the assumption of idealized outer-layer similarity and the near-wall mean velocity profile [91], the ΔU^+ can be estimated with the skin-friction coefficient through

$$\Delta U^+ = \frac{1}{\kappa}\ln\sqrt{\frac{C_{f,r}}{C_{f,s}}} + \sqrt{\frac{2}{C_{f,s}}} - \sqrt{\frac{2}{C_{f,r}}} \ , \tag{2.47}$$

where $C_{f,s}$ and $C_{f,r}$ denote the skin-friction coefficients of smooth and rough surface at corresponding Reynolds number, respectively. The computation of ΔU^+ from these global flow quantities is of great convenience and is frequently used especially for experimental practices [19, 90]. As an intuitive approach, the value of ΔU^+ for interested roughness can be directly obtained through evaluating the inner-scaled mean velocity offset compared to the smooth channel with consistent Re_τ. Busse *et al.* [56] and Thakkar *et al.* [92] computed ΔU^+ as the difference between the mean centreline velocities of the rough- and smooth-wall cases at the same Re_τ. While this approach is clear from its definition, the measurement is potentially subject to a degree of uncertainty introduced by this one-point observation. In order to mitigate this uncertainty of measurement, the value of ΔU^+ is computed as the averaged offset of velocity profiles across the logarithmic layer throughout the present thesis. To this end, the virtual origin of the log-law over the roughness is determined by the zero-plane displacement d, aiming to attain parallel logarithmic mean velocity profiles. Some resulting mean velocity profiles over example roughness are displayed in figure 2.4 (a).

Analogous to the velocity field, the temperature profile over smooth and rough surfaces also exhibits a region with logarithmic profile [93–95]. Similarly to the definition of roughness function ΔU^+, the increase in the heat transfer capability of a surface due to roughness manifests a retardation of the logarithmic inner-scaled temperature profile $\Delta\Theta^+$ [96]. Therefore, the log-law temperature profile over roughness reads:

$$\Theta_R^+ = \frac{1}{\kappa_h}\log(y-d)^+ + A_h(Pr) - \Delta\Theta^+ \ , \tag{2.48}$$

where the offset A_h depends on the *Prandtl number* $Pr = \nu/\alpha$ of the flow. The slope coefficient $k_h \approx \kappa/Pr_t \approx 0.46$. Pr_t denotes *turbulent Prandtl number* typically in the range of 0.85-0.9. In the present thesis, the calculation of $\Delta\Theta^+$ closely follows the procedure employed for ΔU^+.

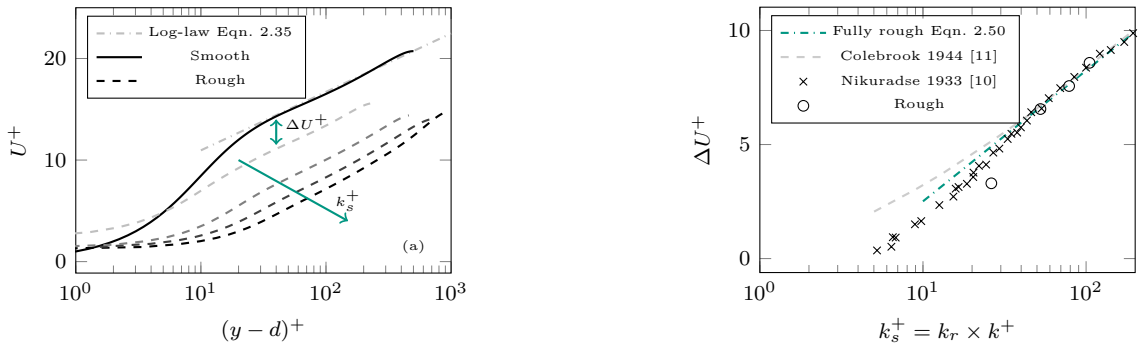


Figure 2.4: (a): Exemplary mean velocity profiles over roughness, line colour gradually changes from grey to black with increasing inner-scaled equivalent sand-grain roughness size k_s^+. (b): Roughness functions from (a) plotted against k_s^+. Data from Nikuradse's uniform sand grain roughness and Colebrook relation provide for industrial pipes are added for comparison. Exemplary roughness data adopted from [Yang et al. 2022]

Equivalent Sand-Grain Size

The roughness function ΔU^+ can be correlated with the roughness topography k^+ by subtracting rough-wall log-law (Eqn.2.45) from its smooth-wall counterpart (Eqn.2.35), which yields:

$$\Delta U^+ = \frac{1}{\kappa} \ln k^+ + A - B(k^+) \ . \tag{2.49}$$

In the fully rough regime function where viscous effect can be neglected, for considerably high roughness Reynolds number $k^+ \gg 1$, $B(\infty)$ asymptotically approaches constant and is consequently solely a function of roughness topography. As previously discussed, it is a common practice to replace the arbitrary roughness size k to the equivalent sand-grain roughness size k_s to seek correspondence to Nikuradse's work [10], such that Eqn. 2.49 can be recast as

$$\Delta U^+ = \frac{1}{\kappa} \ln k_s^+ + A - B_s(\infty) \ , \tag{2.50}$$

where $B_s(\infty) = 8.5$ is a constant fitting to the uniform sand-grain roughness [10]. In common practice, the ΔU^+ value as a function of increasing k^+ is obtained for interested roughness. The k_s value is obtained by fitting the ΔU^+ curve in the fully rough regime with Eqn.2.50. Graphically, this procedure is equivalent to multiplying the characterizing roughness size k with a factor, denoted as the *relative equivalent sand-grain size* (k_r) in the present thesis. Therefore, the collapse of roughness function onto the fully rough asymptote is achieved through adjusting k_r.The equivalent sand-grain size of the roughness under investigation is thus calculated as $k_s = k_r \times k$. One example is shown in figure 2.4(b). As a final remark, it is clear that k_s for any given roughness profile represents a hydraulic property rather than a direct physical scale derived from the roughness topography itself. The equivalent sand-grain size k_s acts as a universal metric across various roughness types, facilitating its correlation with the resulting surface skin friction [97].

2.2.4 Existing Drag Prediction Methodologies

In recent years, research on the impact of roughness on turbulent flow has been greatly bolstered by the ever expanding database acquired through continuous advancements in experimental and numerical methodologies. Within the broad spectrum of studies investigating roughness effects, considerable endeavors are devoted to predicting roughness-induced drag based on various roughness topographical features. The earliest attempts, dating back to Bettermann [98], Dvorak [99], and Dirling [100], entailed the development of numerous empirical correlations for ΔU^+ concerning rods and bars of varying heights and densities. Nevertheless, these correlations faced challenges in extrapolating their performance across a diverse spectrum of roughness types. This limitation is primarily attributed to the significant influence of roughness shapes, as discussed in studies such as [51, 101]. Additionally, the use of the above 1-D regular roughness models

proved inadequate in accurately resolving the intricate topographical characteristics of naturally occurring 2-D irregular roughness. Having this in mind, 2-D irregular roughnesses furnishing different topographical features are frequently explored to derive predictive models that may align more closely with practical applications. Given the random characteristics of two-dimensional irregular roughness, their surface features are commonly characterized using statistical metrics. Over the past few decades, researchers have developed numerous empirical correlations to predict k_s or ΔU^+ utilizing various statistical parameters of the roughness. Some of these correlations are summarized in table 2.2. As evident from the table, several roughness statistics – e.g., Sk and ES – are frequently employed across various models, underscoring their significant impact on the roughness drag. Nevertheless, despite the overlap of the employed roughness parameters, the discernible distinctions across the presented formulations can be attributed to the potential variations in their influence on the resulting drag within different sets of roughness. Such non-linear mechanisms of the roughness drag understandably complicates the current objective of developing a universally applicable predictive model.

Table 2.1: Examples of the empirical correlations for k_s or ΔU^+.

Equation	Investigated data	Literature
$k_s = k_{\mathrm{rms}}(1 + 0.5\mathrm{ES})(1 + 0.2SkKu)$	Experiments on naturally occurring roughness	Musker[41]
$k_s = 2.48k_{\mathrm{rms}} \times (1 + Sk)^{2.24}$, $Sk > 0$ $k_s = 2.11k_{\mathrm{rms}}$, $Sk = 0$ $k_s = 2.73k_{\mathrm{rms}} \times (2 + Sk)^{-0.45}$, $Sk < 0$	Experiments on commercial & artificial roughness	Flack *et al.* [19]
$k_s =$ $1.07k_z(1 - e^{-3.5\mathrm{ES}})\times$ $(0.67Sk^2 + 0.93Sk + 1.3)$	Simulations on randomly distributed roughness elements	Forooghi *et al.* [14]
$k_s = 7.3k_a \times \mathrm{ES}^{0.45}$	Simulations on 2-D sinusoidal roughness	Chan *et al.* [20]
$\Delta U^+ =$ $\ln(S_f/S)\left[1 + 0.09\ln(L_x^{cor}/k_z)\right]\times$ $(4k_{\mathrm{rms}}/k_z)^{-0.44}e^{0.074Sk}$	Simulations on filtered realistic roughness scans	Thakkar *et al.* [37]
$k_s =$ $k_{\mathrm{rms}}(-7.65 + -0.0013Sk +$ $2.9\mathrm{ES} + 9.4e^{0.705Sk\mathrm{ES}})$	Experimetal data of 2-D roughness in literature	Abdelaziz *et al.* [102]

Table 2.2: Summary of some exemplary roughness predictive correlations for ΔU^+ or k_s.

Despite numerous attempts, it is important to recognize the ongoing challenge of establishing a universally applicable correlation that can effectively address a broad range of roughness features. This challenge primarily arises from the multifaceted properties of realistic roughness, rendering the necessity of developing of a multivariate correlation model. Developing such a model through conventional analysis for this purpose is a daunting task – if not an impossible one. Confronted with this challenge, researchers have shifted their focus to data-driven approaches. Jouybari*et al.* [21] employed a *multi-layer perceptron* model (MLP) and *Gaussian process regression* model (GPR) to predict k_s values using a set of 17 roughness parameters, including their dot products. Lee *et al.* [22] applied transfer learning to augment available training data by leveraging the "coarse" physical knowledge from empirical equations to pre-train the MLP model. Subsequently, the model was fine-tuned using a set of DNS simulations on irregular roughness, serving as high-fidelity data. Additionally, Sanhueza *et al.* [103] explored the application of *convolutional neural networks* (CNN) to directly resolve raw roughness maps. This sophisticated CNN model was proven

capable of accurately recovering the local drag force as well as local heat transfer distribution, thereby enabling precise prediction of the global skin-friction coefficient and Stanton number. An introduction to the operational principles of neural networks is provided in section 2.4.

2.3 Numerical Methods

While experiments may approach higher Re values that closely resemble real-world applications, numerical investigations are typically limited to lower achievable Reynolds numbers, even with the state-of-the-art high-performance computing system (HPC). However, numerical simulations can provide enhanced capabilities to investigate small-scale turbulent motions and their derivatives, particularly in regions very close to the wall. Ever since Kim *et al.* [104] initially demonstrated the agreement between *direct numerical simulation* (DNS) and experimental measurements for fully developed channel flow, DNS of channel flow has been increasingly carried out at continuously higher Re_τ values, facilitated by rapid advancements in computational power [105, 106].
In this dissertation, DNS are performed as the main means to investigate the fully developed turbulent channel flow subjected to various roughness topographies and Reynolds numbers. The flow is driven by CPG. A schematic representation of the simulation setup is illustrated in Figure 2.5. As depicted, roughness elements are installed on both the upper and lower walls of the channel. Periodic boundary conditions are applied on the boundaries in the streamwise and spanwise directions. The channel full height 2H is measured between the deepest troughs in the roughness on the bottom and top walls.

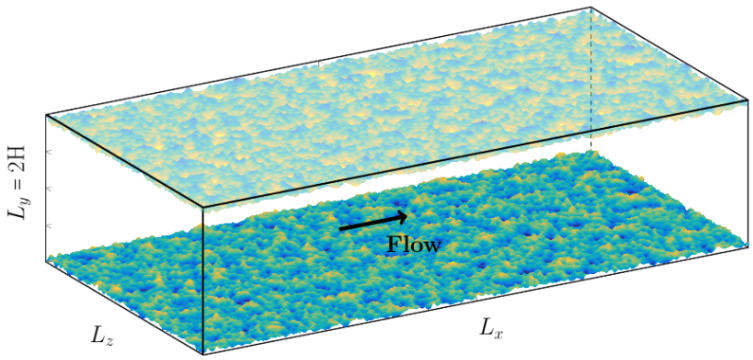


Figure 2.5: Schematic representation of simulation domain with an example pseudo-realistic surface mounted.

2.3.1 Direct Numerical Simulation (DNS)

In this thesis, the mass, momentum, as well as energy conservation equations governing the flow are solved numerically using *direct numerical simulation* (DNS). DNS is performed using the pseudo-spectral solver SIMSON [23], wherein the wall-parallel directions are discretized in Fourier space, while Chebyshev discretization is employed in the wall-normal direction.
In the current DNS, the mass conservation of incompressible flow corresponding to Eqn. 2.14 is solved. The roughness structures are modeled using the *Immersed Boundary Method* (IBM). The utilization of IBM allows retention of Cartesian numerical grids even in the presence of complex boundary conditions. One exemplary representation of the IBM resolved roughness structure is depicted in figure 2.6, sinusoidal stretching of the grid spacing in the wall-normal direction is employed to fit the Chebyshev discretization. In the framework of IBM, the no-slip boundary condition on the roughness is realised by applying force to the fluid within and around the roughness to enforce a velocity of zero. This is achieved by implementing a *Proportional-Integral* (P-I) controller to model the forces acting on the fluid within the numerical grids located inside the roughness and on the roughness-fluid borders, following the approach outlined by Goldstein [24]. The P-I controller of applied force $\mathrm{f}_{i,\mathrm{IBM}}$ is a function of instantaneous and time-integral velocity:

$$\mathrm{f}_{\mathrm{IBM},i}(x_r, y_r, z_r) = \alpha \int_0^t u_i \mathrm{d}t + \beta u_i \tag{2.51}$$

where $(x_r,\ y_r,\ z_r)$ represents grid points inside roughness. α and β are negative constants so that the force acts opposite to the flow. With the incorporation of the IBM forcing term and neglecting gravitational

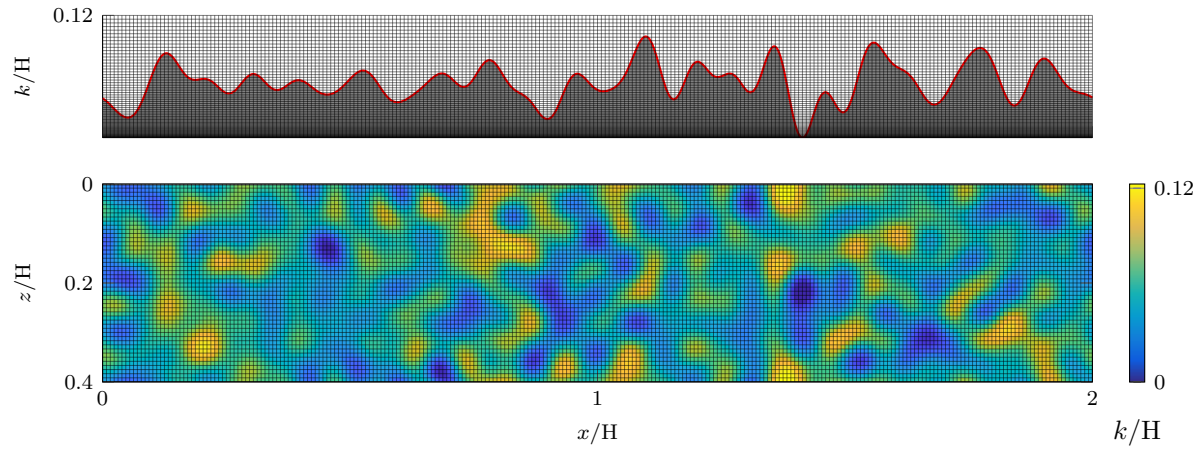


Figure 2.6: Mesh visualization in the near-wall layer with an exemplary rough surface. Upper: mesh in wall-normal direction along $z = 0.2$H, roughness structures represented by IBM are coloured with dark grey. Lower: mesh in wall-parallel direction. Figure adopted from [Yang et al. 2022].

force, the Navier-Stokes equation is modified as:

$$\frac{\partial u_j}{\partial t} + u_i \frac{\partial u_j}{\partial x_i} = -\frac{1}{\rho}\frac{\partial p}{x_j} + \nu \frac{\partial^2 u_j}{\partial x_i \partial x_i} - \frac{1}{\rho}\Pi\delta_{1j} + \mathrm{f}_{\mathrm{IBM},j} \ , \tag{2.52}$$

The constant pressure gradient Π is prescribed to achieve the desired Re_τ. The relation between Π and Re_τ can be derived through combining and reformulating Eqn. 2.43 and 2.44 as

$$\Pi = \frac{\mathrm{Re}_\tau^2 \nu^2 \rho}{(\mathrm{H} - k_{\mathrm{md}})^3} \ . \tag{2.53}$$

As elaborated in the previous sections, the MT thermal boundary condition for temperature field is employed, wherein the constant temperature $\theta_w = 0$ is applied on the lower and upper wall as well as the roughness structures, a source term Q_Θ introduced into the energy conservation equation to achieve the constant temperature gradient in the streamwise direction IBM is used to impose the constant temperature boundary condition $\theta_w = 0$ to the rough surface by introducing external source term to the energy (f_θ) equation. The constant Prandtl number of the fluid is set to $Pr = 0.71$. Consequently, the energy conservation equation with the incorporation of IBM source term f_θ is given as:

$$\frac{\partial \theta}{\partial t} + u_i \frac{\partial \theta}{\partial x_i} = \alpha \frac{\partial^2 \theta}{\partial x_i \partial x_i} + Q_\Theta + \mathrm{f}_\theta, \tag{2.54}$$

where $Q_\Theta = -\rho c_p u(\mathrm{d}\Theta/\mathrm{d}x)$.Based on the energy conservation, the time-averaged wall heat flux q_w for MT thermal boundary condition can be calculated through Eqn. 2.29:

$$q_w = \rho c_p \int_{y=0}^{\mathrm{H}} -U \frac{\mathrm{d}\Theta}{\mathrm{d}x} \mathrm{d}y \ . \tag{2.55}$$

2.3.2 Minimal Channel DNS

Direct numerical simulation of turbulent flow are prohibitively expensive. The computational cost of DNS increases with Reynolds number, approximately proportional to Re^3. This is attributed to the requirement of exhaustive fine mesh of DNS to capture flow dynamics from the bulk scale down to the smallest turbulence scales [63]. Moreover, DNS of turbulent flow over directly resolved roughness structures often requires a finer mesh resolution compared to that of smooth channel flow, owing to the fine structures

of the irregular roughness. Despite the rapid advancements of computational power in recent years, the computational cost associated with conventional DNS remains considerably high, making it impractical to explore a substantial parameter space of irregular roughness topographies.
Jiménez and Moin [107] proposed the concept of *minimal channel* to simulate exclusively the near-wall flow dynamics to spare unnecessary computational power on the flow dynamics in the outer layer. This concept is currently receiving renewed attention in various contexts of understanding wall-bounded turbulence, including in roughness research. The exclusion of outer-layer flow is realized by dramatically reduced spanwise extent of simulation domain. The size of the minimal channel is thus designed to only accommodate near wall turbulent structures. As for the investigation of the skin friction of the rough surface the central quantities is the offset of velocity profile compared to smooth channel, i.e., ΔU^+. Based on the outer-layer similarity hypothesis, this offset can vary in the vicinity of the wall while maintaining constant at a higher level, e.g., in logarithmic layer and outer-layer. Therefore, Chung *et al.* [26] proposed the utilization of minimal channel for roughness research. These authors showed that, thanks to outer-layer similarity of the wall-bounded turbulence [62], ΔU^+ can be accurately predicted by minimal rough channels. In their demonstration, the values of ΔU^+ of 2-D sinusoidal waves are accurately recovered through minimal channel DNS. The concept of minimal channel was extended to reduced streamwise channel length [108], high aspect ratio transverse bars [109] and also for the passive scalar calculations [13]. Based on a series of subsequent research on rough minimal channels, the following criteria for the size of a minimal channel based on simulations with 2-D sinusoidal roughness are established:

$$L_z^+ \geq \max(100, \frac{\tilde{k}^+}{0.4}, \lambda_{\text{sin}}^+) \ , L_x^+ \geq \max(1000, 3L_z^+, \lambda_{\text{sin}}^+) \ , \tag{2.56}$$

where L_z and L_x are the spanwise and streamwise extents of the minimal channel, respectively, λ_{sin} is the wavelength of sinusoidal roughness, $\tilde{k}$ is the characteristic roughness height. Here, the first criterion for the spanwise dimension ($L_z^+ \geq 100$) stems from the fact that the computational domain must accommodate the near-wall cycle of turbulence and is independent of the roughness topography. The third criterion, namely $L_z^+ \geq \lambda_{\text{sin}}^+$, is set to accommodate the roughness structure in the minimal channel, thus this criterion is relevant to the roughness topography. As a consequence of the limited channel size, the physically resolved turbulent structures in the minimal channel are restricted in the near-wall region. It is demonstrated in [26] that the wall-normal extent of this *healthy turbulence* region can be approximated by the channel spanwise width L_z through

$$y_c^+ \approx 0.4 L_z^+ \ . \tag{2.57}$$

Here, y_c is often denoted as the *critical height*, which defines the region up until which the turbulent flow field can be accurately simulated. Therefore, the second criterion for the spanwise dimension ($L_z^+ \geq \tilde{k}^+/0.4$) ensures that roughness can be included in the healthy turbulent zone under the critical height. As can be seen, the spanwise width of the minimal channel only requires multiple hundreds of viscous units, this is in meaningful contrast to the conventional DNS which usually requires multiple thousands of viscous units in the spanwise direction.
Nevertheless, it is important to note that while the studies mentioned above have demonstrated the potential of minimal channels in evaluating roughness-induced drag with an outstandingly enhanced efficiency, a formal extension of this concept to random, irregular roughness is yet to be made. Despite successful applications of the minimal channel concept to random roughness reported in the literature, e.g., in [21, 110], Alves and Sandham [27] recently highlighted the need for a comprehensive assessment of the capability of minimal channels for simulating realistic roughness. In their study, a series of hybrid DNS/URANS simulations on realistic roughness patches in minimal channels and full-span channels are conducted, revealing certain limitations of minimal channels and underscoring the importance of further investigation of this concept, especially in terms of simulating irregular roughness.

2.4 Data-Driven Artificial Neural Networks

As one of the most recognized sub-domains of *Artificial Intelligence* (AI), *machine learning* (ML) methodologies are distinguished by their ability to enhance their performance dynamically through statistical analysis [111]. In contrast to models created through traditional static programming instructions or empirical mathematical expressions, machine learning models are designed to automatically uncover patterns in the data that enable the mapping of multiple input features to their corresponding target values with minimal error [112]. Various categories of machine learning models have been developed over the past few decades, and several of them have found successful applications in practical problems, for instance *support vector machine* [113], *random forest* [114], *kernel regression* [115], etc. Among the broad variation of machine learning models, the *artificial neural network* (ANN) stands out as one of the most popular solutions for machine learning problems, owing to its outstanding flexibility and its capability to detect underlying patterns and non-linear relationships within the data [116]. The original concept of the feed-forward NN was proposed by Rosenblatt [117], drawing inspiration from the functionality of the biological neural network in the human brain. However, this initial concept exhibits limited success due to the challenges associated with efficiently tuning the parameters of the neural network. In later years, the technique of *backpropagation* was introduced [118, 119], which enables the propagation of error gradients backward through the network, thereby allowing for automatic adjustment of the weights assigned to each neuron in ANN in accordance with the network's performance. In recent decades, a new wave of ML-prosperity has been sparked by remarkable achievements across various ML fields. These achievements are facilitated by the utilization of sophisticatedly designed ANN with an unprecedentedly large number of trainable parameters, thanks to the rapid advancements in the computational capacity. Examples including applications in *computer vision* (CV) [120] and *natural language processing* (NLP) [121].
Typically, the ANNs are represented as directed graphs. An illustrative example of a simplified ANN model is presented in figure 2.7 (a). Additionally, a detailed illustration of the signal processing within a single neuron is displayed in figure 2.7 (b). An introduction to the components and operational principles of ANNs is provided in this section.

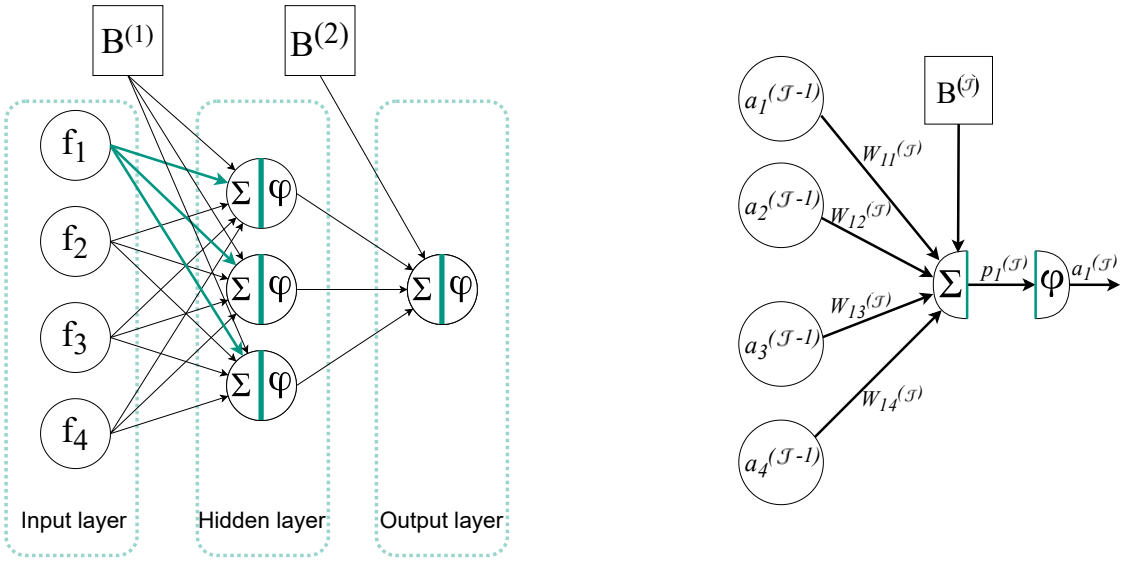


Figure 2.7: (a) Illustration of an artificial neural network. (b) Detailed illustration of the operational principle of a single neuron in layer $\mathcal{J}$.

2.4.1 Elements and Architectures of Neural Networks

An artificial neural network normally consists of a number of *neurons* which are represented by the circular nodes in figure 2.7. The *synapses*, that bridge the signal transmission between neurons, are depicted by the arrows. The transmission of signals along these synapses is weighted by real number weights denoted as

$W_{ij}^{(\mathcal{J})}$. In the context of ANN, the *signals* or *activations* refer to individual real numbers transmitted and processed within the neurons. The activation from the j-th neuron in layer $\mathcal{J}$ is denoted as $a_j^{(\mathcal{J})}$. The set of synapses originating from the same neuron is termed the *axon*. An exemplary axon is highlighted in green in figure 2.7 (a). Following which, the axons are wired together, resulting in a structured – often layered – system, collectively referred to as an artificial neural network.
As depicted in figure 2.7 (a), an ANN consists of three categories of layers, namely the *input layer*, the *hidden layer*, and the *output layer*. The input layer receives sample features. The input features are normally formulated into a 1-D vector $\mathbf{f}_{\text{in}} = \mathrm{f}_{i\in\mathcal{I}}$, where the number of features is correspondent to the number of neurons in the input layer $\mathcal{I}$. These input features are used as initial activation for the input neurons then processed through the hidden layer(s) and finally reach the output layer $\mathcal{O}$ to formulate the model output. The final model output $\mathcal{F}_{o\in\mathcal{O}}$ corresponds to the activation of the o-th neuron in the output layer $\mathcal{O}$, i.e., $\mathcal{F}_{o\in\mathcal{O}} = a_o^{(\mathcal{O})}$ In figure 2.7 (b), the operational principle of a single neuron in the exemplary layer $\mathcal{J}$ is illustrated. The operating procedure of a single neuron can be divided in two stages, namely the summation (Σ) and the activation function(φ). During the summation procedure, activations $a_i^{(\mathcal{J}-1)}$ from the preceding layer $(\mathcal{J}-1)$ are aggregated to form the *pre-activation* p through $p_j^{(\mathcal{J})} = \sum_{i\in\mathcal{J}-1} W_{ji}^{(\mathcal{J})} a_i^{(\mathcal{J}-1)} + B^{(\mathcal{J})}$. Here the notation $p_j^{(\mathcal{J})}$ represents the pre-activation of the j-th neuron at $\mathcal{J}$-th layer. $B^{(\mathcal{J})}$ is the bias term of the current layer $\mathcal{J}$. Subsequently, the pre-activation $p_j^{(\mathcal{J})}$ is fed into the activation function φ to calculate the activation of the neuron through $a_j^{(\mathcal{J})} = \varphi^{(\mathcal{J})}(p_j^{(\mathcal{J})})$. Finally, the activation values $a_j^{(\mathcal{J})}$ are transferred to the following layer $(\mathcal{J}+1)$ or they serve as model output when $\mathcal{J} = \mathcal{O}$. It is noteworthy that the incorporation of an activation function φ is crucial as it introduces non-linearity, thereby enabling even a two-layer ANN to operate as a universal function approximator [122]. By constructing a deep ANN which contains multiple layers with multiple neurons in each layer, a *multi-layer perceptron* model (MLP) is constructed. Finally, following the sequential signal processing procedure depicted in figure 2.7, the outputs of an arbitrary L-layer MLP can be computed through *forward propagation*:

$$\begin{aligned}\mathcal{F}_o(\mathbf{f}_{\text{in}}, \mathbf{W}) = \varphi^{(\mathcal{O})}\Big(\sum_{i\in L-1} W_{oi}^{(\mathcal{O})} \varphi^{(L-1)}\Big(\sum_{i\in L-2} W_{ji}^{(L-1)} \varphi^{(L-2)}\big(\dots \\ \varphi^{(1)}\big(\sum_{i\in I} W_{ji}^{(1)} \mathrm{f}_i + B^{(1)}\big)\dots\big) + B^{(L-2)}\Big) + B^{(L-1)}\Big) + B^{(\mathcal{O})}\Big) \,.\end{aligned} \tag{2.58}$$

In this study, we will omit the subscript $o = 1$ for the current single-output regression problem concerning the prediction of roughness drag.
Historically, the *binary step function* was initially employed as the activation function, inspired by the threshold-like response observed in biological neurons. However, the *vanishing gradient problem* and the lack of differentiability of the binary step function present challenges that can hinder the weights optimization of the model. To facilitate *gradient-based weight optimization*, which will be discussed in the following section, the *Sigmoid function* $\varphi^{(\mathcal{J})}(p_j^{(\mathcal{J})})_{\text{Sigmoid}} = 1/(1+e^{-p_j^{(\mathcal{J})}})$ is introduced as a differentiable alternative to the step function. Additionally, the *rectified linear unit* (ReLU) $\varphi^{(\mathcal{J})}(p_j^{(\mathcal{J})})_{\text{ReLU}} = max\{0, p_j^{(\mathcal{J})}\}$ has been widely adopted in training due to its superior performance in regression problems. However, a potential issue known as *dying ReLU* may arise when a neuron consistently outputs zero during training, rendering it inactive. This problem arises from the negative pre-activation given to the neuron activation function, which prevents weight update due to the zero-gradient of ReLU at $p_j^{(\mathcal{J})} < 0$. To address this issue, the *leaky rectified linear unit* (leaky ReLU) $\varphi^{(\mathcal{J})}(p_j^{(\mathcal{J})})_{\text{Leaky ReLU}} = max\{c \cdot p_j^{(\mathcal{J})}, p_j^{(\mathcal{J})}\}$ is proposed, where $c \leq 1$. Furthermore, it is important to acknowledge the existence of further activation functions. However, for the sake of maintaining simplicity and focus of the present thesis, these additional activation functions are not introduced here.
Finally, it is worth mentioning that while the weights $W_{ij}^{(\mathcal{J})}$ are trainable parameters of the model during the training process, the type of activation functions, the number of hidden layers as well as the number of neurons in each layer are predefined prior to the training step and are kept constant throughout the training process. These predefined parameters are denoted as *hyperparameters*.

2.4.2 Backpropagation

A MLP typically comprises thousands to hundreds of billions of trainable weights. Each weight requires individual tuning during the training process. To efficiently optimize these weights and enhance the model performance, the technique of *backpropagation* is employed. This technique assesses the impact of the variations in each weight on the overall performance of the model. In the context of machine learning, the performance of the model is quantified using a predefined loss function $\mathscr{L}$, which is designed to reflect the discrepancy between the model predictions and the target values. In general, the loss function $\mathscr{L}$ of a regression problem can be formulated as an attempt to maximize the likelihood between predictions and target values. In this context, it is assumed that the target value y_t can be expressed by superposing noise ϵ to the model prediction $\mathcal{F}(\mathbf{f}_{\text{in}}, \mathbf{W})$, i.e., $y_t = \mathcal{F}(\mathbf{f}_{\text{in}}, \mathbf{W}) + \epsilon$. Here, the noise is assumed to follow Gaussian distribution, i.e., $\epsilon \sim f_{\text{G}}(\epsilon|0, \sigma_\epsilon^2)$, where σ_ϵ^2 is the variance of the noise. The probability distribution of y_t can thus be expressed as

$$P(y_t|\mathbf{f}_{\text{in}}, \mathbf{W}, \sigma_\epsilon) = f_{\text{G}}(y_t|\mathcal{F}(\mathbf{f}_{\text{in}}, \mathbf{W}), \sigma_\epsilon^2) = \frac{1}{\sigma_\epsilon\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\mathcal{F}(\mathbf{f}_{\text{in}}, \mathbf{W}) - y_t}{\sigma_\epsilon})^2} \ . \ , \tag{2.59}$$

With a set of N testing samples characterized by the feature sets $\mathbf{F} = \{\mathbf{f}_{\text{in}}^{(1)}, \mathbf{f}_{\text{in}}^{(2)}, ..., \mathbf{f}_{\text{in}}^{(N)}\}$ and the corresponding target values $\mathbf{y}_t = \{y_t^{(1)}, y_t^{(2)}, ..., y_t^{(N)}\}$, the log likelihood $\log \mathbf{L}$ of the predicted values $\mathcal{F}(\mathbf{F}, \mathbf{W})$ with respect to the target values $\mathbf{y}_t$ is given as:

$$\log \mathbf{L} = \log \frac{N}{\sigma_\epsilon\sqrt{2\pi}} - \sum_{i=1}^{N} \left[\frac{(\mathcal{F}(\mathbf{f}_{\text{in}}^{(i)}, \mathbf{W}) - y_t^{(i)})^2}{2\sigma_\epsilon^2} \right] . \tag{2.60}$$

Therefore maximizing the likelihood is equivalent to minimizing $\sum_{i=1}^{N}(\mathcal{F}(\mathbf{f}_{\text{in}}^{(i)}, \mathbf{W}) - y_t^{(i)})^2$. Which can be normalized to *root mean squared error*, RMSE $= \sqrt{\sum_{i=1}^{N}(\mathcal{F}(\mathbf{f}_{\text{in}}^{(i)}, \mathbf{W}) - y_t^{(i)})^2/N}$. Typically, RMSE is used as the standard loss function for training regression ML models. However, it's important to note that the selection of the loss function is not limited to RMSE but can differ based on the specific requirements of the problem and the intended training strategy. Therefore, the choice of the most appropriate loss function – primarily guided by human expertise – is one of the essential factors that influence the ultimate performance of the trained model. Further examples of loss functions include the *mean absolute error* (MAE$= \sum_{i=1}^{N} |\mathcal{F}(\mathbf{f}_{\text{in}}^{(i)}, \mathbf{W}) - y_t^{(i)}|/N$) and the *mean absolute percentage error* (MAPE$= \sum_{i=1}^{N} |\mathcal{F}(\mathbf{f}_{\text{in}}^{(i)}, \mathbf{W}) - y_t^{(i)}|/(y_t^{(i)}N) \times 100$).
Moreover, it is worth mentioning that the definition of $\mathscr{L}$ can be tailored to integrate human expertise [123] or physical insight [124]. For instance, *L2-regularization* is usually appended to the loss function, which represents the summation of the weights squared. The loss function $\mathscr{L}$ based on RMSE with L2-regularization incorporated can be written as:

$$\mathscr{L} = \text{RMSE} + \underbrace{\lambda_{\text{L2}}\Sigma\mathbf{W}^2}_{\text{L2-regularization}} \ , \tag{2.61}$$

where the hyperparameter λ_{L2} is referred to the regularization parameter. The incorporation of the summation of the weights squared is aimed to penalize the concentration of weights on a single neuron, thus mitigating potential over-fitting problem. Apart from this, further possible extensions of $\mathscr{L}$ are proposed by the researchers, such as *L1-regularization* and *elastic net regularization*. In addition to the aforementioned regularization loss terms, physical information – e.g. partial derivative equations – can be incorporated in training process by means of appending additional loss term as well. This aspect of research has drawn growing interest within the domain of physics-informed neural networks (PINN).
The backpropagation is realised by utilizing the chain rule of differentiation to compute the gradient of the loss function $\mathscr{L}$ with respect to the weights $\mathbf{W}$. Firstly, we focus on a single neuron as illustrated in Figure 2.7 (b). Here we assume the loss function $\mathscr{L}$ is based on the current neuron activation a_j. The gradient of $\mathscr{L}$ *w.r.t* W_{ji}, where $i \in \mathcal{J} - 1$ and $j \in \mathcal{J}$, writes

$$\frac{\partial \mathscr{L}}{\partial W_{ji}^{(\mathcal{J})}} = \frac{\partial \mathscr{L}}{\partial a_j^{(\mathcal{J})}} \frac{\partial a_j^{(\mathcal{J})}}{\partial p_j^{(\mathcal{J})}} \frac{\partial p_j^{(\mathcal{J})}}{\partial W_{ji}^{(\mathcal{J})}} \ . \tag{2.62}$$

Wherein, it is noteworthy that

$$\frac{\partial p_j^{(\mathcal{J})}}{\partial W_{ji}^{(\mathcal{J})}} = \frac{\partial(\sum_i W_{ji}^{(\mathcal{J})} a_i^{(\mathcal{J}-1)} + B^{(\mathcal{J})})}{\partial W_{ji}^{(\mathcal{J})}} = a_i^{(\mathcal{J}-1)} \ . \tag{2.63}$$

Therefore, Eqn.2.62 can be simplified as

$$\frac{\partial \mathscr{L}}{\partial W_{ji}^{(\mathcal{J})}} = \frac{\partial \mathscr{L}}{\partial a_j^{(\mathcal{J})}} \frac{\partial a_j^{(\mathcal{J})}}{\partial p_j^{(\mathcal{J})}} a_i^{(\mathcal{J}-1)} = \theta_j^{(\mathcal{J})} a_i^{(\mathcal{J}-1)} \ , \tag{2.64}$$

where $\theta_j^{(\mathcal{J})}$ is defined as

$$\theta_j^{(\mathcal{J})} = \frac{\partial \mathscr{L}}{\partial a_j^{(\mathcal{J})}} \frac{\partial a_j^{(\mathcal{J})}}{\partial p_j^{(\mathcal{J})}} \ . \tag{2.65}$$

In a more general notation, by propagating the gradient from the output layer $\mathcal{O}$ backwards until the input layer, a joint gradient map of $\mathscr{L}$ with respect to $\mathbf{W}$ – denoted as $\nabla_{\mathbf{w}}\mathscr{L}$ – can be generated. Specifically, $\nabla_{\mathbf{w}}\mathscr{L}$ in $\mathcal{J}$-th layer of a L-layer MLP is expressed as nested function

$$\nabla_{\mathbf{w}}\mathscr{L}^{(\mathcal{J})} = \theta_j^{(\mathcal{J})} a_i^{(\mathcal{J}-1)}, \text{with } \theta_j^{(\mathcal{J})} = \begin{cases} \dfrac{\partial \mathscr{L}}{\partial \mathcal{F}} \dfrac{\mathrm{d}\varphi(p^{(\mathcal{O})})}{\mathrm{d}p^{(\mathcal{O})}} a_i^{(L-1)} & \text{in output layer } \mathcal{O} \\ (\sum\limits_l W_{lj}^{(\mathcal{J}+1)} \theta_l^{(\mathcal{J}+1)}) \dfrac{\mathrm{d}\varphi(p_j^{(\mathcal{J})})}{\mathrm{d}p_j^{(\mathcal{J})}} a_i^{(\mathcal{J}-1)} & \mathcal{J} \in \{1, 2, ..., L-1\} \end{cases} \tag{2.66}$$

where the subscript l indicates the index of the neuron in the subsequent layer $\mathcal{J}+1$ linked with the j-th neuron in the current layer $\mathcal{J}$, i is the index of the neuron in the preceding layer $\mathcal{J}-1$. It is worth noting that in the output layer, $\varphi^{(\mathcal{O})}(p^{(\mathcal{O})}) \equiv \mathcal{F}$.

2.4.3 Gradient-Based Weights Optimization

The optimization task for the weights $\mathbf{W}$ can be formulated as

$$\mathbf{W}_{\text{opt}} = \underset{\mathbf{W}}{\operatorname{argmin}} \ \mathscr{L}(\mathcal{F}(\mathbf{F}, \mathbf{W}), \mathbf{y}_t) \ . \tag{2.67}$$

The problem can be iteratively solved by *gradient descent* approach:

$$\mathbf{W}^{(t)} = \mathbf{W}^{(t-1)} - \alpha \nabla_{\mathbf{w}}\mathscr{L} \ , \tag{2.68}$$

here the hyperparameter α is a constant *learning rate*, the superscript t represents the number of optimization iteration. In order to estimate $\nabla_{\mathbf{w}}\mathscr{L}$, the *mini-batch stochastic gradient descent*(SGD) method is utilized. This technique involves computing the average gradient over a subset of N randomly selected samples from the training dataset through

$$\nabla_{\mathbf{w}}\mathscr{L} \approx \frac{1}{N} \sum_{i=1}^{N} \nabla_{\mathbf{w}}\mathscr{L}(\mathcal{F}(\mathbf{f}_{\text{in}}^{(i)}, \mathbf{W}), y_t^{(i)}) \ . \tag{2.69}$$

Nonetheless, it is crucial to acknowledge that the optimization problem 2.67 is often non-convex and can encompass numerous local optima, frequently leading to premature convergence [125]. To address this difficulty, *momentum-based optimization* methods have been developed and extensively utilized in various ML applications [125]. The SGD with momentum can be formulated as

$$v^{(t)} = \eta v^{(t-1)} - \alpha \nabla_{\mathbf{w}}\mathscr{L} \ , \tag{2.70a}$$

$$\mathbf{W}^{(t)} = \mathbf{W}^{(t-1)} + v^{(t)} \ . \tag{2.70b}$$

Here the hyperparameter $\eta \in (0,1)$ is the coefficient of momentum. However, due to the constant momentum coefficient, oscillations can occur during the search of optima, thereby reduce the training efficiency. Bearing this in mind, the *adaptive moment estimation* (Adam) algorithm is proposed [126]:

$$v^{(t)} = \beta_1 v^{(t-1)} - (1-\beta_1)\nabla_{\mathbf{w}}\mathscr{L} \ , \tag{2.71a}$$

$$s^{(t)} = \beta_2 s^{(t-1)} - (1-\beta_2)\nabla_{\mathbf{w}}\mathscr{L}^2 \ , \tag{2.71b}$$

$$\mathbf{W}^{(t)} = \mathbf{W}^{(t-1)} - \eta \frac{v^{(t)}}{\sqrt{s^{(t)}+\varepsilon}}\nabla_{\mathbf{w}}\mathscr{L} \ . \tag{2.71c}$$

Here β_1 and $\beta_2 \in (0,1)$ are the hyperparameters of Adam, $\varepsilon \to 0$ is a small number that prevent numerical instability.

2.4.4 Model Interpretation

Typically, neural networks are considered as "black box" models. This is because the processing of data within the ANN model is accomplished with wired neuron activities that are not comprehensible for humans. However, in light of the widespread success of machine learning models across various applications in recent years, researchers have shifted their focus from mere prediction to understanding the reasoning behind these predictions. The importance of model interpretability is underscored by its profound implications for safety in fields closely related to everyday life, such as autonomous driving [127] or healthcare [128]. On the other hand, understanding a model through interpretations is essential for validating its correctness, identifying biases, and uncovering patterns in the data. Therefore, despite the remarkable advancements of machine learning methodologies in the recent years, the increasing complexity of ML models – e.g., in CNN [120, 129], – has raised ever increasing concern in their interpretability [130]. To fulfil the growing demand for transparency and comprehension in ML models, the field of *explainable AI* (XAI) has emerged. XAI aims to develop models and methodologies that enhance the interpretability and transparency of ML models, thereby assessing the reliability of model predictions and facilitating the discovery of underlying mechanisms of the observed problem from a data-driven perspective.
The interpretability of the model refers to the human's ability to make sense, or derive meaning, from a given model stimulus [131], e.g., from the model output. Model interpretation methods can generally be categorized into two groups, namely global-scope interpretations and local-scope interpretations. The global-scope interpretations, such as *sensitivity analysis* (SA) [132], examine the overall behaviour of the model beyond individual data points. In contrast, local-scope interpretations – also known as instance-based interpretations – such as *layer-wise relevance propagation* (LRP) [133] and *local interpretable model-agnostic explanations* (LIME) [134], focus on explaining the reasons behind the specific predictions on individual instances. In the following content, detailed descriptions of SA and LRP will be provided.

Sensitivity Analysis

As a simple, yet powerful technique of analysing the global feature sensitivity, SA is frequently used in data-driven research, including studies in fluid mechanics [135]. The sensitivity analysis is a model interpretation technique based on the analysis of the gradient of model output *w.r.t.* the model input features. The sensitivity of the model output $\mathcal{F}$ in regard of a set of m features $\mathbf{f}_{\text{in}}^{*} = \{\mathrm{f}_1, \mathrm{f}_2, ..., \mathrm{f}_m\}$ can be expressed as *Jacobian matrix*:

$$J_i(\mathbf{f}_{\text{in}}^{*}) = \left.\frac{\partial \mathcal{F}(\mathbf{f}_{\text{in}}^{*}, \mathbf{W})}{\partial \mathrm{f}_i}\right|_{\mathbf{f}_{\text{in}}^{*}} = \begin{pmatrix} \left.\frac{\partial \mathcal{F}(\mathbf{f}_{\text{in}}^{*}, \mathbf{W})}{\partial \mathrm{f}_1}\right|_{\mathbf{f}_{\text{in}}^{*}} \\ \vdots \\ \left.\frac{\partial \mathcal{F}(\mathbf{f}_{\text{in}}^{*}, \mathbf{W})}{\partial \mathrm{f}_m}\right|_{\mathbf{f}_{\text{in}}^{*}} \end{pmatrix} \ . \tag{2.72}$$

Consequently, the global sensitivity of the output $\mathcal{F}$ with respect to the input feature f_i – denoted as $\mathcal{S}_i$ – is estimated through averaging the Jacobian matrix across the database [22], namely:

$$\mathcal{S}_i = \frac{1}{N}\sum_{n=1}^{N} |J_i(\mathbf{f}_{\text{in}}^{(n)})| = \frac{1}{N}\sum_{n=1}^{N} \left| \left.\frac{\partial \mathcal{F}(\mathbf{f}_{\text{in}}^{(n)}, \mathbf{W})}{\partial \mathrm{f}_i}\right|_{\mathbf{f}_{\text{in}}^{(n)}} \right| \ . \tag{2.73}$$

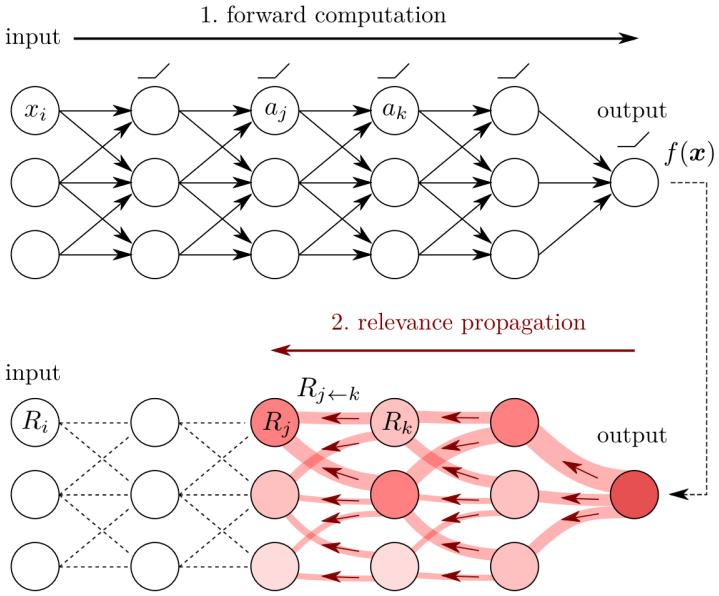


Figure 2.8: Illustration of LRP procedure, Figure adopted from [30].

where N denotes the number of samples in database and $\mathbf{f}_{\text{in}}^{(n)}$ represents the set of features corresponding to the n-th sample in the database.

Layer-Wise Relevance Propagation

The LRP technique was firstly proposed in the field of computer vision in search of pixel-wise explanations of the model in image classification tasks [133]. This technique has previously proven successful in other contexts as a way to interpret decisions of NN models [136, 137]. In clear contrast to SA, LRP is an instance-based technique which focuses on explaining contribution of each input features (or pixels) to the final prediction [30]. A sketch of the procedure of the LRP method is depicted in figure 2.8. As the first step, an instance sample is provided to the trained model. The activation of each hidden neuron as well as the output neuron are calculated through forward propagation (upper row of the figure). Subsequently, the contribution of each neuron in regard of the currently provided sample to the resulting output are calculate through propagating the output activation backward from the output neuron till the input layer [138]. In this manner, each neuron within the input layer receives a share of the *contribution score* (or *relevance*), representing the relevance of each input feature of the given instance sample with respect to the resulting model prediction. This backward propagation process of the contribution score is based on the conservation principle of activation throughout the NN.
According to the above description, the contribution score of neuron j at each layer of the deep neural network can be expressed as

$$R_j = \sum_l \left(\frac{a_j W_{jl}}{\sum_j a_j W_{jl}} \right) R_l, \tag{2.74}$$

where R_l is the contribution score of neuron l in the subsequent layer. In equation 2.74, w and a are the weight and activation of the neuron that are obtained when the model is used to predict one instance (here the k_r for the roughness sample of interest). For better interpretability the contribution score of the output neuron is assigned with the value of unity. As a result, the sum of contribution scores of all inputs must be unity. It is, however, worth mentioning that the contribution score of any single neuron is not bounded. As a result, the input neurons which receives highest scores are regarded as the most relevant features.

3 Characterization of Irregular Roughness in Minimal Channels

Following the overarching objective of developing a data-driven roughness predictive model, the training data is to be produced by simulating turbulent flows over various artificial irregular roughness topographies in minimal channels. Preceding studies have demonstrated the concept of minimal channels on smooth [77] or regular 2-D sinusoidal wave roughness [26]. Despite the improved computational efficiency offered by the concept of minimal channels, a comprehensive evaluation of their capability in the context of irregular roughness is still missing. As reported in the work by Alves Portela *et al.* [27], where the failure of minimal channel can be attributed to inappropriate configuration of minimal channel size to represent a converged roughness statistics. Having this in mind, it is essential to evaluate the accuracy and robustness of this data-acquisition process prior to its application. To this end, the minimal channel DNS outcomes for twelve types of roughness – distinguished by their different PS and p.d.f. configurations – are investigated and compared with those obtained from the conventional full-span DNS counterparts.
In this chapter, a systematic exploration of the proper minimal channel size configurations concerning different roughness properties is carried out. The assessment of the applicability of minimal channel on irregular roughness is performed across a range of Reynolds number spanning from the transitionally rough to the fully rough regime. Furthermore, the statistical convergence as well as the impact of the random nature of the reduced-sized artificial irregular roughness on turbulent flow is investigated. This analysis is performed based on a series of independently generated roughness patches, each exhibiting deterministically distinct realization but sharing identical PS and p.d.f.. In order to shed further light on the interaction of roughness structures and turbulent flow, the distribution of roughness surface force is investigated. The coherence between roughness structures at various length scales and the resultant surface force is evaluated. Finally, utilizing the the same dataset, the impact of different roughness statistical parameters on the roughness skin friction is investigated. Three existing empirical correlations for predicting roughness skin friction based on roughness statistical parameters are assessed.
This chapter is based on the publication *Direct numerical simulation-based characterization of pseudo-random roughness in minimal channels* [Yang et al. 2022] and the contribution to *13th International ERCOFTAC symposium* (ETMM13) titled *study of irregular roughness in minimal channels* [Yang et al. 2021]. Compared to the publication and proceeding, the numbering (of the sections, figures, citations and tables) as well as the mathematical symbols are adapted to comply with the present thesis but no further editing was made.

3.1 Case Description

Utilizing the pseudo-random roughness generation algorithm introduced in section 2.1.2, multiple roughness samples are generated. These roughness are defined by systematically combining varied p.d.f. with self-affine power-law PS of different slope θ_{ps}. Specifically, the power-law PS is expressed as $E_k(\mathbf{q}) = C_0(\|\mathbf{q}\|/q_0)^{\theta_{ps}}$, where $\mathbf{q}$= $(q_x, q_z)^T$ is the wavenumber vector. An overview of the considered configurations of p.d.f. and PS is illustrated in figure 3.1. In figure 3.1 (b,c,d) the PS density normalized by the k_{rms} are compared in pairs. "The upper and lower cutoff wavelengths λ_0 and λ_1 are transformed to cutoff wavenumbers $q_0 = 2\pi/\lambda_0$ and $q_1 = 2\pi/\lambda_1$, which are represented by the red dashed lines in figure 3.1 (b-d) on the left and right sides of the figures respectively. The lower cutoff wavelength [...] $\lambda_1 = 0.08$H is applied for all roughness topographies in the present work. Consequently, with an isotropic roughness and a fixed λ_1, the PS is determined by two remaining parameters, λ_0 and θ_{ps}. In the present work, two values of θ_{ps} ($\theta_{ps} = -1$ and $\theta_{ps} = -2$) are examined, the PS of which are shown in figure 3.1 (b). For the selection of θ_{ps} values we seek similarity to previous works [55, 139, 140]. Moreover, two different upper cutoff wavelengths

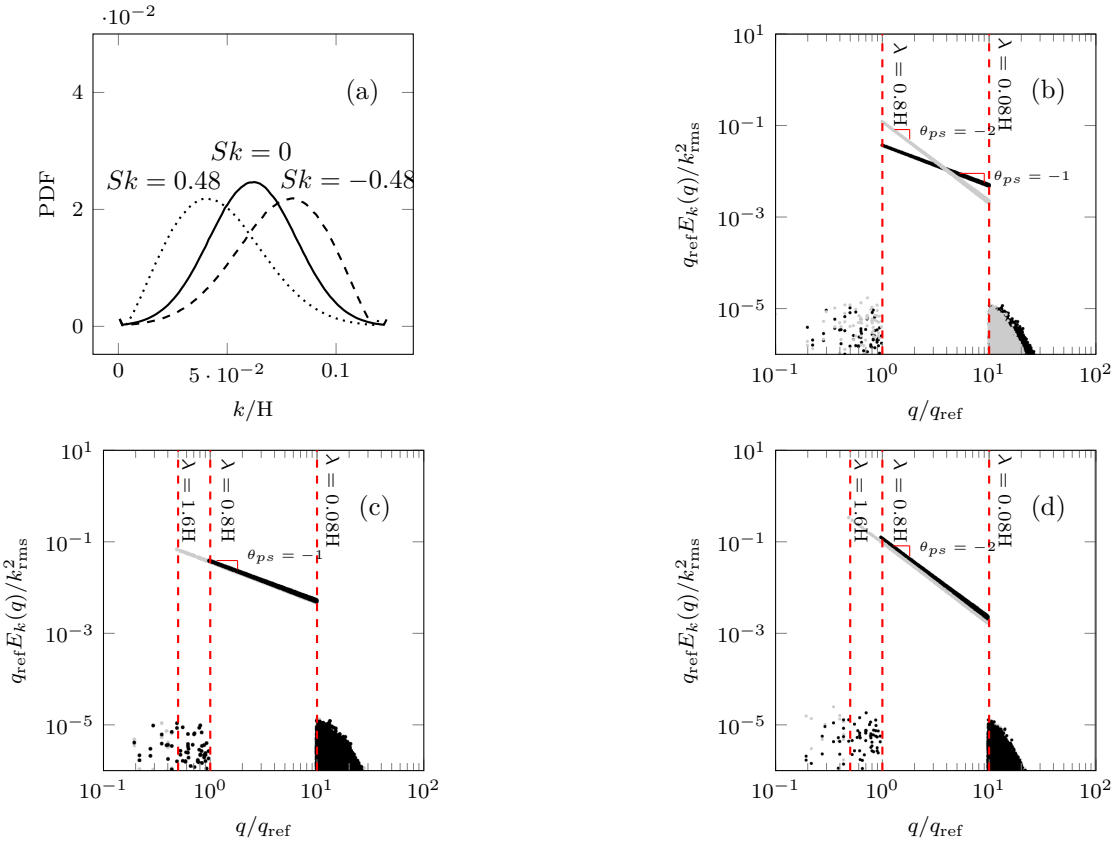


Figure 3.1: Statistical representation of the studied roughness. (a): p.d.f. of roughness, (b): Normalized PS density with different θ_{ps}, (c): Normalized PS density with different λ_0, $\theta_{ps} = -1$, (d): Normalized PS density with different λ_0, $\theta_{ps} = -2$. In (b,c,d) wavenumber q is normalized by the reference wavenumber $q_{\text{ref}} = 2\pi/(0.8\text{H})$. Vertical dashed lines are high-pass filtering and low-pass filtering, corresponding to λ_0 & λ_1 respectively. Figure adapted from [Yang et al. 2022]

($\lambda_0 = 0.8$H and $\lambda_0 = 1.6$H) of the roughness PS are investigated. The PS with $\lambda_0 = 0.8$H and 1.6H with identical slopes θ_{ps} are compared in figure 3.1(c,d), where wavenumber q is normalized by referencing wavenumber $q_{\text{ref}} = 2\pi/(0.8\text{H})$"[Yang et al. 2022]
Three types of p.d.f.s distinguished by their Sk values are examined. Among these types of p.d.f., the zero skewness p.d.f. is obtained by applying Gaussian distribution following Eqn. 2.3. Furthermore, to ensure that the results can be generalized to a wide spectrum of naturally occurring roughness in different applications, a relatively large range of Sk is considered. For this purpose, a Weibull distribution following Eqn. 2.4 is employed to generate skewed p.d.f.. "The Sk value is always adjusted to the value of 0.48. Similar to the Gaussian distribution, the kurtosis of the Weibull distribution is always equal to 3. A negatively skewed p.d.f. is obtained by flipping the p.d.f. of a positively skewed Weibull p.d.f.. Here $Sk = -0.48$ is prescribed. In the present work, the 99% confidence interval of roughness height p.d.f., k_{99}, is used as the characteristic size of roughness, i.e. $k = k_{99}$. This measure is related to the standard deviation of the roughness, and hence can be directly prescribed. We used a fixed value of $k = 0.1$H in all cases. Indeed, k is a statistical measure of the maximum peak-to-trough roughness size, which, unlike the absolute peak-to-trough size k_z, is not deteriorated by extreme events. These types of p.d.f.s are illustrated in figure 3.1. Moreover, in order to avoid extreme high roughness elevations in the simulations, roughness heights outside 1.2 times the 99% confidence interval of the p.d.f. are excluded." [Yang et al. 2022] The current choice of the roughness characteristic height k is made to ensure a substantial scale separation of $\text{H}/k \approx 10$, aiming to attain outer-layer similarity and thereby facilitate the calculation of ΔU^+. Following the above described definition of the roughness height, a peak-to-trough height of $k_z = 0.12$H is achieved for all the considered roughness topographies. "Combining the three types of p.d.f. (with different values of Sk) with the four types of PS (two values of each of θ_{ps} and λ_0 each) 12 different roughness topographies are studied in the present work, which are summarized in table 3.1. Selected patches of all 12 roughness topographies on surfaces with size $2.4\text{H} \times 0.8\text{H}$ are displayed in figure 3.2. Above each roughness map, the 1-D roughness profile at $z = 0.4$H along the streamwise direction is shown.
For each roughness sample, simulations in full-span and minimal channels are carried out. For minimal channels, the spanwise size L_z is the main subject of the study. As the log-layer flow structures are set by the spanwise dimension L_z [141], that is often most critical in terms of reducing the computational cost. The spanwise sizes of the present minimal channels are designed to fulfil the three criteria set by inequalities (Eqn.2.56). [...] We translate the criteria for a realistic roughness by replacing $\tilde{k}$ for the sinusoidal roughness by the characteristic roughness height k for any arbitrary roughness. " [Yang et al. 2022] Furthermore, it is worth mentioning that the third criterion for the spanwise extent of minimal channel delineated in Eqn. 2.56, namely $L_z^+ \geq \lambda_{\text{sin}}^+$, states that the minimal channel should contain the entire sinusoidal roughness roughness structure in order to fully capture its impact on the turbulent flow. " However, for realistic surfaces, a single characteristic wavelength is not naturally determined. A conservative choice for this limit of the channel width can be the largest in-plane length scale, which is λ_0 in the present study. This ensures that all wavelengths present in the roughness topography are included in the spanwise domain. Recalling the aim of reducing the cost of the roughness simulation, however, we seek a less conservative choice, in which some of the larger wavelengths are excluded. Particularly, for simulations of engineering roughness, it is often impracticable to include extremely large roughness scales. To formalize our choice, we denote the largest spanwise wavelength that a domain can accommodate as λ^*, and calculate the portion of surface energy that larger wavelengths contribute to the original roughness as

$$\Phi_c\left(\frac{2\pi}{\lambda^*}\right) = \frac{\int_{2\pi/\lambda^*}^{2\pi/\lambda_1} E_k(q)\mathrm{d}q}{\int_{2\pi/\lambda_0}^{2\pi/\lambda_1} E_k(q)\mathrm{d}q} \ . \tag{3.1}$$

where λ is the discrete wavelength. If the spanwise domain size is λ^*, the simulation resolves a roughness with Φ_c portion of the original surface variance.
In the current research, we examine a choice of spanwise channel size corresponding to half the size of the largest length scale, i.e. $\lambda_0/2$. With the adopted power-law PS, this leads to the values of Φ_c equal to or larger than 90% for all samples under investigation. Hence the third criterion is replaced by $L_z \geq \lambda_{1/2}$, where $\lambda_{1/2} = \lambda_0/2$, which means that the simulations resolve at least 90% of the original surface variance. The new criterion leads to a minimum channel width of $L_z = 0.8$H for half of the investigated roughness topographies (those with $\lambda_0 = 1.6$H) and $L_z = 0.4$H for the other half (those with $\lambda_0 = 0.8$H). We label the

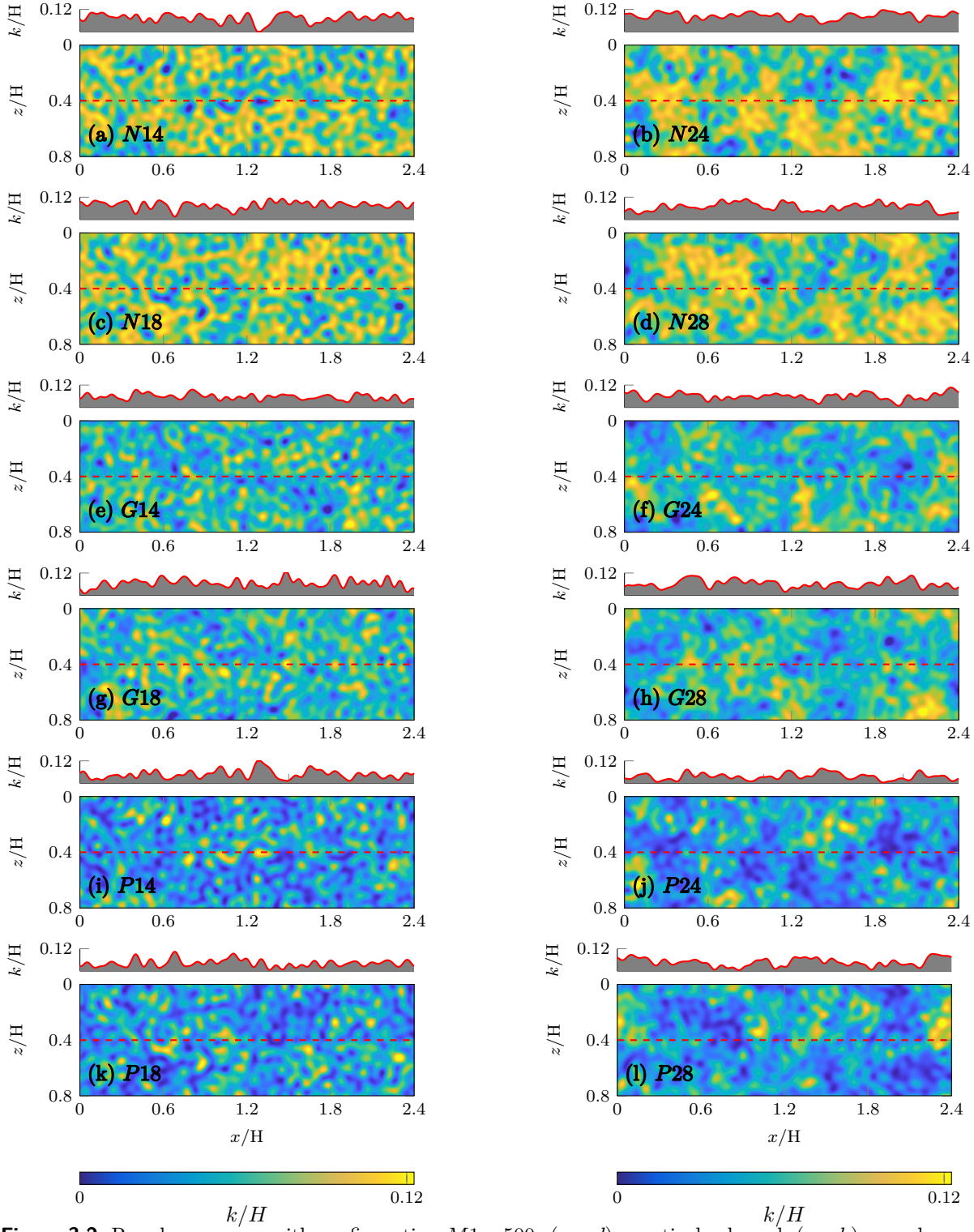


Figure 3.2: Roughness maps with configuration $M1-500$: $(a-d)$ negatively skewed, $(e-h)$ zero skewness and $(i-l)$ positively skewed; for (a, c, e, g, i, k) $\theta_{ps} = -1$ and (b, d, f, h, j, l) $\theta_{ps} = -2$. Colour indicates height. The 1.D roughness profiles at $z = 0.4$H (red dashed line) is shown above each roughness map. Figure adapted from [Yang et al. 2022]

minimal channels with the former (larger) and latter (smaller) spanwise sizes as $M1$ and $M2$, respectively. For roughness samples with $\lambda_0 = 0.8\mathrm{H}$, simulations at both $M1$ and $M2$ channels are carried out. To complete the investigation, some simulations with a further reduced channel size $M3$ ($L_z = 0.3\mathrm{H}$) are carried out. It is worth mentioning that, since $k = 0.1\mathrm{H}$ holds for all topographies, $M1$, $M2$ and $M3$ channels satisfy the $L_z \geq k/0.4$ criterion. For all simulation configurations the streamwise channel size L_x is set according to the inequalities 2.56 once L_z is known."[Yang et al. 2022]
"Simulations are carried out at four different friction Reynolds numbers, $\mathrm{Re}_\tau = 250$, 500, 750 and 1000, at fixed $k/\mathrm{H} = 0.1$, leading to $k^+ = 25$, 50, 75, and 100."[Yang et al. 2022] However, the parametric study on the impacts of roughness topography on both turbulent flow and temperature fields is only conducted at $k^+ = 50$. "Apart from minimal-channel simulations, conventional full-span channel simulations with the size $L_x \times L_y \times L_z = 8\mathrm{H} \times 2\mathrm{H} \times 4\mathrm{H}$, labelled as F, are also carried out for all roughness topographies. For the two highest Reynolds numbers, however, such large simulations are costly. Consequently, for these Reynolds numbers, the largest investigated channels are smaller than the F channel (but still larger than $M1$, $M2$ and $M3$). These channels are labelled as $M0$. Table 3.2 summarizes the details of all simulations carried out for rough channels. "[Yang et al. 2022] Additionally, in order to provide a reference for determining the roughness function ΔU^+ and temperature roughness function $\Delta\Theta^+$, smooth-wall simulations with identical boundary conditions are also performed in $M2$-, $M1$- and F-sized channels at $\mathrm{Re}_\tau = 500$ (not shown in the table). "Overall, each rough-wall simulation case is defined by a combination of roughness topography and simulation configuration (channel size and Reynolds number). Throughout the article, the following naming convention is used to describe the cases:

$$\overbrace{\underbrace{\boxed{G}}_{\text{p.d.f.}}\ \underbrace{\boxed{2}}_{-\theta_{ps}}\ \underbrace{\boxed{4}}_{10\lambda_{1/2}/\mathrm{H}}}^{\text{Topography}} \ \Big|\ \overbrace{\underbrace{\boxed{F}}_{\text{ch. size}} - \underbrace{\boxed{500}}_{10k^+}}^{\text{Simulation configuration}} \quad . \tag{3.2}$$

- The first character indicates the type of p.d.f.: G for Gaussian distribution, P for positively skewed ($Sk \approx 0.48$) and N for negatively skewed ($Sk \approx -0.48$).
- The second digit indicates the PS spectral slope: 1 for $\theta_{ps} = -1$ and 2 for $\theta_{ps} = -2$.
- The third digit represents half of the large cutoff wavelength $\lambda_{1/2}$, with which the channel width is determined: 4 for $\lambda_{1/2} = 0.4\mathrm{H}$ and 8 for $\lambda_{1/2} = 0.8\mathrm{H}$.
- The following character(s) indicate(s) the channel spanwise size: F is for the full channel ($L_z = 4\mathrm{H}$); $M1$ and $M2$ are for the larger ($L_z = 0.8\mathrm{H}$) and smaller ($L_z = 0.4\mathrm{H}$) minimal channels, respectively; $M3$ utilizes the spanwise width $L_z = 0.3\mathrm{H}$; and $M0$ is introduced to the cases in which $k^+ = 75$ and 100 with $L_z = 2\mathrm{H}$ and 1H, respectively
- The last number denotes $10k^+$, or, equivalently, Re_τ."[Yang et al. 2022]

Topography	Sk	θ_{ps}	λ_0/H	k_a/H	k_md/H	k_rms/H	ES	ΔU^+	$\Delta\Theta^+$	d/k
$P14$	0.48	-1	0.8	0.017	0.046	0.0208	0.57	7.33	4.17	0.81
$P18$	0.48	-1	1.6	0.017	0.046	0.0208	0.54	7.23	4.65	0.81
$P24$	0.48	-2	0.8	0.017	0.046	0.0208	0.44	6.99	3.82	0.78
$P28$	0.48	-2	1.6	0.017	0.046	0.0208	0.39	6.57	3.16	0.76
$G14$	0	-1	0.8	0.016	0.061	0.0200	0.54	6.67	3.93	0.95
$G18$	0	-1	1.6	0.016	0.061	0.0200	0.53	6.56	3.92	0.95
$G24$	0	-2	0.8	0.016	0.061	0.0200	0.43	6.30	3.65	0.92
$G28$	0	-2	1.6	0.016	0.061	0.0200	0.37	5.94	3.52	0.90
$N14$	-0.48	-1	0.8	0.017	0.074	0.0208	0.57	6.14	3.87	1.06
$N18$	-0.48	-1	1.6	0.017	0.074	0.0208	0.54	5.84	3.80	1.06
$N24$	-0.48	-2	0.8	0.017	0.074	0.0208	0.44	6.09	3.54	1.03
$N28$	-0.48	-2	1.6	0.017	0.074	0.0208	0.39	5.51	3.32	1.01

Table 3.1: Summary of roughness topographical properties studied in the present chapter; The values of ΔU^+, $\Delta\Theta^+$ and d/k are computed in full channels at $\mathrm{Re}_\tau = 500$. Data adopted from [Yang et al. 2022] and [Yang et al. 2021].

Topographies	Configuration	Re_τ	L_x/H	L_z/H	N_x	N_z	Δ_x^+	Δ_z^+	$\Delta_{y,k}^+$	FTT
$G24$	$M2-250$	250	4.0	0.4	512	48	1.95	2.08	0.88	1200
$G24$	$M1-250$	250	5.0	0.8	576	96	2.17	2.08	0.88	300
$G24$	$F-250$	250	8.0	4.0	900	480	2.22	2.08	0.88	100
$G24$	$M3-500$	500	2.0	0.3	256	48	3.91	3.13	1.74	1000
$**4$ & $G28$	$M2-500$	500	2.0	0.4	256	48	3.91	4.17	1.74	2000
all	$M1-500$	500	2.4	0.8	256	96	4.69	4.17	1.74	500
all	$F-500$	500	8.0	4.0	900	480	4.44	4.17	1.74	80
$G24$	$M2-750$	750	1.4	0.4	288	96	3.65	3.13	2.59	1200
$G24$	$M1-750$	750	2.4	0.8	480	160	3.75	3.75	2.59	300
$G24$	$M0-750$	750	4.0	2.0	640	320	4.69	4.69	2.59	100
$G24$	$M2-1000$	1000	1.2	0.4	288	96	4.17	4.17	3.53	500
$G24$	$M1-1000$	1000	2.4	0.8	576	192	4.17	4.17	3.53	300
$G24$	$M0-1000$	1000	3.0	1.0	720	240	4.17	4.17	3.53	100

Table 3.2: Summary of all simulation cases including roughness topography and simulation configurations. For all cases $L_y/\mathrm{H} = 2$,$N_y = 401$. Moreover, $**4$ indicates all roughness topographies with $\lambda_{1/2} = 0.4\mathrm{H}$, $\Delta_{y,k}^+$ indicates the grid size at the roughness height i.e. at $y = 0.1\mathrm{H}$, flow through time (FTT= TU_b/L_x) for statistics collection duration are shown in the last column, where T is the total integral time. Data adopted from [Yang et al. 2022]

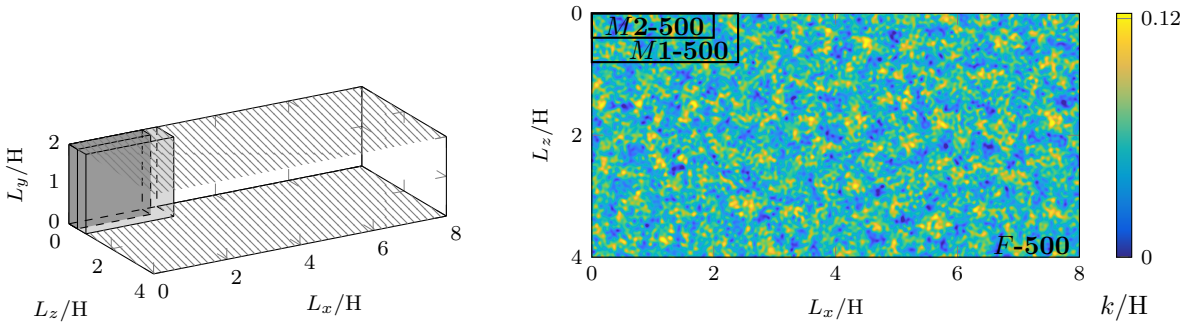


Figure 3.3: Comparison of channel sizes. Left: Schematic simulation domain of $F-500$, $M1-500$ and $M2-500$, hatch pattern represents roughness. Right: Roughness map of $G24F-500$, black frames indicate minimal channels $M1-500$ & $M2-500$. Figure adapted from [Yang et al. 2022].

3.2 Evaluation of Minimal Channel for Irregular Roughness

The capability of the minimal channels DNS in simulating irregular roughness is evaluated in the present section. Particularly to the interest of this dissertation, the performance of minimal channel DNS in terms of recovering ΔU^+ as well as k_s values is assessed. "In order to comprehensively assess the applicability and limits of the minimal channel, first the simulations of the roughness topography $G24$ with variation of channel size at matched $k^+ \approx 50$ are discussed in 3.2.1. This is followed by the results for all different roughness topographies in 3.2.2. As mentioned before, the roughness generation process in the present research has a random nature, where only statistical properties are prescribed. To understand if a mere 'statistical representation' can lead to a unique flow response, simulations are carried out for eight random realizations of the roughness topography $G24$ with $k^+ \approx 50$. In studying different random realizations with identical p.d.f. and PS, we believe we could have contributed to answering this fundamental question. The results are presented in 3.2.3. Finally, in 3.2.4, one roughness topography is studied in a wide range of roughness Reynolds numbers ($k^+ = 25-100$) in order to assess the prediction of the minimal channel in different rough regimes. Here the roughness topography $G24$ is evaluated using minimal channels $M2$ and $M1$ and (pseudo-) full-span channels F and $M0$."[Yang et al. 2022]

3.2.1 Comparison of Minimal-Channel and Full-Sized DNS

"The 3-D schematic representations of minimal channels $M2$ and $M1$ as well as the full-span channel F used for simulations at $k^+ \approx 50$ are shown in figure 3.3 (left); $M3$ is not shown for simplicity. The hatched pattern indicates where the roughness is mounted. Roughness topography $G24$ in the full-size simulation is shown in figure 3.3 (right). For a direct comparison, the boundaries of minimal channels $M1$ and $M2$ are represented by the black frames. The pseudo-random surfaces for each configuration are generated independently. That is, for a specific topography, the surface height map in each simulation is unique, but they all share identical statistical properties."[Yang et al. 2022]
The inner-scaled velocity and temperature profiles obtained from roughness topography configuration $G24$ with $k^+ \approx 50$ are shown in figure 3.4 (left) and figure 3.5 (left), respectively. The solid lines of different colours represent rough channels of different sizes, whereas dashed lines of corresponding colours indicate smooth channels with respective dimensions. The critical heights of minimal channels $M2-500$ and $M1-500$, i.e. $y_c^+ = 0.4 \times L_z^+ = 80$ and 160 are depicted by black vertical dashed lines in both figures, respectively. The minimal channel $M3-500$, along with its critical height $y_c^+ = 60$ is exclusively shown in figure 3.4 (left). "It can be observed from the figure, that the minimal-channel cases $M2-500$ and $M1-500$ successfully reproduce the velocity profile [and temperature profile] of a conventional full-span channel under the critical height y_c. The velocity profiles [and temperature profiles] deviate above the critical height y_c due to the nature of minimal channels. For $G24M3-500$, however, some discrepancy of the profile can be observed even under its critical height. In figure 3.4 (right), the velocity offset profiles for F, $M1$ and $M2$ channels are displayed. The velocity offset profiles are obtained by subtracting the

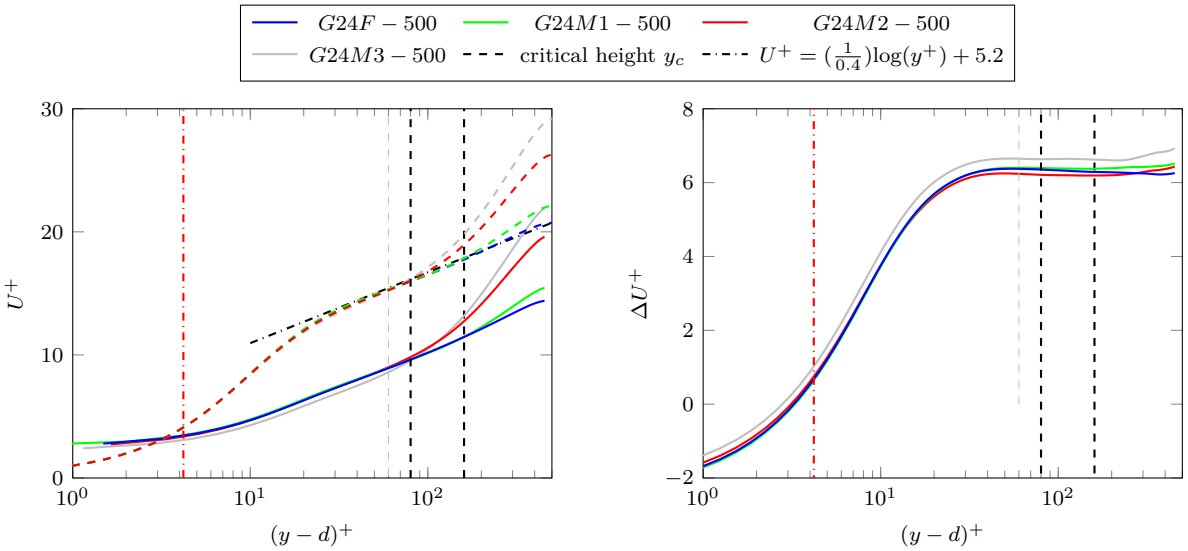


Figure 3.4: Simulation results of roughness type $G24$. Left: mean velocity profiles (——: Rough, - - -: Smooth), right: velocity offset profiles. The red vertical line indicates roughness height measures from the zero-plane displacement $(k-d)^+$. Figure adapted from [Yang et al. 2022].

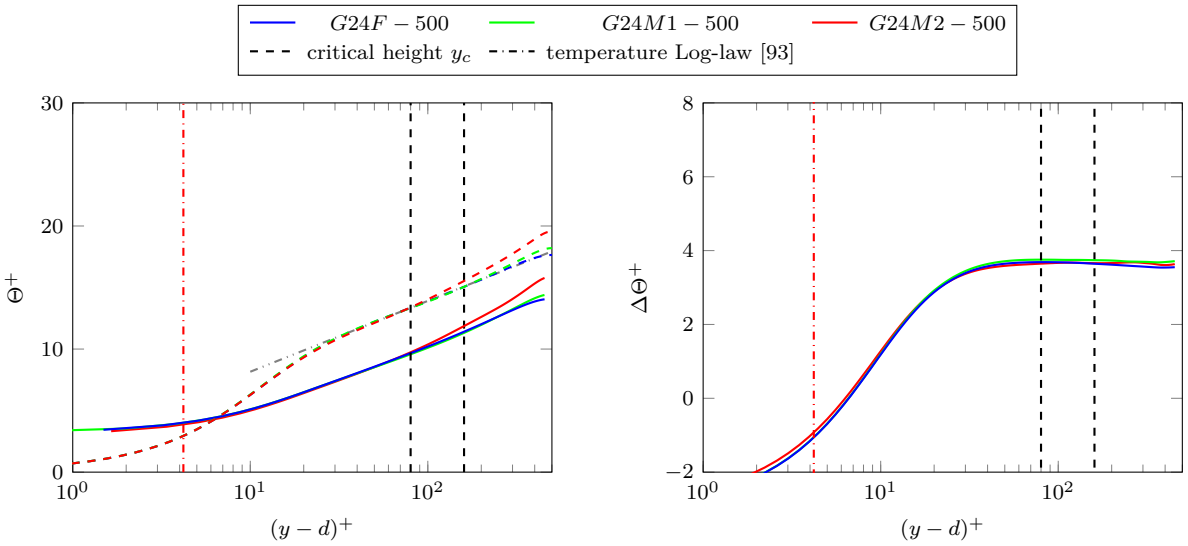


Figure 3.5: Simulation results of roughness type $G24$. Left: mean temperature profiles (——: Rough, - - -: Smooth), right: temperature offset profiles. The red vertical line indicates roughness height measures from the zero-plane displacement $(k-d)^+$. The temperature log-law is based on Kader [93]. Data adopted from [Yang et al. 2021].

rough channel velocity profile from each corresponding smooth channel velocity profiles."[Yang et al. 2022] Additionally, the temperature roughness function are obtained through identical procedure and are displayed in figure 3.5 (right). As can be seen from the figures, the velocity and temperature offset profiles from minimal channels $M1$, $M2$ and full-span channel F on the right panel show excellent agreement. "Consequently, it seems like the velocity [and temperature] offset is not meaningfully affected by the absence of the large wavelengths in the spanwise direction with a small contribution to the roughness height PS. However, this does not hold for channel $M3$, where $\Phi_c(2\pi/L_z) = 82\%$. A similar parameter study performed for roughness topography $G28$ (not shown here) revealed that the velocity offset starts to deviate for channel $M2$ ($\Phi_c(2\pi/L_z) = 84\%$). In both cases, the deviation of the velocity profiles starts when the contribution of excluded large wavelengths in the roughness height spectrum is larger than 10%. In contrast to the minimal domain guidelines Eqn. 2.56 that were based on the single sinusoidal wavelength, the generalization of this idea is not entirely clear for irregular roughness, which contains a wide range of wavelength. Following the findings that the large, undulating scales do not contribute to drag [55], the present paper demonstrates that an *a priori* rule of thumb is 90% of the surface variance. This is checked *a posteriori* by comparing the coherence spectra between surface variance and drag to show that the drag-carrying physics are resolved which will be discussed in 3.3."[Yang et al. 2022]
In the present section, ΔU^+ and $\Delta\Theta$ of the minimal channels are obtained by averaging the mean velocity and temperature offset profiles from their respective critical height y_c^+ to half of channel half height $0.5H^+$, which is equivalent to an inner-scaled height of $0.5\mathrm{Re}_\tau$. Similarly, for full-span channel DNS, the averaging region is chosen as $y^+ = 80-250$ to seek consistency with the minimal channel cases. This gives $\Delta U^+ = 6.3$ and $\Delta\Theta^+ = 3.62$ for case $G24F-500$, $\Delta U^+ = 6.4$ and $\Delta\Theta^+ = 3.72$ for case $G24M1-500$ and $\Delta U^+ = 6.2$ and $\Delta\Theta^+ = 3.68$ for $G24M2-500$

3.2.2 Minimal Channels for Different Roughness Topographies

Applying the same analysis procedure to all considered roughness topographies at $k^+ \approx 50$, the roughness functions ΔU^+ and temperature roughness function $\Delta\Theta^+$ are calculated and illustrated in figure 3.6 (left) and figure 3.7, respectively. In these figures, ΔU^+ and $\Delta\Theta^+$ values predicted by the minimal channels are compared with the prediction by the full-span channels with matched topographical property. It should be mentioned that, for case $G24M2-500$, multiple simulations are carried out for the purpose that will be discussed in 3.2.3. Therefore, ΔU^+ and $\Delta\Theta^+$ values of $G24M2-500$ are the respective mean values over the repeated $G24M2-500$ simulations.
In figure 3.6 (left) and figure 3.7, the $\pm5\%$ disagreement intervals are illustrated by the green shadows around the respective predicted values by the full-span channel DNS(red line). "Another key quantity widely discussed in the framework of roughness studies is the zero-plane displacement d. Similar to ΔU^+, d is often used as input to roughness models, and therefore its prediction is of practical value. The predicted zero-plane displacements d in minimal channels are compared with full-span channels in figure 3.6 (Right). It can be observed that minimal-channel predictions show excellent agreement with those for the conventional full-span channel, the discrepancy being less than 5%."[Yang et al. 2022]
Consistent predictions of ΔU^+ and $\Delta\Theta^+$ indicate the capability of the minimal channels to reproduce these values of the irregular pseudo-realistic roughness even if a certain range of larger wavelengths are excluded. Obviously, in minimal channel simulations, large turbulent structures in the outer layer cannot be resolved; therefore an non-physical wake behaviour is observed in the outer layer of minimal channels mean velocity and temperature profiles (see figure 3.4 and 3.5). However, as the present results suggest, capturing the near wall turbulence in the minimal channel is adequate for the prediction of ΔU^+ and $\Delta\Theta^+$ as the quantification of momentum and heat transfer across the rough wall. "This, obviously, cannot be generalized to all aspects of the turbulent flow."[Yang et al. 2022]

3.2.3 Randomness of Irregular Roughness in Minimal Channel

"In the present work, roughness is generated following a pseudo-random process with prescribed p.d.f. and PS. As a result, individually generated rough surfaces with identical statistics are not deterministically identical."[Yang et al. 2022] This randomness can be a source of uncertainty when pseudo-random roughness is used as a surrogate of realistic roughness (and possibly explaining the scatter observed in figure 3.6

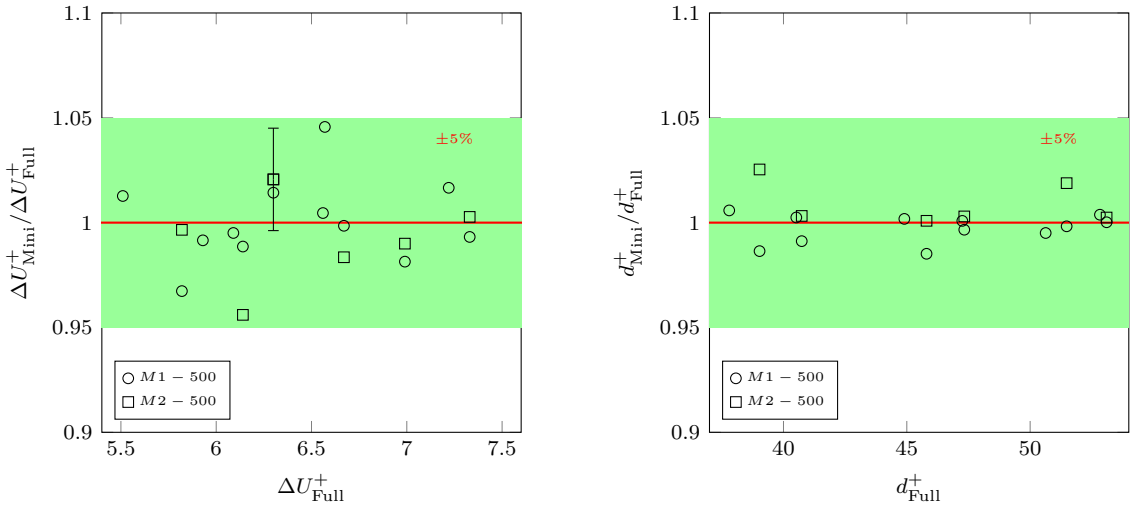


Figure 3.6: Roughness function (left) and zero plane displacement (right) predicted by minimal channels normalized with full-span channel prediction, ○: $M1-500$, □: $M2-500$. The green shading indicates prediction error interval of ±5%. In left panel, the 99% confidence interval for the different simulations of case $G24M2-500$ is shown as an error bar. Figure adapted from [Yang et al. 2022]

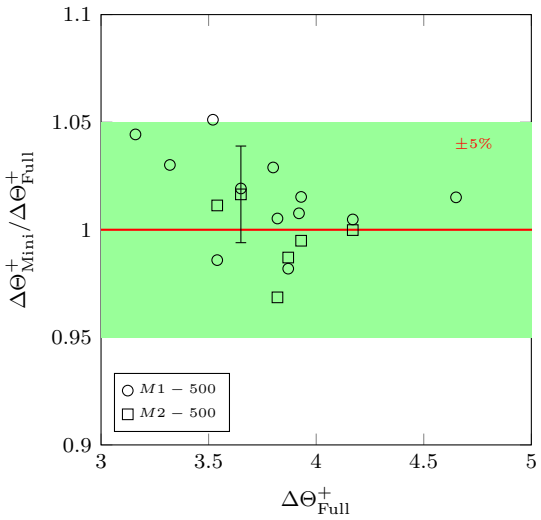


Figure 3.7: Temperature roughness function $\Delta\Theta^+$ predicted by minimal channels normalized with full-span channel prediction, ○: $M1-500$, □: $M2-500$. The green shading indicates prediction error interval of ±5%. In left panel, the 99% confidence interval for the different simulations of case $G24M2-500$ is shown as an error bar. Data adopted from [Yang et al. 2021]

and 3.7). “The pseudo-random roughness generation process can also be considered an imitation of random roughness formation processes in the Nature or industry. Hence it can be used to shed light on whether a statistical representation of stochastic roughness is adequate to predict the flow response. To this end, eight rough surfaces corresponding to $G24$ topography are generated independently. The realization of each randomly generated surface is unique, while the statistical properties are virtually identical. ”[Yang et al. 2022]
The averaged values of the roughness function and temperature roughness function over the eight samples calculated at $k^+ \approx 50$ in minimal channel $M2$ are $\overline{\Delta U^+}_{G24M2-500} = 6.43$ and $\overline{\Delta\Theta^+}_{G24M2-500} = 3.71$, respectively. The 99% uncertainty interval of distribution of ΔU^+ and $\Delta\Theta^+$ across the eight statistically equivalent samples are 0.31 and 0.16 respectively. This averaged value is shown in figure 3.6 along with the uncertainty interval. One can observe that the uncertainty bars well encompasses the ΔU^+ and $\Delta\Theta^+$ in the full channel. “This can be taken as an indication that the minimal-channel prediction can approximately converge to the exact value if the main uncertainty due to randomness is ruled out. Nevertheless, as stated before, the error associated with one random realization is still considerably low. Additionally, one cannot rule out a minor influence due to other factors, e.g. the nature of turbulence in the minimal channel, but the present data suggest those influences to be minor, if present. It is observed in figure 3.6 that the 99% confidence bar lies in the green shaded area – the 5% error range.”[Yang et al. 2022] Overall, it can be demonstrated that DNS in minimal channels is an accurate tool for predicting ΔU^+ $\Delta\Theta^+$ values of irregular roughness, apart from a small discrepancy, which is arguably linked to the effect of randomness.
At this point, it is important to mention that p.d.f. and PS are effectively ‘reduced-order’ characterization of the roughness topography. The current results suggest that this characterization approach can almost uniquely determine the behaviour of the turbulent flow subjected to the studied types of roughness topographies as far as ΔU^+ or $\Delta\Theta^+$ is concerned. However, since this statistical characterization approach does not convey any local details, minor differences in ΔU^+ among surfaces with the same p.d.f. and PS is not an unexpected observation. “In other words, while the generated roughness is controlled in a global sense by p.d.f. and PS, the local distribution of its features can be affected by randomness. For example, the occurrence of clustered or streamwise-aligned roughness peaks can lead to attenuation of drag due to the sheltering effect, which will be discussed in section 3.3.”[Yang et al. 2022]

3.2.4 Minimal Channel in Transitionally and Fully Rough Regimes

“The values of ΔU^+ reported for the simulations with $k^+ \approx 50$ suggest that the flow probably lies in the border between the transitionally and fully rough regimes. In order to ensure that minimal channels deliver acceptable predictions in a wide range of scenarios including both regimes, in this section we study one roughness topography ($G24$) in a range of roughness Reynolds numbers $k^+ \approx 25 - 100$. Both minimal-channel simulations $M2$ and $M1$ as well as large-span channel simulations F and $M0$ are carried out. The simulation set-ups are summarized in table 3.2. Mean velocity profiles are shown in figure 3.8 (left), while roughness functions ΔU^+ plotted against k_s^+ are shown in figure 3.8 (right). The inner-scaled equivalent roughness height k_s^+ on the abscissa of the latter is obtained by scaling the calculated k_s (as explained below) with the viscous length scale δ_ν at different Re_τ. One can observe from figure 3.8 (left) that each velocity profile deviates above the respective critical height $y_c^+ = 0.4 \times L_z^+$, which is not shown for clarity. Based on these velocity profiles, roughness functions are obtained by calculating the velocity difference at each critical height relative to the log-law $U^+ = (1/0.4)\log(y^+) + 5.2$. The equivalent sand-grain roughness k_s is calculated by fitting the roughness function to the asymptotic roughness function in the fully rough regime of Nikuradse sand-grain roughness as shown in figure 3.8 (right). In doing so, we obtain an equivalent sand-grain roughness size of $k_s \approx 1.05k$ for roughness topography $G24$. Notably, the calculated values of the roughness function from both minimal and full channels show an excellent agreement. Furthermore, it can be observed that ΔU^+ asymptotically approaches the fully rough regime at $\Delta U^+ \approx 6$ for both minimal and full channels.”[Yang et al. 2022]

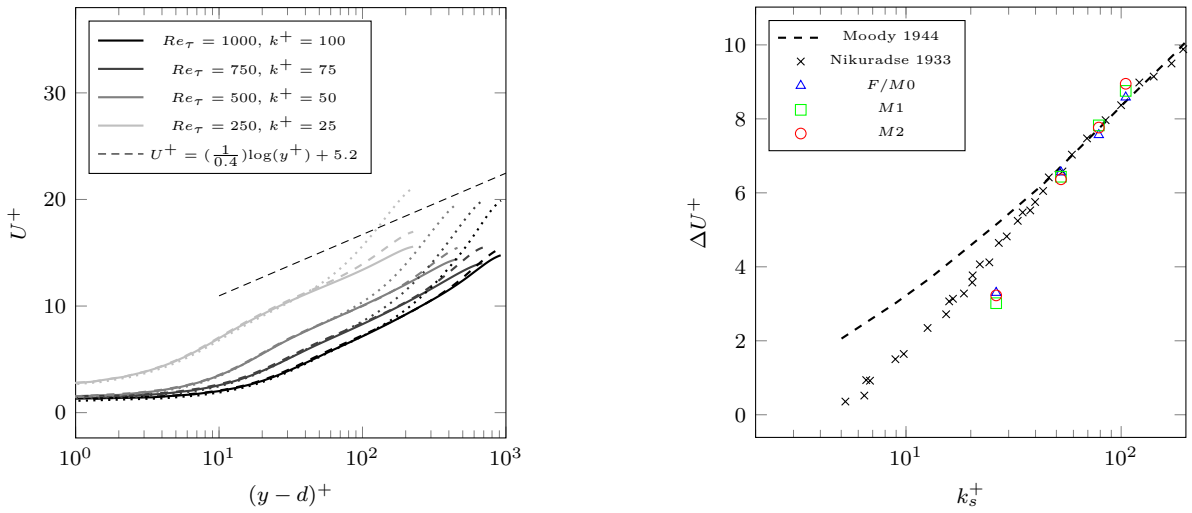


Figure 3.8: Simulations at different k^+ from 25 to 100 at fixed k/H. Left: Mean velocity profiles, line colour gradually changes from grey to black with increasing Re_τ; (——: F and $M0$, - - -: $M1$, ·····: $M2$). Right: Roughness function. Data from Nikuradse's uniform sand-grain roughness and the Colebrook relation provided for industrial pipes are added for comparison. Figure adapted from [Yang et al. 2022]

3.3 Roughness Surface Force

In the preceding sections, the capability of minimal channel in resolving near-wall turbulent flow across various types of irregular roughness is demonstrated. It can be well understood that the successful reproduction of near wall turbulence in minimal channels is predominantly grounded upon a precise reflection of the local flow response to wall structures in minimal channels. Therefore, the distribution of the roughness surface force obtained from minimal channels is deemed plausible and are utilized in the subsequent sections to explore the relationship between different characterizations of roughness structures and the surface skin friction.

"In the present simulations, the IBM introduces a volume force within the solid area imposing zero velocity and hence represents the action of pressure and viscous drag force. One of the advantages of the IBM is the explicit representation of the localized hydrodynamic force exerted by roughness [12], here referred to as 'surface force'.

Previous studies on irregular roughness report that a certain range of roughness scales is dominant in generation of skin friction [55]."[Yang et al. 2022] A deeper insight into the contribution of different roughness scales to the drag force is the aim of this section. To this end, the local forcing map $f_i(x, z)$ is obtained by time-averaging the IBM force field $f_{\mathrm{IBM},i}(x, y, z, t)$ and integrating the force in wall-normal direction y:

$$f_i(x, z) = \frac{1}{T} \int_0^H \int_0^T f_{\mathrm{IBM},i}(x, y, z, t) \mathrm{d}t \mathrm{d}y, \tag{3.3}$$

$f_i(x, z)$ is the force vector component in i-direction. " One should note that, precisely speaking, $\mathbf{f}$ equals force per unit density and wall-projected area. Nevertheless, as we are interested in its trend rather than its absolute value, this quantity always appears in a normalized form; hence, for brevity we refer to it as 'force'.

The visualization of the normalized forcing map for all $M2-500$ cases with their roughness distribution maps are shown in figure 3.9. The entire set of surface force distributions show spanwise-elongated coherent areas of negative forcing.

Furthermore, force distributions in the streamwise direction at $z = 0.2$H are displayed for three cases, $N24M2{-}500$ (negative skewness), $G24M2{-}500$ (Gaussian) and $P24M2{-}500$ (positive skewness), in figure 3.10 along with the surface height functions at the same location. In this figure, the solid blue line represents the normalized negative force profile, $-f_x/f_{x,rms}$, while the dashed red line represents the corresponding normalized roughness profile, k/k_{rms}. Comparing different roughness topographies, it can be observed that the Gaussian surface demonstrates a larger number of extreme force peaks than the surfaces with asymmetric p.d.f.. As expected, the peaks in the surface force are mostly collocated with the peaks in roughness height. A sudden rise in the negative force is expected when the mean flow impinges on the windward side of the roughness element followed by a rapid drop on the leeward side, which is also observed in the figure. Interestingly, the force peaks are much narrower than the surface height peaks, which can be attributed to the separation of flow behind the roughness peak. Another notable observation is that the pronounced force peaks show a much longer streamwise separation than the height peaks. Such an observation can be linked to the sheltering effect, which causes a significant reduction in the flow momentum in the wake of a tall roughness element.

Yang *et al.* [48] investigated the sheltering effect on a surface roughened by rectangular-prism roughness elements and argued that, once the region sheltered by the upstream roughness element covers the neighbouring elements, the surface drag decreases. They suggested that an attenuation parameter for skin friction should incorporate the 'shadowed area'. To provide further insight into the present observations, we apply the wake expansion model proposed by these authors – with some simplification – to the roughness profiles in figure 3.10. According to Yang *et al.* [48], the streamwise slope of the sheltered region from the separation point down to the ground can be calculated from the wake expansion rate by $\tan\theta = C_\theta u_\tau/U_h$, where u_τ is the friction velocity, U_h is the velocity at roughness element height, $C_\theta = 1 - \frac{2}{3}(1 - h/w)$ is the shape parameter of the roughness and h/w denotes the aspect ratio of the rectangular prisms. With the aim to investigate the underlying physical mechanism of the sheltering model, we expand the use of the model – which is obtained based on rectangular-prisms roughness – to more realistic irregular roughness. To approximate the expansion rate of irregular roughness in the present work, we use the double-averaged mean velocity at each roughness peak height as U_h. We also replace h with the characteristic roughness

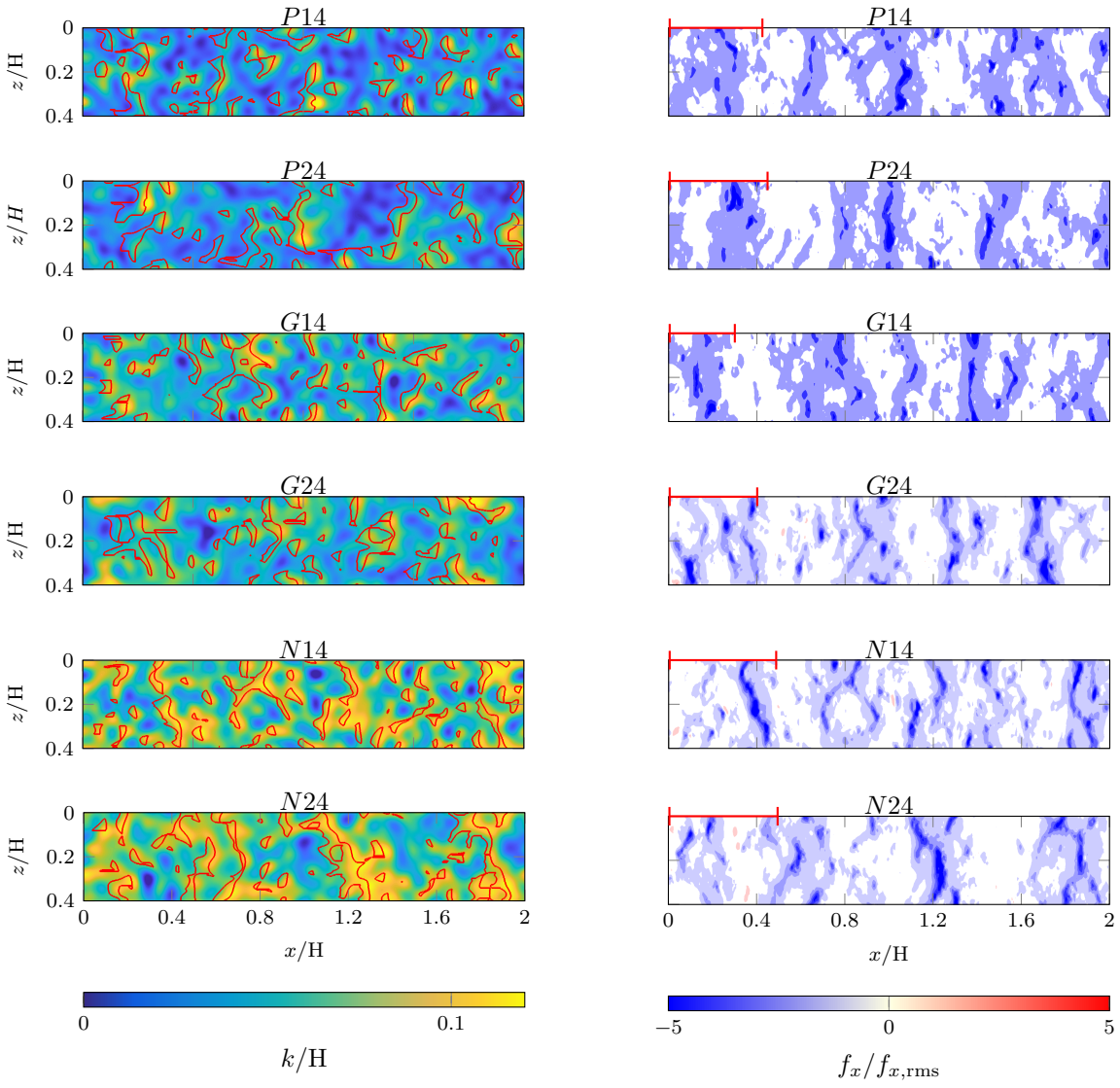


Figure 3.9: Roughness (left column) and surface force distribution (right column) pairs in $M2$. The exposed surface derived from the 1-D sheltering model is marked by red contour lines on the roughness distribution maps. The separation lengths of the force peaks obtained from the autocorrelation are represented by the red bars in the upper left corner of the surface force maps. Figure adapted from [Yang et al. 2022].

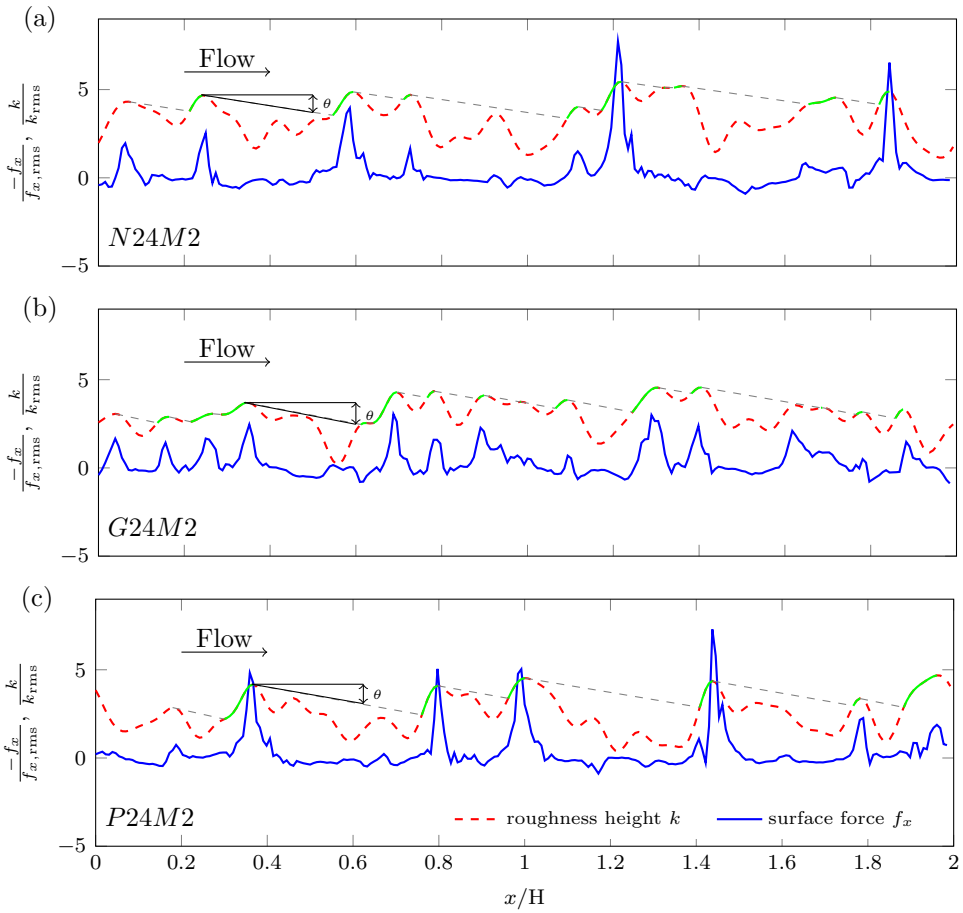


Figure 3.10: Normalized force f_x and roughness distribution profile for $N24M2$ (upper), $G24M2$ (middle) and $P24M2$ (lower) at $z = 0.2$H. The sheltering effect modelled by Yang et al. [48] is illustrated by grey dashed lines, and the unsheltered surfaces are marked by the green profile. Figure adapted from [Yang et al. 2022]

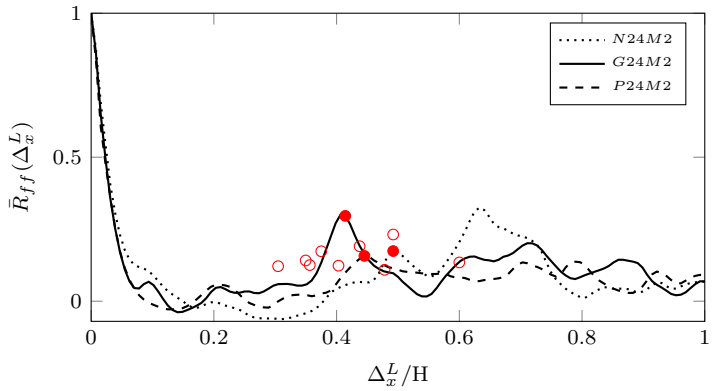


Figure 3.11: Autocorrelation function of the streamwise surface force component as a function of streamwise separation Δ_x^L. The first correlation peaks of the three autocorrelation functions are marked by red filled points, whereas the peaks of the remaining cases studied are marked by open red circles. Figure adapted from [Yang et al. 2022]

height $k_{99} = 0.1$H, and replace w with the spanwise integral length scale of the roughness $L_{k,z} \approx 0.05$H, which will be defined in the following section. Using these values, $C_\theta = 1.7$ is obtained and the resulting shadowed area in figure 3.10 is indicated by the grey dashed lines. In the same figure, the 'exposed' areas on the roughness peaks are highlighted by green lines. It is clear that the extreme force peaks coincide almost exclusively with the green areas, and the shadowed areas rarely produce any significant local force. This can be an indication of the applicability of the wake expansion model to irregular roughness. The 1-D sheltering model is applied to the 2-D roughness distribution in figure 3.9. The exposed roughness surface is outlined by the red contour lines. One can observe that the exposed surface contours match well with the localization of the surface force. Notably, the spanwise-elongated patterns of the surface force and their streamwise separation can be well reproduced with the help of the sheltering model. This finding can also be an indication of the feasibility to predict the local drag with a knowledge of roughness structure, which has a predictive potential for more complex problems, e.g. inhomogeneous and anisotropic roughness structures.
To shed further light on the surface force patterns and sheltering effect, the streamwise autocorrelation functions of the streamwise surface force $\bar{R}_{ff}(\Delta_x)$ for $N24M2$, $G24M2$ and $P24M2$ cases are shown in figure 3.11. The streamwise autocorrelation function of the surface force is defined as

$$\bar{R}_{ff}(\Delta_x) = \frac{1}{L_x L_z} \int_0^{L_z} \int_0^{L_x} \frac{f_x(x,z)}{f_{x,\mathrm{rms}}} \frac{f_x(x+\Delta_x,z)}{f_{x,\mathrm{rms}}} \mathrm{d}x\mathrm{d}z \ . \tag{3.4}$$

A rapid drop of autocorrelation in the vicinity of zero separation – a result of the narrow peaks in the surface force distribution – is observed in figure 3.11. Additionally, a mild but clear positive peak in autocorrelation function, as marked by red circles, is observed at a separation of approximately 0.3-0.6H. This value is likely to be related to the streamwise distance between the force peaks. Locations of the second autocorrelation peaks for the rest of studied cases are marked by hollow red circles in the same figure without showing the autocorrelation functions for better clarity. For visual comparison, we also indicate these values by red bars on the upper left corner of the respective surface force maps in figure 3.9. Here it can be confirmed that, roughly speaking, the lengths of the bars are similar to the separation between the spanwise-elongated areas with high surface force."[Yang et al. 2022]

3.3.1 Correlation Between Surface Force and Roughness Height

"In this section, the link between the streamwise component of the surface force, $f_x(x,z)$, and the roughness height distribution, $k(x,z)$, is analysed by means of the correlation function between the two quantities. The correlation function $R_{kf}(\Delta_x)$ is calculated along the streamwise direction, followed by averaging in the spanwise direction:

$$\bar{R}_{kf}(\Delta_x^L) = \frac{1}{L_x L_z} \int_0^{L_z} \int_0^{L_x} \frac{(k(x,z) - k_{\mathrm{md}})}{k_{\mathrm{rms}}} \frac{\left(f_x(x+\Delta_x^L,z) - \bar{f}_x\right)}{f_{x,\mathrm{rms}}} \mathrm{d}x\mathrm{d}z, \tag{3.5}$$

where the subscript rms and overbar indicate the root mean square value and the mean value, respectively.
The calculated values of the correlation coefficient $\bar{R}_{kf}(\Delta_x^L = 0)$ for the studied topographies are summarized in Table 3.3. A negative sign of the correlation indicates that high roughness peaks are correlated with negative surface force (force directed against the streamwise mean flow), as expected. It is observed that the negatively skewed topographies show lower correlation coefficients, which can be linked to the fact that this type of roughness is rather prone to generation of recirculation and separation zones in the surface valleys and indentations, so the responding force is less localized in those areas. In contrast, for the positively skewed topographies, higher correlation coefficients are observed due to the peak-dominated structures, in which the protruding parts of roughness are directly responsible for the generation of a localized drag force.
The correlation coefficients of the exposed surface with the surface force distribution, $\bar{R}_{kf,\mathrm{exp}}$, and are also estimated and shown in table 3.3. Hereby only the exposed surfaces, i.e. the surface areas that are marked by red contour lines in figure 3.9, are kept on the height map, while the sheltered surfaces are replaced by zero elevation. A noticeable increase in the correlation coefficient can be observed for all cases, especially for negatively skewed roughness, where the correlation is increased by approximately 35%, whereas the

Case	$\bar{R}_{kf}$	$\bar{R}_{kf,\exp}$	Case	$\bar{R}_{kf}$	$\bar{R}_{kf,\exp}$
$P14M2-500$	-0.46	-0.51	$P18M1-500$	-0.48	-0.54
$P24M2-500$	-0.48	-0.56	$P28M1-500$	-0.47	-0.56
$G14M2-500$	-0.44	-0.56	$G18M1-500$	-0.46	-0.56
$G24M2-500$	-0.44	-0.51	$G28M1-500$	-0.45	-0.59
$N14M2-500$	-0.38	-0.54	$N18M1-500$	-0.42	-0.55
$N24M2-500$	-0.41	-0.55	$N28M1-500$	-0.43	-0.57

Table 3.3: Cross-correlation coefficient $\bar{R}_{kf}(\Delta=0)$ and $\bar{R}_{kf,\exp}(\Delta=0)$. Data adopted from [Yang et al. 2022]

increases for Gaussian and positively skewed roughness are 20% and 16%, respectively. This, once again, highlights the importance of the sheltering effect and exposed roughness areas for the generation of surface force.
Furthermore, we calculate the integral length scales of the streamwise surface force (L_f) and the roughness height distribution (L_k). The integral length scales are calculated in a similar way to that proposed by Quadrio *et al.* [142] using the following expression for the integral length scale of roughness height:

$$L_k = \int_{\Delta=0}^{L^{\mathrm{corr}}} \bar{R}_{kk}(\Delta)\mathrm{d}\Delta. \tag{3.6}$$

Here Δ is the separation in either the streamwise or spanwise direction and L^{corr} is the separation at which the autocorrelation function drops below the arbitrary value of 0.2. The integral length scale for surface force L_f is computed in a similar fashion.
The calculated values can be regarded as a scale for the width of the roughness elements or force peaks. Both integral length scales are calculated for different cases and plotted in figure 3.12 as functions of topographical properties, i.e. PS slope θ_{ps} and λ_0. The streamwise integral length scales are plotted on the upper row in figure 3.12, while spanwise integral length scales are plotted on the lower row, grouped by Sk. Square symbols represent L_k while circles represent L_f. It is observed that the surface force has a smaller streamwise integral length scale than the surface height, which is in line with the qualitative observation of the very narrow force peaks in figure 3.10. As expected, topographical parameters show a clear impact on the integral length scales of roughness height. In contrast to this, the streamwise integral length scale of force $L_{f,x}$ does not show a strong sensitivity to the considered roughness variation. A notable observation is that the spanwise integral length scale of force $L_{f,z}$ is more sensitive to the topographical changes than the streamwise length scale. The value of $L_{f,z}$ is comparable to the surface height integral length scale $L_{k,z}$. Based on the limited data points in our dataset, the two quantities show similar trends with θ_{ps} and (to some extent) λ_0: a higher value of $L_{f,z}$ is obtained for the roughness with $\theta_{ps}=-2$ and $\lambda_0 = 1.6$H. Unlike the isotropic behaviour of the roughness height function, clearly illustrated by the comparable streamwise and spanwise integral length scales, the distribution of surface force is observed to be strongly anisotropic. The fact that the integral length scale of surface force is different in the x- and z-directions is the quantitative manifestation of spanwise-elongated coherent areas of surface force observed in figure 3.9."[Yang et al. 2022]

3.3.2 Coherence Function of Surface Force and Roughness Height

"To further understand the correlation between the force and height distributions at different roughness length scales, the coherence function between the two distributions is calculated. The coherence function represents the correlation of the force distribution with the roughness height distribution as a function of wavenumber:

$$\gamma^2_{kf}(q) = \frac{|E_{kf}(q)|^2}{E_f(q)E_k(q)}, \tag{3.7}$$

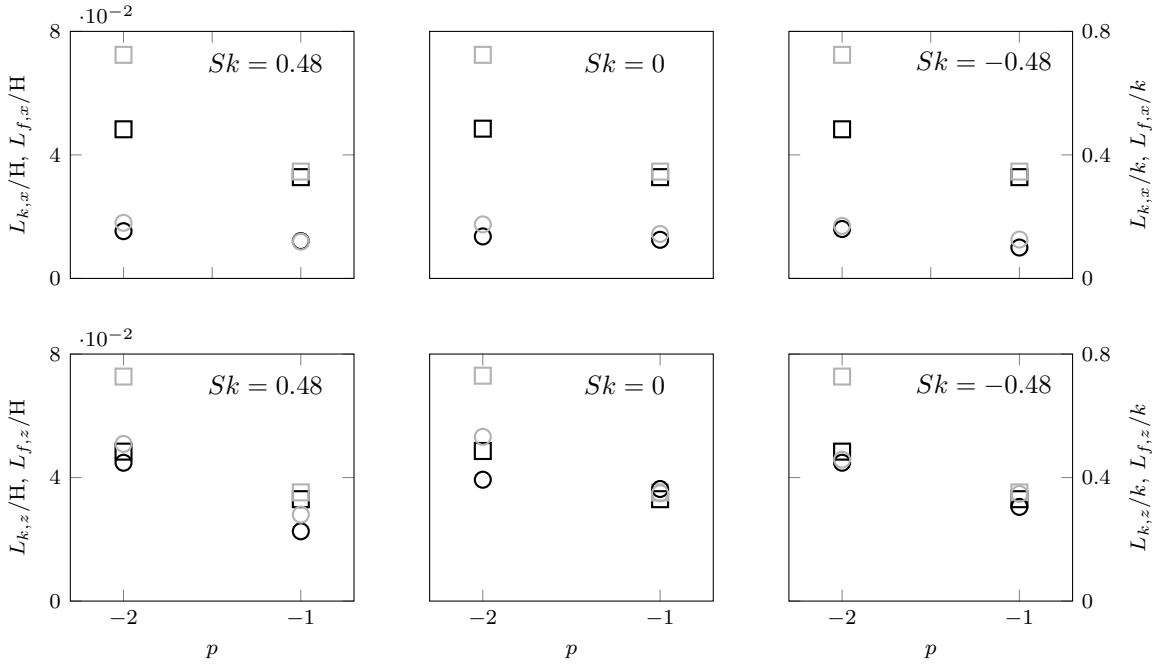


Figure 3.12: Integral length scales L_k and L_f as functions of θ_{ps}, grouped by Sk. Left axis: normalized by H, right axis: normalized by k. The squares represent L_k while circles represent the L_f. Black: $\lambda_0 = 0.8$H, grey: $\lambda_0 = 1.6$H. Upper row: streamwise integral length, lower row: spanwise integral length. Figure adapted from [Yang et al. 2022]

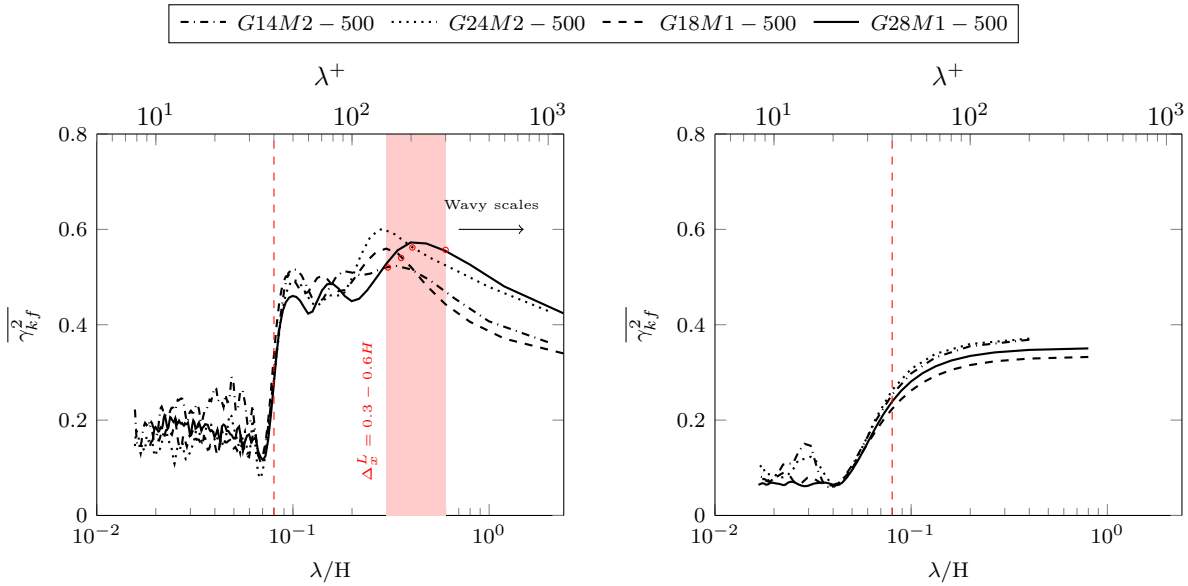


Figure 3.13: Mean coherence function $\overline{\gamma^2_{kf}}$ as a function of $\lambda = 2\pi/q$ normalized by H. Left: streamwise, right: spanwise. $\lambda_1 = 0.08H$ is marked by vertical dashed lines on the left and right side, respectively. λ^+ for the current cases $\mathrm{Re}_\tau \approx 500$ is shown on the upper axis. The length scale of force separation obtained from its autocorrelation functions are marked by red circles on each coherence function, respectively. Figure adapted from [Yang et al. 2022]

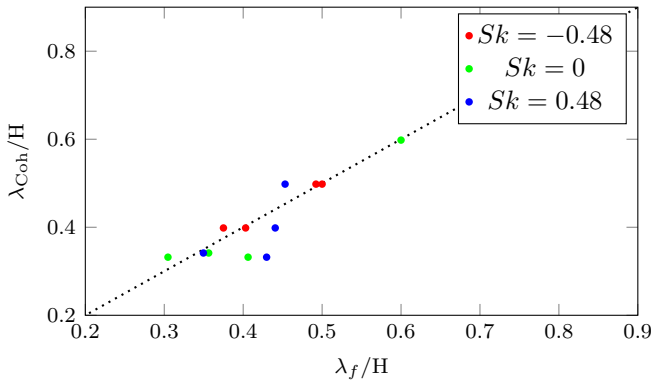


Figure 3.14: The length scale of force peak separation detected from the autocorrelation functions λ_f compared with the coherence dropping wavelength λ_{Coh}. Figure adapted from [Yang et al. 2022]

where $E_{kf}(q)$ represents the cross-PS of roughness topography $k(x,z)$ and force map $f_x(x,z)$, while $E_f(q)$ represents the PS of $f_x(x,z)$. Power spectra are calculated based on 1-D distribution profiles along the streamwise and spanwise directions, and the mean coherence function $\overline{\gamma_{kf}^2}$ is obtained by averaging each parallel signal pair. Figure 3.13 shows the mean coherence function of Gaussian surfaces in (a) streamwise direction and (b) spanwise direction as a function of wavelength $\lambda = 2\pi/q$; the upper axis shows inner-scaled wavelength λ^+ at $\mathrm{Re}_\tau = 500$. It is worth reminding that, for all topographies, the smallest in-plane roughness scale is prescribed to be $\lambda_1 = 0.08\mathrm{H}$. This is clearly related to the observation that coherence functions are significantly smaller below $\lambda \approx 0.08\mathrm{H}$ ($\lambda^+ \approx 40$). Above this threshold, coherence functions increase and retain high values until a certain wavelength, which is roughly at $\lambda \approx (0.3-0.6)\mathrm{H}$ ($\lambda^+ \approx 150-300$) for the studied cases. With further evolution of the coherence function to larger wavelengths, the coherence decreases monotonically. Similar observations are made for negatively and positively skewed roughness. The force becomes less correlated with the surface features at very large scale, or, in other words, very large wavelengths in the streamwise direction do not contribute to the generation of surface force. These length scales might be related to the 'wavy roughness' concept stemming from the observations by Schultz *et al.* [15]. Barros *et al.* [55] stated that these length scales can be filtered out in regard to determining the skin friction.

Notably, the streamwise wavelength at which the coherence starts to drop has a similar value to the streamwise separation distance of the surface force peaks. A comparison between the coherence dropping wavelength λ_{Coh} and the length scale of force peak separation λ_f is conducted in figure 3.14. It can be observed that all data points are clustering around $\lambda_{\mathrm{Coh}} = \lambda_f$ (dotted line), indicating a clear correspondence of these length scales. The significant coherence at relatively small length scales might be linked to the interaction of roughness structure and sheltering. As discussed before, the occurrence of extreme force peaks is strongly determined by the roughness areas that are exposed in the high-momentum flow, or outside the sheltering. Thus, streamwise recurring force peaks caused by sheltering can be found in figure 3.9. Less prominent force peaks can be found between two successive extreme peaks, which contributes to the coherence function at small wavelength. Furthermore, the roughness structures whose length scales are comparable to the distance between two successive extreme peaks, i.e. λ_f, show significance in the coherence function at corresponding wavelength. Beyond this length scale, no larger force peak separation can be found, and the coherence function keeps decreasing into the long-wavelength region. Figure 3.13(b) demonstrates that, unlike the tortuous behaviour exhibited by the streamwise mean coherence function, the spanwise mean coherence function shows a monotonically increasing trend before a plateau at larger wavelengths. "[Yang et al. 2022]

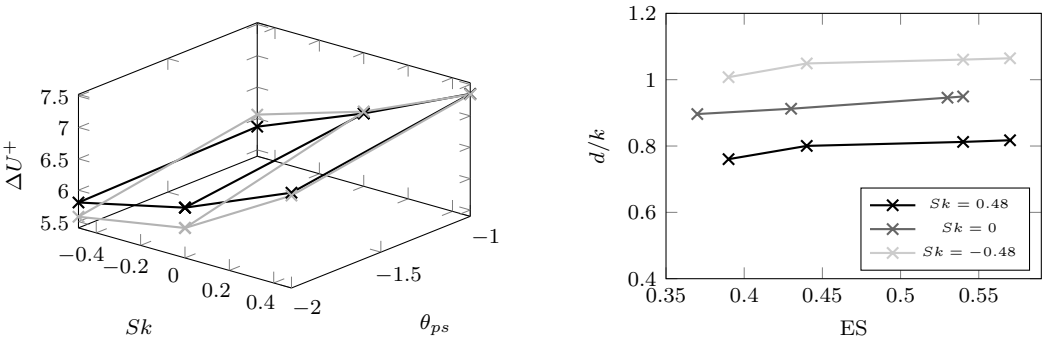


Figure 3.15: Effect of roughness topographical properties to the ΔU^+ prediction (left) from minimal channels, black: $\lambda_0 = 0.8$H, grey: $\lambda_0 = 1.6$H; d predictions (right) as the function of ES, grouped by Sk. Figure adapted from [Yang et al. 2022].

3.4 Impact of Roughness Topography

"In 3.2, we discussed the applicability of minimal-channel concept for characterization of realistic roughness. In doing so, we examined a relatively large number of roughness topographies, which provide a basis for studying the effect of roughness topography on the hydrodynamic properties of the surface, which is discussed in the present section. It has been already shown that p.d.f. and PS can be considered as a reduced representation of roughness topography (almost) uniquely reproducing the hydrodynamic response. The common practice in the literature is, however, to parameterize roughness in terms of a few statistical parameters – an even further reduced representation. Aiming to establish a link with the existing literature, we adopt and examine this approach in section 3.4. We first show the trends of ΔU^+ and zero-plane displacement with some statistical parameters in section 3.4.1. Then some existing roughness correlations are assessed based on the present results in section 3.4.2."[Yang et al. 2022]

3.4.1 Effect of Roughness Statistical Parameters on ΔU^+ and d

"An overview of the roughness function ΔU^+ for all topographies (see 3.1) is plotted in figure 3.15 (left), where the roughness function ΔU^+ is shown as a function of two of the investigated roughness parameters, i.e. skewness Sk and PS slope θ_{ps}. As investigated by Flack *et al.* [19], positively skewed rough surfaces give higher skin friction than non-skewed or negatively skewed roughness. In general, ΔU^+ reaches a higher value with $\theta_{ps} = -1$. As illustrated in figure 3.1(b), at $\theta_{ps} = -2$ larger wavelengths contribute more to the roughness and *vice versa* at $\theta_{ps} = -1$. These findings agree with the study by Barros *et al.* [55] and the results in 3.3 highlighting that larger horizontal length scales contribute less to the hydrodynamic drag. It should be noted that the effective slope of roughness is larger for $\theta_{ps} = -1$ compared to the corresponding cases with $\theta_{ps} = -2$. Furthermore, it can be observed that the surfaces with $\lambda_0 = 1.6$H show stronger sensitivity to the change of PS slope θ_{ps} than those with $\lambda_0 = 0.8$H.
Although the observed trends with selected statistical parameters can, to some extent, justify the use of these parameters in predictive correlations, it is also observed that predicting skin friction based on a few roughness statistics is incomplete. Considering the roughness statistics and the values of roughness function in table 3.1, it is observed that different types of roughness with similar statistical properties, e.g. roughness with same values of Sk and k_{rms} in table 3.1, show meaningful variation of their ΔU^+ values. A better correlation for ΔU^+ is observed when ES is added to the prediction. However, inclusion of ES is still not necessarily expected to yield unique predictions. A simple illustrative example could be that roughness formed by staggered and aligned roughness elements with identical statistics can lead to significantly different values of k_s, as shown in the study by Forooghi *et al.* [14].
As stated before, for a rough surface, the logarithmic layer of the flow is shifted upwards with respect to the bottom plane. Thus, the origin of the wall-normal coordinate cannot be defined *a priori*. As a result, it is necessary to use a physically justified virtual origin for the logarithmic law of the wall. The virtual origin lies above the $y = 0$ plane at a distance equal to the zero-plane displacement d. The value of the

zero-plane displacement d/k following Jackson's method [81] is documented in table 3.1. To summarize the effect of roughness topography on the zero-plane displacement, d/k is plotted as a function of effective slope ES on the right panel of figure 3.15, while the data points are grouped by Sk. Even though ES is not explicitly prescribed in the present work, it is indirectly controlled by the two PS parameters θ_{ps} and λ_0. It can be observed that the value of d/k increases with an increase in ES and a decrease in Sk, while the skewness effect is more dominant."[Yang et al. 2022]

3.4.2 Assessment of Existing Empirical Correlations

"In this section, results from previously introduced topographies at $Re_\tau \approx 500$ are used to assess some of the existing roughness correlations. In fact, existing roughness correlations are developed based on a limited number of data points covering a certain region of the parameter space [8]. In this section, we are particularly interested to shed light on the generalization of these correlations outside their original parameter space, which is key for a correlation to work across the wide range of rough surfaces encountered in different applications.

In the following we assess three relatively recent correlations by Chan *et al.* [20], Forooghi *et al.* [14] and Flack *et al.* [19], each predicting k_s based on a few roughness statistical parameters.""[Yang et al. 2022]

The mathematical formulation of these correlations can be found in table 2.2. The selection of these three empirical correlations is primarily motivated by their representativeness in terms of their incorporated roughness parameters. "These correlations are applicable in the fully rough regime. Figure 3.16 visualizes the selected correlations, where the parameter space covered by the original fitting data of each correlation is represented by a red frame. In each panel the data points from the present work are depicted as symbols. Different symbol colours are used to make a distinction between the data points lying inside and outside the parameter space originally used for development of the correlation. Here the parameter space is expressed in terms of the two widely used parameters Sk and ES."[Yang et al. 2022]

"The required roughness statistics of the present roughness topographies are listed in Table 3.1. First we examine the correlation proposed by Chan *et al.* [20],[...] which is developed based on 2-D sinusoidal roughness. In figure 3.16 (a) the data points from the present work are located in the range of fitting data except for the topographies with $Sk < 0$. Obviously, since Sk is not used as a predictive parameter in this correlation, it returns the same predictions for different values of Sk. Furthermore, two correlations developed by Forooghi *et al.* [14], [...] and by Flack *et al.* [19][...] are examined. The original fitting data, from which these correlations are extracted, are also only partly cover the current data as illustrated in figure 3.16.

In order to directly evaluate the models, k_s^+ predicted by the correlations are normalized and plotted against the full-span DNS results in figure 3.17. The equivalent sand-grain sizes k_s^+ of roughness from the simulations are obtained by fitting the roughness function to the fully-rough asymptote, i.e. Eqn. 2.50. One should recall that the present data points can cover both transitionally and fully rough regimes, while the correlations in question are to be examined for the latter regime. Based on the result in section 3.2.4, and similar to the approach adopted by Jouybari *et al.* [21], an approximate value of $\Delta U^+ \approx 6$ is regarded as the threshold of the fully-rough regime and only the data points in the fully-rough regime are shown in figure 3.17. It has to be mentioned that this value is an approximate criterion to estimate the roughness regime. The exact ΔU^+ criterion for different types of roughness can only be achieved through comprehensive experiments.

In figure 3.17 different error intervals in prediction of k_s^+ are illustrated by various green shades. In general, a similar range of error can be observed among the predictions of all correlations in the figure where a limited number of data points lie outside the 30% k_s error area. The fact that none of the correlations are able to perfectly reproduce the effect of topography on k_s – as already pointed out by other authors [19] – can be acknowledged in figure 3.17. Among all the correlations, those by Chan *et al.* and Flack *et al.* incorporate less geometrical information by taking one parameter related to the topography each, namely the effective slope of roughness in the former and the skewness in the latter. Forooghi *et al.* combined both approaches. While the latter correlation delivers better predictions for some data points, an obviously superior accuracy cannot be established. One notable observation from all the correlations is that their prediction does not deteriorate particularly for data outside their original fitting range. This can be an indication that either of the models can be used with a similar level of reliability in a wider parameter

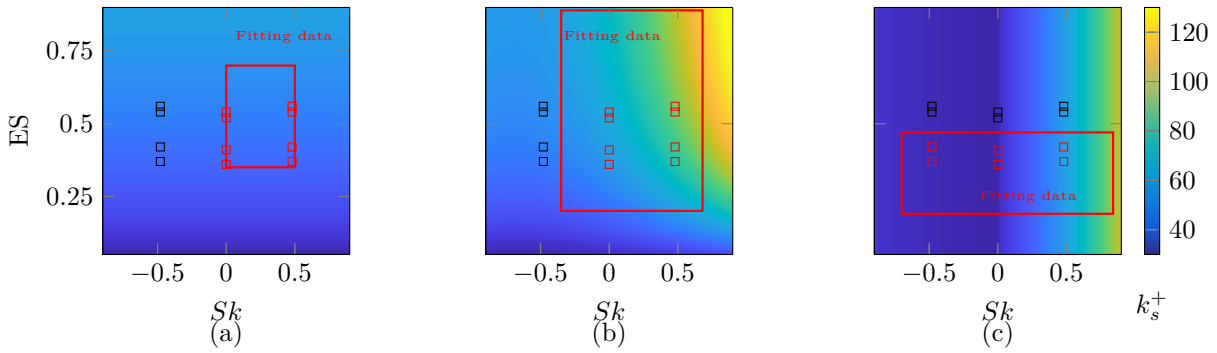


Figure 3.16: Predictive correlations for k_s^+. Squares indicate data points from the present work while the red frame represents the fitting data from the literature. (a): Correlation by Chan *et al.* [20], (b): Correlation by Forooghi *et al.* [14], (c): Correlation by Flack *et al.* [19]. Figure adapted from [Yang et al. 2022]

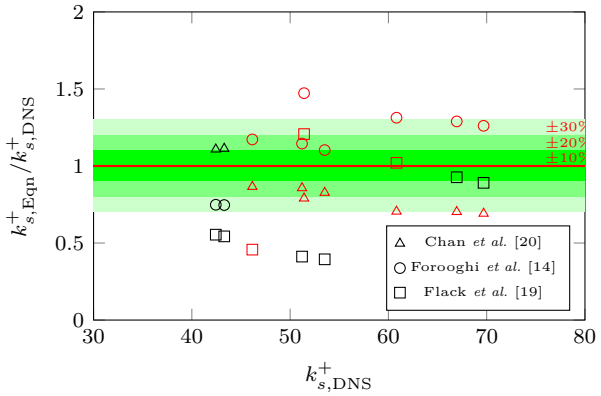


Figure 3.17: Prediction of k_s^+ compared to DNS. Red symbols indicate the present data points located within the fitting dataset. The 10%, 20% and 30% error intervals are represented by green shading. Figure adapted from [Yang et al. 2022]

space than it was originally designed for.
To summarize, this section highlights the need for a more general model, e.g. by taking advantage of advances in data-driven methods. For the future development of roughness models/correlations, it is suggested by the present work that roughness statistics that contain both the height distribution and horizontal scales of roughness need to be incorporated for generalizable predictions."[Yang et al. 2022]

3.5 Summary

In this chapter, the viability of an efficient data acquisition framework for constructing a massive ML database is examined. Within this framework, roughness topographies are mathematically generated using the method proposed by Pérez-Ràfols and Almqvist [17], enabling precise prescription of p.d.f. and PS of the roughness. Subsequently, DNS is carried out for turbulent flow over the generated roughness in plane channels with reduced stream- and spanwise extents – referred to as minimal channels. The obtained results are then compared with those derived from conventional full-channel DNS.
Specifically, simulations are run for 12 types of roughness topographies at $k^+ = 50$. These different types of roughness topographies are characterized and manipulated in terms of three parameters, namely the skewness Sk, the PS power-law slope θ_{ps} and the largest in-plane roughness wavelength λ_0. The turbulent flow over a selected topography is simulated across $k^+ = 25 - 100$, covering the transitionally and the fully rough regimes. For all cases, both full-channel DNS and one or more minimal-channel DNS are carried out. It is systematically demonstrated, that the values of ΔU^+ and $\Delta\Theta^+$ for the examined types of irregular roughness exhibiting stochastic characteristics can be recovered from minimal channel with a margin of error mostly within $\pm 5\%$. "This can be achieved as long as the minimal channel dimensions follow a relaxed version of the criteria suggested for regular sinusoidal roughness by previous authors [26, 108]. The relaxation concerns the condition that the channel should contain all horizontal scales of roughness. Current data suggest that accurate prediction can be achieved as long as the size of channel is large enough to accommodate more than 90% of original roughness height spectral energy based on the area under 2-D PS. This finding is particularly relevant in DNS-based characterization of realistic rough surfaces that may contain very large wavelengths with limited contribution to the root mean square roughness height." [Yang et al. 2022]
"For one topography, multiple rough surfaces are generated. Owing to the random nature of the roughness generation process, these surfaces are deterministically different while being statistically identical"[Yang et al. 2022]. Simulations are performed for these surfaces at $k^+ = 50$ and slightly scattered values of ΔU^+ and $\Delta\Theta^+$ are obtained. Notably, these values obtained from the full channel reside within 99% uncertainty interval of these scattered predictions. "The results indicate that, at fixed p.d.f. and PS, randomness in roughness generation can lead to a small uncertainty, which is also likely to be the origin of the observed $\pm 5\%$ discrepancy between the predictions of the minimal and full channels. This can be an indication that one can consider a combination of p.d.f. and PS as a reduced-order representation of roughness topography leading to nearly unique dynamic flow responses."[Yang et al. 2022]
"In addition to global flow properties, local surface forces for different types of roughness are calculated and their correlations with respective roughness height functions are studied. It is observed that not all roughness height peaks generate force peaks. Applying the sheltering model proposed by Yang *et al.* [48] with some assumptions, we are able to show that only 'exposed' (unsheltered) roughness peaks generate prominent peaks in the surface force. Notably, the spanwise-elongated patterns of the surface force and their streamwise separation can be well reproduced with the help of the sheltering model. This can be taken as a clear indication of the relevance of the sheltering effect in flow over irregular roughness – e.g. for complex terrains.
To shed light on the contributions of different roughness scales to global drag, we also studied the coherence function of roughness height and surface force power spectra as a function of streamwise and spectral scales. In the streamwise direction, it was observed that coherence starts dropping beyond a certain length. This observation can be interpreted as a smaller contribution of very large roughness wavelengths to the drag force. These large roughness length scales might be related to the 'wavy roughness' concept stemming from the previous studies [15].
Our analysis of surface force reveals certain previously unattended facts about roughness-induced skin friction, e.g. reduced coherence between friction and roughness height at large scales. Notably, the wavelength at which the coherence starts dropping is shown to be related to the separation between the peaks of surface force, which is linked to the sheltering effect itself. Unlike the streamwise direction, the coherence function does not drop in the spanwise direction for the cases studied in this paper."[Yang et al. 2022]
"As stated above, the present results suggest that an accurate yet computationally economical framework for the characterization of irregular, realistic rough surfaces is in hand. Such a framework can, for example,

be a basis for generation of the large databases required for future 'data-driven' roughness correlations. "[Yang et al. 2022] Furthermore, leveraging the results from the 12 simulated roughness topographies, we conducted analysis to study dependence between ΔU^+ as well as the zero-plane displacement d on some key statistical parameters. "Notably, it was shown that the normalized zero-plane displacement d/k is most sensitive to the skewness of the roughness distribution (larger at smaller values of skewness), and it also mildly increases with effective slope."[Yang et al. 2022]
Finally, a number of widely cited roughness correlations from the literature are assessed. These include the correlations developed by Chan *et al.* [20], Flack *et al.* [19], and Forooghi *et al.* [14]. "While some correlations show a certain level of success in reproducing the roughness function or equivalent sand-grain roughness compared to the DNS results (see figure 3.17), there is an obvious need for improvement. An interesting observation is that none of the assessed correlations show a dramatic loss of accuracy when used outside the parameter space of its original fitting data. However, even the most successful correlations can reproduce the DNS data only within $\pm 30\%$ accuracy. This can arguably be the ground for a paradigm shift in the development of future roughness correlations" [Yang et al. 2022]. As previously discussed and will be further elucidated in the subsequent chapters, employing a data-driven approach that is capable of capturing the stochastic multi-scale characteristics of roughness and its interaction with near-wall turbulence presents a promising path towards addressing this challenge.

4 DNS-Based Evaluation of Artificial Surrogates for Realistic Rough Surfaces

In pursuit of a machine learning model capable of accurately predicting any arbitrary roughness encountered in practical scenarios, it is crucial to derive the model using roughness samples that closely resemble naturally-occurring roughness topographical characteristics. Bearing this in mind, we assess the capability of the chosen roughness generation algorithm in terms of replicating realistic roughness statistics, and simultaneously maintaining consistent hydrodynamic and thermal characteristics. In line with the selected roughness generation algorithm, these roughness surrogates are generated by emulating realistic p.d.f. and PS. It is noteworthy that the artificially generated roughness are subject to the randomness introduced by the generation algorithm. Consequently, each rough surface possesses unique realizations while sharing identical statistics. As demonstrated in the previous chapter, the randomness introduced by the present pseudo-random roughness exhibits negligible impact on the flow. Hence, the reproduced roughness surrogates in the present work are considered as instances representing the roughness with the target p.d.f. and PS.

In this chapter, multiple realistic roughness topographies originating from various practical applications are incorporated. "Special attention is paid to ice accretion roughness, as an example of realistic roughness in which the morphology can be very different from one case to the other. Atmospheric ice accretion can be regarded as a type of surface degradation leading to serious deterioration of equipment performance. For instance, the accumulation of ice on wind turbines can lead to the reduction of their power production [143]. The aerodynamic performance, and consequently the flight safety, of an aircraft can be lowered due to ice accretion, making deep understanding of roughness effect imperative. In an aircraft, the intake lip of the engine is particularly exposed to icing events during take-off or climb [6]."[Yang et al. 2023b] Laboratory experiments are carried out to acquire natural ice accretion surfaces on an aero-engine nacelle [144]. Additionally, the respective surrogates of a combustion chamber roughness sample that is available in our institute– originally reported in [4] – and a P60 sandpaper roughness scan[1] are reproduced. Subsequently, the flow response to both the actual and reproduced samples through DNS in a turbulent channel flow at otherwise similar conditions are compared. Additionally, the impact of the roughness p.d.f. on the resulting hydrodynamic and thermal properties is illustrated by substituting the realistic p.d.f. with a Gaussian p.d.f. In the end, laboratory experimental results on the P60 sandpaper is adopted for comparison with the simulation results on its numerical counterpart. This comparative analysis is intended to illustrate the capability of the current numerical procedure to faithfully replicate the actual experimental measurements.

In this chapter, we firstly outline the methodology for the reproduction of realistic roughness surrogates. The analysis initiates with an examination of mean velocity and temperature profiles over sandpaper and combustion chamber roughness surrogates using full-channel DNS and minimal-channel DNS. The artificial surrogates for the ice accretion roughness are exclusively investigated in minimal channels. Subsequently, higher-order statistics, such as Reynolds stresses, over the realistic surfaces and their reproduced surrogates are compared. Finally, the challenges and methodologies associated with numerically reproducing laboratory roughness experiments are outlined. The mean velocity profiles obtained through DNS are compared with the measurements obtained from the *hot-wire anemometry* (HWA).

This chapter is based on the publications *A comparison of hydrodynamic and thermal properties of artificially generated against realistic rough surfaces* [33] and *Flow resistance over heterogeneous roughness made of spanwise-alternating sandpaper strips* [34]. Compared to the publications, numbering of the figures, equations and tables, as well as citation style have been adapted to be consistent with this thesis.

[1]The sandpaper roughness topography is acquired using white light interferometry by Dr.-Ing Lars von Deyn.

4.1 Numerical Procedure

4.1.1 Realistic Roughness Acquisition

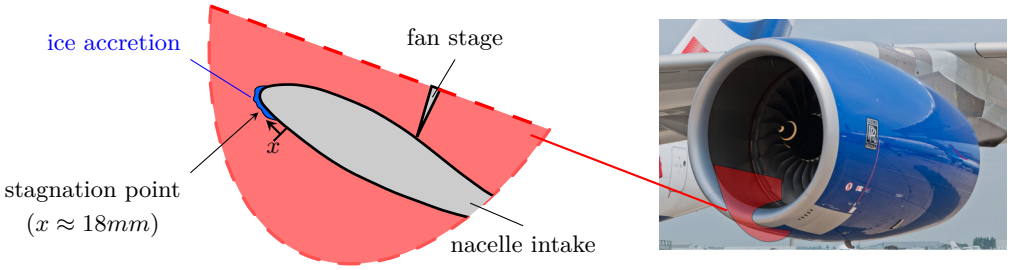


Figure 4.1: Sketch of the ice accretion and location of the stagnation line ($x \approx 18mm$) on the aero-engine nacelle. Figure adapted from [Yang et al. 2023b].

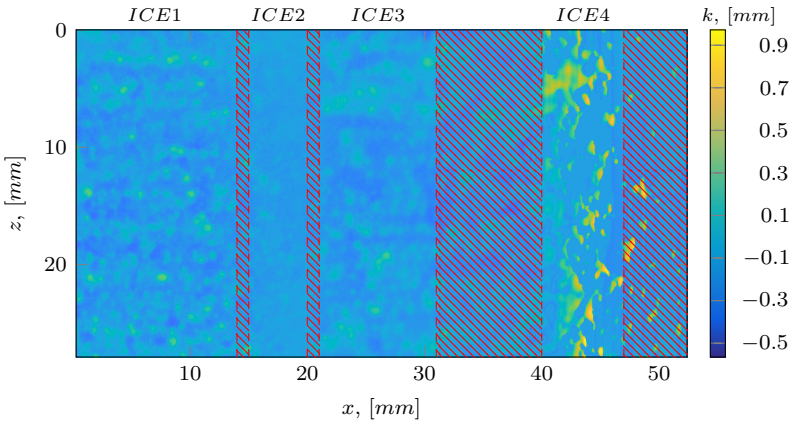


Figure 4.2: Ice accretion roughness overview, the section $ICE2$ corresponds to the stagnation region of the surface. Hatched patterns indicate transitional roughness structures. Figure adapted from [Yang et al. 2023b].

"As a starting point, the sandpaper roughness of ISO Grit P60 is adopted as a generic representation of realistic 'naturally-occurring' roughness, the corresponding roughness statistical metrics are specified in the table 4.1. In the following content, roughness associated with sandpaper is labelled with the keyword *Sand*. Ice accretion roughness on a simplified aero-engine nacelle were generated in the Braunschweig Icing Wind Tunnel [144]. The condition details of the tests set-up can be found in [145]. Testing reproduces the ice accretion process on a complex geometry in take-off conditions during glaze ice formation, which implies that the physics of drawing away a water film influences ice formation. A sketch of the ice geometry is shown in figure 4.1. The geometry of the nacelle leads to different local boundary conditions for the accretion process (such as local pressure, local air velocity and water film thickness). Such differences produce an inhomogeneous surface with different roughness properties. The surface is unfolded considering the form of the nacelle model and the x-coordinate is used in the streamwise direction. Once unfolded, the surface can be divided into a relatively smooth region, on the stagnation line of the model and three different rough regions, away from the stagnation line. This is known as a roughness zonal model in the icing community [146]. Recalling the aim of the present work, the evaluation of roughness reproduction framework is carried out on the homogeneous roughness whose properties are invariant to the position. Therefore, homogeneous roughness patches are sought from the ice accretion roughness for the present study. The ES in the z-direction of figure 4.2 was used to differentiate these regions and produce different

roughness patches with almost homogeneous characteristics. These surfaces are illustrated in figure 4.2. The differences in the surfaces characteristics respond to different in the physics of the ice accretion process. A relatively smooth section can be found at $x \approx 18\ mm$ located at the stagnation region. The growth of ice accretion is visible towards both suction side and pressure side. At the right end of the ice accretion surface spiky ice can be found. This zone is only reached by remaining of a water film or accumulates isolated water beads, which explains the formation of spiky structures. The homogeneous roughness section on the left end of the surface is named $ICE1$ corresponding to the pressure side of the test model. $ICE2$ corresponds to the stagnation area of the test model. $ICE3$ and $ICE4$ both lay on the suction side of the model but present different roughness characteristics. The last sample studied in this work is the CCD surface, which is extracted from the piston head of a single-cylinder research internal combustion engine under realistic operating condition. The roughness is adopted from the previous research by Forooghi *et al.* [4]. More detailed description of these surfaces can be found in the original references."[Yang et al. 2023b]

4.1.2 Artificial Roughness Reproduction

"Artificial roughness is generated based on the p.d.f. and PS associated with the realistic surfaces in the present study. As an example, the p.d.f. and PS of the *Sand* surface scan are shown in figure 4.3 with grey marks while its artificial surrogate is represented by black marks. It is worth noting that, the PS of a 2-D signal is a surface in 2-D frequency space. The 1-D PS is obtained by radially accumulating the PS data points around its origin, i.e. around $(q_x, q_z) = (0, 0)$, as a function of wavenumber $q = \sqrt{q_x^2 + q_z^2}$, where q_x and q_z represent the wavenumber in x and z direction, respectively. This 1-D PS is simply referred to as PS in the following text. The scatter of the PS in figure 4.3(b) is the evidence of the slight anisotropic nature of the realistic surface. In order to achieve isotropic roughness distribution, the PS is subjected to a radial averaging process around the origin [58]. Subsequently, the moving average is applied along the PS in order to represent the realistic PS with a reduced level of complexity. Furthermore, low- and high-pass filters are applied to the surface PS to eliminate roughness structures with extremely large or small wavelengths. The selection of these values will be elaborated in the next section. The largest and the smallest roughness wavelength bounds are denoted as λ_0 and λ_1 in the following text, respectively. Eventually, the simplified PS of the artificial surface is shown with the black dots in the figure 4.3(b)."[Yang et al. 2023b]
The influence of p.d.f. and PS on the resulting hydrodynamic and thermal properties are investigated in an isolated manner. This entails including artificial surfaces with matched PS but different p.d.f. into the present comparison campaign. For these artificial surfaces, a Gaussian p.d.f. with its k_{rms} matching that of the target surface is prescribed. "Finally, the p.d.f. of realistic and artificial surfaces (including Guassian p.d.f.) are shown in the figure 4.3(a). The artificial rough surfaces along with the original sandpaper scan are visualized in figure 4.4(a). In the following content, the realistic surfaces and the artificial surfaces utilizing realistic p.d.f. are denoted by the abbreviations -R and -A, respectively. The Gaussian distributed artificial surfaces are labelled with -G.
Using the same procedure as above, the $ICE1$-4 and the CCD are reproduced. The substitution of Gaussian p.d.f. is performed for $ICE1$-4 as well. These artificial surfaces are visualized in figure 4.4 (b-f) along with their original realistic surface scans. As is reflected in the figure 4.4, characteristic roughness structures, such as the sediment-like roughness elements on the CCD-R, are successfully reproduced by the generation method. The statistical metrics for roughness are quantitatively compared in Table 4.1[...]. It is noticeable that there is a limited range of statistics values available in the current roughness data set. As discussed in the previous section, the roughness properties of a surface strongly depend on its background and forming mechanism. For this reason, no negatively skewed realistic roughness is currently available at our institute. In spite of this, the currently included surfaces already exhibit a considerable range of roughness properties in terms of Sk and ES that covers a considerable parts of the previous studies [4, 8, 53, 147, 148]. However, it must be stressed that over-generalization of the present results should be avoided. Among these statistics, it is clear that the values of k_{md}, k_{rms}, k_{t} as well as Sk and Ku are directly related to the configuration of p.d.f.. The PS, however, is more complex since it accounts for the spatial distribution of roughness at various length scales. The value of ES is thus closely related to the PS. But it can be appreciated from the present data set and literature [31], that while the ES value is largely determined by PS, the impact of p.d.f. cannot be ignored. The value of L^{Corr}, on the other hand, can be

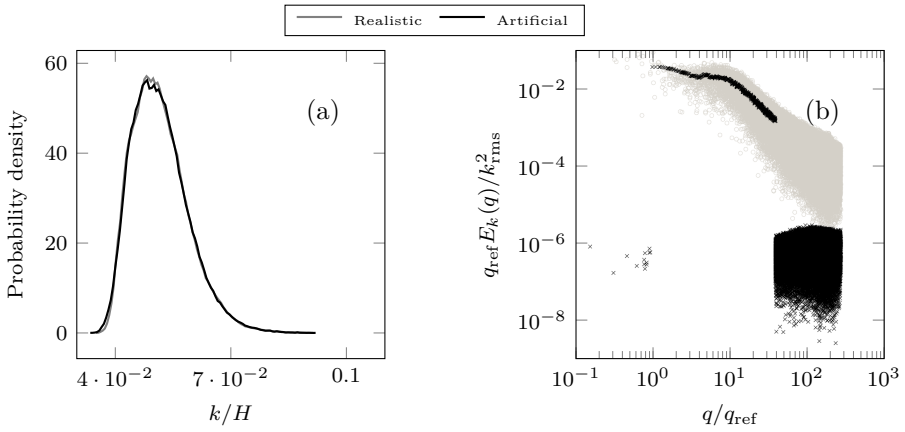


Figure 4.3: Comparison of realistic and artificial sandpaper roughness p.d.f. (a) and PS (b). Where $q_{\mathrm{ref}} = 2\pi/\lambda_0$. Figure adapted from [Yang et al. 2023b].

uniquely defined by PS because the auto-correlation function can be directly derived from PS through reverse Fourier transformation. However, as a matter of fact, the realistic roughness typically consists of a wide range of length scales which cannot be adequately described by a set of roughness statistics. The use of PS has advantage of reflecting the behaviour of each roughness length scale. Therefore, PS is expected to provide a more comprehensive representation of roughness spatial distribution than the aforementioned roughness statistics. Based on this comparison, one can conclude that roughness topographical properties can be successfully reproduced by imitating the PS and p.d.f. of the original surfaces. A slight deviation in the statistics can be observed, which might be attributed to the random nature of the generation process [31] as well as the averaging process of PS. In clear contrast to the different degrees of anisotropy exhibited by realistic surfaces – as is reflected by SAR [60] – the artificial surfaces exhibit nearly perfect isotropic distribution with SAR$\approx$ 1. Because of the strongly anisotropic roughness distribution of $ICE4$-R (SAR$\approx$1.7), this results in the departure of the generated surface topography, for this surface. It can be seen from figure 4.4(e) that the anisotropic roughness elements of the surface $ICE4$-R are deformed into isotropic roughness elements through the roughness reproduction process. Likewise, the strip-like anisotropic valleys of CCD-R are transformed into isotropic pits in CCD-A.
Furthermore, the impact of the Gaussian p.d.f. is evident if one compares the roughness statistics. In the present work, the p.d.f.s of the considered roughness are all positively skewed, i.e. the majority of the roughness height is located below the roughness median height $k_z/2$. A symmetric height distribution around $k_z/2$ is obtained by substituting Gaussian p.d.f., which results in a lift of the roughness melt-down height k_{md}. Moreover, it is worth noting that while the other statistical measurements appears to be influenced by the variation in p.d.f. configuration, ES is less susceptible to this variation."[Yang et al. 2023b]

4.1.3 Simulation Configurations

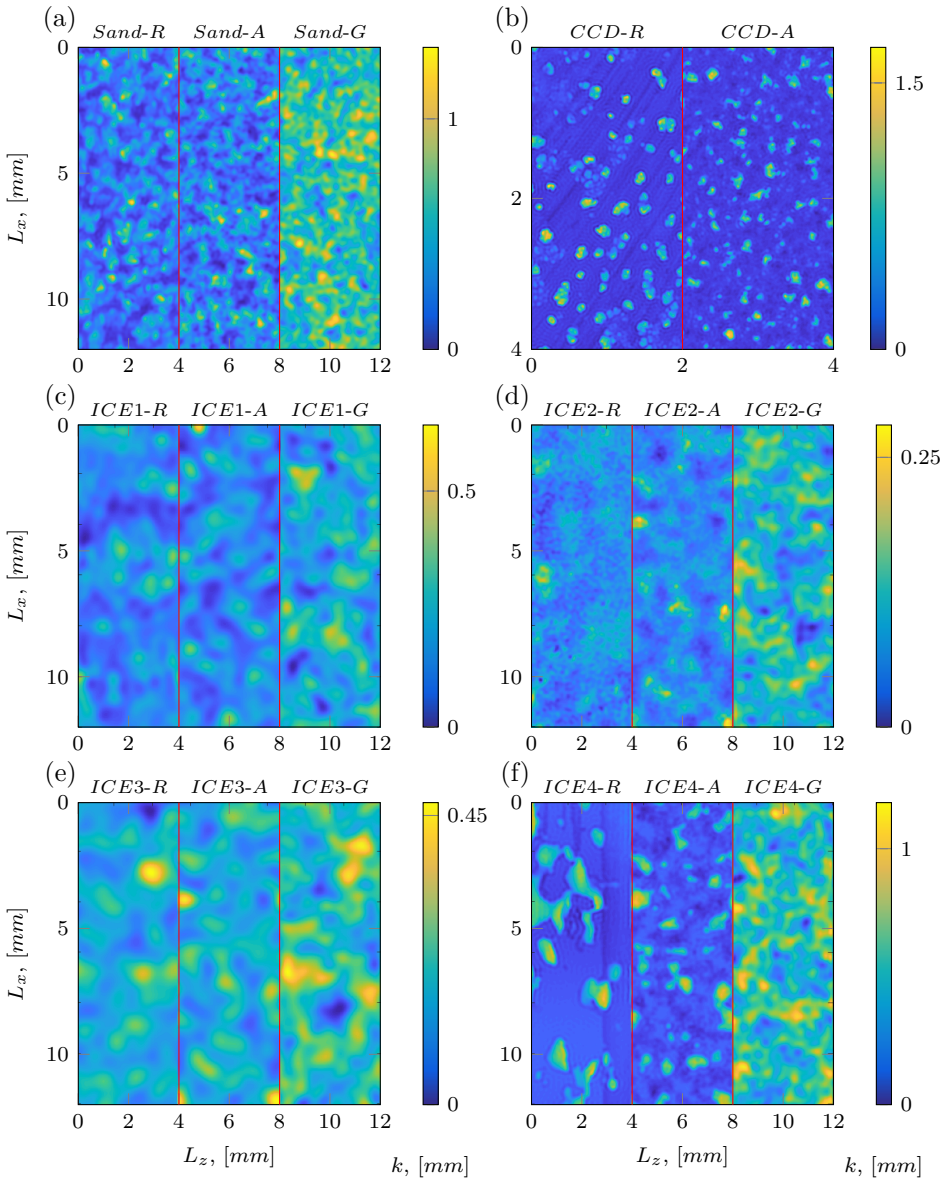


Figure 4.4: Overview of the realistic rough surfaces along with their artificial surrogates. Figure adapted from [Yang et al. 2023b].

Table 4.1: Summary of surface geometrical statistics. Units: mm. Data adopted from [Yang et al. 2023b].

Surface	Configuration	k_{md}	k_{rms}	k_z	Sk	Ku	ES	L_x^{Corr}	L_z^{Corr}	SAR
$ICE1$-R	Realistic	0.19	0.07	0.63	0.76	4.72	0.16	0.60	0.68	0.88
$ICE1$-A	Realistic p.d.f.+PS	0.19	0.07	0.63	0.76	4.75	0.17	0.54	0.54	1.00
$ICE1$-G	Gaussian p.d.f.+PS	0.30	0.07	0.64	0.00	3.00	0.18	0.54	0.54	1.00
$ICE2$-R	Realistic	0.09	0.03	0.28	1.17	6.39	0.13	0.64	0.56	1.14
$ICE2$-A	Realistic p.d.f.+PS	0.09	0.03	0.28	1.18	6.46	0.11	0.50	0.52	0.96
$ICE2$-G	Gaussian p.d.f.+PS	0.13	0.03	0.27	0.00	2.98	0.12	0.50	0.52	0.96
$ICE3$-R	Realistic	0.20	0.05	0.47	0.49	4.42	0.10	0.74	0.92	0.80
$ICE3$-A	Realistic p.d.f.+PS	0.20	0.05	0.47	0.50	4.46	0.12	0.58	0.58	1.00
$ICE3$-G	Gaussian p.d.f.+PS	0.23	0.05	0.44	0.01	3.00	0.11	0.58	0.58	1.00
$ICE4$-R	Realistic	0.26	0.17	0.98	1.87	5.73	0.40	0.64	0.38	1.68
$ICE4$-A	Realistic p.d.f.+PS	0.25	0.17	0.97	1.88	5.74	0.34	0.40	0.40	1.00
$ICE4$-G	Gaussian p.d.f.+PS	0.61	0.17	1.18	0.00	2.88	0.42	0.50	0.52	0.96
Sand-R	Realistic	0.65	0.09	1.16	0.77	3.80	0.48	0.29	0.30	1.03
Sand-A	Realistic p.d.f.+PS	0.65	0.09	1.16	0.74	3.74	0.47	0.31	0.31	1.00
Sand-G	Gaussian p.d.f.+PS	0.84	0.09	1.29	0.00	3.00	0.45	0.28	0.28	1.01
CCD-R	Realistic	0.23	0.20	1.72	3.02	12.3	0.32	0.86	0.76	1.13
CCD-A	Realistic p.d.f.+PS	0.24	0.20	1.71	3.02	12.3	0.31	0.78	0.78	1.00

Building upon the successful application in the previous chapter, minimal channel DNS – termed *Mini* – is employed to compute ΔU^+ and $\Delta\Theta^+$ for both realistic roughness scans and their corresponding artificial surrogates. To complete the investigation, full-span DNSs – referred to as *Full*– is additionally carried out on *Sand* and *CCD* roughness. "The physical scaling of the channel half height H is selected for each case individually to achieve a clear separation of length scales H and k while maintaining a significant roughness effect within the current range of Re_τ. Among other cases, the H for the *Sand* cases is selected to be consistent with the experiment facility at the Institute of Fluid Mechanics, KIT [149]. The largest and smallest roughness wavelengths λ_0 and λ_1 of each type of roughness are selected individually to resolve as much roughness length scales that are influential for the roughness effect to the flow as possible. The simulation grid size criteria with $\Delta_x \approx \Delta_z \lessapprox \lambda_1/4$ is applied for all simulations, where Δ_x and Δ_z represent wall-parallel grid sizes in streamwise and spanwise direction, respectively. The extent of the simulation domain is selected to accommodate for the largest roughness wavelength, i.e. with $L_z \geq \lambda_0$. A summary of the currently investigated cases along with the their corresponding simulation configurations is provided in the table 4.2. "[Yang et al. 2023b]

Table 4.2: Simulation configurations. * is the wildcard that represents R (realistic), A (artificial) and G (Gaussian) case. The word in the brackets indicates the domain size configuration – Mini and Full stand for minimal channel and full-span DNS, respectively. Data adopted from [Yang et al. 2023b]

Roughness	Re_τ	H, $[mm]$	L_x/H	L_z/H	N_x	N_z	N_y	Δ_x^+	Δ_z^+	λ_0/H	λ_1/H
*ICE*1-* (Mini)	500	5	4.0	0.8	512	96	401	3.9	4.2	0.8	0.06
	750	5	4.0	0.8	800	160	401	3.8	3.8	0.8	0.06
	900	5	4.0	0.8	960	192	401	3.8	3.8	0.8	0.06
	1000	5	4.0	0.8	1024	192	401	3.9	4.2	0.8	0.06
*ICE*2-* (Mini)	500	5	4.0	0.8	1024	192	401	2.0	2.1	0.8	0.02
	1000	5	4.0	0.8	1024	192	401	3.9	4.2	0.8	0.02
	1500	5	4.0	0.8	1200	240	401	5.0	5.0	0.8	0.02
	2000	5	4.0	0.8	1200	240	401	6.7	6.7	0.8	0.02
*ICE*3-* (Mini)	500	5	4.0	0.8	512	96	401	3.9	4.2	0.8	0.06
	750	5	4.0	0.8	800	160	401	3.8	3.8	0.8	0.06
	1000	5	4.0	0.8	1024	192	401	3.9	4.2	0.8	0.06
	1200	5	4.0	0.8	1152	256	401	4.2	3.8	0.8	0.06
	1500	5	4.0	0.8	1200	240	401	5.0	5.0	0.8	0.06
*ICE*4-* (Mini)	500	5	4.0	0.8	1024	192	401	2.0	2.1	0.8	0.02
Sand-R (Full)	500	12.8	5.5	3.9	768	384	385	3.6	5.0	0.8	0.02
Sand-A (Full)	500	12.8	6.3	3.1	768	384	385	4.2	4.2	0.8	0.02
Sand-R (Mini)	500	12.8	4.0	0.8	540	108	385	3.7	3.7	0.8	0.02
Sand-A (Mini)	500	12.8	4.0	0.8	540	108	385	3.7	3.7	0.8	0.02
Sand-G (Mini)	500	12.8	4.0	0.8	540	108	385	3.7	3.7	0.8	0.02
CCD-R (Full)	500	10	6.0	4.0	864	576	401	3.5	3.5	4	0.05
CCD-A (Full)	500	10	6.0	4.0	864	576	401	3.5	3.5	4	0.05

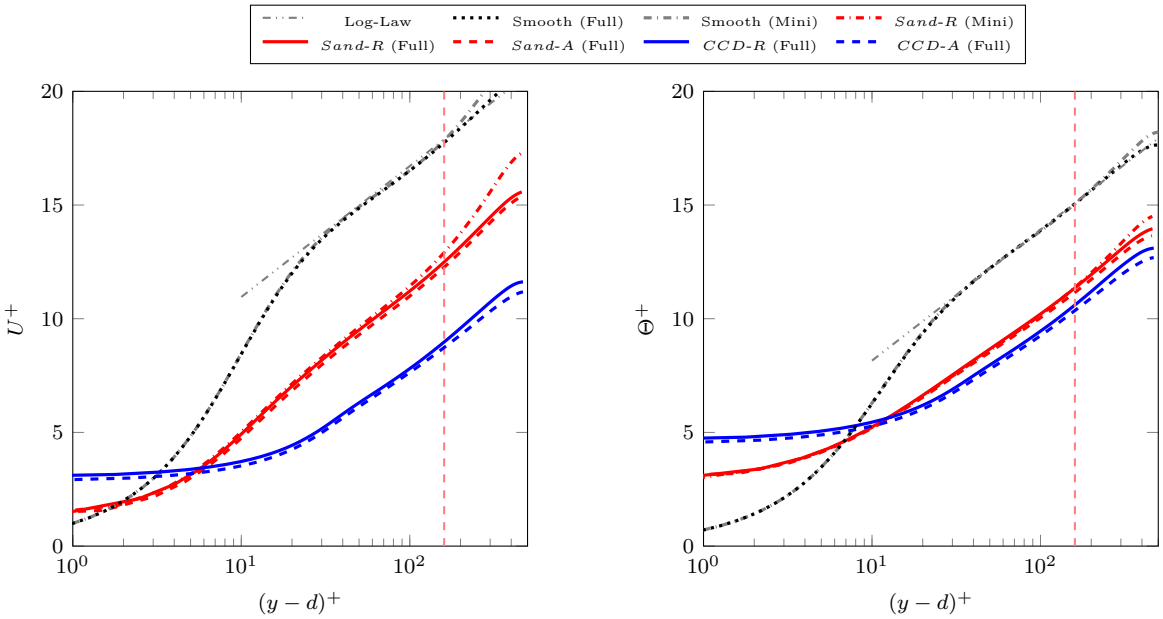


Figure 4.5: Mean velocity and temperature profiles at $\mathrm{Re}_\tau = 500$. The critical height y_c^+ for the minimal channel is marked with red vertical dashed line. d represents zero-plane displacement following jackson's definition [81]. The temperature log-law is based on Kader [93]. Figure adapted from [Yang et al. 2023b].

4.2 Examination of Hydrodynamic and Thermal Properties of Artificial Roughness Surrogates

4.2.1 Global Flow Statistics

This section focuses on the analysis of the skin friction coefficient C_f and the Stanton number St of both the artificial roughness surrogates and their corresponding target roughness. The C_f value is calculated using Eqn. 2.27. Analogous to the physical interpretation of the skin friction coefficient, the heat transfer ability of the roughness can be quantified by St according to Eqn. 2.30. As can be seen, the precise estimation of these global flow properties necessitates an accurate representation of the outer layer flow dynamics. It is acknowledged in the previous chapters that minimal channels do not directly resolve the outer layer flow; thus, these properties can exclusively be assessed through full-channel DNS.

Table 4.3: Global flow statistics obtained from full-channel simulations. Data adopted from [Yang et al. 2023b]

Flow statistics	*Sand-R*	*Sand-A*	*CCD-R*	*CCD-A*
C_f	0.0138	0.0134	0.0266	0.0268
St	0.0068	0.0069	0.0036	0.0036

"The mean velocity and temperature profiles over these rough surfaces are shown in figure 4.5. It can be seen from the figure, that the profiles of the artificial surfaces collapse well to their target realistic counterparts. As a result, the maximum error of C_f and St lower than 3% is achieved for the artificially generated roughness *Sand-A* and *CCD-A* as indicated in the table 4.3. As mentioned before, minimal channel DNS is employed for study of ice accretion samples in this work. To demonstrate how this approach performs, we additionally conducted a minimal channel simulation for *Sand-R* as well as a smooth minimal channel, which are further included in the figure 4.5 with dash-dotted line pattern. The critical height of the minimal channel $y_c^+ = 160$ is marked with the vertical red dashed line. It is evident that the mean flow and temperature profile in the minimal channel agree well with the conventional full-span channel results

up until the critical height y_c^+ in logarithmic layer. The departure of the mean velocity as well as the mean temperature profiles for minimal channels can be observed in the outer region. This can be attributed to limited spanwise width of the minimal channel. The structures larger than the spanwise size of the channel are not captured in the outer layer, leading to a build up of undissipated kinetic energy [107]. This translate to the overshoot of the mean profiles in the outer region beyond y_c."[Yang et al. 2023b] This observation, in line with the results already reported in chapter 3, shows that the proposed minimal channel configurations are capable for recovering realistic roughness-induced mean velocity and temperature profile offset. In the following sections, the comparison of artificial roughness surrogates and realistic roughness scans are based on minimal channels.

4.2.2 Mean Velocity and Temperature Profiles

"The inner-scaled mean velocity and temperature profiles of all artificial roughness in minimal channels, i.e. $ICE1$-A/G to $ICE4$-A/G and $Sand$-A/G, at $\mathrm{Re}_\tau = 500$ are depicted in figure 4.6. The target roughness, i.e. $ICE1$-R to $ICE4$-R and $Sand$-R, are included in the same plots with solid lines. As can be seen in the figure, the artificial surfaces successfully reproduce the mean velocity and temperature profiles associated with realistic roughness for all roughness except $ICE4$, which will be discussed separately. It is remarkable that all the Gaussian artificial surfaces do not show a considerable deviation from the original non-Gaussian realistic surfaces beyond the region very close to the wall. This results in an almost identical ΔU^+ value between the non-Gaussian and Gaussian artificial surfaces with the same PS. The $Sand$ roughness is the exception to this observation. This observation, while puzzling at the first sight, can be attributed to the low ES of the first three samples (see table 4.1). It has already been shown by Forooghi et al. [14] that at low ES, the roughness function is much less sensitive to the moments of p.d.f.. The fact that the Gaussian and non-Gaussian profiles of $Sand$ roughness, which has a considerably larger ES, do not collapse agrees with this explanation too. Further evidence of this conjugated correlation can be found in the previous research by Lee *et al.* [22]. In their work, the predominant role of the surface statistics $Sk\times$ES in determining the roughness skin friction is reported through the sensitivity analysis on a machine learning model trained for predicting the roughness equivalent sand-grain size. In light of the observations by Barros *et al.* [55], it is possible to draw a physical explanation: roughness with low ES value falls under the so-called 'wavy' regime [15], in which regime the roughness induced pressure drag is less pronounced compared to the frictional drag. Thus, the pressure drag contribution by the roughness peaks is less dominant to the surface drag. As a consequence, the resultant roughness skin friction is less sensitive to the change of the roughness height distribution - which is represented by the shape of p.d.f.. As mentioned, for all samples, however, the mean profiles very close to the wall do not agree. This can be due to the significantly altered volume occupation of the Gaussian artificial roughness.
Despite the excellent performance of other artificial roughness, $ICE4$-A is unable to reproduce the realistic mean velocity and temperature profiles with a significantly larger offset in the mean velocity profile than $ICE4$-R. An obvious explanation can be the strong anisotropic nature of this roughness that cannot be reproduced by the present isotropic approach. Busse et al. [60] reported that an isotropic roughness aggravates the roughness disturbance by breaking the "streamwise-channeling" effect comparing to the streamwise elongated structures of realistic surface."[Yang et al. 2023b]

4.2.3 Reynolds Stresses

"The Reynolds stresses of all investigated roughness at $\mathrm{Re}_\tau = 500$ are visualized in figure 4.7, roughness crest height k_z and the critical height of minimal channel y_c^+ are indicated with grey and red vertical dash lines, respectively. Note that the CCD surfaces are simulated in a full-span channel with no critical height. The corresponding Gaussian artificial surface is shown in each plot as well for comparison. It is clear that for all cases, except $ICE4$, the artificial roughness (dashed lines) successfully reproduced the Reynolds stress profiles, while Reynolds stresses of Gaussian artificial surfaces (dotted lines) fails to reproduce the peak value and shows considerable deviation below the roughness crest $y < k_z$. This is arguably attributed to the flow blockage effect of the surface which is manipulated by the p.d.f. configuration. Note that the Gaussian artificial surface has an increased k_{md} values (measured from the bottom wall), which possibly contributes to this deviation and the fact that the peaks of the Reynolds stresses are moved away from

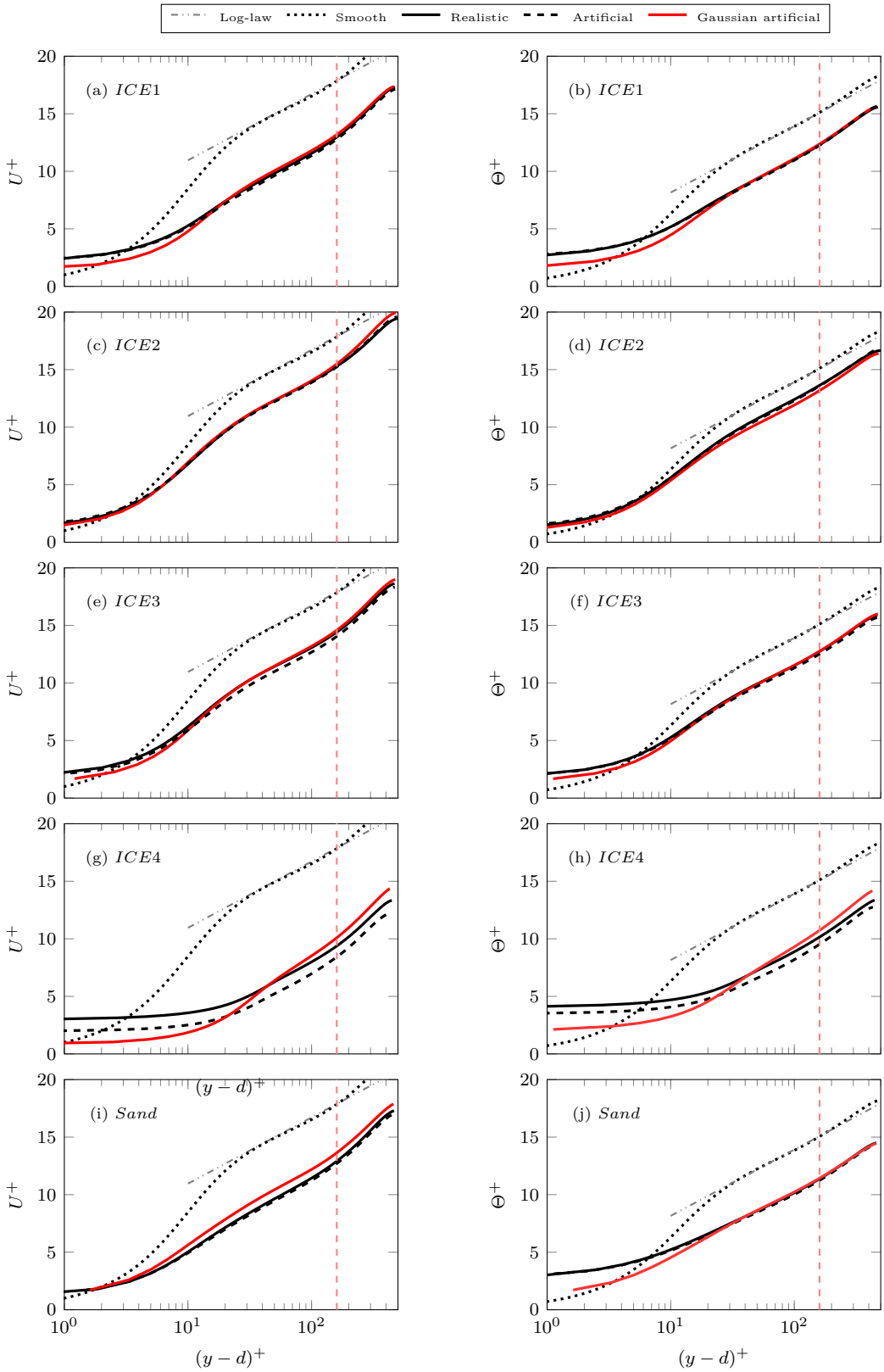


Figure 4.6: Mean velocity and temperature profiles at $\mathrm{Re}_\tau = 500$ in minimal channels. The vertical dashed line represents the critical height of the minimal channel $y_c^+ = 160$. (a,c,e,g,i): mean velocity profiles ΔU^+, (b,d,f,h,j): mean temperature profiles Θ^+. The temperature log-law is based on Kader [93]. Figure adapted from [Yang et al. 2023b].

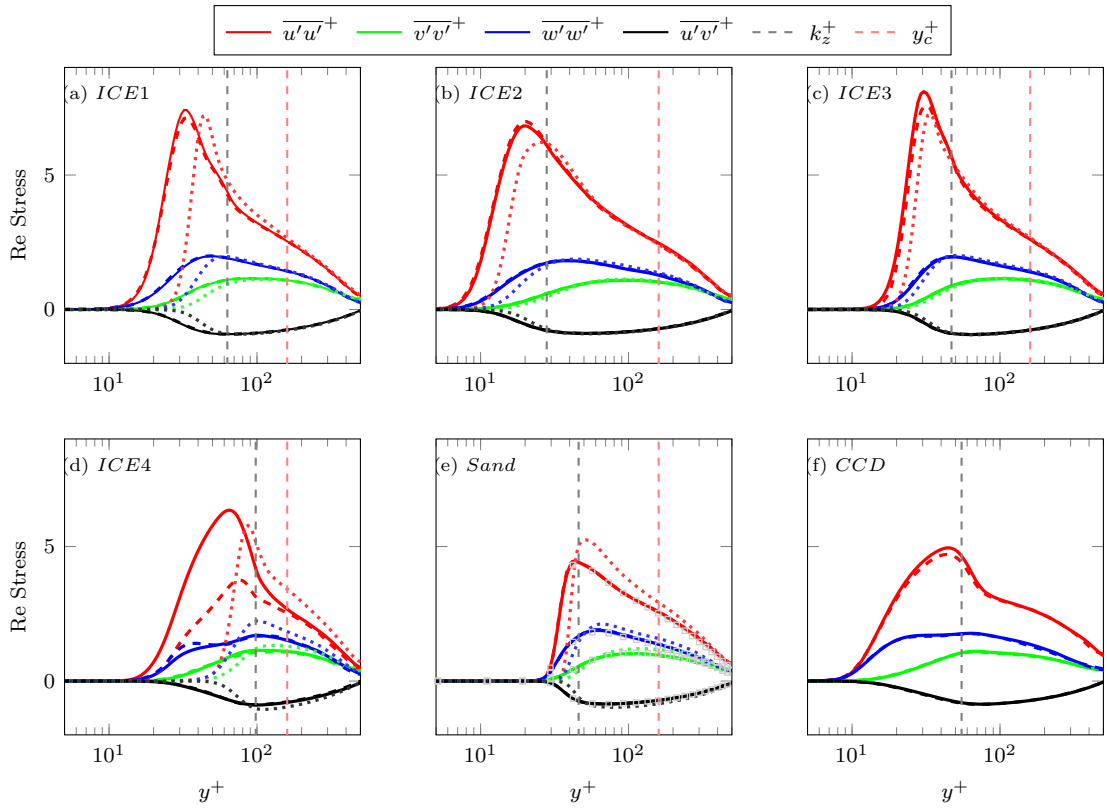


Figure 4.7: Reynolds stresses of ICE-1 to ICE-4, $Sand$ and CCD at $Re_\tau = 500$. The grey vertical dashed line represents k_z^+ of realistic surfaces, the red vertical dashed represents y_c^+ of minimal channel simulations. Solid line: realistic, dashed line: artificial, dotted line: Gaussian. In (e), Full-span channel DNS results of $Sand$-R are included in the plot with grey squares. Figure adapted from [Yang et al. 2023b].

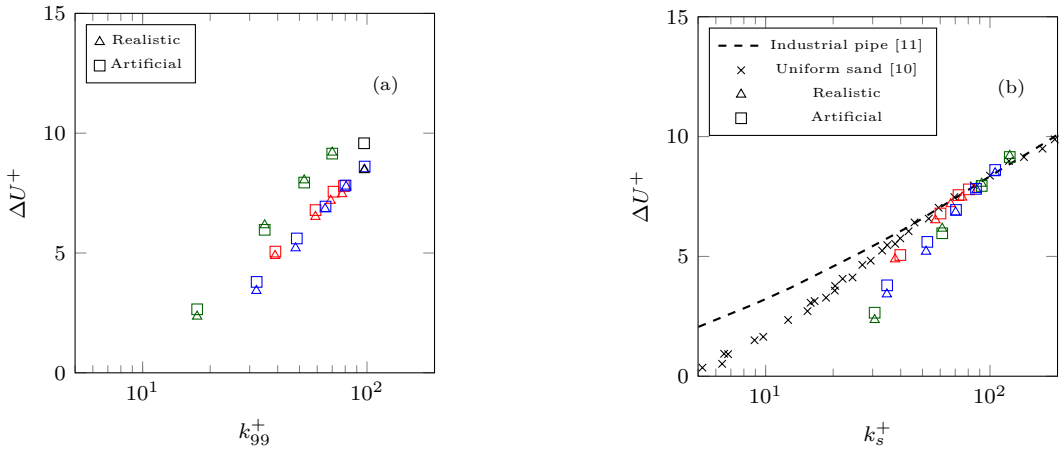


Figure 4.8: Roughness functions of different roughness with varying Re_τ as functions of k_{99}^+ (a) and k_s^+ (b). Red: $ICE1$ ($k_s = 0.38mm$ vs. $0.40mm$), green: $ICE2$ ($k_s = 0.30mm$), blue: $ICE3$ ($k_s = 0.35mm$) and black: $ICE4$. It should be noted that the black triangle overlaps with the blue triangle at $k_{99}^+ \approx 100$. Figure adapted from [Yang et al. 2023b].

Table 4.4: Equivalent sand-grain size of realistic and artificial surfaces, unit: mm.Data adopted from [Yang et al. 2023b].

k_s	$ICE1$	$ICE2$	$ICE3$
Realistic	0.38	0.30	0.35
Artificial	0.40	0.30	0.35

the bottom wall. The Reynolds stresses from full span channel simulation of *Sand-R* are included in figure 4.7(e) with grey squares. While the Reynolds stresses show excellent agreement among *Sand-R/A* surfaces below the critical height y_c^+, slight departure of the Reynolds stress profiles can be observed for the *Sand-R* between minimal channel and full-span channel simulations beyond the critical height y_c due to the nature of minimal channel. For all the cases considered in the present work, the peaks of streamwise Reynolds stresses are located below the roughness crest k_z.
Similar to the mean profiles, the artificial roughness $ICE4$-A fails to reproduce the Reynolds stresses. As discussed in the previous section, the considerable discrepancies of the roughness effect is due to the isotropic assumption of the artificial roughness. As one can see from the profile of streamwise normal Reynolds stress $\overline{u'u'}^+$, realistic roughness yields a higher $\overline{u'u'}^+$ peak value compared to the artificial isotropic roughness. This is likely due to the fact that realistic roughness shows surface anisotropy and are prone to form quasi-streamwise vortices. This finding is in agreement with the work by Busse *et al.* [60]. In contrast, roughness anisotropy shows less impact on other Reynolds stresses. Among others, spanwise Reynolds stress $\overline{w'w'}^+$ shows slight increase in the near wall region for artificial isotropic roughness. This could be the consequence of the increased spanwise interference in comparison to the realistic rough surface."[Yang et al. 2023b]

4.2.4 Equivalent Sand-Grain Size

"As indicated in the previous sections, satisfactory results are obtained for the reproduced roughness at $\mathrm{Re}_\tau = 500$. However, the roughness likely lies in transitionally rough regimes at the present Re_τ. In order to ensure that the artificial surfaces hold comparable performance to their target surfaces in fully rough regime too, Re_τ is extended up to the fully rough regime for successfully reproduced cases in minimal channels, i.e. $ICE1$ to $ICE3$. Using the data at fully rough regime, it is possible to calculate the equivalent sand-grain size k_s for different realistic and artificial samples. In order the calculate the k_s values, roughness function ΔU^+ for each roughness at different Re_τ is evaluated. The ΔU^+ values in the present work are

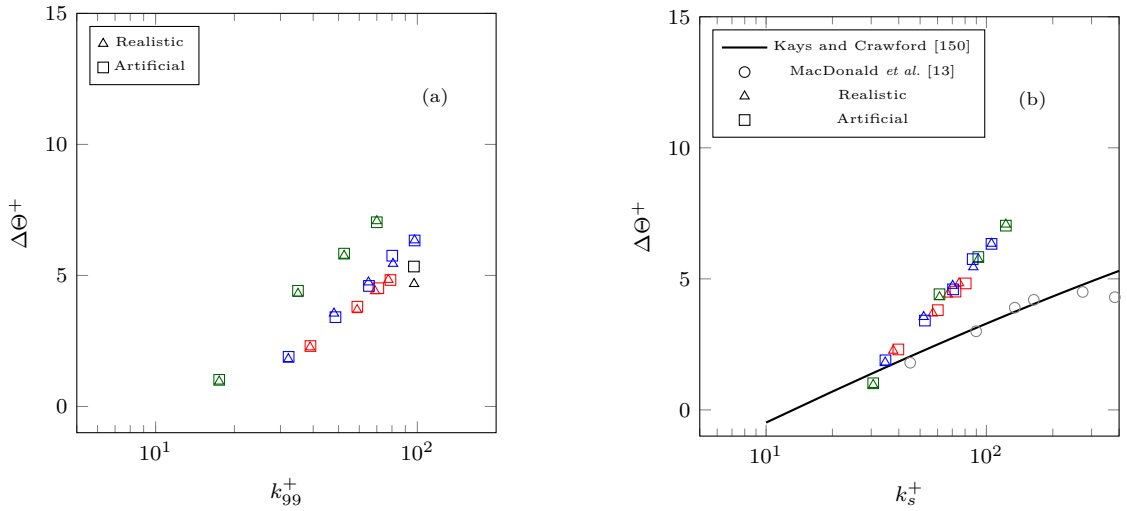


Figure 4.9: Temperature roughness functions of different roughness with varying Re_τ as functions of k_{99}^+ (a) and k_s^+ (b). Red: $ICE1$, green: $ICE2$, blue: $ICE3$ and black: $ICE4$. Figure adapted from [Yang et al. 2023b].

calculated by subtracting the mean inner-scaled velocity profile over roughness from the value obtained from equation 2.35 (smooth wall) at $y_c^+ = 160$.
The values of roughness function ΔU^+ are plotted in figure 4.8(a) as functions of roughness height k_{99}^+ for the case $ICE1$-A to $ICE3$-A with squares.[...] The realistic surfaces $ICE1$-R to $ICE3$-R are represented by triangles. All the surfaces exhibit fully-rough behaviour at adequately high Re_τ. They collapse to a single logarithmic curve similar to but with an offset from the Nikuradse roughness function [10] for uniform sand grain. To complete the comparison, the ΔU^+ values of $ICE4$-R and $ICE4$-A at $\mathrm{Re}_\tau = 500$ are additionally included in figure 4.8(a) with black triangle and square, respectively. The $ICE4$-A surface yields a disagreement in ΔU^+ of 13%. "[Yang et al. 2023b] The equivalent sand-grain size k_s of the $ICE1$ to $ICE3$ surfaces are calculated by fitting the roughness function in fully rough regime with the empirical equation 2.50. "The roughness function against calculated k_s^+ values are shown on the right panel of figure 4.8, also included are the experiments by [10] and [11]. The figure shows that each pair of artificial and realistic roughness produce very similar values of ΔU^+. It is also observed that, in the transitionally rough regime the variation of ΔU^+ against k_s^+ for $ICE1$ to $ICE3$ roughness do not follow that of the Nikuradse sand-grain roughness [10], and steeper transition to the fully rough regime can be observed. The identical transition behaviour between each pair of realistic and artificial surfaces indicates the consistency of the roughness flow properties. The considered $ICE1$ to $ICE3$ surfaces reach fully rough at approximately $60 < k_s^+ < 80$. Overall the results in figure 4.8 indicate that the pseudo-random roughness with matched p.d.f. and PS can be a suitable surrogate for realistic roughness when it comes to the investigations of equivalent sand-grain size k_s.
The temperature roughness functions $\Delta\Theta^+$ are calculated by subtracting the mean inner-scaled temperature profile over roughness from the value obtained from the empirical log-law of the smooth wall for temperature profile [93],

$$\Theta^+ = \frac{Pr_t}{\kappa}\ln(y^+) + \beta(Pr), \tag{4.1}$$

where $Pr_t = 0.85$ is the turbulent Prandtl number and the empirical function $\beta(Pr)$ is given by

$$\beta(Pr) = (3.85Pr^{1/3} - 1.3)^2 + \frac{Pr_t}{\kappa}\ln Pr. \tag{4.2}$$

In figure 4.9, temperature roughness functions for ICE samples are plotted against k_{99}^+ and k_s^+ on the left and right panels, respectively. The first observation is a very good agreement between realistic and artificial temperature roughness functions for $ICE1$-3. Considerable discrepancy($\approx 14\%$) for $ICE4$-A in terms of

$\Delta\Theta^+$ is observed, which can be attributed to the surface anisotropy as discussed above. On the right panel of figure 4.9, we also observe a surprising collapse of data when $\Delta\Theta^+$ is plotted against k_s^+." [Yang et al. 2023b] It is important to note that, unlike ΔU^+, the value of $\Delta\Theta^+$ is not necessarily proportional to k_s^+. The roughness topography can introduce a certain degree of scatter in the data [8]. Specifically, the flow recirculation zones induced by roughness structures can create adverse pressure gradients, significantly affecting heat transfer [151]. However, the roughness samples currently studied, namely $ICE1$ to $ICE3$, exhibit very low effective slopes and positive skewness values. These roughness parameters suggest that these surfaces are inert at generating flow recirculation zones, thus leading to a closer analogy between skin friction and heat transfer [54]. As a result, a similarity in the evolution of $\Delta\Theta^+$ and ΔU^+ with respect to k_s^+ is observed.

"Additionally, in figure 4.9(b) the empirical correlation proposed by Kays and Crawford [150] as well as DNS data of MacDonald *et al.* [13] are plotted for comparison. Remarkably, the slope of $\Delta\Theta^+$ in the current data is closer to that of ΔU^+ than the empirical correlation for $\Delta\Theta^+$. Also the saturation reported by MacDonald *et al.* is absent here. This is also likely attributed to the low effective slope of the current samples leading to a closer resemblance in the trends of ΔU^+ and $\Delta\Theta^+$. Furthermore, the current realistic samples span a range of scales. Due to the delayed onset of the transitionally rough regime for the temperature field [13] more smaller roughness structures start to behave as thermodynamically rough and contribute to the heat transfer as the Re_τ increases. This possibly translates to the steeper increase of $\Delta\Theta^+$ compared with the sinusoidal roughness. Moreover, note that the slope predicted by Kays and Crawford correlation is for roughness in the 'fully rough' regime while, as observed in figure 9(b), the current cases reach the fully rough asymptote only marginally at the largest Reynolds numbers. Finally the saturation of $\Delta\Theta^+$ is observed by MacDonald *et al.* for a sinusoidal roughness in the fully rough regime. As discussed above, compared to a sinusoidal roughness, the realistic roughness may include length scales smaller than k_s, which can delay $\Delta\Theta^+$ reaching an asymptotic constant value. Considering the fact that the saturation of $\Delta\Theta^+$ takes place with a slight delay at $k_s^+ \approx 250$ while the flow reaches fully rough regime at $k_s^+ \approx 150$ for the sinusoidal roughness, the saturation behaviour for the ICE surfaces is thus expected at higher k_s^+ value with higher saturation constant value. "[Yang et al. 2023b]

4.3 Reproduction of Laboratory Experiment on Sandpaper Roughness through DNS

The above comparison of realistic roughness and artificial surrogates is primarily carried out within the numerical context. Nevertheless, it's well understood that a significant disparity may exist between the observations in real-world applications and the aforementioned numerical investigations. This gap can be attributed to various factors, including imprecise rough surface measurements and inaccuracies in representing boundary conditions during experiments. The disparities are clearly demonstrated when comparing laboratory experiments conducted on the P60 sandpaper with DNS of the roughness scan, i.e. *Sand-R*. This comparative analysis is carried out at the Institute of Fluid mechanics, KIT, utilizing a high precision blower-type wind tunnel to generate turbulent channel flow over P60 sandpaper. Pressure drop measurements as well as HWA are carried out to assess skin friction and mean velocity profile, respectively. Details will be given in the following sections. These experimental results are then compared with DNS results. A discernible departure in C_f value of more than 10% is observed (not shown in the dissertation). This notable inconsistency raises significant concerns regarding the accuracy of the planned data-driven model in real-world applications, particularly considering that the model is predominantly developed based on a series of numerical simulations of roughness surfaces.
To address this concern, it is important to recognize that the white light interferometry scan of the sandpaper sample is subjected to significant light reflections, resulting in distorted roughness topography. As demonstrated in this section, the aforementioned discrepancy can be effectively mitigated by employing appropriate roughness scanning techniques. To this end, we propose a precise technical process for accurately measuring realistic roughness properties. This process involves acquiring high-fidelity p.d.f and PS information of the roughness topography separately through respectively suitable measurement techniques. Consequently, DNS investigations can be carried out on the artificial roughness surrogate created according to the high-fidelity p.d.f. and PS. It is further emphasized that the associated laboratory experiment on sandpaper roughness as well as the measurement of sandpaper geometry are performed by the co-author Lars von Deyn. Thus, their acquisition is not part of this thesis and is only briefly described in this dissertation.
This process involves acquiring high-fidelity p.d.f and PS information of the roughness separately through individual appropriate measurement techniques, namely tactile measurements and 3-D photogrammetry, respectively.

4.3.1 Acquisition of Topographical Information of Sandpaper Sample

Recognizing the limitations of white-light interferometry measurements due to its sensibility to light reflections, alternative measurement techniques must be nominated. Nevertheless, bearing in mind the agreement between experiment and simulation, instead of a direct measurement technique to reconstruct the 2-D roughness scan, reproducing roughness surrogates through a combination of multiple well-established measurement techniques may be an accessible solution to this task. Following this idea, roughness statistics, namely p.d.f. and PS, are be separately obtained through respective accurate approaches. "Tactile measurements (five 1-D line scans of 5 cm length each with perthometer Mahr MarSurf PCV®) and three-dimensional reconstruction of highly resolved photographs (five surface samples of size 2 cm × 3 cm) using photogrammetry [152] were used to obtain the PS and the p.d.f. of the sandpaper surface height, respectively."[Frohnapfel et al. 2024][1]. The acquired p.d.f. and PS are displayed in appendix A.1. Subsequently, having demonstrated the satisfactory representativeness of these two aspects of topographical information in uniquely determining the roughness impact on the flow, roughness surrogate is generated based on the case-averaged p.d.f. and PS across the each five measurement through the pseudo-random roughness generation algorithm.

[1]The measurements of sandpaper roughness and experiments are carried out by Dr.-Ing. Lars von Deyn and Dr.-Ing. Alexander Stroh.

Table 4.5: Dimensions of the experimentally investigated sandpaper configurations. Data adopted from [Frohnapfel et al. 2024]

Case	k_{md} [mm]	$\mathrm{H_{eff}}$ [mm]	$h_{\mathrm{md}}/\mathrm{H_{eff}}$ [%]	C_f at $\mathrm{Re}_b = 1.8 \times 10^4$ DNS	Experiment
Smooth	0	12.60	0	0.0061	0.0064
Sandpaper	0.67	12.15	5.5	0.01238	0.01236

Table 4.6: DNS case overview of turbulent flow configurations. Data adopted from [Frohnapfel et al. 2024]

Case	Re_b	$L_x \times L_y \times L_z$	Δx^+	Δy^+_{min}	Δy^+_{max}	Δz^+	Re_τ
Smooth	1.8×10^4	$8\mathrm{H} \times 2\mathrm{H} \times 3.72\mathrm{H}$	5.17	0.017	4.06	4.80	496
Rough	1.8×10^4	$6.25\mathrm{H} \times 2\mathrm{H} \times 3.13\mathrm{H}$	4.44	0.024	5.81	4.44	700

4.3.2 Experimental and Numerical Methodologies

On the experiment aspect, the homogeneous rough channel is achieved by completely covering the P60 grit sandpaper on the smooth top and bottom surfaces. Table 4.5 provides an overview of the experimental configurations of both smooth and rough cases. The experiments are carried out in a blower tunnel facility depicted in figure 4.10. "The flow is driven by a radial fan and progresses through a supply pipe into a large settling chamber. The air is blown towards the back wall of the settling chamber. On the opposite side of the settling chamber the flow is directed through a nozzle into the actual test section. The air flows through five grids embedded in wooden frames and a honeycomb flow straightener on its way through the settling chamber towards the turbulent channel flow test section. The facility allows to vary the bulk Reynolds number of the channel flow in the range of $5 \times 10^3 < Re_b < 8.5 \times 10^4$ [153, 154].
Measurements of the streamwise pressure gradient can be carried out by evaluating the static pressure at 21 pairs of pressure taps located along both side walls of a 314 H long channel test section with an aspect ratio of 1:12. The channel test section is divided into three segments of 76 H, 119 H and 119 H streamwise extent. As in previous investigations with the same facility [90, 91, 149, 155], the flow is tripped at the inlet of the first segment. "[Frohnapfel et al. 2024]
The average skin-friction is going to be determined from the measurement of pressure gradient Π and volumetric flow rate $\dot{V}$. For statistically two-dimensional and fully developed turbulent channel flow, the wall-shear stress τ_w depends on Π through Eqn. 2.43. The mean velocity profile is measured through HWA, detailed introduction of the experimental procedure can be found in [34].
The simulation configurations strive for precise replication of the experimental geometries outlined in table 4.5. The numerical surrogate of the sandpaper is generated according to the averaged p.d.f. and PS as outlined in section 4.3.1. "Turbulent channel flow simulations are carried out with constant flow rate at $Re_b = 1.8 \times 10^4$ for the smooth, rough cases. The simulation domain size and resolution are summarized in table 4.6. "[Frohnapfel et al. 2024]

4.3.3 Comparison of Experimental and DNS Results

Figure 4.11 shows the skin-friction coefficient C_f curves of smooth and rough walls, derived from the pressure drop in experiments. In this figure, C_f is plotted agains Re_b, which is derived from the flow rate measurements. The DNS data of both smooth and rough cases at $\mathrm{Re}_b = 18000$ are included in the figure. Furthermore, quantitative comparison of C_f for the experiment and DNS results is provided in the table 4.5. As can be observed from the table, "the smooth wall DNS data point is slightly below the experimentally obtained $C_{f,0}$, in agreement with classical literature data [157]. This difference reflects the drag contribution of the duct side walls. The improved agreement between DNS and experimental data for

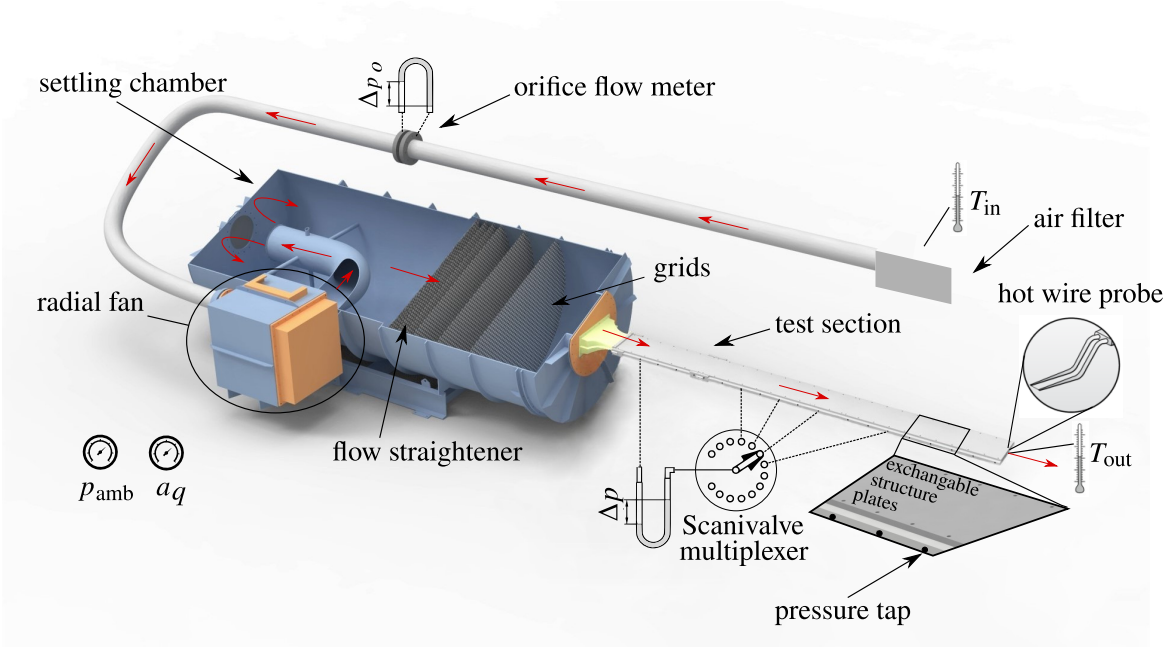


Figure 4.10: Schematic of the utilised wind tunnel including measurement instrumentation. Figure adapted from [Frohnapfel et al. 2024]

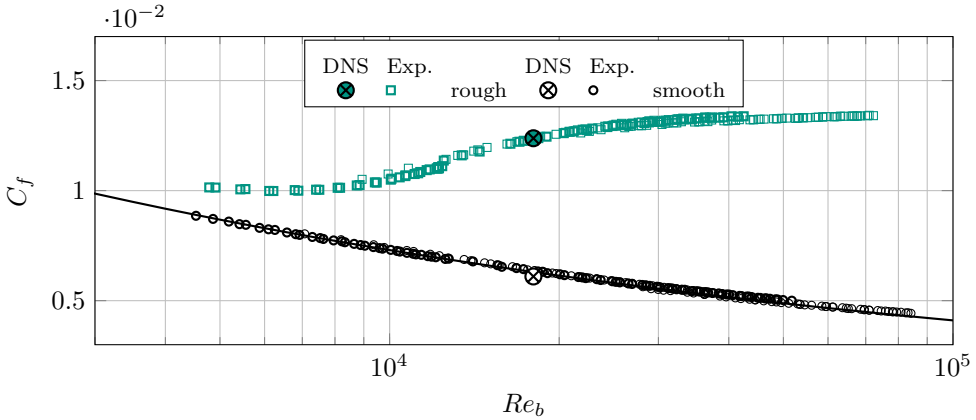


Figure 4.11: C_f versus Re_b for smooth and homogeneous rough surfaces. The experimental rough wall data. The complementary DNS results are included. The correlation proposed by [156] for turbulent duct flows of large aspect ratio is shown as a black line. Data adopted from [Frohnapfel et al. 2024]

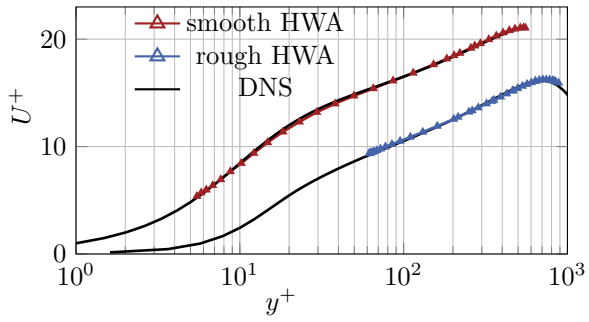


Figure 4.12: Comparison of experimental measurements and DNS results of smooth and rough channel flow. Data adopted from [Frohnapfel et al. 2024].

the rough surfaces can be explained by the smaller relative contribution of the (smooth) side wall to the total friction drag when the main duct walls are rough. "[Frohnapfel et al. 2024] Moreover, the inner-scaled mean velocity profiles over the smooth and sandpaper rough walls are plotted in figure 4.12. The mean flow velocities measured by HWA are included in the figure for comparison. An excellent agreement of the mean velocity profile in the outer layer is achieved through present DNS on the sandpaper surrogate generated through high-fidelity p.d.f. and PS measured from sandpaper.

4.4 Summary

The selected pseudo-random roughness generation algorithm is employed to produce realistic roughness surrogates by emulating realistic p.d.f. and PS. The impact of these roughness surrogates on turbulent flow is compared with that of the corresponding original surfaces. "The p.d.f.s and PSs are extracted from 6 types of industrial rough surfaces including ice accretion on aero-engine (ICE), sandpaper ($Sand$) as well as combustion chamber deposit roughness (CCD), and fed into the roughness generation method. The generated artificial rough surfaces are assessed using DNS in both full-span channels and minimal channels at a wide range of $\mathrm{Re}_\tau = 500 - 2000$ to cover both transitionally and fully rough regimes.
The artificial counterparts of the presently considered realistic rough surfaces successfully reproduce both the hydrodynamic and thermal properties of their corresponding roughness as long as the roughness can be considered isotropic. The maximum error in the values of C_f and St is estimated to be less than 3% for the artificial surfaces $Sand$-A and CCD-A compared to the original surfaces. The artificial surfaces, particularly $Sand$-A, CCD-A as well as $ICE1$-A to $ICE3$-A, are demonstrated to be capable of accurately predicting the mean velocity and temperature profiles as well as the Reynolds stresses. The equivalent sand-grain size k_s^+ of $ICE1$-A to $ICE3$-A are calculated, these values are in excellent agreement with their original surfaces. The satisfactory results highlight the promising applicability of the employed roughness generation framework for the future systematic investigation of flow over reality-based irregular roughness surfaces. Significant deviations in drag and heat transfer is observed for reproducing the strongly anisotropic roughness $ICE4$-R (SAR≈1.7), as is reflected by the 13% increase of ΔU^+ and 14% increase of $\Delta\Theta^+$ of the artificial surface $ICE4$-A at $\mathrm{Re}_\tau = 500$. Furthermore, the decrease of the peak value in the streamwise normal Reynolds stress $\overline{u'u'}^+$ is observed for the artificial roughness $ICE4$-A compared to the realistic roughness $ICE4$-R. In light of the present observations, it is estimated that the threshold for a significant streamwise surface anisotropy effect is located at SAR$<$ 1.7. This value may, understandably, be altered depending on the interplay of other roughness parameters. A detailed examination of this aspect should be conducted in the future studies.
When the artificial surfaces are generated by only matching the PS and using a Gaussian p.d.f., considerable departures of the mean flow and temperature profiles are identified in the vicinity of roughness. However, at small Effective Slope values (ES$<$0.2), it is possible to obtain values of roughness function ΔU^+ with a good acuracy using Gaussian artificial surfaces: however, a considerable decrease of ΔU^+ is observed for the Gaussian artificial surface with ES$>$0.4, which translates to a lower roughness-induced drag." [Yang et al. 2023b]
Moreover, the present roughness reproduction procedure is utilized for replicating laboratory experiment of turbulent channel flow over P60 sandpaper roughness. In this exercise, measurement uncertainty associated with roughness scanning is taken under the scope. This challenge is addressed by simulating artificial roughness surrogate exhibiting high-fidelity p.d.f. and PS of sandpaper roughness extracted through respective accurate measuring techniques, i.e. through 3-D photogrammetry and perthometer, respectively. By striving close consistency of the topographical features between the experimental and numerical roughness instances, accurate prediction of C_f values with a discrepancy lower than 1% can be achieved through DNS when compared to the experimental pressure drop measurement. According to the current comparative analysis, it is demonstrated that with accurate representation of roughness topographical features in the numerical testing roughness patch, the flow behaviour– including skin-friction coefficient and mean velocity profile– obtained through DNS closely resembles the experimental observations. Therefore, the present data-acquisition framework is demonstrated to be able to convey high-fidelity physics of the roughness effect of the naturally-occurring roughness topographies.

5 Data-Driven Prediction of Equivalent Sand-Grain Size and Identification of Drag-Relevant Roughness Scales

The present skin-friction prediction task is framed as a regression problem. The skin friction exerted by an irregular roughness can be jointly influenced by a broad range of its topographical properties, these factors including the skewness Sk [43], effective slope ES [53], spatial arrangement of roughness elements [14]. This multi-variant feature of the present problem poses difficulty in deriving generalizable prediction through mathematical correlations. To address this challenge, Jouybari *et al.* [21] firstly employed neural network and *Gaussian progress regression* models to map a set of 17 single-valued roughness statistical parameters to k_s. Having acknowledged the stressed scarcity of realistic roughness data in practical engineering applications, Lee *et al.* [22] employed transfer learning technique for training the model. The authors initially pre-train the neural network model with various empirical correlation functions, and subsequently fine-tune the model with a small set of high-fidelity data acquired from (numerical) experiments. This approach is demonstrated to improve model training efficiency given a restricted amount of data, thus holding great practical implication.

While these pioneering works deliver promising results, the data-driven approach arguably has the potential to realize truly universal models, which can generalize beyond a certain class of roughness. As another attempt to explore this potential, the present ML model is trained through sophisticatedly designed model architecture and training strategy. The previously evaluated roughness data-generation tool is utilized to construct training database. In the present chapter, the calculation of k_s through minimal-channel DNS referred to as 'labelling' of those samples borrowing the term from the ML terminology. Moreover, each roughness topography along with its k_s value is called a training 'data point'. One should note that labelling is a computationally expensive process due to the need to perform DNS. While the computational efficiency is enhanced by employing the concept of minimal-channel DNS, it must be noted that a massive number of data points is typically required, that are passively received from an already existing training database. Nevertheless, the useful information for the training process provided by each training data point can be improved by providing the model liberty to pick most curious samples. This training strategy, referred to as *active learning* (AL), is utilized in the present work to enhance the efficiency of the database construction while simultaneously maximizing the learning effectiveness of the model, thereby minimizing computational burdens. However, conducting AL process necessitates model feedback during the training process, involving the estimation of prediction uncertainties on the candidate roughness topographies. In the present work, the quantification of the prediction uncertainty is realized through employing a specific type of machine learning model, namely the *ensemble neural network* (ENN). The ENN model encompasses 50 individual MLPs. Each MLP-member in the ENN is furnished with an identical set of roughness features, i.e. roughness height p.d.f. and PS, as input. The training efficiency provided by AL is demonstrated through the comparison against a baseline training strategy. The performance of the final AL-ENN model is subsequently evaluated using realistic and artificial roughness samples, drawn from both the present work and external literature. Moreover, the rationale grounding the prediction of the AL-ENN model is assessed through the layer-wise relevance propagation (LRP)technique, thereby providing a deeper understanding of the mechanism of roughness skin friction from a data-driven statistical perspective. Leveraging LRP, the relevance scores of the roughness structures of different length scales to the resulting skin friction are evaluated. Further simulations are carried out on the filtered roughness, on which drag-irrelevant structures identified by LRP are excluded. Detailed investigations on the local drag-inducing phenomena over both original and filtered rough surfaces are conducted.

Moreover, the respective advantages of utilizing different roughness characterization methods, namely using

p.d.f. and PS or single-valued roughness parameters as model inputs are examined and elaborated. To this end, the present AL database is employed for training two additional ENN models, each based on different sets of roughness parameters as input features. The hyperparamters of each model are individually optimized through *Bayesian optimization* (BO) to attain optimal performance. These models are assessed using the same testing databases. Finally, the response of the models to the variations in their input roughness parameters concerning skin friction prediction is evaluated through SA.
The present chapter is based on the publications *Prediction of equivalent sand-grain size and identification of drag-relevant scales of roughness – a data-driven approach* [35] and *Assessment of roughness characterization methods for data-driven predictions* [36]. Compared to the publications, numbering of the figures, equations and tables, as well as citation style have been adapted to be consistent with this thesis.

5.1 Methodology

5.1.1 Roughness Repository

The (unlabelled) roughness 'repository' $\mathcal{U}$ is constructed by a collection of 4200 artificially generated rough surfaces leveraging the chosen pseudo-random roughness generation algorithm. "For creation of the present repository, p.d.f. and PS are parameterized, as described shortly, and their parameters are varied randomly within a realistic range to generate a variety of roughness samples while imitating the random nature of roughness formation in practical applications.
In total, three types of p.d.f. – namely, Gaussian, Weibull and bimodal – are used, and for each new roughness added to the repository, one type is randomly selected."[Yang et al. 2023a] Among these p.d.f.s, the Weibull distribution follows Eqn. 2.4. The shape parameter $0.7 < K_W < 1.7$ is selected randomly with $\beta_W = 1.0$. Moreover, the bimodal distribution is obtained following Eqn. 2.5. In the equation, the Gaussian distribution term $f_{\mathrm{G}}(x|\mu_{\mathrm{G}}, \sigma_{\mathrm{G}})$ obeys randomized mean $0 < \mu_{\mathrm{G}} < 0.5$ and randomized standard deviation $0 < \sigma_{\mathrm{G}} < 0.5$. The p.d.f. variable k is then scaled from 0 to the roughness peak-to-trough height k_z, whose value is randomly determined in the range $0.06 < k_z/\mathrm{H} < 0.18$.
"The PS of the roughness samples in the repository is controlled by two randomized parameters, namely the roll-off length L_r [58] and the power-law decline rate θ_{PS} [59], whose values are selected in the range $0.1 < L_r/(\log(\lambda_0/\lambda_1)) < 0.6$ and $-3 < \theta_{\mathrm{PS}} < -0.1$. Here, λ_0 and λ_1 represent the upper and lower bounds of the PS, or the largest and smallest wavelengths forming the roughness topography. Random perturbations are added to the PS to achieve higher randomness in PS. The lower bound of the roughness wavelength is set to $\lambda_1 = 0.04\mathrm{H}$ to ensure that the finest structures can be discretized by an adequate number of grid points. The upper bound of the roughness wavelength λ_0 is selected randomly in the range $0.5\mathrm{H} < \lambda_0 < 2\mathrm{H}$. As will be discussed later, the roughness sample size as well as the simulation domain size should both be adjusted to accommodate this wavelength.
Eventually, 4200 separate pairs of p.d.f. and PS are generated using the described random process, each leading to one rough surface added to the repository $\mathcal{U}$. A representation of the parameter space covered by these samples is illustrated in section 5.3. Moreover, examples of the generated samples can be seen in Appendix A.2."[Yang et al. 2023a]

5.1.2 Numerical Method

In order to approach fully rough regime while maintaining affordable computational cost, $\mathrm{Re}_\tau = 800$ is selected for the simulations. Due to the high computational demand of many DNS, the concept of minimal channel[26, 158] is utilized. The analysis conducted in the previous chapters suggests that the minimal channel size for each roughness should be able to accommodate the largest in-plane roughness structure λ_0, which imposes the most stringent constraint. As a result, the corresponding minimal channel sizes are primarily determined by λ_0 of each roughness. Accordingly, table 5.1 provides an overview of the simulation setups based on their respective λ_0 values. "Due to the different sizes of the simulation domains, the chosen numbers of grid points differ according to the mesh size, but in all cases $\Delta^+_{x,z} \leq 4$. In wall-normal directions, cosine stretching mesh is adopted for the Chebyshev discretization. The mesh independence is confirmed in a set of additional tests.

Table 5.1: Simulation setups. Data adapted from [Yang et al. 2023a]

λ_0/H	L_x/H	L_z/H	L_y/H	N_x	N_z	N_y	Δ_x^+	Δ_z^+	$\Delta_{y,min}^+$	$\Delta_{y,max}^+$
$0.5 \geq \lambda_0/\mathrm{H} \geq 0.6$	1.8	0.6	2	400	160	451	3.6	3.0	0.02	5.7
$0.6 > \lambda_0/\mathrm{H} \geq 0.7$	2.1	0.7	2	576	144	451	2.9	3.9	0.02	5.7
$0.7 > \lambda_0/\mathrm{H} \geq 0.8$	2.4	0.8	2	480	192	451	4.0	3.3	0.02	5.7
$0.8 > \lambda_0/\mathrm{H} \geq 0.9$	2.7	0.9	2	640	256	451	3.4	2.8	0.02	5.7
$0.9 > \lambda_0/\mathrm{H} \geq 1.0$	3.0	1.0	2	640	256	451	3.8	3.1	0.02	5.7
$1.0 > \lambda_0/\mathrm{H} \geq 1.1$	3.3	1.1	2	720	288	451	3.7	3.1	0.02	5.7
$1.1 > \lambda_0/\mathrm{H} \geq 1.2$	3.6	1.2	2	720	288	451	4.0	3.3	0.02	5.7
$1.2 > \lambda_0/\mathrm{H} \geq 1.3$	3.9	1.3	2	800	320	451	3.9	3.3	0.02	5.7
$1.3 > \lambda_0/\mathrm{H} \geq 1.4$	4.2	1.4	2	960	384	451	3.5	2.9	0.02	5.7
$1.4 > \lambda_0/\mathrm{H} \geq 1.5$	4.5	1.5	2	960	384	451	3.8	3.1	0.02	5.7
$1.5 > \lambda_0/\mathrm{H} \geq 1.6$	4.8	1.6	2	960	384	451	4.0	3.3	0.02	5.7
$1.6 > \lambda_0/\mathrm{H} \geq 1.7$	5.1	1.7	2	1080	432	451	3.8	3.1	0.02	5.7
$1.7 > \lambda_0/\mathrm{H} \geq 1.8$	5.4	1.8	2	1200	480	451	3.6	3.0	0.02	5.7
$1.8 > \lambda_0/\mathrm{H} \geq 1.9$	5.7	1.9	2	1200	480	451	3.8	3.2	0.02	5.7
$1.9 > \lambda_0/\mathrm{H} \geq 2.0$	6.0	2.0	2	1200	480	451	4.0	3.3	0.02	5.7

For each investigated roughness, ΔU^+ is determined from the offset in the logarithmic velocity profile comparing corresponding rough and smooth DNS. Notably, when plotting mean velocity profiles, zero-plane displacement d is applied in order to achieve parallel velocity profiles in the logarithmic layer, where d is determined [...] following Jackson's method [81]. It is worth noting that in the extensive literature on rough wall-bounded turbulent flows, various definitions of d have been proposed, and furthermore, the choice of virtual wall position can affect the predicted rough-wall shear stress τ_w and thus the resulting k_s value [12]. Therefore, it is important to recognize this as a possible source of uncertainty, and take into account the definitions of τ_w and d when comparing data from different sources.
It is also important to determine if the flow has reached the fully rough regime in each simulation. To this end, ΔU^+ is combined with Eqn. 2.50 to yield a testing value of k_s^+. Then following the threshold adopted by Jouybari *et al.* [21] a roughness with $k_s^+ \geq 50$ is deemed to be in the fully rough regime, and all samples not matching this criterion are excluded from the training or testing process. The selected threshold $k_s^+ \geq 50$ is somewhat lower than the common threshold of $k_s^+ \geq 70$ [42] and thus may introduce into the database into the database some data points with limited transitionally rough behaviour. This threshold is, however, chosen deliberately as a trade-off to maximize the number of training data given the limited computational resources. One should note that an increase in the threshold value of k_s^+ while maintaining the same parameter space would be possible by increasing Re_τ. This would, however, lead to an obvious compromise in the final performance of the model by reducing the number of training data points at a given computational cost.
Overall, 85 roughness samples are DNS-labelled and eventually included in the labelled data set $\mathcal{L}$ to train the final AL-based model. The procedure for selection of these training samples is explained in detail in the following. Eight out of the 85 labelled samples are located in the range of $50 \leq k_s^+ \leq 70$. We observe that incorporating these samples into the training process improves model performance. This improvement in the model performance can be attributed both to the incorporation of more informative samples according to AL as well as to the regularization effect of data diversity introduced by including transitionally rough training samples, which makes the model more robust and mitigates over-fitting [159, 160]."[Yang et al. 2023a] Additionally, having acknowledged that the present evaluation process relies on the existence of outer layer similarity, careful examination of the outer layer similarity is necessitated. To this end, deficit velocity profiles of the considered roughness simulations are illustrated in the appendix A.3, grouped by their minimal-channel sizes.

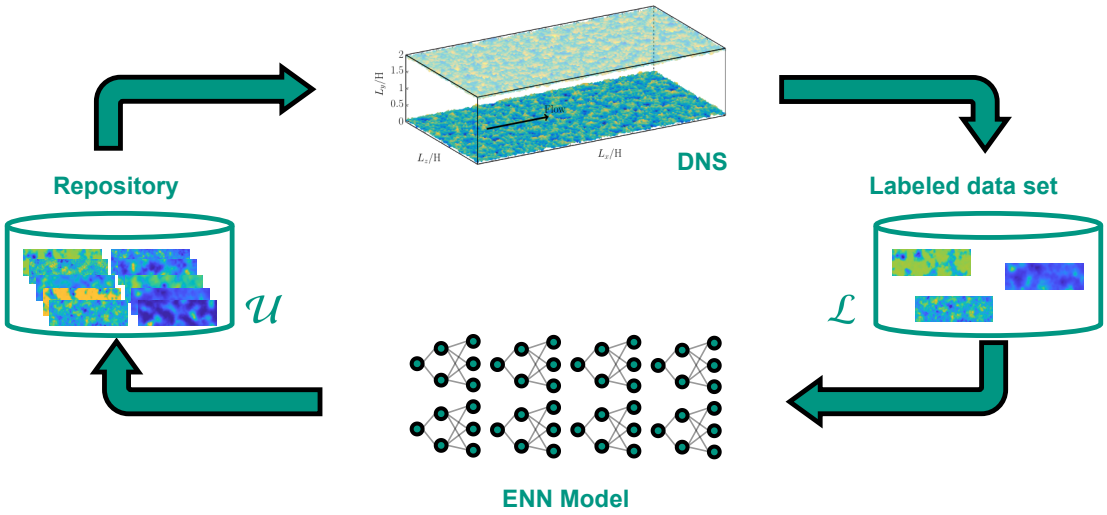


Figure 5.1: Schematic of the AL-framework. Figure adapted from [Yang et al. 2023a]

5.1.3 Active Learning

In order to train the data-driven roughness model, k_s for several roughness samples should be labelled. "One should note that labelling is a computationally expensive process due to the need to perform DNS" [Yang et al. 2023a] – even by means of minimal-channel DNS, given the substantial number of data points necessary for training data-driven models. "In dealing with such scenarios, ML methods classified under active learning– also known as query-based learning [161] or optimal experimental design [162] in different contexts – have been proven particularly advantageous [29, 163, 164]. In AL, selection of the training data is navigated in a way such that the information gain from a certain amount of available data is maximized [165]. The 'informativeness' of a potential data point is commonly measured by the uncertainty in its prediction, which needs to be determined without labelling, e.g. through the standard deviation of the predictive distribution of a Bayesian model [166] or the variation of the predictions among a number of individual models [167].
Two major AL categories can be identified in the literature [168–170]. The methods based on membership query synthesis expand an existing data set by creating and labelling new samples that the model is most curious about. In contrast, the methods based on pool-based sampling utilize a 'bounded' unlabelled data set (also called a repository) $\mathcal{U}$, select and label the most informative samples from $\mathcal{U}$ and include them in the labelled training data set $\mathcal{L}$. In the present work, pool-based sampling is deemed more suitable as it can prevent creating unrealistic samples [168]. Moreover, identification of the most informative samples follows a *query-by-committee* (QBC) strategy [171], in which variance in the outputs of an ensemble of individual models (the committee)" [Yang et al. 2023a]
In order to facilitate AL, "the ML model in the present work is constructed in a QBC fashion by building an ensemble neural network model consisting of 50 independent NNs with identical architecture as the 'committee' members. Similar to the methods proposed by Raychaudhuri and Hamey [167] and Burbidge *et al.* [172], the prediction uncertainty of the ENN model is defined as the variance of the predictions among the members, σ_{k_r}.
The workflow of the AL framework is sketched in the figure 5.1. Two collections of roughness samples are included in the framework. These are the (unlabelled) repository $\mathcal{U}$ and the (labelled) training data set $\mathcal{L}$. As a starting point in the AL framework, 30 samples are randomly selected from the repository, labelled (i.e. their k_s is calculated) through DNS, and used to train a first ENN model, which is referred to as the 'base model'. This preliminary 'base model' is subsequently improved throughout multiple AL iterations." [Yang et al. 2023a]

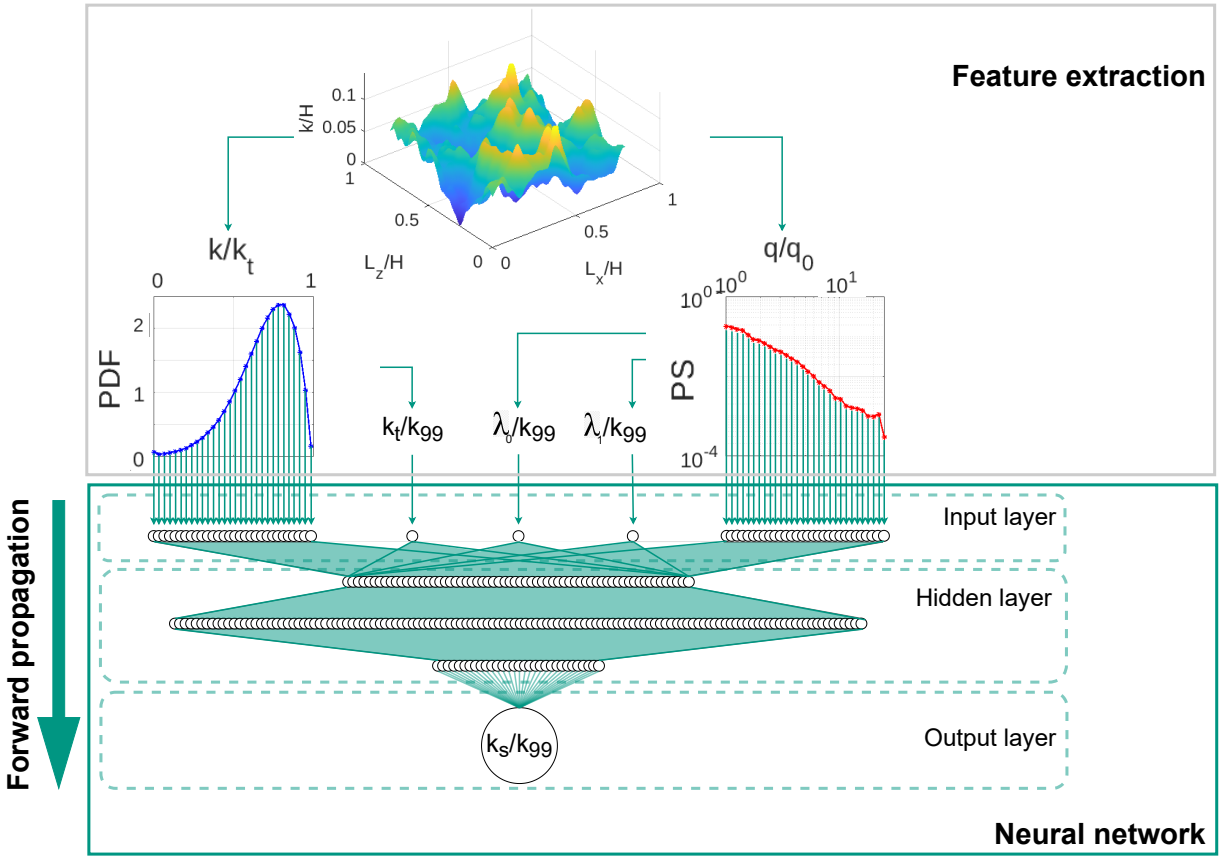


Figure 5.2: Schematic of a single NN in an ENN. Figure adapted from [Yang et al. 2023a]

5.1.4 Ensemble Prediction

"The function of the ENN model is to regress the (dimensionless) equivalent sand-grain roughness $k_r = k_s/k_{99}$, and to calculate the variance of the predictions σ_{k_r} as a basis for QBC [...]. The ENN is composed of multiple NNs with similar structures that is shown in figure 5.2. The input vector $\boldsymbol{I}$ of the NN contains the discretized roughness p.d.f. and PS along with three additional characteristic features of the rough surface, i.e. k_t/k_{99} and the normalized largest and smallest roughness wavelength $\lambda_0^* = \lambda_0/k_{99}$ and $\lambda_1^* = \lambda_1/k_{99}$, respectively. The input elements in $\boldsymbol{I}$ that represent the roughness p.d.f. and PS are obtained by discretizing equidistantly the roughness p.d.f. and PS each into 30 values within the height range $0 < k < k_t$ and the wavenumber range $2\pi/\lambda_1 > 2\pi/\lambda > 2\pi/\lambda_0$. Each NN in the ensemble is constructed with one input layer with 63 (3+30+30) input elements, three hidden layers with 64, 128 and 32 nonlinear neurons with ReLU activation, and one linear neuron in the output layer. The optimal number of neurons at each layer is determined through a grid search of a range of numbers that achieves the lowest model prediction error on $\mathcal{T}_{\text{inter}}$. The L2-regularization is applied to the loss function. Adam is employed to train the model. The final prediction of the ENN is defined as the mean prediction over the 50 NNs, namely $\mu_{k_r} = \sum_{i=1}^{50} \hat{k}_{r,i}/50$, where $\hat{k}_r$ represents the prediction of a single NN, and the index i indicates the index of the NN. The prediction variance is calculated as $\sigma_{k_r} = \sqrt{\sum_{i=1}^{50}(\hat{k}_{r,i} - \mu_{k_r})^2/50}$. It is worth noting that each NN in the ENN model is trained individually based on 90% of the randomly selected samples in the labelled data set $\mathcal{L}$, while the rest of the samples are used for validation. The initial weights of the neurons in each NN are assigned randomly at the beginning of the training process. In such a way, the diversity among the QBC members is ensured, which is an important factor in determining the generalization of the ENN model [173]. It is important to note that the current ensemble members used in the model are deterministic NNs, and

the uncertainty of the training data from DNS is assumed to be minimal. However, when considering experimental training data, where (aleatoric) uncertainties arise from possible measurement errors, the performance of the current ENN approach may be compromised due to its limited capability in handling such uncertainties. In these scenarios, the utilization of probabilistic models – such as Bayesian NNs – may be more suitable as they allow for the explicit incorporation of measurement uncertainties." [Yang et al. 2023a]

5.1.5 Testing Dataset

"In the present work, three distinct testing data sets are introduced to evaluate the model performance and its universality. The difference among the data sets lies in the nature and origin of the samples that they contain. The first data set, $\mathcal{T}_{\text{inter}}$, is composed of 20 samples chosen randomly from $\mathcal{U}$ that have never been seen by the model during the training process.
Despite the fact that the employed roughness generation method can generate irregular, multi-scale surfaces resembling realistic roughness, we test the model separately for additional rough surfaces extracted from scanning of naturally occurring roughness, which form the second testing data set $\mathcal{T}_{\text{ext},1}$. There are five samples in this 'external' data set. These include roughness generated by ice accretion [6], deposit in internal combustion engine [4], and a grit-blasted surface [37]. In addition to that, we test the model against a second external data set, $\mathcal{T}_{\text{ext},2}$, which contains irregular roughness samples from the database provided by Jouybari *et al.* [21]. In this data set, many roughness samples are generated by placing ellipsoidal elements of different sizes and orientations on a smooth wall, making them rather distinct from the type of roughness used to train the model. We separate this testing data set from the other two as it contains a specific type of artificial roughness." [Yang et al. 2023a]

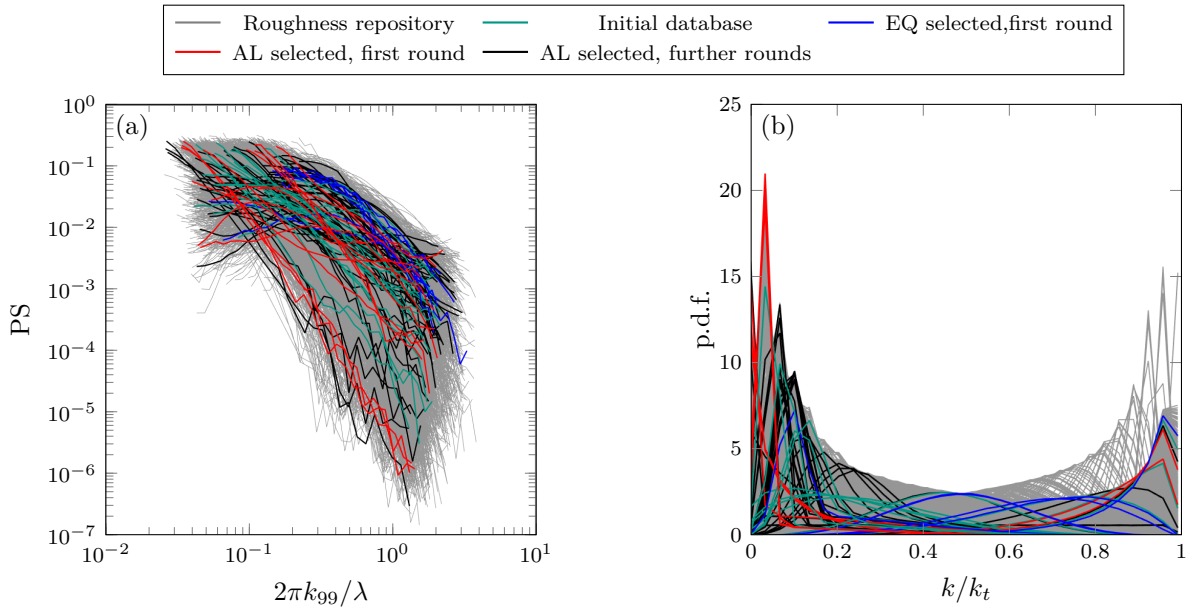


Figure 5.3: PS (a) and p.d.f. (b) of 4200 roughness samples in the roughness repository (grey). The samples selected for training are distinguished with different colours. While AL model tends to explore the PS and p.d.f. domain, the EQ model contains samples that are placed closely to the known initial database. Figure adapted from [Yang et al. 2023a]

5.2 Assessment of Active Learning Framework

"In this subsection, we explore if the AL framework enhances the training behaviour of the model. To do so, we compare a model trained with AL-selected data points to one trained with an arbitrary selection of data points. To avoid the computational cost of running many eventually unused DNSs, the comparison is made for only one AL iteration. Figure 5.3 shows all p.d.f. and PS pairs contained in the repository $\mathcal{U}$ (grey) and those randomly selected for the initial base model (green), as well as those selected for further training (other colours). The wide range of available roughness can be understood from the area covered by grey curves. As explained before, once the base model is trained using the initial randomly selected data set, it is used to determine which samples from the repository $\mathcal{U}$ should be selected for the next round of training. In figure 5.4(a), the green line shows the prediction variance σ_{k_r} of all roughness samples in the repository based on the base model. Here, the abscissa is the sample number sorted from high to low σ_{k_r} values. According to the AL framework, the samples selected for the next round are the ones with the largest σ_{k_r}. These are shown in red in figure 5.3. For comparison, a second sampling strategy (denoted as EQ) is employed in which the same number of samples as in AL are selected, but they are distributed equidistantly along the abscissa of figure 5.4(a). These samples are shown in blue in figure 5.3. It is observed clearly in figure 5.3 that the AL model explores surfaces that are least similar to those in the initial data set (green) and tend to cover the entire repository, with a higher weight given to the marginal cases. Furthermore, the parameter distribution as well as the corresponding k_r values of the selected roughness by means of AL and EQ is compared in the insets of figure 5.4. It can be seen that both the AL and EQ models generally prioritize selecting samples within the waviness regime i.e. effective slope ES<0.35 [53]. This preference may arise from the fact that the resulting drag in the waviness regime (ES<0.35) is sensitive to changes in ES [15]. Conversely, beyond this regime (ES >0.35), the resulting ΔU^+ saturates in relation to increasing ES, making these samples less interesting for both labelling strategies. On the other hand, the AL model particularly tends to sample the roughness with positive skewness and low correlation length. This can similarly be a result of the roughness effect being highly sensitive to the variations in roughness statistics within these ranges of parameters, which is in line with previous findings [15, 44].
Subsequently, two separate models are trained based on the AL and EQ strategies. These models are

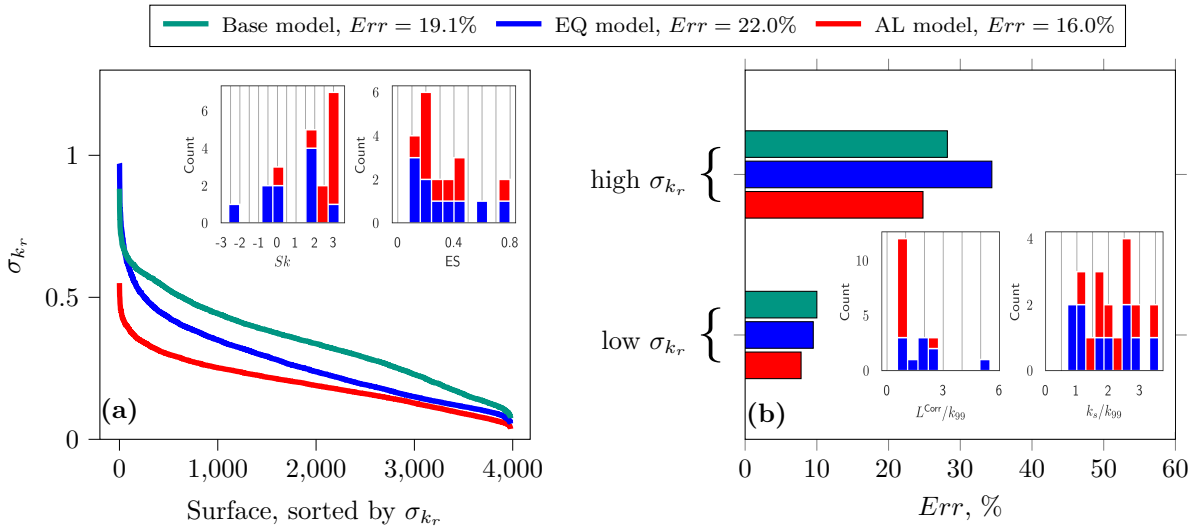


Figure 5.4: (a) Prediction variance σ_{k_r} obtained by three different models for all the samples in repository $\mathcal{U}$. (b) The average error obtained by the three models for 10 high-variance samples and 10 low-variance samples in $\mathcal{T}_{\text{inter}}$ (sorted based on the variance of the base model). The total averaged errors are displayed in the legend. Insets: The distribution of the statistical parameters as well as the corresponding k_r of the new samples with AL and EQ sampling strategies with identical colour code. Figure adapted from [Yang et al. 2023a]

applied separately to determine the variance of prediction for roughness in the repository, and the results are depicted in figure 5.4(a) using red and blue lines. It is evident from the results that both the AL and EQ models generally reduce the prediction variance. However, a more substantial decline in the values of σ_{k_r} is achieved by the AL model. This is the expected behaviour as AL is designed to reduce the prediction uncertainty by targeting regions of the parameter space where the uncertainty is the largest. Interestingly, some increase in σ_{k_r} of the EQ model can be observed for a number of samples with very high σ_{k_r}, which can be a sign that the performance of the EQ model in the 'difficult' tasks deteriorates as it is not trained well for those tasks due to ineffective selection of its training data. Moreover, the prediction errors (calculated based on correct k_s values of testing data set $\mathcal{T}_{\text{inter}}$ obtained by DNS) are illustrated in figure 5.4(b). The averaged prediction errors, Err, achieved by the base model, the AL model and the EQ model for the entire $\mathcal{T}_{\text{inter}}$ are 19.1%, 16.0% and 22.0%, respectively. While the AL model yields a meaningful reduction in Err, the overall performance of the EQ model deteriorates, possibly due to the over-fitting, which in our case refers to the condition where the model is trained to fit a limited number of relatively similar data points so precisely that its ability to extrapolate on dissimilar testing data is degraded [174]. To better analyse this observation, the testing data set $\mathcal{T}_{\text{inter}}$ is split evenly into two subsets according to their σ_{k_r}, namely the high- and low-variance subsets. The Err values for both high- and low-variance subsets are illustrated in the figure. It is clear that while the EQ strategy improves the model performance for the already low-variance test data, its error increases for high-variance test data, which can be taken as an indication of over-fitting as described above. The AL sampling strategy, in contrast, seems to protect the model from over-fitting – especially in the circumstance of a small training data set – hence the error is reduced for both high- and low-certainty test data as a result of effective selection of training data."[Yang et al. 2023a]

5.3 Evaluation of Final Model

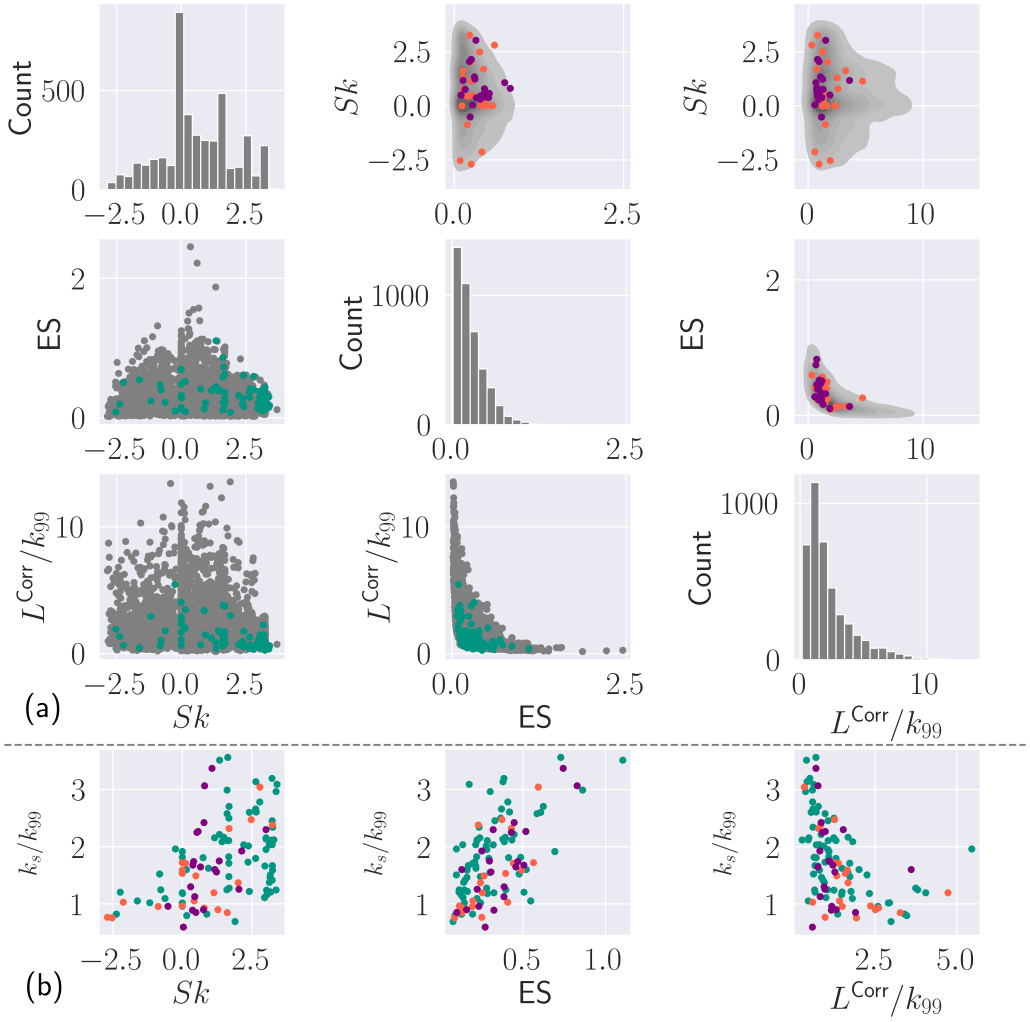


Figure 5.5: (a) Pair plots of roughness statistics. Lower left: The distribution of the samples in $\mathcal{U}$ (grey) and $\mathcal{L}$ (green). Diagonal: Histogram of single roughness statistics in $\mathcal{U}$. Upper right: Joint probability distribution of statistics overlaid by test data in $\mathcal{T}_{\mathrm{inter}}$ (orange) and $\mathcal{T}_{\mathrm{ext,1\&2}}$ (purple). (b) Values of $k_r = k_s/k_{99}$ obtained from DNS (ground truth) as a function of the selected statistics. colour code is same as in (a). Figure adapted from [Yang et al. 2023a]

"Having demonstrated the advantage of AL over random sampling, three additional AL iterations are carried out. The distributions of the PS and p.d.f. of the selected roughness from the second to the fourth AL iterations are displayed in figure 5.3 with black lines. A number of roughness maps from each AL round are also displayed in Appendix A.2.
The total number of data points for training of the model after four iterations adds up to 85; these are the data that form $\mathcal{L}$. The scatter plots of some widely investigated roughness parameters in $\mathcal{L}$ as well as in the unlabelled repository $\mathcal{U}$ are displayed in the lower left part of figure 5.5(a).[...] An inverse correlation can be observed between L^{Corr} and ES, which is expected as roughness with larger dominant wavelength tends to have lower mean slope. The distribution of other statistics in $\mathcal{U}$ appears to be reasonably random.
For the sake of comparison, additionally the test data are represented in the upper right part of the plots with orange (for $\mathcal{T}_{\mathrm{inter}}$) and purple (for $\mathcal{T}_{\mathrm{ext,\ 1\&2}}$) symbols. It is worth noting that only the roughness samples that locate in the fully rough regime at the currently investigated Re_τ are included in $\mathcal{L}$ and

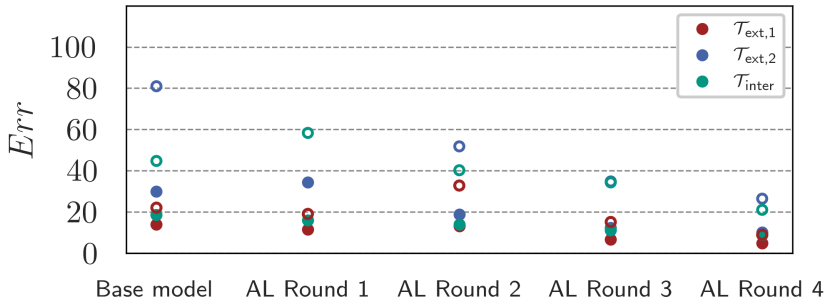


Figure 5.6: The arithmetically averaged Err [%] as well as maximum Err of the model after different training rounds on each of the testing data sets $\mathcal{T}_{\text{inter}}$, $\mathcal{T}_{\text{ext,1}}$, $\mathcal{T}_{\text{ext,2}}$. The mean Err is represented with closed circle while the maximum Err is displayed with open circle with corresponding colour. The maximum Err for $\mathcal{T}_{\text{ext,2}}$ at AL Round 1 is out of the plot range. Figure adapted from [Yang et al. 2023a]

shown in the figure. Figure 5.5(b) shows the values of k_r (from DNS) against the three roughness statistics for all labelled data in the training and testing data sets. As can be observed clearly in the figure, while equivalent sand-grain roughness shows some general correlation with each of these statistics (increasing with Sk and ES, decreasing with L^{Corr}), the collapse of data is far from perfect. Clearly, no roughness statistics can capture entirely the effect of an irregular multi-scale roughness topography on drag, which is essentially a motivation behind seeking an NN-based model to find the functional relation between k_s and a higher-order representation of roughness (here p.d.f. and PS).

Eventually, the final model is trained on the entire labelled data set $\mathcal{L}$. The mean and maximum error values achieved by this model on all three testing data sets, as well as those errors after each training round, are displayed separately in figure 5.6. The figure shows a generally decreasing trend in both mean and maximum error as the model is trained progressively for more AL rounds, despite some exceptions to the general trend in the first two rounds when the number of data points is low. It is notable that the AL model is particularly successful in bringing down the maximum error, and hence can be considered reliable over a wide range of scenarios.

One should mention that the model performs consistently well for three different testing data sets with different natures. While the data set $\mathcal{T}_{\text{inter}}$ covers an extensive parameter space – hence containing more extreme cases – it is generated employing the same method as the training data. Therefore, to avoid a biased evaluation of the model, two 'external' testing data sets from literature are also included. The data set $\mathcal{T}_{\text{ext,2}}$ is believed to be particularly challenging for the model, since it is formed by roughness generated artificially using discrete elements [21], which is fundamentally different from the target roughness of this study. Nevertheless, the final model yields very similar errors for all data sets; what can be taken as an indication of its generalizability. The averaged errors of the final model within the data sets $\mathcal{T}_{\text{inter}}$, $\mathcal{T}_{\text{ext,1}}$, and $\mathcal{T}_{\text{ext,2}}$ are approximately 9.3%, 5.2% and 10.2%, respectively.

It is crucial to acknowledge that the present model is developed under the assumption of statistical surface homogeneity. However, when reaching beyond this assumption, the presence of surface heterogeneity introduces additional complexity to the problem that cannot be represented adequately by the current training samples. As a consequence, the effect of heterogeneous roughness structures [25, 175] cannot be accounted for adequately by the current model."[Yang et al. 2023a]

5.4 Data-Driven Exploration of Drag-Relevant Roughness Scales

5.4.1 Layer-Wise Relevance Propagation

"The fact that naturally occurring roughness usually has a multi-scale nature with continuous spectrum is well established [176]. How spectral content of roughness affects skin-friction drag and whether a certain range of length scales dominates it, are, however, questions receiving attention more recently [55, 139, 177, 178]. In this sense, Busse *et al.* [56] applied low-pass Fourier filtering to a realistic roughness and observed no significant effect on skin-friction drag when the filtered wavelengths were lower than a certain threshold. On the other hand, Busse *et al.* [55] used high-pass filtering and suggested that very large length scales may not contribute significantly to drag. Alves Portela *et al.* [57] examined three filtered surfaces, each maintaining one-third of the original spectral content associated with large, intermediate or small scales. In all cases, the filtered scales were shown to include 'drag-relevant' information. While both lower and higher limits of drag-relevant scales (if they exist) can be a matter of discussion, the present study focuses mainly on the latter. Possibly related to that question, Schultz and Flack [15] documented the equivalent sand-grain size of pyramid-like roughness with wavelengths higher (hence lower effective slopes) than a certain value not to scale in the same way as those with smaller wavelengths. These authors coined the term 'wavy' for the high-wavelength roughness behaviour. Later, Yuan and Piomelli [179] revealed that the wavy regime may emerge at a different threshold (in terms of effective slope) in a multi-scale roughness compared to the single-scale pyramid-like roughness."[Yang et al. 2023a] Furthermore, as shown in chapter 3 that the spectral coherence of roughness topography and time-averaged roughness surface force drops with increasing streamwise wavelengths, which, in line with the finding of Barros *et al.* [55], suggests decreasing drag relevance of large scales.

In this section, we aim at explore the possibility of extracting the drag-relevant scales from the knowledge embedded in the data-driven model leveraging LRP technique. Following the description of LRP in section 2.4.4, the input roughness information, i.e. p.d.f. and PS, are passed through each MLP-member, the contribution score is subsequently assigned to unity at the output neuron and propagated backwards until the input neurons according to Eqn. 2.74. " Note that the contribution scores shown in this section are averaged over the 50 NN members.

In order to extract drag-relevant scales, we consider the following idea. A wavelength that does not affect k_s (which is a measure of added drag) still contributes to an increasing variance of the roughness height, and hence k_{99}. Therefore, the related output of the NN, which is the ratio k_s/k_{99}, is decreased. An input that decreases the output shows a negative LRP contribution score. With that in mind, figure 5.7 shows three exemplary roughness samples (named A, B and C) and their discretized PS. Each discrete wavenumber in a PS is an input to the model, thus has a contribution score, which is indicated using the specified colour code. The spectra are shown in pre-multiplied form, and the p.d.f. of each roughness is also displayed. Samples with both Gaussian and non-Gaussian p.d.f.s are included. It is observed in figure 5.7 that the small wavenumbers (i.e. large wavelengths) generally have more negative contribution scores, which is in accordance to the suggestion of Barros *et al.* [55]. Indeed the most negative contributions belong consistently to the largest wavelengths for all samples. On the other hand, smaller wavelengths generally show larger contribution scores, but the trend is not monotonic. This might indicate that drag-relevant scales reside within a certain range of the spectral content."[Yang et al. 2023a]

"In order to examine whether or not negative LRP contribution score indeed indicates drag irrelevance, we apply high-pass filtering to the samples in figure 5.7, and examine the resulting roughness using DNS under the same conditions as for the original roughness. The position of the filter is chosen to be the largest wavelength with non-positive contribution score (a three-point moving average is applied to smooth the LRP scores beforehand). Figure 5.8 shows the height map of original versus filtered samples, the spectra with filter positions, and the inner-scaled mean velocity profiles before and after filtering for samples A, B and C. Some statistical properties of all original and filtered samples are also displayed in table 5.2. It is clear from figure 5.8 that the velocity profiles of original and filtered samples collapse very well in the logarithmic region and beyond, which obviously leads to similar values of the roughness function and the drag coefficient. This observation lends support to the hypothesis that the large roughness scales beyond a threshold do not have a meaningful contribution to the added drag, and that LRP analysis can be a data-driven route to identifying those scales *a priori*. One obvious application of this finding can be in

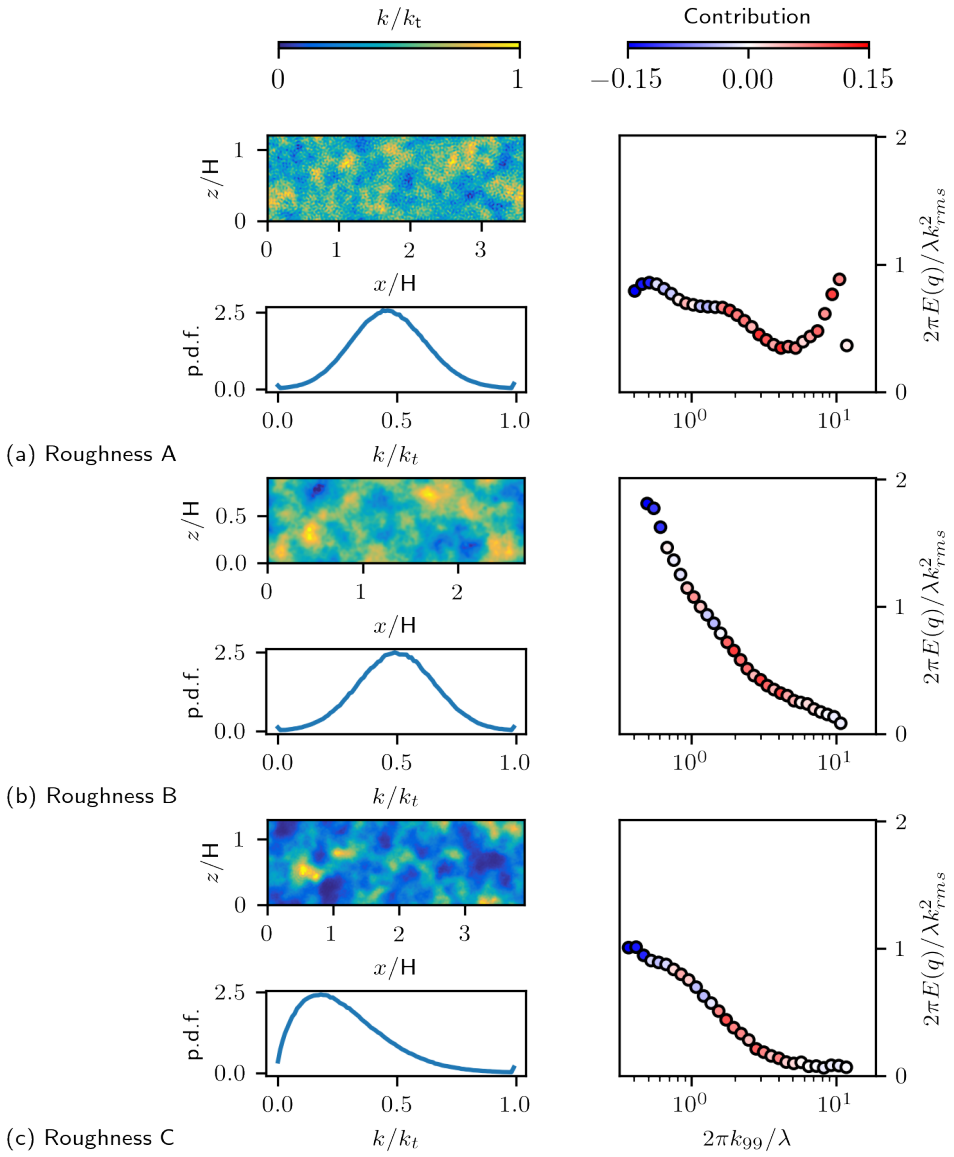


Figure 5.7: Height maps, p.d.f.s and discretized colour-coded pre-multiplied roughness height PS of three exemplary samples A (a), B (b), and C (c). The PS are coloured by the LRP contribution scores. Figure adapted from [Yang et al. 2023a]. $E(q) = E_k(q)$ is roughness height PS.

	k_{md}/H	k_{99}/H	Sk	ES	L^{Corr}/H
Roughness A original	0.043	0.076	0.149	0.520	0.149
Roughness A filtered	0.052	0.065	0.115	0.519	0.092
Roughness B original	0.041	0.069	0	0.146	0.198
Roughness B filtered	0.024	0.041	0.170	0.128	0.102
Roughness C original	0.026	0.078	0.885	0.129	0.234
Roughness C filtered	0.034	0.056	0.307	0.122	0.126

Table 5.2: Statistical properties of selected surfaces A, B and C. Data adopted from [Yang et al. 2023a]

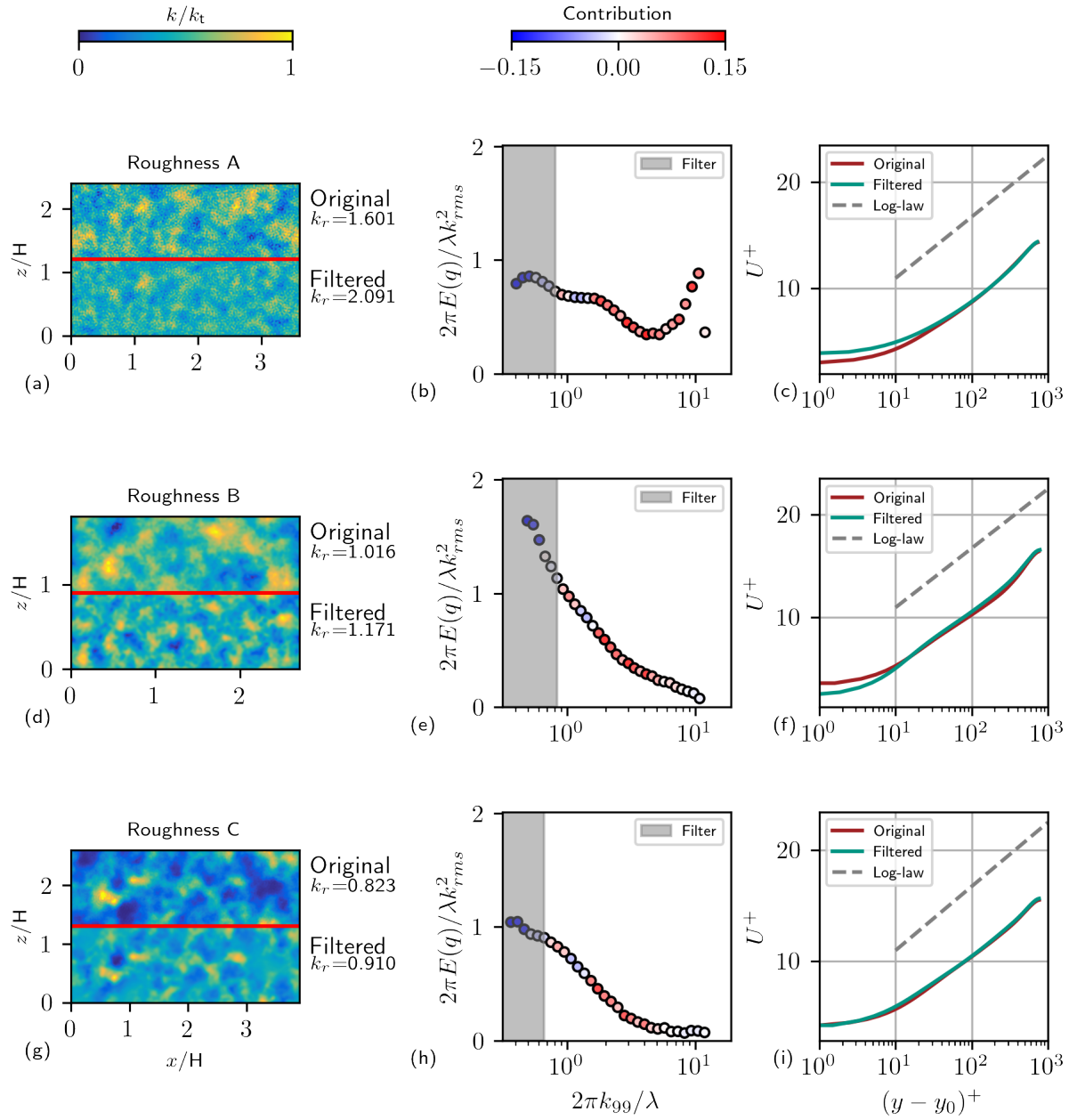


Figure 5.8: The original and high-pass filtered roughness (a, d, g), the pre-multiplied roughness height PS with the filtered scales indicated by grey shade (b, e, h), and inner-scaled mean velocity profiles out of DNS on the original and filtered roughness (c, f, i). Note that the DNS are carried out in minial channels. Figure adapted from [Yang et al. 2023a]. In this figure, $y_0 = d$ represents zero-plane displacement. $E(q) = E_k(q)$ is roughness height PS.

selection of sampling size for the investigations of roughness effect. In practice, it is not always possible to obtain roughness samples that are large enough to encompass the full spectrum of scales. However, once the range of drag-relevant scales is covered completely by a roughness sample, a miscalculation due to a limited sample size can be avoided.
Interestingly, in all samples shown in figure 5.8, the filtered scales have a significant contribution to the roughness height variance based on the pre-multiplied roughness spectra. This is also reflected in the significant decrease in roughness height k_{99} and L^{Corr} in table 5.2, as anticipated. Additionally, based on the three observed cases, the reductions in the ES values are found proportional to the filtered fraction of the PS. Other statistical parameters also undergo changes due to filtering, while obviously none of these changes is relevant in determining the drag. It is worth noting that in addition to the roughness height k_{99} and ES, other drag-determining quantities, such as Sk, undergo a general reduction for roughness C. According to some existing empirical correlations (e.g. the correlations proposed by Chan *et al.* [20], Forooghi *et al.* [14] and Flack *et al.* [19]) the simultaneous reduction in these statistics should lead to a lower k_s. This is, however, not the case in reality based on the DNS results which can be reminiscent of the suggestion by Barros *et al.* [55] that a high-pass filtering is necessary if predictive correlations are to capture the correct trend between k_s and the roughness statistics. This also provides an indication for the hypothesis that while statistical parameters can correlate the equivalent sand-grain size of irregular roughness to some degree, only a combined statistical-spectral approach can fully capture the physics of roughness-induced drag.
Furthermore, it is observed in figure 5.8 that the mean velocity profiles of original and filtered can exhibit some deviation very close to the wall. These deviations can be attributed to the altered volume occupied by roughness close to the wall, as reflected by their k_{md} and Sk values. However, these do not seem to have a significant influence beyond the region occupied by roughness. "[Yang et al. 2023a]

5.4.2 Time-Averaged Flow over Drag-Relevant Structures

"To shed further light on why the filtered large scales do not contribute to added drag, exemplary x-y planes of the time-averaged streamwise velocity field are examined in figure 5.9. The overlaid white contour lines are iso-contours of streamwise mean velocity $\overline{u} = 0$, which mark the regions of reversed flow. As expected, roughness A exhibits relatively frequent flow recirculation due to larger local surface slope. In contrast, the occurrences of flow separation over roughnesses B and C seem to be less frequent, which could be linked to the waviness characteristics [15] and less dominant form drag as a result of the low surface slope. When comparing the flow fields over filtered and original roughness, it is evident that the locations of flow recirculation are the same, and filtering has a minimal impact on the extent of reversed flow regions. Moreover, in figure 5.9, red contours are used to show the blanketing layer, which following [54], is defined as the flow region confined by iso-surfaces of $\overline{u}^+ = 5$. On a smooth wall, the blanketing layer would be identical to the viscous sub-layer, while on a rough wall, it can be an indication of how the near-wall flow adapts to the roughness topography. Similarly to the observations in [54], one can observe in figure 5.9 that the blanketing layers in the cases shown do not follow the small roughness scales and steep roughness patterns. This behaviour can be recognised better if the 'depth' of the blanketing layer, i.e. $\Delta D_{\overline{u}^+=5}(x,z) = y_{\overline{u}^+=5}(x,z) - k(x,z)$, is considered.
The maps of $\Delta D_{\overline{u}^+=5}(x,z)$ are shown in figure 5.10, where a visual inspection reveals relative insensitivity of the blanketing layers to the smaller scales of roughness topography (which appear when the roughness height is subtracted from the $\overline{u}^+ = 5$ iso-contour height). Interestingly, in the same figure, a fair level of similarity is observed between the $\Delta D_{\overline{u}^+=5}(x,z)$ maps of the corresponding original and filtered cases. This can be a hint that the blanketing layer has adapted to the filtered scales. "[Yang et al. 2023a]
Subsequently, the relation between filtered scales and blanketing layer is examined through a spectral analysis of blanketing layer depth maps $\Delta D_{\overline{u}^+=5}(x,z)$ for cases A-C. Specifically, "we plot pre-multiplied PS of the blanketing layer depth maps over the original rough surfaces in figure 5.11. For more clarity, the two-dimensional spectra are azimuthally averaged and plotted in green colour. Moreover, the locations of LRP-identified filters and the pre-multiplied roughness PSs are also added to the plots. Interestingly, all spectra show a significant decrease in the contribution of wavelengths larger than (wavenumbers smaller than) the filter. Note that all plots in figure 5.11 belong to the original cases with no influence from the filtering. A wavelength that is present in the roughness topography but absent in the blanketing layer

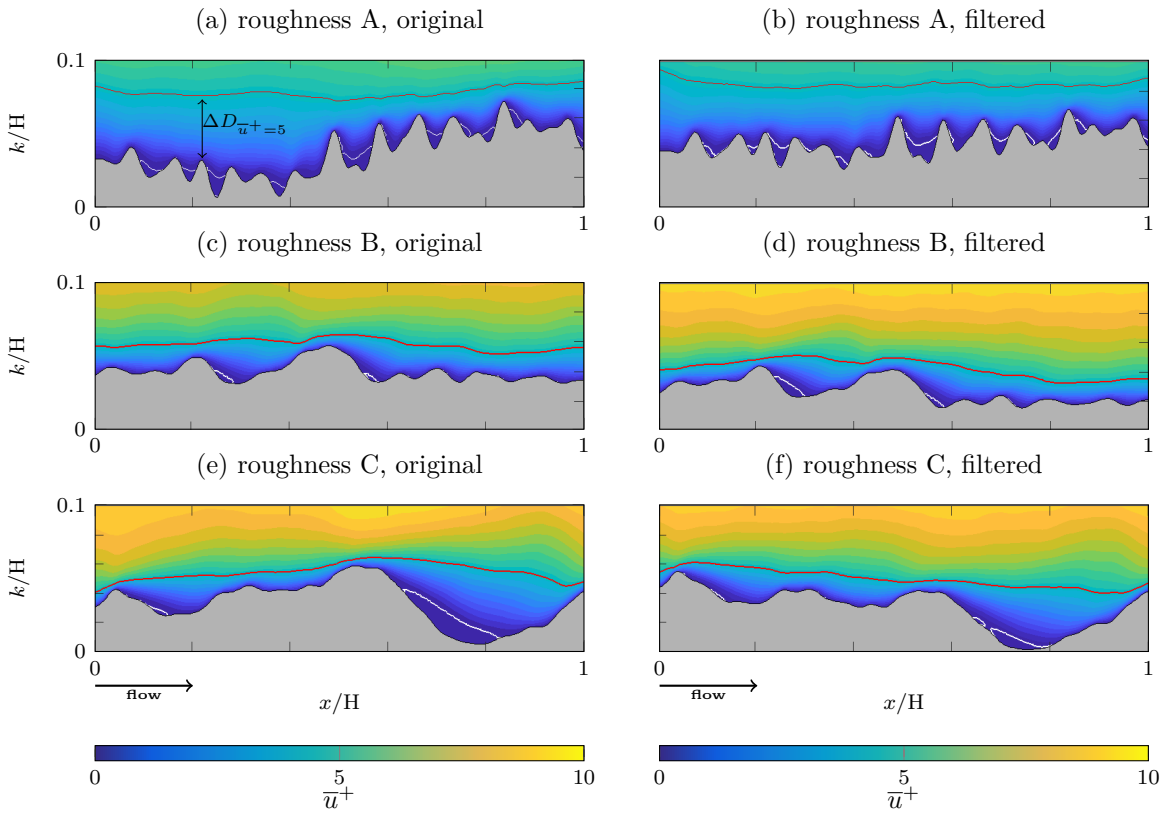


Figure 5.9: Time averaged streamwise velocity distribution $\overline{u}^+$ in selected z-normal plane for the original and filtered cases A-C. The overlaid white contour lines mark the regions of reversed flow ($\overline{u} < 0$). Blanketing layer (iso-contours of $\overline{u}^+ = 5$) is displayed with red contour lines. The grey colour represents the rough structures. The calculation of blanking layer depth $\Delta D_{\overline{u}^+=5}$ is schematically illustrated in (a). Figure adapted from [Yang et al. 2023a].

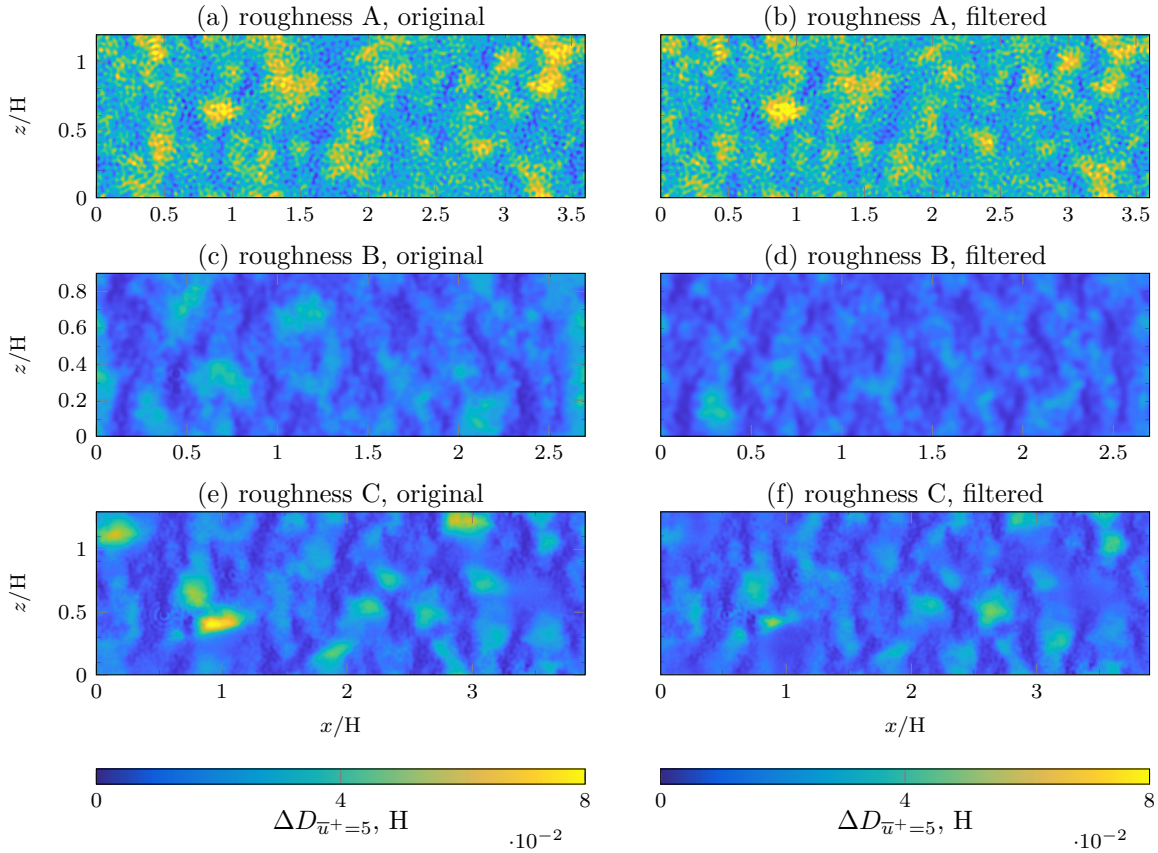


Figure 5.10: Blanketing layer depth $\Delta D_{\overline{u}^+=5}(x,z) = y_{\overline{u}^+=5}(x,z) - k(x,z)$ measured from rough surface for the original and filtered cases A-C. Figure adapted from [Yang et al. 2023a].

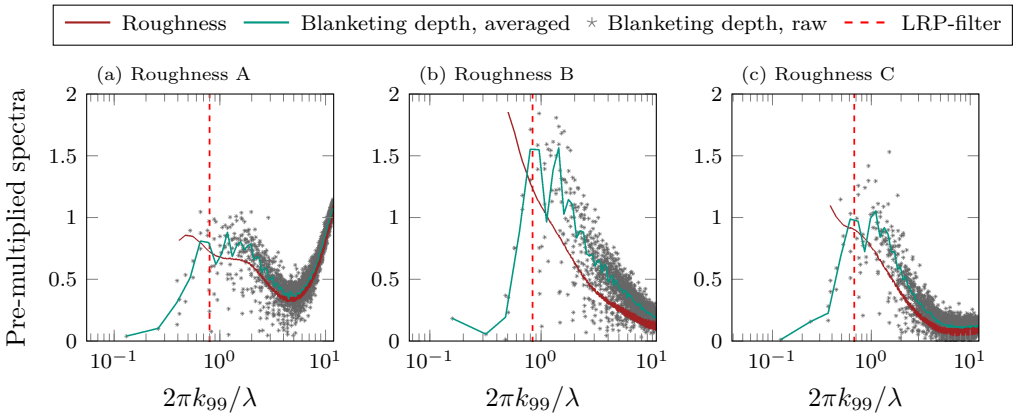


Figure 5.11: Pre-multiplied spectra of blanketing layer depth $\Delta D_{\overline{u}^+=5}$ overlaid with that of the corresponding roughness topography. Symbols indicate the spectrum in different directions while green line show the azimuthal average. The scatter of the symbols indicates the anisotropic characteristics of the map. Structures smaller than the smallest in-plane roughness wavelength λ_1 are omitted. Figure adapted from [Yang et al. 2023a].

depth is one to which the layer has adapted. Therefore, the fact that the spectrum drops for drag-irrelevant scales might suggest that those scales are the ones to which the blanketing layer can adapt."[Yang et al. 2023a]

However, despite the above discussion of the spectral analysis of the blanketing layer depth map, "further systematic investigations are required to establish a conclusive evidence as the present study covers limited ranges of roughness scales and Reynolds numbers. Ideally data on surfaces with more 'drag-irrelevant' large scales and at a much wider range of Reynolds numbers are required to establish a solid hypothesis. Additionally, one should bear in mind that the current results are obtained in the fully-rough regime, and as discussed by Busse *et al.* [54], blanketing layer can behave differently at different regimes."[Yang et al. 2023a]

"As a final remark, a relation between the drag and blanketing layer depth is physically plausible as a change in this depth is generally accompanied by modifications in local flow phenomena (flow separation, strong changes in local velocity gradient on the wall, etc.) that can be linked to added drag. "[Yang et al. 2023a]

5.5 Turbulent Statistics over Original and Filtered Roughness

"In the previous subsection, we used an LRP analysis of the trained model to identify which roughness scales contribute to the added skin-friction drag. While ΔU^+ is arguably the most important flow statistic in the practical sense, due to its relation to drag, roughness also affects higher-order flow statistics, particularly in the so-called 'roughness sub-layer' [8]. In the present study, we focus specifically on comparing the turbulent and dispersive stresses over pairs of unfiltered and filtered samples A, B and C from section 5.4 as the main means of momentum transfer away from the wall."[Yang et al. 2023a]

The velocity fluctuations in rough channels can be decomposed into turbulent and time-averaged spatial fluctuations following the triple decomposition of the velocity field (Eqn. 2.37). "Extrinsic plane-averaging is utilized in the present calculation of statistics, i.e. the solid regions are included in the averaging procedure with zero velocity (similar to e.g. [25, 147]). Based on the triple decomposition, local turbulent stresses $\overline{u_i' u_j'}(x, y, z)$ can be interpreted as measures of momentum transfer due to turbulent fluctuations. Analogous to the local turbulent stresses, one can define the dispersive stresses $<\tilde{u}_i \tilde{u}_j>(y)$ as the momentum transfer due to roughness-induced spatial fluctuations."[Yang et al. 2023a] Furthermore, DA turbulent stresses are calculated through spatial averaging of the local turbulent stresses, i.e. $<\overline{u_i' u_j'}>(y)$.

"The comparison of turbulent stresses for the three considered rough surfaces A, B and C in filtered and unfiltered states are shown in the near-wall region $(y-d)^+ < 200$ in figure 5.12 (red colour). Only a minor difference between original and filtered roughness can be observed for the turbulent Reynolds stresses. For the $<\overline{u'u'}>^+$ component, the peak values are comparable, although, for samples B and C, filtering increases the peak value slightly. Roughness has been shown previously to damp inner-scaled streamwise turbulent stress that can be related to the suppression of elongated near-wall turbulent structures [61, 147]. This effect seems not to be affected significantly by elimination of drag-irrelevant large roughness scales. Moreover, the agreement can be observed for the other two normal turbulent stresses as well as the shear stress ($<\overline{w'w'}>^+$ is not shown for the sake of brevity). Excellent agreement of wall-normal turbulent stresses is reminiscent of the suggestion by Orlandi and Leonardi [180] that the roughness function is related to this component of turbulent stress. Furthermore, the collapse of shear stress profiles is an indication of similarity in the vertical mean momentum transport due to turbulence. The agreement of these components thus contributes to the concordance of the mean velocity profiles in the log-layer.

For the dispersive stresses, it is apparent that the only component affected by filtering of roughness is the streamwise normal component $<\tilde{u}\tilde{u}>^+$, for which the peak values are reduced by filtering. It is worth mentioning that same trend (reduction of the $<\tilde{u}\tilde{u}>^+$ peak values, and agreement of other dispersive stresses) can be observed if an intrinsic averaging approach is used (not shown for brevity). Despite the possible shift of the zero-plane d after filtering, the peak of $<\tilde{u}\tilde{u}>^+$ is observed consistently at the vicinity of the respective zero-planes, i.e., at $(y-d) \approx 0$. The discernible reduction in this peak value suggests a less pronounced inhomogeneity of the mean streamwise velocity when larger wavelengths are filtered. Arguably, the large-scale undulations present in the original roughness lead to large-scale variations in mean velocity, resulting in greater flow inhomogeneity, as also pointed out by Yuan and Jouybari [181].

Despite the fact that the values of dispersive shear stress are small in all cases, a comparison among the three cases shown can provide certain insight into the roughness-flow interactions. As depicted in figure 5.12 (c, f, i), roughness A exhibits a positive $-<\tilde{u}\tilde{v}>^+$ peak, whereas roughnesses B and C display negative peaks. Such a negative sign can be attributed to the 'waviness effect' since a wavy structure (one with relatively low slope) causes an acceleration of the mean flow on the windward side, and a deceleration on the leeward side [57]. Positive $-<\tilde{u}\tilde{v}>^+$, on the other hand, can be linked to recirculation behind steep roughness elements [181]. This is in line with the fact that roughness A has a much larger ES compared to the other two. The collapse of dispersive shear stress profiles in figure 5.12 shows that none of these behaviours is affected by the applied filtering.

The results shown so far indicate that the streamwise dispersive stress is the only second-order one-point velocity statistic affected by filtering of drag-irrelevant scales. This, however, does not modify the shear stress profile as discussed above. To elaborate this finding further, joint p.d.f.s of local dispersive motions in the wall-parallel plane $y = d$ are calculated for all the three samples, along with their filtered counterparts, and shown in figure 5.13. Here, intrinsic averaging is used, meaning that the areas inside roughness are excluded for calculation of the dispersive velocities shown in the joint p.d.f. The subscript in denotes intrinsic averaging. Following the idea of quadrant analysis [182], the $\tilde{u}_{in}^+$–$\tilde{v}_{in}^+$ plane is divided into four

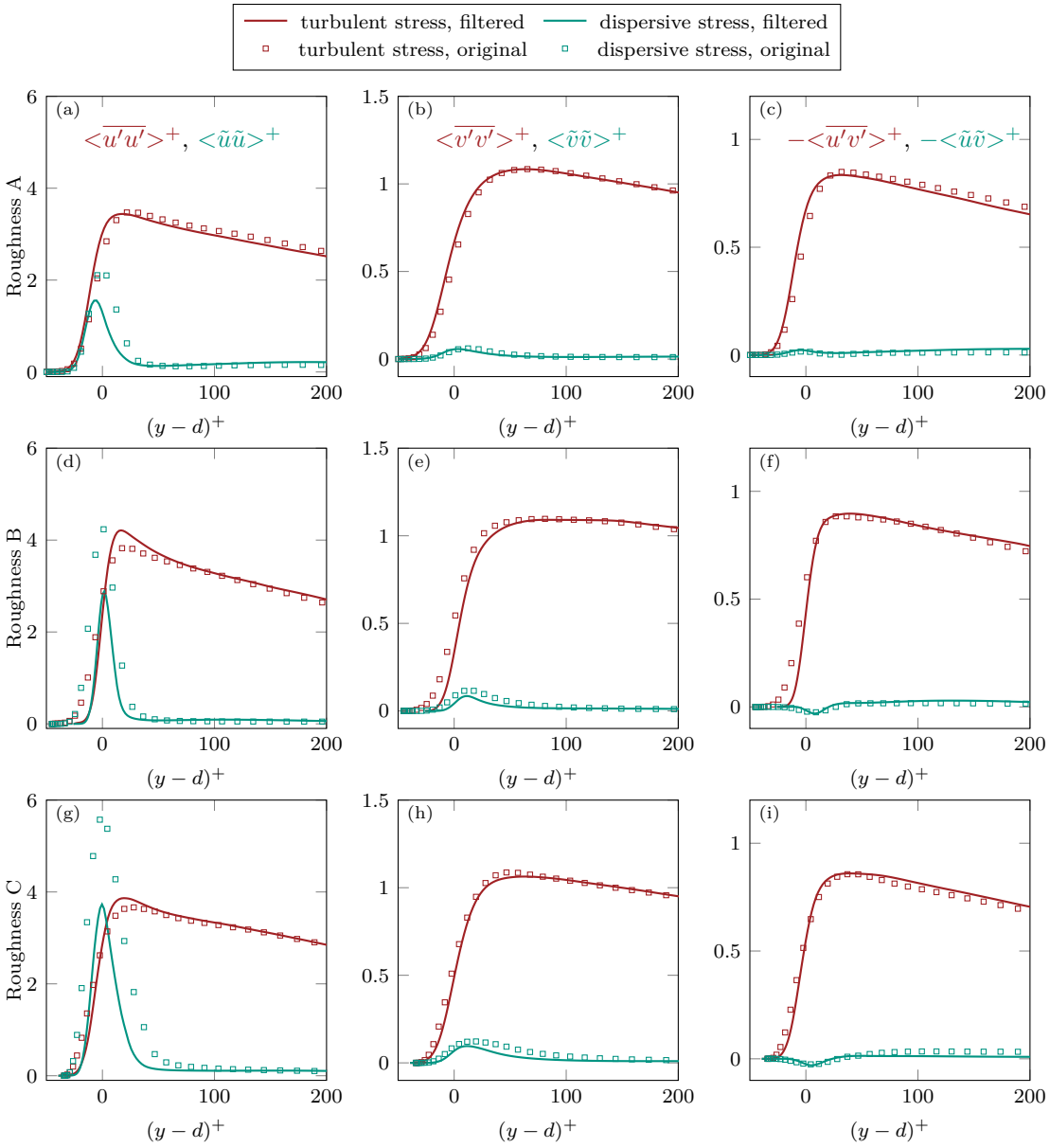


Figure 5.12: DA turbulent and dispersive stresses for Roughness A, B and C. Figure adapted from [Yang et al. 2023a].

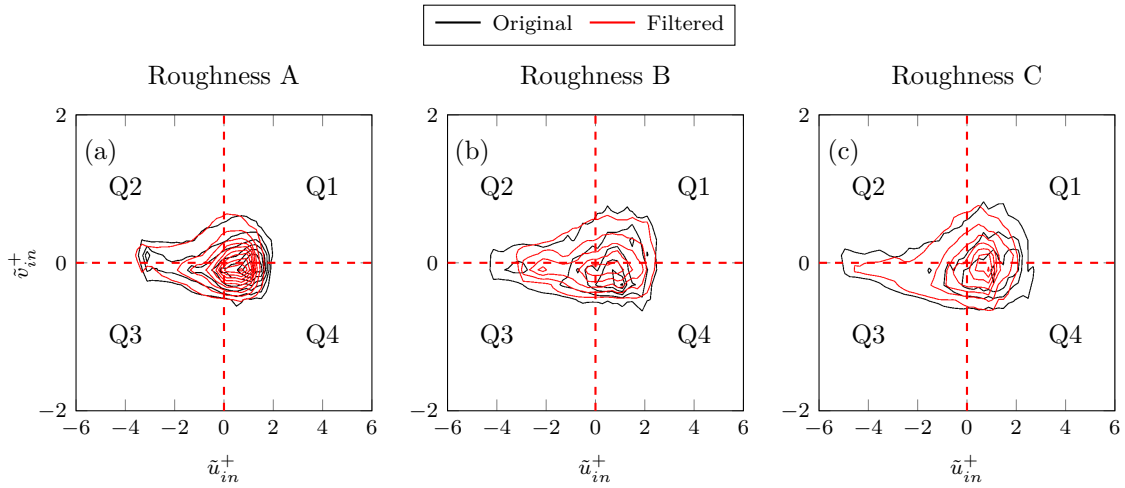


Figure 5.13: Joint p.d.f. of $\tilde{u}_{in}^+$ and $\tilde{v}_{in}^+$ at plane $y = d$, values in roughness excluded. Contour lines range from 0.05 to 1.55 with step 0.1. Subscript in indicates being a result of intrinsic averaging. Figure adapted from [Yang et al. 2023a].

quadrants, Q1-Q4, based on the signs of $\tilde{u}_{in}^+$ and $\tilde{v}_{in}^+$. While the joint p.d.f.s look relatively similar before and after high-pass filtering, it is observed that filtering results in contours shrinking along the $\tilde{u}_{in}^+$-axis. This is in line with the reduction of peak values of streamwise dispersive components discussed before. Notably, the joint p.d.f. retains its near-symmetry with respect to the $\tilde{v}_{in}^+$-axis, which means that reduction of extreme $\tilde{u}_{in}^+$ fluctuations shows no preference in the direction of momentum transfer. This results in the similar shape of the contours apart from horizontal stretching. An obvious outcome is that the shear stress profiles are unaffected by modifications in $\tilde{u}_{in}^+$."[Yang et al. 2023a]

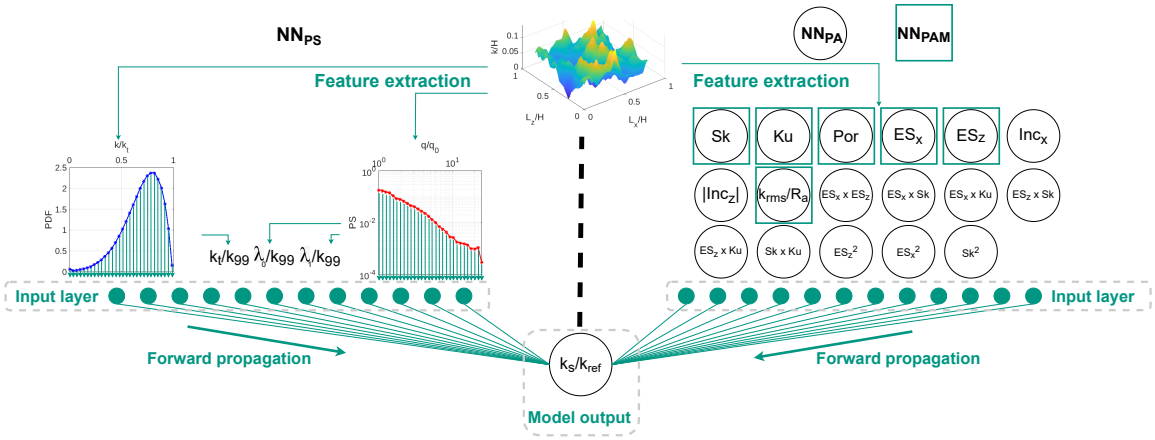


Figure 5.14: Schematic illustration of the prediction process of considered models. Left: $\mathrm{NN_{PS}}$; right: $\mathrm{NN_{PA}}$ and $\mathrm{NN_{PAM}}$. The different input roughness features incorporated in $\mathrm{NN_{PA}}$ and $\mathrm{NN_{PAM}}$ are marked with circles and squares, respectively. The specific architectures of each model are documented in table 5.3. Here R_a is the averaged absolute deviation of roughness from k_{md}. *Por* denotes porosity (the fraction of fluid volume to the entire volume under roughness crest height k_z). $\mathrm{Inc_{x,z}} = \tan^{-1}\{Sk(\partial k/\partial x, z)/2\}$ are the inclinations in x and z directions, respectively. Figure adapted from [Yang et al. 2024]

5.6 Assessment of Different Roughness Statistics as Model Input

While the roughness characterization methods utilized so far (namely p.d.f + PS) are capable of accurate predictions, it's important to acknowledge that obtaining these representation metrics is considerably more challenging practically compared to retrieving conventional single-valued statistical parameters As will be elucidated in the following content, achieving converged yet accurate representations of roughness p.d.f. and PS demands higher measurement accuracy and a larger sampling area. "This difficulty may constrain the industrial deployment of these models compared to the empirical symbolic regression or data-driven models that employ statistical parameters. Moreover, roughness parameters inherently reflect human-understandable roughness properties from various perspectives. Data-driven models constructed based on these parameters hold significant potential of developing explainable white-box models, aligning with the current trend in research known as Explainable Artificial Intelligence (XAI)."[Yang et al. 2024] In light of these considerations, the choice of a predictive model, particularly in terms of roughness characterization statistics, necessitates thorough examination of various factors, extending beyond mere prediction accuracy. Therefore, a comprehensive assessment of different types of statistical characterization metrics as inputs to data-driven models is carried out in this section. The training and testing databases are adopted from the previous section, i.e. $\mathcal{L}$, $\mathcal{T}_{\mathrm{inter}}$ and $\mathcal{T}_{\mathrm{ext,1\&2}}$.

5.6.1 Investigated Machine Learning Models

"As described earlier, the present study compares machine learning models based on two distinct roughness characterization methods. The working principles of the models are displayed in figure 5.14."[Yang et al. 2024] The input layer of $\mathrm{NN_{PS}}$ (left) is adopted from the neural network described in previous section which receives roughness p.d.f. and PS features along with the three additional parameters to account for roughness scaling. Consequently, 63 quantities in total are transferred to the $\mathrm{NN_{PS}}$ model. "On the other hand, $\mathrm{NN_{PA}}$ incorporates a set of 17 roughness parameters along with their product as model input as described in [21]. As demonstrated in the following results, a deterioration in the model performance is achieved for $\mathrm{NN_{PA}}$. The sensitivity analysis reveals that the model complexity, stemming from the high dimensionality of the input vector, results in inaccurate model responses. Having this in mind, a simplified model $\mathrm{NN_{PAM}}$ is developed in the present work. The roughness parameters incorporated by $\mathrm{NN_{PAM}}$ are highlighted by green squares in the right panel of figure 5.14. Notably, all the parameter dot products are excluded in $\mathrm{NN_{PAM}}$. Additionally, the omission of $\mathrm{inc}_{x,z}$ is motivated by the fact that these parameters in

the current training database are consistently zero. Consequently, $\mathrm{NN_{PAM}}$ incorporates only 6 roughness parameters from the 17 input variables of $\mathrm{NN_{PA}}$.
The three models are constructed in the fashion of fully connected MLP. L2 regularization is applied to the loss function to mitigate over-fitting. In the figure, the non-dimensionalization of the k_s value is performed with respect to the reference length scale k_{ref}. To maintain consistency between the input and output of each model, the choice of k_{ref} aligns with the non-dimensionalization procedure applied to the input quantities. Consequently, k_{99} is utilized as k_{ref} for $\mathrm{NN_{PS}}$ while Ra is employed for $\mathrm{NN_{PA}}$ and $\mathrm{NN_{PAM}}$. The non-dimensionalized k_s values are denoted as $k_s^* = k_s/k_{\mathrm{ref}}$. The architectures of each model type are individually determined through BO. The technical details of BO are provided in the subsequent section."[Yang et al. 2024] Following the framework described above, "ensemble prediction technique is applied for each type of models: the ENN models are formed by training 50 NNs members with identical BO-optimized architecture but different combination of training and validation data samples. In the end, The final prediction of the ensemble model is the arithmetic averaged predictions among the 50 NN members. These ensemble models, depending on the type of their NN members, are referred to as $\mathrm{ENN_{PS}} = \{\mathrm{NN_{PS,1}}, \mathrm{NN_{PS,2}}, ..., \mathrm{NN_{PS,50}}\}$, $\mathrm{ENN_{PA}} = \{\mathrm{NN_{PA,1}}, \mathrm{NN_{PA,2}}, ..., \mathrm{NN_{PA,50}}\}$ or $\mathrm{ENN_{PAM}} = \{\mathrm{NN_{PAM,1}}, \mathrm{NN_{PAM,2}}, ..., \mathrm{NN_{PAM,50}}\}$. "[Yang et al. 2024]

5.6.2 Bayesian Optimization of Model Architecture

Table 5.3: Summary of considered hyperparameters during BO process as well as the finally selected hyperparameters for $\mathrm{NN_{PS}}$ and $\mathrm{NN_{PA}}$. Here N_i indicates the number of neurons in i-th hidden layer. Data adopted from [Yang et al. 2024]

Hyperparameter	Range	$\mathrm{NN_{PS}}$	$\mathrm{NN_{PA}}$	$\mathrm{NN_{PAM}}$
N_1	1-256	71	177	256
N_2	1-256	43	252	51
N_3	1-256	215	19	19
Activation function	sigmoid, linear, relu, leaky relu	relu	leaky relu	relu
Learning rate	0.0001-0.1	0.005	0.037	0.05
L2 regularization	0.0001-0.1	0.002	0.1	0.0001
Trainable weights	-	1.7×10^4	5.3×10^4	1.5×10^4

Given the distinct data processing requirements of each considered ENN model, it is unlikely that a single model architecture can ensure optimal performance for all ENNs. The architectures of the MLP members in each ENN including the number of neurons in each layer, and the choice of activation functions, as well as training parameters such as the learning rate and L2 regularization, are thus critical hyperparameters that must be carefully selected / optimized prior to training the model. "The tuning of these hyperparameters is pivotal for realizing the full potential of the model. Nevertheless, the multitude of potential hyperparameter combinations poses challenges in determining the optimal configuration for each model.
The optimization task for the hyperparameters is formulated as:

$$\mathbf{H}_{\mathrm{opt}} = \mathrm{argmin} f_{\mathrm{BO}}(\mathbf{H}) \ , \tag{5.1}$$

where $\mathbf{H} = (H_1, H_2, ..., H_n)$ represents the set of $n > 0$ hyperparameters that are to be designed. The objective function f_{BO} denotes the mean absolute percentage error (MAPE) of the trained model in validation set. In the present work, f_{BO} is calculated by averaging the MAPE values in the fashion of 10-fold validation. In traditional approaches, such as random optimization or grid search optimization, the tuning of hyperparameters is typically performed without considering the knowledge gained from previous trials. As a result, these methods are often considered inefficient. In this study, BO is utilized to overcome these limitations and efficiently explore the admissible space of $\mathbf{H}$. This is done by intelligently exploring the hyperparameter configurations based on past evaluations, leading to faster convergence towards the

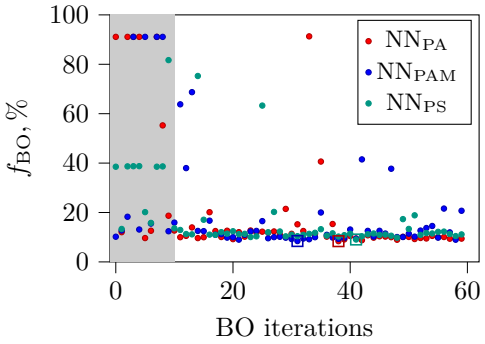


Figure 5.15: History of BO objective function f_{BO}. The BO selected models are marked by open square marks. The grey background indicates random search of BO process corresponding to step 0. Figure adapted from [Yang et al. 2024]

optimal solution. The Bayesian optimization is based on Bayes' Theorem:

$$P(f_{BO}|\mathbf{X}) = (P(\mathbf{X}|f_{BO})P(f_{BO}))/P(\mathbf{X}) \ . \tag{5.2}$$

The data set $\mathbf{X}$ consists of the collection of n observed hyperparameters along with their corresponding f_{BO} values, denoted as $\mathbf{X} = \{\mathbf{H}_1, f_{BO,1}, \mathbf{H}_2, f_{BO,2}, ..., \mathbf{H}_n, f_{BO,n}\}$. $P(f_{BO}|\mathbf{X})$ is the posterior probability or probabilistic surrogate model within the framework of Bayesian optimization, $P(\mathbf{X}|f_{BO})$, $P(f_{BO})$ and $P(\mathbf{X})$ are likelihood, prior probability for f_{BO} and $\mathbf{X}$ respectively. The BO process can be summarized as follows:

0. Initiating BO by randomly testing 10 combinations of hyperparameters and calculate their corresponding f_{BO} values to form the initial data set $\mathbf{X}$;

1. Constructing surrogate function $P(f_{BO}|\mathbf{X})$ using Gaussian process regression (GPR) based on known data set $\mathbf{X}$;

2. Choosing global best hyperparameters $\mathbf{H}_{best}$ of the surrogate function according to a acquisition function, which is randomly selected from lower confidence boundary (LCB), expected improvement (EI) and probability of improvement (PI) for each BO iteration [183];

3. Calculating $f_{BO}(\mathbf{H}_{best})$ and augmenting data set $\mathbf{X}$ by $\mathbf{H}_{best}$, return to step 1.

A total of 60 iterations of BO are performed for each type of model. The optimized hyperparameters, denoted as $\mathbf{H}_{opt}$, are determined by selecting the data point that yields the minimum value of the objective function $f_{BO}(\mathbf{H}_{opt})$ among the 60 iterations. The considered hyperparameters as well as their admissible ranges are summarized in the table 5.3."[Yang et al. 2024]
The following BO process as well as training process of the model are carried out utilizing the training database $\mathcal{L}$. "The evolution of the objective function f_{BO} during the BO processes are displayed in the figure 5.15. As can be seen, after 10 random searching steps, the newly queried f_{BO} values are mostly maintained in a relatively low level, indicating the efficiency of the BO framework in nominating competitive hyperparameter combinations. The respective minimum f_{BO} values for NN_{PS}, NN_{PA} as well as NN_{PAM} are marked with open squares in the figure with corresponding colour. The optimized hyperparameters for each model type are summarized in Table 5.3. Substantial differences in model architectures may stem from variations in processing different types of input quantities as well as from the difficulty of mapping the input quantities to the objective values. Additionally, the markedly higher number of trainable weights required by the optimized NN_{PA} underscores the necessity for increased complexity in recovering predictions from the provided inputs. This observation may suggest the challenge of recovering surface drag given the high dimensionality of the input variables. In clear contrast, the number of trainable weights for NN_{PAM} is considerably smaller. This implies that the modified model necessitates lower model complexity for drag prediction due to the diminished dimensionality of the selected input parameters."[Yang et al. 2024]

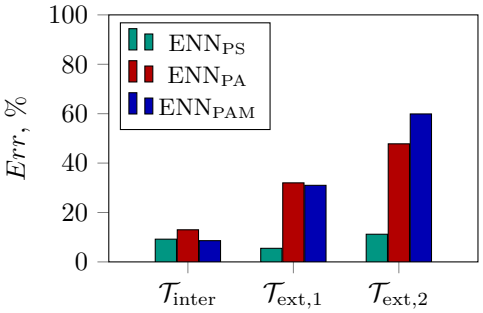


Figure 5.16: Comparison of the MAPE for all the models in different testing data sets. Data adopted from [Yang et al. 2024].

5.6.3 Comparison of Model Performance

"The final ensemble models ENN_{PS}, ENN_{PA} and ENN_{PAM} are built with the respective optimized architectures as documented in table 5.3. These models are tested with the testing data sets $\mathcal{T}_{\text{inter}}$, $\mathcal{T}_{\text{ext,1}}$ and $\mathcal{T}_{\text{ext,2}}$. The averaged percentage prediction errors Err within each testing set are displayed in the figure 5.16. Acceptable performance is achieved by all the considered models in $\mathcal{T}_{\text{inter}}$. While ENN_{PS} achieves an averaged percentage error of 9.2%, ENN_{PAM} exhibits similar performance of 8.6%. The full-sized ENN_{PA} model, though featured with the highest model complexity, achieves the highest averaged error of 13.0%. Interestingly, while the full-sized ENN_{PA} is expected to exhibit performance at least in line with ENN_{PAM} since they share the same subset of input parameters, its actual performance noticeably differs from which of the simplified counterpart. This could be attributed to the diminished training efficiency of ENN_{PA} due to the increased input dimensionality."[Yang et al. 2024]
As can be observed from distribution of roughness parameters within the training and (internal and external)testing databases in Figure 5.5, the training database adequately spans the parameter domain of the external testing database ($\mathcal{T}_{\text{ext,1\&2}}$) in relation to the observed roughness parameters "Consequently, it was expected that the parameter-based models perform similarly for the internal and external test data-sets, as long as the utilized parameters contain adequate information about roughness. However, the results in figure5.16 show that the performance of ENN_{PA} and ENN_{PAM} deteriorate significantly in these testing databases, exhibiting errors higher than 30%. This deterioration in model performance is most likely due to the dissimilarity of the external samples. This can be an indication that rough surfaces with identical values of such parameters but different generation natures may not have same k_s values. In contrast, ENN_{PS} achieves a more consistent performance with an average error of around 10% – despite the underlying dissimilarities in the testing databases. The remarkable generalizability of ENN_{PS} aligns with the fact that roughness p.d.f. and PS provide a more comprehensive representation of topographical features at various length scales compared to the roughness parameters.
In summary, despite the relatively limited generalizability of the both models based on the statistical roughness parameters, it is evident that a similar performance can be achieved by these models within $\mathcal{T}_{\text{inter}}$. This signifies the potential of the parameter-based models in interpolating predictions while lacking the capability of extrapolating their performance to diverse roughness topographies. However, the aforementioned advantages of the characterization method utilizing p.d.f. and PS are accompanied with inherent drawbacks compared to conventional single-valued parameters, particularly in terms of their accessibility. To attain a statistically converged single-valued roughness parameter, the evaluation of a substantial area of roughness samples is necessitated. Similarly, achieving a converged height p.d.f. necessitates calculating of numerous converged frequencies of height within a number of finite small height intervals (bins) along the height p.d.f. axis, i.e. $k/k_t \in [0, 1]$. It is crucial to recognize that these discrete p.d.f. statistics, unlike the independent parameters utilized in the parameter-based models [or empirical correlations], are correlated. Consequently, a larger testing section is necessary to obtain a converged p.d.f. profile. On the other hand, the challenge in measuring PS across a range of wavenumbers critical for drag prediction lies in the necessity of exploring a sufficiently large area of the roughness sample with a reasonably

fine resolution. These practical challenges are concretely reflected in [section 4.3][...] The difficulties in obtaining accurate p.d.f. and PS from a realistic sandpaper stem from issues such as light reflection in white-light interferometry measurements, slow convergence of p.d.f. in perthometer measurement and limited accuracy of PS measurement in 3-D photogrammetry measurements. In conclusion, the utilization of a combination of measurement techniques may be essential for general application scenarios when employing p.d.f. and PS characterization in reality. Therefore, it is essential to appreciate the significant convenience of acquiring the single-valued parameters in a practical sense and to take this aspect into account when selecting suitable type of models. "[Yang et al. 2024]

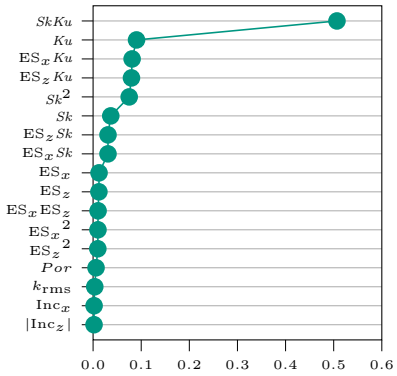
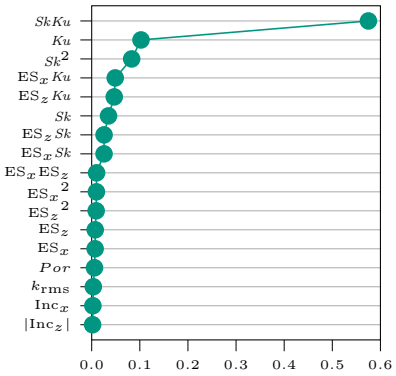
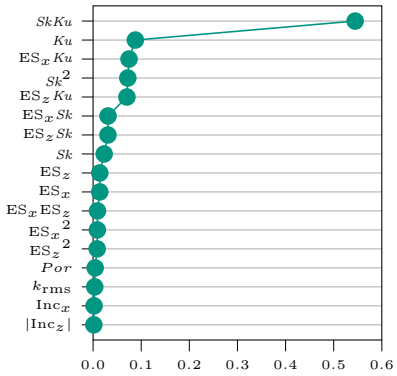
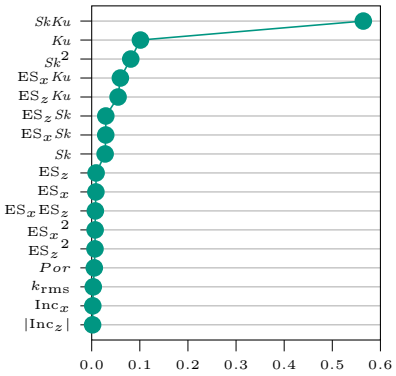
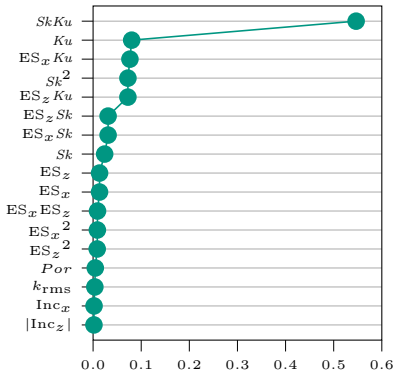
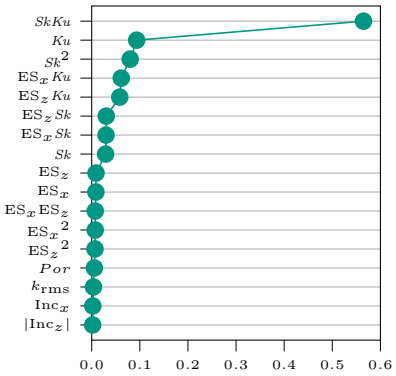
5.6.4 Zonal Sensitivity Analysis

"As previously discussed, the acquisition of single-valued roughness parameters in realistic applications is easier compared to their p.d.f. and PS. For the sake of simplicity, the single-valued roughness statistical parameters are referred to as roughness parameters in the following. Furthermore, the practical utility and convenience of these parameter-based models present a compelling rationale for continuing the study of this type of model. The model evaluation in the previous section reveals different performances between the two parameter-based models. Additionally, the distinct model architectures – as indicated in table 5.3 – suggest variations in the underlying perceptron processes for mapping roughness information into their predictions. A detailed investigation of these differences thus allows to gain insights into the physics of roughness. To this end, the following content focuses on the investigation of the present parameter-based models, i.e. $\mathrm{ENN_{PA}}$ and $\mathrm{ENN_{PAM}}$, by means of sensitivity analysis."[Yang et al. 2024] The sensitivity matrix of each MLP member concerning the input variables is computed following Eqn. 2.73. Subsequently, the sensitivity matrix of the ENN models is arithmetically averaged over their MLP-members. "As is illustrated by the equation, sensitivity analysis requires concrete sample points for calculating the Jacobian matrix. In order to include as large the parameter space as possible, roughness topographies in the repository $\mathcal{U}$ compromising 4200 roughness topographies are utilized. The symbol $\langle\cdot\rangle$ represents the averaging operation over different sample points. The derivative $\partial k^*_{s,i}/\partial I_j$ is calculated using an automatic differentiation method. Subsequently the sensitivity scores are normalized though $S^*_j = S_j/\sum_k S_k$. [...]. This choice is motivated by the fact that all models perform consistently for $\mathcal{T}_{\mathrm{inter}}$.

As reported in the literature, the impact of roughness in distinct morphological regimes may demonstrate varying correlations with roughness parameters [19, 53]. For instance, a significant transition in roughness effect, marked by ES$\approx$ 0.35, delineates the transition in the dominating drag contribution effect from friction drag to form (pressure) drag [15]. Additionally, variations in the drag mechanisms are reflected by Sk, where positive, zero, and negative Sk values exhibit distinct drag exertion mechanisms [43]. To explore these phenomena jointly leveraging the present data-driven perspective, sensitivity analysis is conducted for $\mathrm{ENN_{PA}}$ and $\mathrm{ENN_{PAM}}$ across different ranges of roughness parameters, controlled by the values of ES_x and Sk. This approach facilitates an individual examination of the models' responses to roughness in wavy conditions ($\mathrm{ES}_x < 0.35$) and rough regimes ($\mathrm{ES}_x \geq 0.35$), grouped by the sign of Sk. In this classification, roughness samples are categorized based on their Sk values into three regions, namely positive ($Sk \geq 0.1$), negative ($Sk \leq -0.1$) and near Gaussian ($-0.1 < Sk < 0.1$). The selection of the threshold value, 0.1, is arbitrary and is chosen to encompass a reasonable number of samples in the near Gaussian region. Consequently, the sensitivity analysis in the present work is performed separately for six different zones (2 ES_x zones $\times$ 3 Sk zones) in the parameter space, distinguished based on the values of ES_x and Sk, hence referred to as zonal. It is, however, crucial to acknowledge that the currently proposed parameter criteria for distinguishing different types of roughness stem from the known transition of the roughness behaviour based on these parameters [15, 43]. Obviously, a finer division of the parameter regions based on more parameters can provide a better illustration of the physics of roughness skin friction, especially when extending the analysis to a broader range of roughness types. The exploration of the criteria for classifying similarly behaved roughness based on various parameters can be crucial for developing roughness predictive models, which is called in the future investigations.

The result of zonal sensitivity analysis for $\mathrm{ENN_{PA}}$ is illustrated in table 5.4. The sensitivity analysis outcomes for further ranges of roughness in the database display notable similarities. As can be observed from the figure, the current model identifies only half of the considered roughness parameters as the influential inputs, while the remaining half of the roughness parameters demonstrate nearly negligible

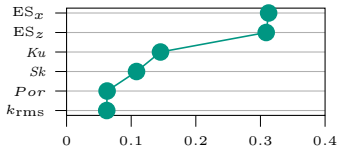
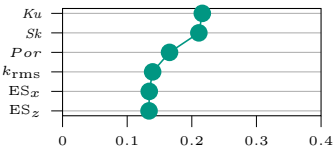
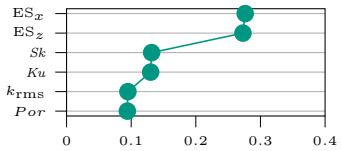
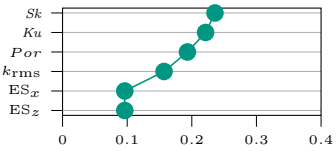
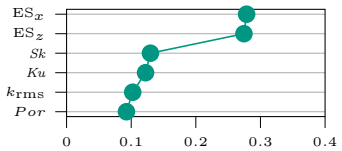
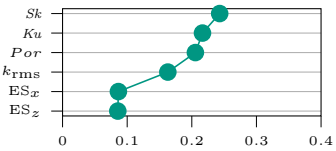
Table 5.4: Zonal sensitivity analysis of model $\mathrm{ENN_{PA}}$. Data adopted from [Yang et al. 2024].

	Wavy roughness $\mathrm{ES}_x < 0.35$	Steep roughness $\mathrm{ES}_x \geq 0.35$
$Sk \geq 0.1$		
$0.1 > Sk > -0.1$		
$-0.1 \geq Sk$		

sensitivity to the prediction. Wherein, markedly high importance is attributed to the input variable $Sk \times Ku$ across different zones of the database. This peculiar behaviour may stem from the reduced training efficiency attributed to the enhanced complexity of the model arising from the high dimensionality of the input. Consequently, a diminished performance of the model within $\mathcal{T}_{\mathrm{inter}}$ is observed.

Table 5.5 presents zonal sensitivity analysis results for $\mathrm{ENN_{PAM}}$, offering insightful indications regarding the different impact of roughness across different roughness types. It is evident that, for roughness with $\mathrm{ES}_x < 0.35$, higher sensitivity in the prediction is associated with the $\mathrm{ES}_{x,z}$ values. However, the significance of ES in this region diminishes with increasing Sk values. A plausible interpretation of this behaviour is that positively skewed roughness corresponds to peak-dominant roughness, and the ES in the waviness

Table 5.5: Zonal sensitivity analysis of model $\mathrm{ENN_{PAM}}$. Data adopted from [Yang et al. 2024].

	Wavy roughness $\mathrm{ES}_x < 0.35$	Steep roughness $\mathrm{ES}_x \geq 0.35$
$Sk \geq 0.1$		
$0.1 > Sk > -0.1$		
$-0.1 \geq Sk$		
	S^*	S^*

regime directly influences the contribution of form drag (pressure drag) downstream of roughness peaks [15]. In contrast, the near-Gaussian and negatively skewed roughness contains an increasing portion of the wall structures in the form of pits, which are prone to generate stable vortices inside the vacancy and thus result in less pressure drag. It is worth mention that the comparable sensitivity of ES_x and ES_z may arise from the inherent constraint of the current isotropic roughness training database. It is anticipated that, with the inclusion of anisotropic surfaces, the model's response to ES_x and ES_z could differ [49, 60]. In contrast, for the roughness in rough regime, i.e. $\mathrm{ES}_x \geq 0.35$, the sensitivity of the model to ES is ranked as the last. This aligns with findings in the literature [15, 53], where it is reported that the contribution of pressure drag saturates with increasing ES within this range of values. As a consequence, the roughness effect exhibit lowest sensitivity to ES for the negatively skewed steep roughness. Furthermore, the variation of effect of roughness properties against the evolution of Sk values can be observed. A discernible decrease in sensitivity as Sk increases can be observed from the figures. This saturation effect of Sk is in line with literature [44]. As a result, steep roughness with negative skewness exhibits the highest sensitivity to Sk, whereas positively skewed wavy roughness attains the lowest sensitivity ranking. These sensible model responses across different database divisions may be a crucial factor contributing to its outstanding performance in $\mathcal{T}_{\mathrm{inter}}$. The successful training of this model can be ascribed to its lower complexity and reduced dimensionality of the input, facilitating efficient training with the current limited training data sets.

In conclusion, the zonal sensitivity analysis reveals that the full-sized $\mathrm{ENN_{PA}}$ demonstrates similar behaviour across the observed database, with sensitivity concentrated on a few input variabales. This observation suggests a potential insufficiency of training samples for such a complex model. In contrast, the simplified model, $\mathrm{ENN_{PAM}}$, exhibit physically explainable model response in different zones, attributed to improved training efficiency resulting from reduced model complexity. However, it is possible that furnishing the full-sized model $\mathrm{ENN_{PA}}$ with an ideally comprehensive training database may enable the model to potentially achieve a physical response. However, given the constraints of the current limited database, which is in close alignment with the challenges in this field of research, the preceding discussion carries significant practical implications for dealing with the scarcity of roughness data. The present analyses under the constraint of limited available data size, suggest the potential for accurate drag predictions on specific types of roughness using low-complexity models trained on similar roughness samples."[Yang et al. 2024]

5.7 Summary

In this chapter, a novel ML ENN model is developed for predicting the normalized equivalent sand-grain height $k_r = k_s/k_{99}$ of homogeneous irregular roughness based on roughness p.d.f. and PS. "The model is developed within the AL framework to effectively reduce the required amount of training data. This framework searches for roughness samples with the highest prediction variances σ_{k_r} in an unlabelled repository $\mathcal{U}$ of 4200 samples. Eventually, a labelled data set $\mathcal{L}$ comprising 85 AL-selected samples is constructed and utilized to derive the ENN model. The significant improvement of the learning efficiency of the model through AL is demonstrated by comparing it with a non-AL approach. Furthermore, it is observed that the employment of AL serves to effectively mitigate the deleterious effects of over-fitting, as evidenced by the observed general drop in prediction error. The mean prediction errors of the final AL-ENN model for an internal testing data set $\mathcal{T}_{\text{inter}}$, as well as two external data sets containing both realistic and artificially generated roughness, $\mathcal{T}_{\text{ext, 1}}$ and $\mathcal{T}_{\text{ext, 2}}$, are 9.3%, 5.2%, and 10.2%, respectively. The consistently good predictions for testing data with different natures can be taken as a sign that a universal model is approached.

Moreover, novel physical insights on the interactions between roughness and turbulent flow are sought by exploring the information embedded in the data-driven model. To this end, the LRP technique is employed to evaluate the contributions of different wave-numbers in the discretized roughness PS towards the predicted value k_r. The PS content identified with positive contribution according to the LRP are interpreted as 'drag-relevant'. Subsequently, high-pass filtering is used to exclude the drag-irrelevant scales, and based on the DNS results for exemplary cases, it is observed that despite the considerable variations in the roughness appearance and statistics, the mean velocity profiles of these high-pass filtered samples collapse well into the original samples in the logarithmic layer, thus having the same k_s values. The LRP-identified drag-irrelevant structures are further studied through an analysis of the behaviour of the blanketing layers over filtered and original roughness. Similarity is observed when maps of blanketing layer 'depth' $\Delta D_{\overline{u}^+=5}$ of filtered and original roughness are compared. This can indicate that the blanketing layer can adapt to the drag-irrelevant scales."[Yang et al. 2023a] Spectral analysis is performed on the blanketing layer depth $\Delta D_{\overline{u}^+=5}$, revealing a remarkable coincidence between the peak of the pre-multiplied power spectrum of $\Delta D_{\overline{u}^+=5}$ and the LRP-filtering length scale. "Furthermore, turbulent and dispersive stresses over original and filtered roughness are compared; it is shown that the turbulent stresses are not affected meaningfully by removal of the drag-irrelevant structures. Agreement is observed for both turbulent and dispersive shear stresses which indicates identical momentum transport pattern in the wall-normal direction over original and filtered roughness. The sole effect of filtering observed on one-point second-order velocity statistics is the the reduced streamwise dispersive stress, which can be an indication of less inhomogeneity of mean flow over the filtered roughness. Finally, the joint p.d.f. of streamwise and wall-normal dispersive velocity components are compared for original and filtered roughness , and it is observed that the probability contours are generally similar, with the streamwise component having a smaller extent in the filtered case. No strong change of preference towards a certain quadrant results from filtering.

In summary, according to the present results, it can be stated that use of roughness height p.d.f and PS as the model inputs, combined with an AL framework for exploring the vast parameter space, has the potential for developing universal roughness predictive models. Additionally, the present work shows a clear potential to extract physical information from the data-driven models through interpretation techniques, here the LRP.

The LRP-based analysis presented in this work is obviously a first step towards utilizing data-driven models beyond merely predictive tools in the context of rough-wall turbulence. Future investigations can explore other avenues to extract knowledge on the roughness-turbulence interactions from such models. The present LRP-based analysis can also be further investigated towards more rigorous criteria for identifying the drag-irrelevant structures. Furthermore, the LRP-based filtering can be a basis for developing more accurate empirical correlations solely incorporating the drag-relevant structures."[Yang et al. 2023a]

Additional ENN models, employing single-valued roughness statistical parameters are developed. The first model, denoted as ENN_{PA}, incorporates a finite set of roughness statistical parameters as well as their products adopted from the work by Jouybari et al. [21]. The second model, ENN_{PAM}, is developed using a subset of 6 roughness parameters carefully chosen from the input variables of ENN_{PA}. This selection involves excluding dot products and constant roughness features. These parameter-based models

are compared with the model utilizing p.d.f. and PS as inputs, which is denoted as ($\mathrm{ENN_{PS}}$). It is noteworthy that the models in the present comparative analysis, including $\mathrm{ENN_{PS}}$, are developed with the hyperparameters subject to the Bayesian optimization (BO) to fully explore potential of each model. "In the comparison of the optimized NN member architectures, it is evident that the $\mathrm{ENN_{PA}}$ model requires higher complexity – reflected by the total number of trainable parameters – to achieve a performance comparable to other models. The optimized models are subsequently evaluated using the testing databases from the current roughness generation algorithm ($\mathcal{T}_{\mathrm{inter}}$) and from literature ($\mathcal{T}_{\mathrm{ext,1\&2}}$)."[Yang et al. 2024]
In the contrast to the consistent performance of $\mathrm{ENN_{PS}}$ across various testing databases, "while the parameter-based models, specifically $\mathrm{ENN_{PA}}$ and $\mathrm{ENN_{PAM}}$, exhibit comparable performance on $\mathcal{T}_{\mathrm{inter}}$, their deteriorated performance on the external databases $\mathcal{T}_{\mathrm{ext,1\&2}}$ underscores the limited capability of the employed roughness parameters to extrapolate the predictive performance across divers roughness types. Based on the current model experiment, it is demonstrated that the multi-scale characterization achieved by employing p.d.f. and PS of roughness height as input variables leads to the outstanding performance generalizability of the model. However, similar performance within $\mathcal{T}_{\mathrm{inter}}$ implies that the parameter-based models work effectively within the same type of roughness encountered in the training data. In the end, a sensitivity analysis on the input parameters of the two parameter-based models is carried out, and it is observed that $\mathrm{ENN_{PAM}}$ reproduces sensitivities more in line with previous physical observations, e.g. it successfully captures saturation effects of both extreme Sk and $\mathrm{ES}_{x,z}$ values on the resulting drag penalty. This potentially contributes to the somewhat better performance of this model compared to $\mathrm{ENN_{PA}}$. "[Yang et al. 2024]

6 Conclusion and Outlook

Throughout this dissertation, the feasibility of developing a universally applicable data-driven model for predicting the skin friction exerted by any arbitrary irregular roughness is investigated. The research starts with a thorough assessment of the tools employed in generating the roughness datasets, encompassing a pseudo-random roughness generation algorithm and the minimal-channel direct numerical simulation (DNS). Particularly, the viability of minimal-channel DNS for accurately predicting the skin friction of irregular roughness is examined. Additionally, the ability of the chosen pseudo-random roughness generation algorithm to reproduce realistic roughness effects is demonstrated through DNS. Having the data-producing tools validated, the machine learning models are constructed in the fashion of an ensemble neural network (ENN). The training of the models is facilitated by means of an active learning framework (AL). Finally, the performance of the trained models is evaluated using the roughness databases stemming from the literature. The model predictions are translated into human-interpretable explanations through multiple model interpretation techniques. Simultaneously with the examination of the aforementioned tools and the trained models, detailed analysis of the available datasets provides valuable insights into the physics of roughness-induced skin friction. While these findings are summarized at the end of each chapter, this chapter provides a narrative of the overall research path that connects these findings to the conclusion of the dissertation as well as the outlook of future works.

In chapter 3, the process of generating roughness data, which involves creating roughness topographies followed by conducting minimal-channel DNS, is examined. Here, the term 'data' refers to the pairs of roughness topographies along with their measurements of skin friction. As a main means of the investigation of roughness effect in the present dissertation, the roughness structures are represented by immersed boundary method (IBM) in the simulations. Initially, the ability of minimal-channel DNS to predict the roughness function ΔU^+ as well as the temperature roughness function $\Delta\Theta^+$ is verified using a series of artificial roughness possessing systematically varied roughness height probability density functions (p.d.f.) and power spectra (PS). The simulations are carried out across a broad range of roughness Reynolds numbers $k^+ = 25 - 100$ to approach the fully rough regime. The flexibility of the present roughness generation algorithm as well as the capability of the minimal-channel DNS for efficiently annotating the equivalent sand-grain size k_s are demonstrated. A modified criterion for minimal channel dimensions, adapted from that for regular sinusoidal roughness [26, 108], is employed. The values of ΔU^+ and $\Delta\Theta^+$ for irregular roughness predicted by minimal-channel DNS exhibit errors within $\pm 5\%$ compared to their full-span DNS counterparts. The origin of this mild disagreement between minimal-channel and full-span DNS is investigated by generating multiple rough surfaces for one of the considered combinations of p.d.f. and PS. Due to the random nature of the roughness generation process, these surfaces, while exhibiting identical statistical properties, are deterministically different in their realizations. Minimal-channel DNS is performed for these surfaces with $k^+ = 50$, and slightly scattered values of ΔU^+ and $\Delta\Theta^+$ are obtained. Notably, these values for the full channel locate within the 99% uncertainty interval of these scattered predictions. The results indicate that, at fixed p.d.f. and PS, randomness in roughness generation can lead to a small uncertainty, which is also likely the origin of the observed $\pm 5\%$ discrepancy between predictions of minimal and full channels. Subsequently, the distributions of the local surface force for the considered types of roughness are extracted, and their correlations with the corresponding roughness height functions are studied. The sheltering model proposed by Yang *et al.*[48], with some assumptions, has been demonstrated to provide a qualitative explanation for the reduction in surface force experienced by roughness peaks located in close proximity downstream of preceding peaks. In the streamwise direction, the coherence function of surface force and roughness structure reveals dropping coherence beyond a certain wavelength, indicating a predominant role of the relatively small-scale structures in the induced surface force, thus contributing to the overall roughness skin friction. These large roughness length scales are suspected to be related to the concept of wavy roughness outlined in previous studies[15]. Furthermore, the dependence of the ΔU^+ values with respect to the investigated roughness properties, namely Sk, the

largest in-plane roughness wavelength λ_0 and PS power-law slope p are investigated. Finally, various widely cited correlations in the literature are assessed using the current datasets; however, even the most successful model exhibits a maximum prediction error of ±30%. This indicates that achieving accurate predictions necessitates a comprehensive representation of the multifaceted impact of various roughness topographical features, including the stochastic multi-scale nature of irregular roughness – a challenge addressed in this dissertation.

Moreover, it is crucial to recognize the fundamental importance of constructing a comprehensive training database that closely resembles realistic roughness characteristics to achieve the optimal performance of the model in real-world applications. As such, the effectiveness of the chosen pseudo-random roughness generation algorithm is examined in its capacity to reproduce naturally occurring roughness features by imitating realistic p.d.f. and PS. To this end, a number of realistic roughness scans originating from various engineering contexts are included to generate their artificial surrogates according to their respective p.d.f. and PS. Moreover, in addition to the comparison of artificial roughness surrogates that strictly mimic realistic p.d.f. and PS, Gaussian surrogates are generated with realistic PS but following a Gaussian p.d.f. This is done to examine the isolated impact of the p.d.f. on the resulting skin friction. Minimal-channel and full-span channel DNS are performed on both roughness surrogates and their target realistic roughness. The roughness surrogate with matched p.d.f. and PS achieves excellent agreement, with approximately 3% discrepancy in their C_f and St values. The collapse of the mean velocity and temperature profiles and consequently agreement in ΔU^+ and $\Delta\Theta^+$ values are achieved by the roughness surrogates of isotropic realistic roughness. Furthermore, higher-order turbulent statistics, such as Reynolds stresses, demonstrate excellent agreement between the realistic roughness and their surrogates. However, discernible deviation of these quantities can be observed for Gaussian surrogates; this disagreement is found to be mostly pronounced for the roughness in the rough regime, whose ES$>$ 0.35. This highlights the contribution of the roughness peaks when the contribution of pressure drag dominates. The present roughness reproduction method is found inadequate to represent the anisotropic properties of the target roughness exhibiting a significant value of surface anisotropy ratio (SAR$\approx$1.7). On the other hand, to further validate the effectiveness of the artificial roughness data in conveying accurate physical information, a laboratory experiment on realistic sandpaper roughness is replicated using the current numerical approach. During the endeavor of capturing the roughness topography from the realistic sandpaper sample, the practical difficulties encountered in attaining accurate roughness scans are illustrated. To overcome these challenges, the artificial roughness surrogate of the sandpaper is utilized in the numerical investigation. Here, the roughness surrogate is generated by combining the p.d.f. and PS measured from respective accurate measurement techniques, namely 3-D photogrammetry and perthometer, respectively. DNS conducted on the surrogate achieves excellent agreement in both the skin-friction coefficient C_f and the velocity profile. As can be observed, this data generation pipeline enjoys outstanding flexibility and efficiency in reflecting accurate flow responses, rendering it appealing for producing a massive roughness database.

Finally, a machine learning model, recovering relative equivalent sand grain height k_r based on roughness p.d.f. and PS, is trained leveraging the aforementioned data generation framework. Initially, a roughness topography repository $\mathcal{U}$ comprising 4200 roughness topographies is created through the chosen roughness generation algorithm with randomly prescribed roughness p.d.f. and PS. During the training phase, the model's learning efficiency is promoted by employing an active learning (AL) strategy to select the most informative roughness topographies from the repository $\mathcal{U}$ along with their corresponding minimal-channel DNS results. Additionally, the presently trained model is constructed in the fashion of ensemble prediction to facilitate the AL process, and is denoted as ENN. The ENN model consists of 50 individually trained multi-layer perceptron models (MLP), and the final prediction of the ENN model is the arithmetic mean value of the predictions of the 50 MLP members. The trained ENN model is assessed with a wide spectrum of types of roughness including those investigated in literature. The mean prediction errors of 9.3%, 5.2%, and 10.2% are achieved for internal testing data set $\mathcal{T}_{\text{inter}}$ and two external data sets $\mathcal{T}_{\text{ext,1\&2}}$, respectively. Based on the same model, layer-wise relevance propagation (LRP) technique is employed to interpret model predictions. Spectral filtering is applied to isolate drag-relevant roughness structures identified by LRP, identical ΔU^+ values are achieved for the LRP filtered roughness indicating the predominant skin friction contribution of the roughness structures that are successfully identified through the model. A further analysis of the blanketing layer over filtered and original roughness is carried out. Meaningful agreement is observed for the blanketing layer depth measured from roughness surface indicates the preservation of

drag-inducing mechanisms through the LRP-filtering. In the end, the impact of the model input, namely the roughness characterization through p.d.f. and PS is analyzed by comparing it with additional type of input using single-valued statistical parameters. Specifically, two parameter-based ENN models, labelled as $\mathrm{ENN_{PA}}$ and $\mathrm{ENN_{PAM}}$, are examined, with $\mathrm{ENN_{PA}}$ incorporating 17 roughness parameters and $\mathrm{ENN_{PAM}}$ utilizing a subset of 6 roughness parameters as input. While these models perform similarly on $\mathcal{T}_{\mathrm{inter}}$, their deteriorated performance on the external databases $\mathcal{T}_{\mathrm{ext,1\&2}}$ highlights the restriction of the utilized roughness parameters to generalize the predictive accuracy across various roughness types. Based on the current model experiment, it is demonstrated that the multi-scale characterization achieved by employing p.d.f. and PS of roughness height as input variables leads to the outstanding performance generalizability of the model. However, similar performance within $\mathcal{T}_{\mathrm{inter}}$ implies that the parameter-based models work effectively within the same type of roughness encountered in the training data. In the end, a sensitivity analysis on the input parameters of the two parameter-based models is carried out, and it is observed that $\mathrm{ENN_{PAM}}$ reproduces sensitivities more in line with previous physical observations, e.g. it successfully captures saturation effects of both extreme Sk and $\mathrm{ES}_{x,z}$ values on the resulting drag penalty. This potentially contributes to the somewhat better performance of this model compared to $\mathrm{ENN_{PA}}$. Moreover, it is found that the training efficiency is closely influenced by the model complexity, determined by the number of inputs in the present investigation.

This dissertation presents a promising framework for developing data-driven predictive models for roughness skin friction. While it is demonstrated that the trained model can accurately predict realistic roughness of diverse types, it should be noted that the current model is not able to resolve the spatial heterogeneity of roughness due to the limitation of the present roughness characterization method based on 1-D PS, as well as the lack of training data for this specific type of roughness. Considering that the mechanism of surface skin friction may differ, e.g., through secondary motion of turbulent flow, this limitation becomes apparent. Moreover, as demonstrated in Chapter 4, the directional preference of roughness formation may lead to a significant alteration of the roughness effect. The present characterization method for the spatial distribution of roughness elements using 1-D PS is not capable of resolving the directional preference of the structure. Therefore, a more complete representation of roughness topography, especially concerning the directional elongation of the roughness structures, should be incorporated in the future development of the model. Future investigations can focus on the construction of an extensive database containing various realistic heterogeneous or anisotropic rough surfaces, or a generation method for creating systematically controlled surface properties in these aspects is necessary to comprehensively investigate their impact on turbulent flow. Furthermore, it is essential to recognize that while the current study focuses on equilibrium turbulent channel flow, real-world applications often involve non-equilibrium flows characterized by adverse or favorable pressure gradients, as well as flow pulsations. These non-equilibrium conditions can significantly impact skin friction and other flow behaviours, potentially introducing different underlying mechanisms. A comprehensive investigation in this direction, although crucial, is relatively rare in the literature.

Additionally, as a field of ever-increasing interest, an additional machine learning model can be developed using the present training framework to predict heat transfer across irregular rough surfaces. The present data generation framework is demonstrated to be applicable for investigating heat transfer of irregular roughness in minimal channels. In light of the investigation of the parameter-based machine learning model in the present work, a reduced-complexity model based on a small set of roughness parameters can potentially be developed for any specific group of roughness, depending primarily on the application background. This approach can alleviate the need for precise p.d.f. and PS measurements while maintaining comparable predictive accuracy. To this end, an application-oriented database is required. On the other hand, sophisticated convolutional neural networks can be developed with state-of-the-art architectures. In this way, the model is subjected to lossless representation and may potentially be applicable for any form of roughness, being furnished with a comprehensive training database. However, in clear contrast to the well-established measurement techniques of roughness parameters, accurate digitization over a reasonably large extent of realistic roughness samples is necessary for successful applications in practical problems. More recently, physics-informed neural networks (PINNs) have emerged as a field of machine learning research, which incorporates physical knowledge into the training process of the model. A machine learning model that is integrated with the already established knowledge of the flow, e.g., the sheltering effect or the mass and momentum conservation, may be a promising path towards not only predicting global flow quantities but also recovering turbulent flow fields over any arbitrary rough surface. Finally, following the

trend of explainable artificial intelligence (XAI), a white-box model, e.g., data-driven symbolic regression, can be developed to map a larger number of roughness parameters to roughness-induced skin friction or heat transfer. Such an XAI model can be used to navigate the optimization of wall structures that exhibit desired hydrodynamics and thermal features.

Bibliography

[1] K. Avila, D. Moxey, A. de Lozar, M. Avila, D. Barkley, and B. Hof. The onset of turbulence in pipe flow. *Science*, 333(6039):192–196, 2011. doi: 10.1126/science.1203223.

[2] O. M. Faltinsen. *SURFACE EFFECT SHIPS*, page 141–164. Cambridge University Press, 2006. doi: 10.1017/CBO9780511546068.006.

[3] J. P. Bons, R. P. Taylor, S. T. McClain, and R. B. Rivir. The many faces of turbine surface roughness. *Journal of Turbomechinery*, 123:739–748, 2001.

[4] P. Forooghi, A. Weidenlener, F. Magagnato, B. Böhm, H. Kubach, T. Koch, and B. Frohnapfel. DNS of momentum and heat transfer over rough surfaces based on realistic combustion chamber deposit geometries. *International Journal of Heat and Fluid Flow*, 69:83 – 94, 2018. ISSN 0142-727X. doi: https://doi.org/10.1016/j.ijheatfluidflow.2017.12.002.

[5] S. Sarakinos and A. Busse. Investigation of rough-wall turbulence over barnacle roughness with increasing solidity using direct numerical simulations. *Phys. Rev. Fluids*, 7:064602, Jun 2022. doi: 10.1103/PhysRevFluids.7.064602.

[6] J. Velandia and S. Bansmer. Topographic study of the ice accretion roughness on a generic aero-engine intake. 01 2019. doi: 10.2514/6.2019-1451.

[7] M. P. Schultz, J. A. Bendick, E. R. Holm, and W. M. Hertel. Economic impact of biofouling on a naval surface ship. *Biofouling*, 27(1):87–98, 2011. doi: 10.1080/08927014.2010.542809. PMID: 21161774.

[8] D. Chung, N. Hutchins, M. P. Schultz, and K. A. Flack. Predicting the drag of rough surfaces. *Annual Review of Fluid Mechanics*, 53:439–471, 2021. doi: 10.1146/annurev-fluid-062520-115127.

[9] A. S. Y. Alsabagh, W.Tiu, Y. Xu, and M. S. Virk. A review of the effects of ice accretion on the structural behavior of wind turbines. *Wind Engineering*, 37(1):59–70, 2013. ISSN 0309524X, 2048402X.

[10] J. Nikuradse. *Stroemungsgesetze in rauhen Rohren*. VDI-Verl., Berlin, 1933.

[11] L.F. Moody. Friction factors for pipe flow. *Trans. ASME*, 66(8):671–677, 1944.

[12] C. Chan-Braun, M. García-Villalba, and M. Uhlmann. Force and torque acting on particles in a transitionally rough open-channel flow. *Journal of Fluid Mechanics*, 684:441–474, 2011. doi: 10.1017/jfm.2011.311.

[13] M. MacDonald, N. Hutchins, and D. Chung. Roughness effects in turbulent forced convection. *Journal of Fluid Mechanics*, 861:138–162, 2019. doi: 10.1017/jfm.2018.900.

[14] P. Forooghi, A. Stroh, F. Magagnato, S. Jakirlić, and B. Frohnapfel. Toward a universal roughness correlation. *Journal of Fluids Engineering*, 139(12), 08 2017. ISSN 0098-2202. doi: 10.1115/1.4037280. 121201.

[15] M. P. Schultz and K. A. Flack. Turbulent boundary layers on a systematically varied rough wall. *Physics of Fluids*, 21(1):015104, 2009. doi: 10.1063/1.3059630.

[16] K. Flack, M. Schultz, J. Barros, and Y.C. Kim. Skin-friction behavior in the transitionally-rough regime. *International Journal of Heat and Fluid Flow*, 61:21–30, 06 2016. doi: 10.1016/j.ijheatfluidflow.2016.05.008.

[17] F. Pérez-Ràfols and A. Almqvist. Generating randomly rough surfaces with given height probability distribution and power spectrum. *Tribology International*, 131:591 – 604, 2019. ISSN 0301-679X. doi: https://doi.org/10.1016/j.triboint.2018.11.020.

[18] T. O. Jelly and A. Busse. Reynolds number dependence of reynolds and dispersive stresses in turbulent channel flow past irregular near-gaussian roughness. *International Journal of Heat and Fluid Flow*, 80:108485, 2019. ISSN 0142-727X. doi: https://doi.org/10.1016/j.ijheatfluidflow.2019.108485.

[19] K.A. Flack, M.P. Schultz, and J.M. Barros. Skin friction measurements of systematically-varied roughness: Probing the role of roughness amplitude and skewness. *Flow, Turbulence and Combustion*, 104(2-3):317–329, 2020. doi: 10.1007/s10494-019-00077-1.

[20] L. Chan, M. MacDonald, D. Chung, N. Hutchins, and A. Ooi. A systematic investigation of roughness height and wavelength in turbulent pipe flow in the transitionally rough regime. *Journal of Fluid Mechanics*, 771:743–777, 2015.

[21] M. A. Jouybari, J. Yuan, G. J. Brereton, and M. S. Murillo. Data-driven prediction of the equivalent sand-grain height in rough-wall turbulent flows. *Journal of Fluid Mechanics*, 912:A8, 2021. doi: 10.1017/jfm.2020.1085.

[22] S. Lee, J. Yang, P. Forooghi, A. Stroh, and S. Bagheri. Predicting drag on rough surfaces by transfer learning of empirical correlations. *Journal of Fluid Mechanics*, 933:A18, 2022. doi: 10.1017/jfm.2021.1041.

[23] M Chevalier, P. Schlatter, A Lundbladh, and D. Henningson. SIMSON–A pseudo-spectral solver for incompressible boundary layer flow. *Tech. Rep. TRITA-MEK 2007:07, Royal Institute of Technology, Stockholm, Sweden*, pages 1–100, 01 2007.

[24] D. Goldstein, R. Handler, and L. Sirovich. Modeling a no-slip flow boundary with an external force field. *Journal of Computational Physics*, 105(2):354–366, 1993.

[25] A. Stroh, K. Schäfer, B. Frohnapfel, and P. Forooghi. Rearrangement of secondary flow over spanwise heterogeneous roughness. *Journal of Fluid Mechanics*, 885:R5, 2020. doi: 10.1017/jfm.2019.1030.

[26] D. Chung, L. Chan, M. MacDonald, N. Hutchins, and A. Ooi. A fast direct numerical simulation method for characterising hydraulic roughness. *Journal of Fluid Mechanics*, 773:418–431, 2015. doi: 10.1017/jfm.2015.230.

[27] F. Alves Portela and N.D. Sandham. A DNS/URANS approach for simulating rough-wall turbulent flows. *International Journal of Heat and Fluid Flow*, 85:108627, 2020. ISSN 0142-727X. doi: https://doi.org/10.1016/j.ijheatfluidflow.2020.108627.

[28] A. Krogh and J. Vedelsby. Neural network ensembles, cross validation, and active learning. In G. Tesauro, D. Touretzky, and T. Leen, editors, *Advances in Neural Information Processing Systems*, volume 7. MIT Press, 1994.

[29] B. Settles and M. Craven. An analysis of active learning strategies for sequence labeling tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, page 1070–1079, USA, 2008. Association for Computational Linguistics.

[30] G. Montavon, W. Samek, and K. Müller. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15, 2018. ISSN 1051-2004. doi: https://doi.org/10.1016/j.dsp.2017.10.011.

[31] J. Yang, A. Stroh, D. Chung, and P. Forooghi. Direct numerical simulation-based characterization of pseudo-random roughness in minimal channels. *Journal of Fluid Mechanics*, 941:A47, 2022. doi: 10.1017/jfm.2022.331.

[32] J. Yang, A. Stroh, S. Jakirlic, B. Frohnapfel, and P. Forooghi. Study of irregular roughness in minimal channels. In *ETMM 13, International symposium on engineering, turbulence, modelling and measurement, 15.09.-17.09.2021*, 2021.

[33] J. Yang, J. Velandia., S. Bansmer, A. Stroh, and P. Forooghi. A comparison of hydrodynamic and thermal properties of artificially generated against realistic rough surfaces. *International Journal of Heat and Fluid Flow*, 99:109093, 2023. ISSN 0142-727X. doi: https://doi.org/10.1016/j.ijheatfluidflow.2022.109093.

[34] B. Frohnapfel, L. von Deyn, J. Yang, J. Neuhauser, A. Stroh, R. Örlü, and D. Gatti. Flow resistance over heterogeneous roughness made of spanwise-alternating sandpaper strips. *Journal of Fluid Mechanics*, 980:A31, 2024. doi: 10.1017/jfm.2024.40.

[35] J. Yang, A. Stroh, S. Lee, S. Bagheri, B. Frohnapfel, and P. Forooghi. Prediction of equivalent sand-grain size and identification of drag-relevant scales of roughness – a data-driven approach. *Journal of Fluid Mechanics*, 975:A34, 2023. doi: 10.1017/jfm.2023.881.

[36] J. Yang, A. Stroh, S. Lee, S. Bagheri, B. Frohnapfel, and P. Forooghi. Asessment of roughness characterization methods for data-driven predictions. *Flow Turbulence and Combustion*, v1:3899648, 2024. doi: https://doi.org/10.21203/rs.3.rs-3899648/v1.

[37] M. Thakkar, A. Busse, and N. Sandham. Surface correlations of hydrodynamic drag for transitionally rough engineering surfaces. *Journal of Turbulence*, 18(2):138–169, 2017. doi: 10.1080/14685248.2016.1258119.

[38] M. Acharya, J. Bornstein, and M. P. Escudier. Turbulent boundary layers on rough surfaces. *Experiments in Fluids*, 4(1):33–47, Jan 1986. ISSN 1432-1114. doi: 10.1007/BF00316784.

[39] K. A. Flack. Moving beyond moody. *Journal of Fluid Mechanics*, 842:1–4, 2018.

[40] M. A. Jouybari, G. J. Brereton, and J. Yuan. Turbulence structures over realistic and synthetic wall roughness in open channel flow at $\mathrm{Re}_\tau = 1000$. *Journal of Turbulence*, 20(11-12):723–749, 2019.

[41] A.J. Musker. Universal roughness functions for naturally-occurring surfaces. *Transactions of the Canadian Society for Mechanical Engineering*, 6(1):1–6, 1980. doi: 10.1139/tcsme-1980-0001.

[42] K. A. Flack and M. P. Schultz. Review of hydraulic roughness scales in the fully rough regime. *Journal of Fluids Engineering*, 132(4), 04 2010. ISSN 0098-2202. doi: 10.1115/1.4001492.

[43] T. O. Jelly and A. Busse. Reynolds and dispersive shear stress contributions above highly skewed roughness. *Journal of Fluid Mechanics*, 852:710–724, 2018. doi: 10.1017/jfm.2018.541.

[44] A. Busse and T. O. Jelly. Effect of high skewness and kurtosis on turbulent channel flow over irregular rough walls. *Journal of Turbulence*, 0(0):e2173761, 2023. doi: 10.1080/14685248.2023.2173761.

[45] K. Pearson. "das Fehlergestetz und seine Verallgemeinerungen durch Fechner und Pearson." A REJOINDER. *Biometrika*, 4(1-2):169–212, 06 1905. ISSN 0006-3444. doi: 10.1093/biomet/4.1-2.169.

[46] W. Peng and B. Bhushan. Modelling of surfaces with a bimodal roughness distribution. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, 214(5):459 – 470, 2000. doi: 10.1243/1350650001543331. Cited by: 15.

[47] M. R. Raupach, R. A. Antonia, and S. Rajagopalan. Rough-Wall Turbulent Boundary Layers. *Applied Mechanics Reviews*, 44(1):1–25, 01 1991. ISSN 0003-6900. doi: 10.1115/1.3119492.

[48] X. I. A. Yang, J. Sadique, R. Mittal, and C. Meneveau. Exponential roughness layer and analytical model for turbulent boundary layer flow over rectangular-prism roughness elements. *Journal of Fluid Mechanics*, 789:127–165, 2016. doi: 10.1017/jfm.2015.687.

[49] T.O. Jelly, A. Ramani, B. Nugroho, N. Hutchins, and A. Busse. Impact of spanwise effective slope upon rough-wall turbulent channel flow. *Journal of Fluid Mechanics*, 951:A1, 2022. doi: 10.1017/jfm.2022.823.

[50] H. Schlichting. *Experimentelle Untersuchungen zum Rauhigkeitsproblem.* INGENIEUR-ARCHIV. 1936.

[51] A. Sigal and James E. Danberg. New correlation of roughness density effect on the turbulent boundary layer. *AIAA Journal*, 28(3):554–556, 1990. doi: 10.2514/3.10427.

[52] J. A. van Rij, B.J. Belnap, and P.M. Ligrani. Analysis and experiments on three-dimensional, irregular surface roughness. *J. Fluids Eng.*, 124(3):671–677, 2002.

[53] E. Napoli, V. Armenio, and M. DeMarchis. The effect of the slope of irregularly distributed roughness elements on turbulent wall-bounded flows. *Journal of Fluid Mechanics*, 613:385–394, 2008. doi: 10.1017/S0022112008003571.

[54] A. Busse, M. Thakkar, and N.D. Sandham. Reynolds-number dependence of the near-wall flow over irregular rough surfaces. *Journal of Fluid Mechanics*, 810:196–224, 2017. doi: 10.1017/jfm.2016.680.

[55] J. M. Barros, M. P. Schultz, and K. A. Flack. Measurements of skin-friction of systematically generated surface roughness. *International Journal of Heat and Fluid Flow*, 72:1 – 7, 2018. ISSN 0142-727X. doi: https://doi.org/10.1016/j.ijheatfluidflow.2018.04.015.

[56] A. Busse, M. Lützner, and N. D. Sandham. Direct numerical simulation of turbulent flow over a rough surface based on a surface scan. *Computers & Fluids*, 116:129 – 147, 2015. ISSN 0045-7930. doi: https://doi.org/10.1016/j.compfluid.2015.04.008.

[57] F. Alves Portela, A. Busse, and N. D. Sandham. Numerical study of fourier-filtered rough surfaces. *Phys. Rev. Fluids*, 6:084606, Aug 2021. doi: 10.1103/PhysRevFluids.6.084606.

[58] T. D. B. Jacobs, T. Junge, and L. Pastewka. Quantitative characterization of surface topography using spectral analysis. *Surface Topography: Metrology and Properties*, 5(1):013001, jan 2017. doi: 10.1088/2051-672x/aa51f8.

[59] I. A. Lyashenko, L. Pastewka, and B. N. J. Persson. On the validity of the method of reduction of dimensionality: Area of contact, average interfacial separation and contact stiffness. *Tribology Letters*, 52(2):223–229, 2013. ISSN 1573-2711. doi: https://doi.org/10.1007/s11249-013-0208-9.

[60] A. Busse and T. O. Jelly. Influence of surface anisotropy on turbulent flow over irregular roughness. *Flow, Turbulence and Combustion*, 104:331 – 354, 2020. ISSN 1573-1987. doi: https://doi.org/10.1007/s10494-019-00074-4.

[61] P. Forooghi, A. Stroh, P. Schlatter, and B. Frohnapfel. Direct numerical simulation of flow over dissimilar, randomly distributed roughness elements: A systematic study on the effect of surface morphology on turbulence. *Phys. Rev. Fluids*, 3:044605, 4 2018. doi: 10.1103/PhysRevFluids.3.044605.

[62] A. A. Townsend. *The structure of turbulent shear flow / A.A.Townsend.* Cambridge University Press Cambridge [Eng.] ; New York, 2nd ed. edition, 1976.

[63] S. B. Pope. *Turbulent Flows.* Cambridge University Press, 2000. doi: 10.1017/CBO9780511840531.

[64] N. Kasagi, Y. Tomita, and A. Kuroda. Direct Numerical Simulation of Passive Scalar Field in a Turbulent Channel Flow. *Journal of Heat Transfer*, 114(3):598–606, 08 1992. ISSN 0022-1481. doi: 10.1115/1.2911323.

[65] Karl Gustafson. Domain decomposition, operator trigonometry, robin condition. *Contemporary Mathematics*, 218:432–437, 01 1998. doi: 10.1090/conm/218/3039.

[66] S. Zaremba. Sur un problème mixte relatif à l'équation de *Laplace.* Krakau Anz. (A), 313-344 (1910)., 1910.

[67] H. Reichardt. Messungen turbulenter schwankungen. *Naturwissenschaften*, 26(24):404–408, Jun 1938. ISSN 1432-1904. doi: 10.1007/BF01772950.

[68] J. Laufer. Investigation of turbulent flow in a two-dimensional channel. 1951.

[69] A. V. Johansson and P. H. Alfredsson. On the structure of turbulent channel flow. *Journal of Fluid Mechanics*, 122:295–314, 1982. doi: 10.1017/S0022112082002225.

[70] S. Pirozzoli, M. Bernardini, and P. Orlandi. Passive scalars in turbulent channel flow at high Reynolds number. *J. Fluid Mech.*, 788:614–639, 2016.

[71] N. R. Wilson and R. H. Shaw. A higher order closure model for canopy flow. *Journal of Applied Meteorology and Climatology*, 16(11):1197 – 1205, 1977. doi: 10.1175/1520-0450(1977)016⟨1197:AHOCMF⟩2.0.CO;2.

[72] M. R. Raupach and R. H. Shaw. Averaging procedures for flow within vegetation canopies. *Boundary-Layer Meteorology*, 22:79–90, 1982. doi: 10.1007/BF00128057.

[73] P. Forooghi, B. Frohnapfel, F. Magagnato, and A. Busse. A modified parametric forcing approach for modelling of roughness. *International Journal of Heat and Fluid Flow*, 71:200–209, 2018. ISSN 0142-727X. doi: https://doi.org/10.1016/j.ijheatfluidflow.2018.03.019.

[74] F. Chedevergne. Modeling rough walls from surface topography to double averaged navier-stokes computation. *Journal of Turbulence*, 24(1-2):36–56, 2023. doi: 10.1080/14685248.2023.2173760.

[75] M. P. Schultz and K. A. Flack. Outer layer similarity in fully rough turbulent boundary layers. *Experiments in Fluids*, 38(3):328–340, Mar 2005. ISSN 1432-1114. doi: 10.1007/s00348-004-0903-2.

[76] K. A. Flack, M. P. Schultz, and J. S. Connelly. Examination of a critical roughness height for outer layer similarity. *Physics of Fluids*, 19(9):095104, 2007.

[77] J. Jiménez. Turbulent flows over rough walls. *Annu. Rev. Fluid Mech.*, 36:173–196, 2004.

[78] J. M. Barros and K. T. Christensen. Observations of turbulent secondary flows in a rough-wall boundary layer. *Journal of Fluid Mechanics*, 748:R1, 2014. doi: 10.1017/jfm.2014.218.

[79] C. Vanderwel and B. Ganapathisubramani. Effects of spanwise spacing on large-scale secondary flows in rough-wall turbulent boundary layers. *Journal of Fluid Mechanics*, 774:R2, 2015. doi: 10.1017/jfm.2015.292.

[80] C. Vanderwel, A. Stroh, J. Kriegseis, B. Frohnapfel, and B. Ganapathisubramani. The instantaneous structure of secondary flows in turbulent boundary layers. *Journal of Fluid Mechanics*, 862:845–870, 2019. doi: 10.1017/jfm.2018.955.

[81] P. S. Jackson. On the displacement height in the logarithmic velocity profile. *Journal of Fluid Mechanics*, 111:15–25, 1981. doi: 10.1017/S0022112081002279.

[82] K. Zhong, N. Hutchins, and D. Chung. Heat-transfer scaling at moderate prandtl numbers in the fully rough regime. *Journal of Fluid Mechanics*, 959:A8, 2023. doi: 10.1017/jfm.2023.125.

[83] F. Secchi, D. Gatti, and B. Frohnapfel. The wall-jet region of a turbulent jet impinging on smooth and rough plates. *Flow, Turbulence and Combustion*, 110(2):275–299, Feb 2023. ISSN 1573-1987. doi: 10.1007/s10494-022-00387-x.

[84] T. Kameda, S. Mochizuki, and H. Osaka. On the virtual origin determined from momentum equation analysis using experimental data within the roughness sublayer. *Experiments in Fluids*, 59(10):146, Sep 2018. ISSN 1432-1114. doi: 10.1007/s00348-018-2600-6.

[85] D. I. Pullin, N. Hutchins, and D. Chung. Turbulent flow over a long flat plate with uniform roughness. *Phys. Rev. Fluids*, 2:082601, Aug 2017. doi: 10.1103/PhysRevFluids.2.082601.

[86] P. Forooghi, M. Stripf, and B. Frohnapfel. A systematic study of turbulent heat transfer over rough walls. *International Journal of Heat and Mass Transfer*, 127:1157–1168, 12 2018. doi: 10.1016/j.ijheatmasstransfer.2018.08.013.

[87] L. Chan, M. MacDonald, D. Chung, N. Hutchins, and A. Ooi. Secondary motion in turbulent pipe flow with three-dimensional roughness. *Journal of Fluid Mechanics*, 854:5–33, 2018. doi: 10.1017/jfm.2018.570.

[88] F.R. Hama. *Boundary-layer Characteristics for Smooth and Rough Surfaces*. 1954.

[89] F. H. Clauser. The turbulent boundary layer. volume 4 of *Advances in Applied Mechanics*, pages 1 – 51. 1956. doi: https://doi.org/10.1016/S0065-2156(08)70370-3.

[90] L. H. von Deyn, D. Gatti, and B. Frohnapfel. From drag-reducing riblets to drag-increasing ridges. *Journal of Fluid Mechanics*, 951:A16, 2022. doi: 10.1017/jfm.2022.796.

[91] D. Gatti, L. von Deyn, P. Forooghi, and B. Frohnapfel. Do riblets exhibit fully rough behaviour? *Experiments in Fluids*, 61:81, 2020. ISSN 14321114. doi: 10.1007/s00348-020-2921-0.

[92] M. Thakkar, A. Busse, and N. D. Sandham. Direct numerical simulation of turbulent channel flow over a surrogate for nikuradse-type roughness. *Journal of Fluid Mechanics*, 837:R1, 2018. doi: 10.1017/jfm.2017.873.

[93] B. A. Kader. Temperature and concentration profiles in fully turbulent boundary layers. *International Journal of Heat and Mass Transfer*, 24(9):1541–1544, 1981. ISSN 0017-9310. doi: https://doi.org/10.1016/0017-9310(81)90220-9.

[94] H. Kawamura, H. Abe, and Y. Matsuo. Dns of turbulent heat transfer in channel flow with respect to reynolds and prandtl number effects. *International Journal of Heat and Fluid Flow*, 20(3):196–207, 1999. ISSN 0142-727X. doi: https://doi.org/10.1016/S0142-727X(99)00014-4.

[95] A. M. Yaglom. Similarity laws for constant-pressure and pressure-gradient turbulent wall flows. *Annual Review of Fluid Mechanics*, 11(1):505–540, 1979. doi: 10.1146/annurev.fl.11.010179.002445.

[96] Y. Miyake, K. Tsujimoto, and M. Nakaji. Direct numerical simulation of rough-wall heat transfer in a turbulent channel flow. *International Journal of Heat and Fluid Flow*, 22(3):237–244, 2001. ISSN 0142-727X. doi: https://doi.org/10.1016/S0142-727X(01)00085-6. Turbulent Heat and Mass Transfer - 3.

[97] P. Bradshaw. A note on "critical roughness height" and "transitional roughness". *Physics of Fluids*, 12(6):1611–1614, 06 2000. ISSN 1070-6631. doi: 10.1063/1.870410.

[98] D. Bettermann. Contribution a l'etude de la couche limite turbulente Ie long de plaques rugueses. *International Journal of Heat and Mass Transfer*, 9(1):153–164, 1966.

[99] F. A. Dvorak. Calculation of turbulent boundary layers on rough surfaces in pressure gradient. *AIAA Journal*, 7(9):1752–1759, 1969. doi: 10.2514/3.5386.

[100] JR. R. Dirling. *A method for computing roughwall heat transfer rates on reentry nosetips*. doi: 10.2514/6.1973-763.

[101] Y. Furuya, M. Miyata, and H. Fujita. Turbulent Boundary Layer and Flow Resistance on Plates Roughened by Wires. *Journal of Fluids Engineering*, 98(4):635–643, 12 1976. ISSN 0098-2202. doi: 10.1115/1.3448434.

[102] M. Abdelaziz, L. Djenidi, M. H. Ghayesh, and R. Chin. On predictive models for the equivalent sand grain roughness for wall-bounded turbulent flows. *Physics of Fluids*, 36(1):015125, 01 2024. ISSN 1070-6631. doi: 10.1063/5.0178798.

[103] R. D. Sanhueza, I. Akkerman, and J. W.R. Peeters. Machine learning for the prediction of the local skin friction factors and nusselt numbers in turbulent flows past rough surfaces. *International Journal of Heat and Fluid Flow*, 103:109204, 2023. ISSN 0142-727X. doi: https://doi.org/10.1016/j.ijheatfluidflow.2023.109204.

[104] J. Kim, P. Moin, and R. Moser. Turbulence statistics in fully developed channel flow at low reynolds number. *Journal of Fluid Mechanics*, 177:133–166, 1987. doi: 10.1017/S0022112087000892.

[105] R. D. Moser, J. Kim, and N. N. Mansour. Direct numerical simulation of turbulent channel flow up to Re_τ=590. *Physics of Fluids*, 11(4):943–945, 04 1999. ISSN 1070-6631. doi: 10.1063/1.869966.

[106] S. Hoyas, M. Oberlack, F. Alcántara-Ávila, S. V. Kraheberger, and J. Laux. Wall turbulence at high friction reynolds numbers. *Phys. Rev. Fluids*, 7:014602, Jan 2022. doi: 10.1103/PhysRevFluids.7.014602.

[107] J. Jiménez and P. Moin. The minimal flow unit in near-wall turbulence. *Journal of Fluid Mechanics*, 225:213–240, 1991. doi: 10.1017/S0022112091002033.

[108] M. MacDonald, D. Chung, N. Hutchins, L. Chan, A. Ooi, and R. Garcia-Mayoral. The minimal-span channel for rough-wall turbulent flows. *Journal of Fluid Mechanics*, 816:5–42, 2017. doi: 10.1017/jfm.2017.69.

[109] M. MacDonald, A. Ooi, R. García-Mayoral, N. Hutchins, and D. Chung. Direct numerical simulation of high aspect ratio spanwise-aligned bars. *Journal of Fluid Mechanics*, 843:126–155, 2018. doi: 10.1017/jfm.2018.150.

[110] S. Pargal, J. Yuan, and G. J. Brereton. Impulse response of turbulent flow in smooth and riblet-walled channels to a sudden velocity increase. *Journal of Turbulence*, 22(6):353–379, 2021.

[111] T.M. Mitchell. *Machine Learning*. McGraw-Hill International Editions. McGraw-Hill, 1997. ISBN 9780071154673.

[112] K.P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012. ISBN 0262018020.

[113] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, Sep 1995. ISSN 1573-0565. doi: 10.1007/BF00994018.

[114] T. K. Ho. Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282 vol.1, 1995. doi: 10.1109/ICDAR.1995.598994.

[115] M. Kac and A. J. F. Siegert. An Explicit Representation of a Stationary Gaussian Process. *The Annals of Mathematical Statistics*, 18(3):438 – 442, 1947. doi: 10.1214/aoms/1177730391.

[116] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, Apr 1980. ISSN 1432-0770. doi: 10.1007/BF00344251.

[117] F. Rosenblatt. (1958) f. rosenblatt, "the perceptron: a probabilistic model for information storage and organization in the brain," psychological review 65: 386-408. In *Neurocomputing, Volume 1: Foundations of Research*. The MIT Press, 04 1988. ISBN 9780262267137. doi: 10.7551/mitpress/4943.003.0010.

[118] A.E. Bryson and Y.C. Ho. *Applied Optimal Control: Optimization, Estimation, and Control*. Blaisdell book in the pure and applied sciences. Blaisdell Publishing Company, 1969.

[119] P. J. Werbos. *The roots of backpropagation: from ordered derivatives to neural networks and political forecasting*. Wiley-Interscience, USA, 1994. ISBN 0471598976.

[120] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, page 1097–1105, Red Hook, NY, USA, 2012. Curran Associates Inc.

[121] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.

[122] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314, Dec 1989. ISSN 1435-568X. doi: 10.1007/BF02551274.

[123] C. Cortes, M. Mohri, and A. Rostamizadeh. L2 regularization for learning kernels. *CoRR*, abs/1205.2653, 2012.

[124] M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2018.10.045.

[125] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Adaptive computation and machine learning. MIT Press, 2016. ISBN 9780262035613.

[126] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[127] D. Sharma. *Evaluation and Analysis of Perception Systems for Autonomous Driving*. PhD thesis, 2020.

[128] A. Bruno, G. Ignesti, O. Salvetti, D. Moroni, and M. Martinelli. Efficient lung ultrasound classification. *Bioengineering (Basel)*, 10(5), May 2023.

[129] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.

[130] P. J. Phillips and M. Przybocki. Four principles of explainable ai as applied to biometrics and facial forensic algorithms, 2020.

[131] W. Kintsch. *The Representation of Meaning in Memory*. European Physics Series. Lawrence Erlbaum Associates, 1974. ISBN 9780470480748.

[132] B. Widrow and M. E. Hoff. Adaptive switching circuits. In *1960 IRE WESCON Convention Record, Part 4*, pages 96–104, New York, 1960. IRE.

[133] S. Bach, A. Binder, G. Montavon, F. Klauschen, and W.h Müller, K.AND Samek. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLOS ONE*, 10 (7):1–46, 07 2015. doi: 10.1371/journal.pone.0130140.

[134] M. T. Ribeiro, S. Singh, and C. Guestrin. "why should I trust you?": Explaining the predictions of any classifier. *CoRR*, abs/1602.04938, 2016.

[135] J. Kim and C. Lee. Prediction of turbulent heat transfer using convolutional neural networks. *Journal of Fluid Mechanics*, 882:A18, 2020. doi: 10.1017/jfm.2019.814.

[136] L. Arras, F. Horn, G. Montavon, K. Müller, and W. Samek. "what is relevant in a text document?": An interpretable machine learning approach. *CoRR*, abs/1612.07843, 2016.

[137] W. Samek, A. Binder, G. Montavon, S. Bach, and K. Müller. Evaluating the visualization of what a deep neural network has learned. *CoRR*, abs/1509.06321, 2015.

[138] W. Landecker, M. D. Thomure, L. M.A. Bettencourt, M. Mitchell, G. T. Kenyon, and S. P. Brumby. Interpreting individual classifications of hierarchical networks. page 32 – 38, 2013. doi: 10.1109/CIDM.2013.6597214. Cited by: 43; All Open Access, Green Open Access.

[139] W. Anderson and C. Meneveau. Dynamic roughness model for large-eddy simulation of turbulent flow over multiscale, fractal-like rough surfaces. *Journal of Fluid Mechanics*, 679:288–314, 2011. doi: 10.1017/jfm.2011.137.

[140] V. I. Nikora, T. Stoesser, S. M. Cameron, M. Stewart, K. Papadopoulos, P. Ouro, R. McSherry, A. Zampiron, I. Marusic, R. A. Falconer, and et al. Friction factor decomposition for rough-wall flows: theoretical background and application to open-channel flows. *Journal of Fluid Mechanics*, 872:626–664, 2019. doi: 10.1017/jfm.2019.344.

[141] O. Flores and J. Jiménez. Hierarchy of minimal flow units in the logarithmic layer. *Physics of Fluids*, 22(7):071704, 2010. doi: 10.1063/1.3464157.

[142] M. Quadrio and P. Luchini. Integral space–time scales in turbulent wall flows. *Physics of Fluids*, 15 (8):2219–2227, 2003. doi: 10.1063/1.1586273.

[143] O. Yirtici, S. Ozgen, and I. H. Tuncer. Predictions of ice formations on wind turbine blades and power production losses due to icing. *Wind Energy*, 22(7):945–958, 2019. doi: https://doi.org/10.1002/we.2333.

[144] S. Bansmer, A. Baumert, S. Sattler, I. Knop, D. Leroy, A. Schwarzenboeck, T. Jurkat-Witschas, C. Voigt, H. Pervier, and B. Esposito. Design, construction and commissioning of the braunschweig icing wind tunnel. *Atmospheric Measurement Techniques*, 11:3221–3249, 2018. doi: 10.5194/amt-11-3221-201.

[145] J. S. Velandia. and S. E. Bansmer. Detailed atmospheric ice accretion surface measurement using micro-computed tomography. *Atmospheric Science Letters*, 21(10):e997, 2020. doi: 10.1002/asl.997.

[146] D. Anderson, D. Hentschel, and G. Ruff. Measurement and correlation of ice accretion roughness. In *36th AIAA Aerospace Sciences Meeting and Exhibit*. American Institute of Aeronautics and Astronautics, jan 1998. doi: 10.2514/6.1998-486.

[147] J. Yuan and U. Piomelli. Roughness effects on the reynolds stress budgets in near-wall turbulence. *Journal of Fluid Mechanics*, 760:R1, 2014. doi: 10.1017/jfm.2014.608.

[148] Y. Kuwata and Y. Kawaguchi. Direct numerical simulation of turbulence over systematically varied irregular rough surfaces. *Journal of Fluid Mechanics*, 862:781–815, 2019. doi: 10.1017/jfm.2018.953.

[149] L. H. von Deyn, M. Schmidt, R. Örlü, A. Stroh, J. Kriegseis, B. Böhm, and B. Frohnapfel. Ridge-type roughness: from turbulent channel flow to internal combustion engine. *Experiments in Fluids*, 63: A18, 2022. ISSN 1432-1114.

[150] W. M. Kays, M. E. Crawford, and B. Weigand. *Convective heat and mass transfer*, volume 4. McGraw-Hill New York, 1980.

[151] J.W.R. Peeters and N.D. Sandham. Turbulent heat transfer in channels with irregular roughness. *International Journal of Heat and Mass Transfer*, 138:454–467, 2019. ISSN 0017-9310. doi: https://doi.org/10.1016/j.ijheatmasstransfer.2019.04.013.

[152] B. Hallert. *Photogrammetry, basic principles and general survey.* McGraw-Hill, 1960.

[153] A. Guettler. High accuracy determination of skin friction differences in an air channel flow based on pressure drop measurements. *Doctoral Thesis, Karlsruher Institut für Technologie (KIT)*, 2015. doi: 10.5445/IR/1000048039.

[154] L. H. von Deyn. Drag of spanwise heterogeneous surface topographies. *Doctoral Thesis, Karlsruher Institut für Technologie (KIT)*, 2023. doi: 10.5445/IR/1000154384.

[155] D. Gatti, A. Güttler, B. Frohnapfel, and C. Tropea. Experimental assessment of spanwise-oscillating dielectric electroactive surfaces for turbulent drag reduction in an air channel flow. *Exp Fluids*, 56: 110, 2015. doi: 10.1007/s00348-015-1983-x.

[156] R. B. Dean. Reynolds number dependence of skin friction and other bulk flow variables in two-dimensional rectangular duct flow. *Journal of Fluids Engineering*, 100:215–223, 6 1978. ISSN 0098-2202. doi: 10.1115/1.3448633.

[157] J. Kim, P. Moin, and R. Moser. Turbulence statistics in fully developed channel flow at low reynolds number. *Journal of Fluid Mechanics*, 177:133–166, 1987. doi: 10.1017/S0022112087000892.

[158] M. MacDonald, D. Chung, N. Hutchins, L. Chan, A. Ooi, and A. García-Mayoral. The minimal channel: a fast and direct method for characterising roughness. *Journal of Physics: Conference Series*, 708:012010, 4 2016. doi: 10.1088/1742-6596/708/1/012010.

[159] R. Reed and R. J. Marks. *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks.* The MIT Press, 02 1999. ISBN 9780262282215. doi: 10.7551/mitpress/4937.001.0001.

[160] C.M. Bishop, P.N.C.C.M. Bishop, G. Hinton, and Oxford University Press. *Neural Networks for Pattern Recognition.* Advanced Texts in Econometrics. Clarendon Press, 1995. ISBN 9780198538646.

[161] A. Naoki and M. Hiroshi. Query learning strategies using boosting and bagging. In *ICML*, 1998.

[162] V.V. Fedorov. *Theory of Optimal Experiments*. Probability and Mathematical Statistics. Academic Press, 1972.

[163] Q. Zhu, A. Stolcke, B. Y. Chen, and N. Morgan. Using MLP features in SRI's conversational speech recognition system. In *Proc. Interspeech 2005*, pages 2141–2144, 2005. doi: 10.21437/Interspeech.2005-695.

[164] P. Bangert, J. O. Moon, H.and Woo, S. Didari, and H. Hao. Active learning performance in labeling radiology images is 90% effective. *Frontiers in Radiology*, 1, 2021. ISSN 2673-8740. doi: 10.3389/fradi.2021.748968.

[165] B. Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

[166] Y. Gal and Z. Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning, 2015.

[167] T. Raychaudhuri and L. Hamey. Minimisation of data collection by active learning. In *Proceedings of ICNN'95-International Conference on Neural Networks*, volume 3, pages 1338–1341. IEEE, 1995.

[168] K. Lang and E. Baum. Query learning can work poorly when a human oravle is used. In *IEEE Intl. Joint Conference on Neural networks*, 1992.

[169] D. D. Lewis and W.A. Gale. A sequential algorithm for training text classifiers. In *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1994.

[170] D. Angluin. Queries revisited. *Theoretical Computer Science*, 313(2):175–194, 2004. ISSN 0304-3975. doi: https://doi.org/10.1016/j.tcs.2003.11.004. Algorithmic Learning Theory.

[171] H. S. Seung, M. Opper, and H. Sompolinsky. Query by committee. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, page 287–294, New York, NY, USA, 1992. Association for Computing Machinery. ISBN 089791497X. doi: 10.1145/130385.130417.

[172] R. Burbidge, Jem J. R., and R. D. King. Active learning for regression based on query by committee. In Hujun Yin, Peter Tino, Emilio Corchado, Will Byrne, and Xin Yao, editors, *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, pages 209–218, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. ISBN 978-3-540-77226-2.

[173] P. Melville and R. J. Mooney. Constructing diverse classifier ensembles using artificial training examples. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, IJCAI'03, page 505–510, San Francisco, CA, USA, 2003. Morgan Kaufmann Publishers Inc.

[174] T. Hastie, R. Tibshirani, and J. Friedman. *Model Assessment and Selection*, pages 219–259. Springer New York, New York, NY, 2009. ISBN 978-0-387-84858-7. doi: 10.1007/978-0-387-84858-7_7.

[175] J. O. Hinze. Secondary Currents in Wall Turbulence. *The Physics of Fluids*, 10(9):S122–S125, 09 1967. ISSN 0031-9171. doi: 10.1063/1.1762429.

[176] R. S. Sayles and T. R. Thomas. Topography of random surfaces (reply). *Nature*, 273(5663):573–573, Jun 1978. ISSN 1476-4687. doi: 10.1038/273573b0.

[177] R. Mejia-Alvarez and K. T. Christensen. Low-order representations of irregular surface roughness and their impact on a turbulent boundary layer. *Physics of Fluids*, 22(1):015106, 2010. doi: 10.1063/1.3291076.

[178] T. Medjnoun, E. Rodriguez-Lopez, M.A. Ferreira, T. Griffiths, J. Meyers, and B. Ganapathisubramani. Turbulent boundary-layer flow over regular multiscale roughness. *Journal of Fluid Mechanics*, 917: A1, 2021. doi: 10.1017/jfm.2021.228.

[179] J. Yuan and U. Piomelli. Estimation and prediction of the roughness function on realistic surfaces. *Journal of Turbulence*, 15(6):350–365, 2014. doi: 10.1080/14685248.2014.907904.

[180] P Orlandi and S Leonardi. Direct numerical simulation of three-dimensional turbulent rough channels: parameterization and flow physics. *Journal of Fluid Mechanics*, 606:399–415, 2008.

[181] J. Yuan and M. A. Jouybari. Topographical effects of roughness on turbulence statistics in roughness sublayer. *Phys. Rev. Fluids*, 3:114603, Nov 2018. doi: 10.1103/PhysRevFluids.3.114603.

[182] J. M. Wallace, H. Eckelmann, and R. S. Brodkey. The wall region in turbulent shear flow. *Journal of Fluid Mechanics*, 54(1):39–48, 1972. doi: 10.1017/S0022112072000515.

[183] M. Hoffman, E. Brochu, and N. de Freitas. Portfolio allocation for bayesian optimization. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, UAI'11, page 327–336, Arlington, Virginia, USA, 2011. AUAI Press. ISBN 9780974903972.

Publications of Jiasheng Yang

(peer-reviewed journal publications cited in this thesis)

[Yang et al. 2022] **Yang J.**, Stroh A., Chung D. and Forooghi P.. Direct numerical simulation-based characterization of pseudo-random roughness in minimal channels. *Journal of Fluid Mechanics*. 2022;941:A47. doi:10.1017/jfm.2022.331 .

[Yang et al. 2023a] **Yang J.**, Velandia J., Bansmer S., Stroh A. and Forooghi P.. A comparison of hydrodynamic and thermal properties of artificially generated against realistic rough surfaces. *International Journal of Heat and Fluid Flow*. 2023; 99:109093. doi:10.1016/j.ijheatfluidflow.2022.109093.

[Yang et al. 2023b] **Yang J.**, Stroh A., Lee S., Bagheri S., Frohnapfel B. and Forooghi P.. Prediction of equivalent sand-grain size and identification of drag-relevant scales of roughness – a data-driven approach. *Journal of Fluid Mechanics*. 2023;975:A34. doi:10.1017/jfm.2023.881

[Frohnapfel et al. 2024] Frohnapfel B., von Deyn L., **Yang J.**, Neuhauser J., Stroh A., Örlü R. and Gatti D.. Flow resistance over heterogeneous roughness made of spanwise-alternating sandpaper strips. *Journal of Fluid Mechanics*. 2024;980:A31. doi:10.1017/jfm.2024.40

[Yang et al. 2024] **Yang J.**, Stroh A., Lee S., Bagheri S., Frohnapfel B. and Forooghi P.. Asessment of roughness characterization methods for data-driven predictions. *Flow Turbulence and Combustion*. 2024. doi: 10.1007/s10494-024-00549-z

(peer-reviewed journal publications not cited in this thesis)

[Lee et al. 2022] Lee S., **Yang J.**, Forooghi P., Stroh A. and Bagheri S.. Predicting drag on rough surfaces by transfer learning of empirical correlations. *Journal of Fluid Mechanics*. 2022;933:A18. doi:10.1017/jfm.2021.1041

(conference contributions cited in this thesis)

[Yang et al. 2021] **Yang J.**, Stroh A., Jakirlic S., Frohnapfel B. and Forooghi P.. Study of irregular roughness in minimal channels. *13th International symposium on engineering, turbulence, modelling and measurement (ETMM13)*, Rhodes, Greece, September 15-17,2021.

(conference contributions related to the content of this thesis)

[Yang et al. 2022b] **Yang J.**, Stroh A., Friederich P., Frohnapfel B. and Forooghi P.. An active learning approach for the prediction of hydrodynamic roughness properties. *12th International symposium on turbulence and shear flow phenomena (TSFP2022)*, Osaka, Japan, July 19-22, 2022.

[Yang et al. 2022c] **Yang J.**, Lee S., Bagheri S., Stroh A. and Forooghi P.. Predicting roughness-induced drag based on active learning. *14th European Fluid Mechanics Conference (EFMC 14)*, Athens, Greece, September 13-16, 2022.

[Yang et al. 2023c] **Yang J.**, Velandia J., Bansmer S., Stroh A. and Forooghi P.. DNS-based thermohydraulic assessment of artificial roughness surrogates. *17th International heat transfer conference (IHTC 2023)*, Cape Town, South Africa, August 14, 2023.

[Yang et al. 2023d] **Yang J.**, Stroh A., Lee S., Bagheri S., Frohnapfel B. and Forooghi P.. Assessing different roughness description methods in skin friction prediction. *14th International symposium on engineering, turbulence, modelling and measurement (ETMM14)*, Barcelona, Spain, September 6-8,2023.

List of Figures

List of Tables

A Appendix

A.1 Reproduction of Sandpaper Roughness through Perthmoeter and 3-D Photogrammetry

PSs and p.d.f.s extracted from each five scans by perthometer and 3-D photogrammetry measurements are displayed in figure A.1 (a) and (b), respectively.

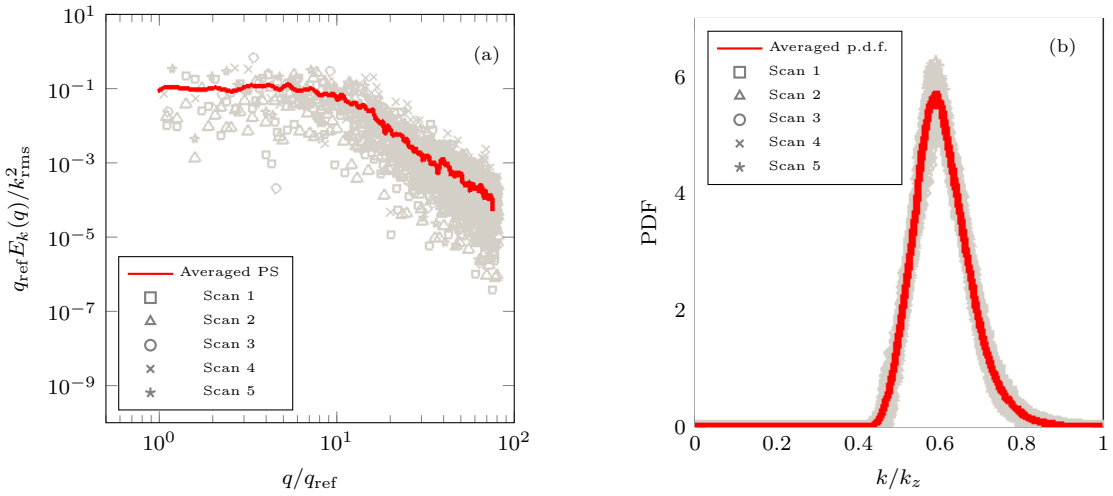


Figure A.1: a)PS extracted from perthometer measurements; b) p.d.f. extracted from 3-D photogrammetry measurements. Red solid lines indicates case-averaged PS and p.d.f..

A.2 Exemplary Roughness at Each Active Learning Round

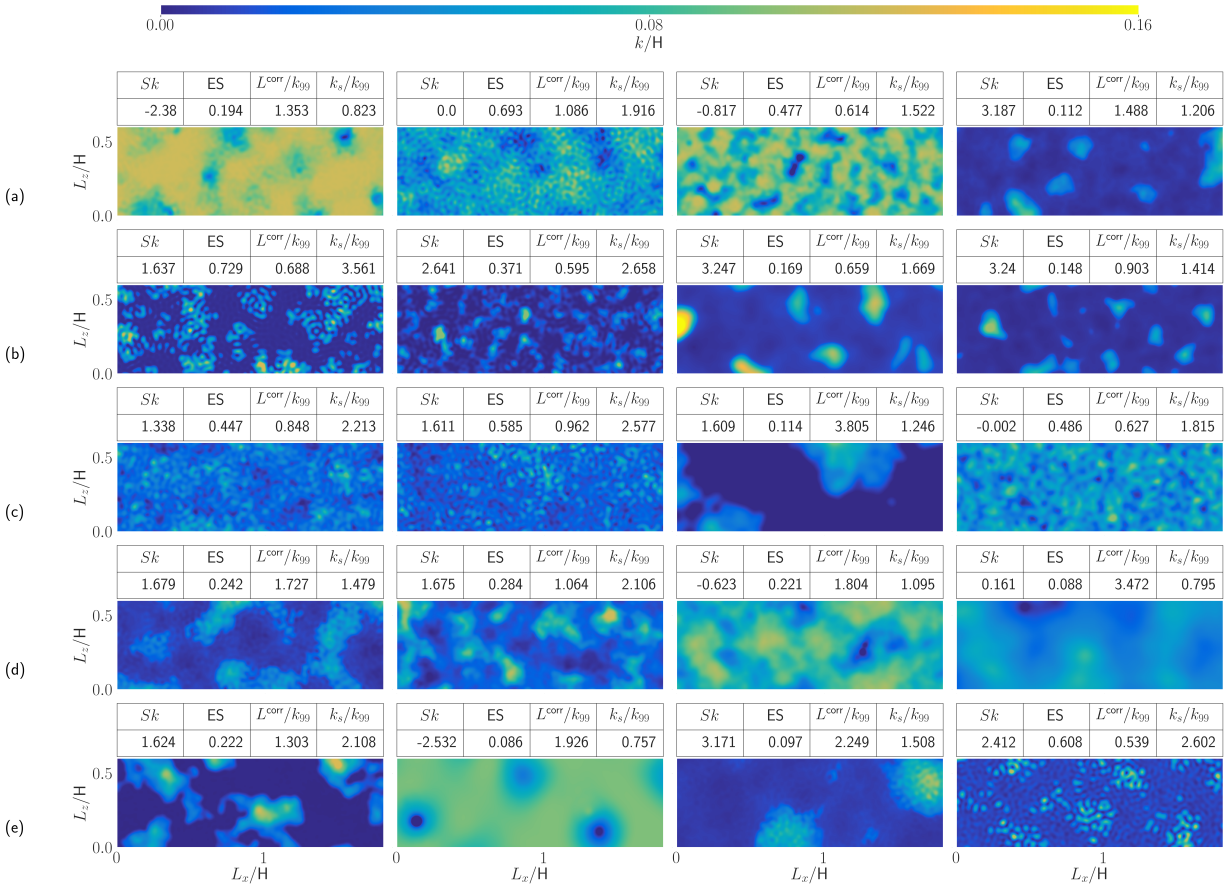


Figure A.2: Examples of roughness samples included in $\mathcal{L}$. Patches of same size extracted from different samples. (a∼e) correspond to initial round and AL round 1∼4, respectively. Figure adopted from [Yang et al. 2023].

Exemplary rough surfaces selected from each iteration along with their statistical parameters and k_s values are shown in figure A.2.

A.3 Assessment of Outer Layer Similarity within the Roughness Database

The calculation of Roughness function ΔU^+ is based on the premise of outer-layer similarity [62], which states that outer-layer flow is unaffected by the near wall events except for the effect due to the wall shear stress, the downward shift of the velocity profile is approximately a constant value in the logarithmic region and possibly beyond if the outer-flow geometry and Reynolds number are matched. Bearing this in mind, outer layer similarity must be validated prior to the calculation of ΔU^+. However, due to the nature of minimal channels, the turbulent flow in outer layer cannot be physically resolved. This results in varied mean velocity profiles in the outer layer for minimal channels of different sizes. In this study, the size of the minimal channel is described by its spanwise width L_z, as the streamwise length L_x is defined by $L_x = 3 \times L_z$. The selected spanwise width L_z is ranged from $L_z = 0.6$H to 2.0H, with increments of 0.1H. The outer layer similarities of the flow over the roughness samples in the present training and testing databases are examined in figure A.3, grouped by the dimension of the minimal channels.

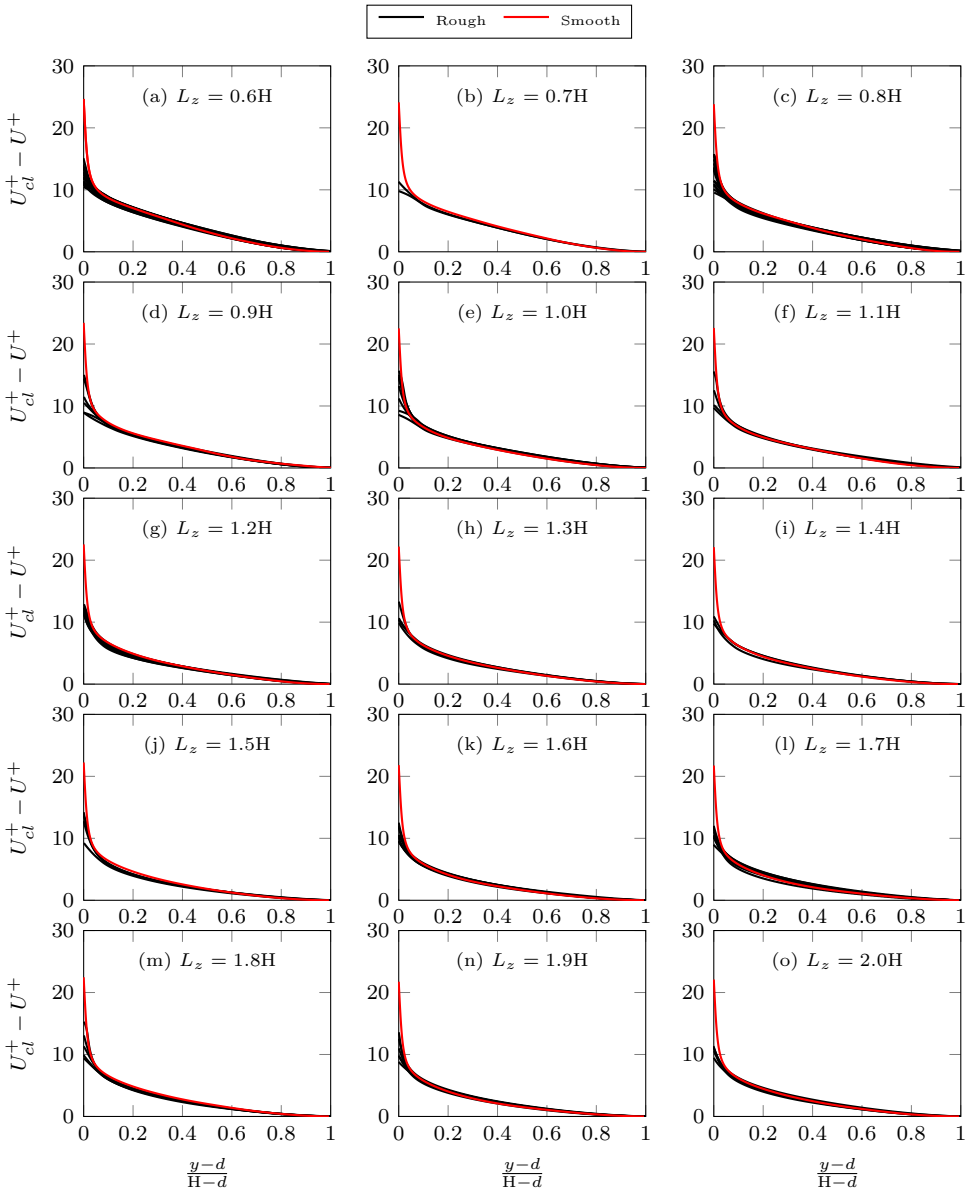


Figure A.3: Mean velocity deficit profiles of rough and smooth minimal channels. (a-o): $L_z = 0.6\text{H} - 2.0\text{H}$ with increments of 0.1H.