12th International Symposium on Turbulence and Shear Flow Phenomena (TSFP12)
Osaka, Japan (Online), July 19-22, 2022

AN ACTIVE LEARNING APPROACH FOR THE PREDICTION OF HYDRODYNAMIC ROUGHNESS PROPERTIES

Jiasheng Yang
Institute of Fluid Mechanics
Karlsruhe Institute of Technology
Karlsruhe, Germany
jiasheng.yang@kit.edu

Alexander Stroh
Institute of Fluid Mechanics
Karlsruhe Institute of Technology
Karlsruhe, Germany
alexander.stroh@kit.edu

Pascal Friederich
Institute of Theoretical Informatics
Karlsruhe Institute of Technology
Karlsruhe, Germany
pascal.friederich@kit.edu

Bettina Frohnapfel
Institute of Fluid Mechanics
Karlsruhe Institute of Technology
Karlsruhe, Germany
bettina.frohnapfel@kit.edu

Pourya Forooghi
Department of Mechanical & Production Engineering
Aarhus University
Aarhus, Denmark
forooghi@mpe.au.dk

ABSTRACT

Realistic surfaces of flow-related equipment are often hydraulically rough due to wear or fouling. Predicting the skin friction exerted by such rough surfaces is a challenging task since the topography of these surfaces is inherently irregular and complex. Recent developments in data-driven methods and increasing affordability of high-fidelity direct numerical simulations (DNS) have created new possibilities for estimation of drag on irregular rough surfaces. In the present work we aim to demonstrate a viable approach to efficiently train a predictive model for the estimation of drag for irregular roughness based on its height probability density function (PDF) and the surface height power spectrum (PS). An active learning (AL) framework is employed to efficiently navigate the construction of a training database. Training data is generated by conducting direct numerical simulations of a flow over artificially generated rough surfaces in minimal channels in order to minimize the computational effort. An ensemble neural network (ENN) model is trained based on the database. The ENN model shows promising potential in predicting the skin friction as well as estimating the epistemic (model) uncertainty. Furthermore, the model – trained on artificial surfaces – is tested on five realistic surface scans, showing that a maximum error of 8.7% between the predicted roughness function ΔU^+ and the DNS results is achieved. Overall, the AL framework shows a great potential as a basis for future research towards a universal predictive tool for any arbitrary roughness.

INTRODUCTION

Roughness is encountered in a variety of engineering applications since many industrial processes inherently include rough surface conditions. The roughness on solid surfaces in flow-related applications can enhance the near-wall momentum transfer, thus it also increases the skin friction drag. This can translate into a significant deterioration of equipment performance. The increase of skin friction is manifested by a downward shift in the mean velocity profile in respect to a smooth surface profile at identical friction Reynolds number Re_τ. This downward shift in the logarithmic layer is known as the (Hama) roughness function ΔU^+. Prediction of the roughness function is of key importance for the estimation of drag force and modeling of turbulent flow over a rough surface. However, ΔU^+ is not known *a priori* for an arbitrary rough surface, and for every new roughness topography at different Re_τ, ΔU^+ needs to be determined using a laboratory experiment or high-fidelity numerical simulation, both of which are costly and time consuming. As an alternative, predictive empirical 'roughness correlations' have been extensively studied by researchers, a review is provided by Chung *et al.* (2021). These correlations predict the roughness function based on statistical properties of the roughness height distribution. Due to the complexity of the roughness geometry, these roughness correlations proposed in the past research have limits in their generalizability to other types of roughness topographies.

It is understood that a comprehensive database is the fundament of an universally valid roughness predictive model. Apart from the difficulties in constructing a massive roughness database that contains realistic surfaces in terms of time consumption and technical difficulties investigating realistic roughness has its disadvantages because of the lack of flexibility for parametric studies. In contrast, regular roughness that contains the specific shape of roughness elements e.g. cubes (Yang *et al.*, 2016), cones (Forooghi *et al.*, 2017), ellipsoids (Jouybari *et al.*, 2021), sinusoidal waves (MacDonald *et al.*, 2016) etc. are frequently used for systematic investigation of roughness effect in respect to single, or a few, roughness parameter(s). However, regular roughness has its shortcomings in representing roughness that are encountered in realistic applications. Recently, mathematical irregular roughness generation methods have received increasing attention in this field of study. The mathematically generated irregular rough surfaces are able to reflect the stochastic nature of realistic roughness while their geometrical statistics can be manipulated. Therefore, the generation method proposed by Pérez-Ràfols & Almqvist (2019) is employed to systematically produce the rough surfaces that can be regarded as surrogates of realistic surfaces.

Under the common goal of predicting roughness-induced drag of an arbitrary roughness based on its geometrical prop-

erties several empirical roughness correlations have been built based on different statistical properties of the roughness, for instance the models based on correlation length L^{corr} (Thakkar *et al.*, 2017), skewness *Sk* (Flack *et al.*, 2020) and effective slope *ES* (Forooghi *et al.*, 2017). As a matter of fact, it is widely reported in the literature that the skin friction of roughness can be partly determined by a wide range of geometrical statistics. In order to incorporate as much aspects of roughness statistical properties as possible into the prediction, multi-variant machine learning approaches have drawn much attention in the field. Jouybari *et al.* (2021) applied neural network to predict the equivalent sand-grain size k_s based on 17 roughness statistics. Lee *et al.* (2022) employed transfer learning method to enhance the performance of the machine learning model with three empirical roughness correlations, significant enhancement of the model is achieved with relative lower amount of training data.

Recently, Yang *et al.* (2021) suggested that the flow properties of an arbitrary roughness can be uniquely determined by a reduced-order representation of roughness, namely the roughness PDF and PS. In the present work, a machine learning framework is employed for the purpose of predicting the roughness function ΔU^+ of homogeneous irregular roughness at $\text{Re}_\tau = 500$ based on its PDF and PS. Active learning (AL) approach is applied to navigate the construction of the training database with the aim of strengthening the model iteratively in an efficient manner. In order to quantify the model uncertainty, ensemble neural network (ENN) is used.

The structure of the paper is as following: in section 2 we present the methodology. The advantage of AL framework as well as the performance of ENN are analysed in section 3.1. In section 3.2 the evaluation of the model is demonstrated with 5 realistic surfaces. Finally in section 3.3 model interpretation is carried out with sensitivity analysis method.

METHODOLOGY

Active learning (AL)

In the present work, we aim at developing an efficient AL framework to predict ΔU^+ at $\text{Re}_\tau = 500$ of given surfaces and quantify the uncertainty for each prediction. For this purpose a roughness topography repository with systematic variation of roughness properties is constructed, containing 400 surfaces. It is prohibitively expensive to conduct DNS investigation for the entire repository. Therefore, it is important to efficiently select the representative roughness samples in the repository with the aim of sparing the computational effort for the roughness with similar properties to the already labeled samples. To this end, AL is proposed to navigate the selection of the most relevant samples based on the model confidence, or model uncertainty (Aggarwal *et al.*, 2014). The workflow of the AL is illustrated on the left panel of figure 1. Within the present AL framework, the model is firstly trained by an initial roughness data set with 20 randomly selected and DNS-annotated surfaces, and subsequently enhanced by a small amount of roughness samples (~ 20) guided by AL.

Ensemble neural network (ENN)

As mentioned in the previous section, the selection of surfaces from the repository for DNS simulations is based on the uncertainty estimation of a ML model. We employ a ENN to predict ΔU^+ values of surfaces and - at the same time - to estimate the uncertainty of the prediction. The ENN is an ensemble of 50 individual densely-connected neural networks (NN) with identical architectures but trained with random initial weights. Each NN is constructed with one input layer with 64 neurons, two hidden layers with 64 and 128 non-linear neurons with rectified linear unit (ReLU) activations, respectively and one linear neuron in the output layer. 10% samples in the database are used for validation while 90% samples are used for training. Diversity of the NNs is ensured by selecting the distribution of the training-validation database independently. L2-regularization is applied to the loss function. Adaptive momentum estimation (Adam) is employed to train the model. A sketch of the NN is shown on the right panel of Figure 1. The input vector $\boldsymbol{I}$ of the NN contains the roughness PDF and PS along with 4 additional characteristic features of the rough surface, i.e. the roughness peak-to-trough height k_t, the 99% confidence interval of the PDF k_{99}, the normalized largest/ smallest roughness wavelength $\lambda_0^* = \lambda_0/k_{99}$, and $\lambda_1^* = \lambda_1/k_{99}$, respectively. The input vectors that represent the roughness PDF and PS are obtained by discretizing the roughness PDF and PS into each 30 values equidistantly within the height range $0 < k < k_t$ and the wavenumber range $2\pi/\lambda_1 > 2\pi/\lambda > 2\pi/\lambda_0$, respectively. The input elements in $\boldsymbol{I}$ are normalized with the standard deviation values of each corresponding element calculated from the repository. The final prediction of the ENN is the averaged prediction over all 50 NNs, namely $\mu_{\Delta U^+} = \sum_{i=1}^{50} \Delta\hat{U}_i^+/50$. Where $\Delta\hat{U}^+$ represents the prediction of a single NN, index i indicates the number of NN. The uncertainty is measured by the standard deviation of the predictions, i.e. $\sigma_{\Delta U^+} = \sqrt{\sum_{i=1}^{50}(\Delta\hat{U}_i^+ - \mu_{\Delta U^+})^2/50}$.

Minimal channel simulation

The rough surfaces from the repository are investigated in closed turbulent channel driven by constant pressure gradient with DNS. The immersed boundary method following Goldstein's method (Goldstein *et al.*, 1993) is employed. The Navier-Stokes equation writes

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

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla\cdot(\mathbf{uu}) = -\frac{1}{\rho}\nabla p + \nu\nabla^2\mathbf{u} - \frac{1}{\rho}P_x\mathbf{e_x} + \mathbf{f}_{\text{IBM}}\,. \tag{2}$$

where $\mathbf{u} = (u,v,w)^\intercal$ is the velocity vector and P_x is the mean pressure gradient in the flow direction added as a constant and uniform source term to the momentum equation to drive the flow in the channel. Here p, $\mathbf{e_x}$, ρ, ν and $\mathbf{f}_{\text{IBM}}$ are pressure fluctuation, streamwise basis vector, density, kinematic viscosity and external body force term due to IBM, respectively. Periodic boundary conditions are applied in the streamwise and spanwise directions. No-slip boundary condition is applied on the rough walls. The friction Reynolds number is defined as $\text{Re}_\tau = u_\tau(\text{H} - k_{md})/\nu$, where $u_\tau = \sqrt{\tau_w/\rho}$ and $\tau_w = -P_x\cdot(\text{H} - k_{\text{md}})$ are the friction velocity and wall shear stress, respectively. Melt-down height denoted by k_{md} is the mean roughness height measured from the deepest trough. In the present work, simulations are performed at $\text{Re}_\tau = 500$.

Due to the unfavorable computational costs for DNS, the concept of minimal channel simulations (MacDonald *et al.*, 2016) is applied. The minimal channel size $L_x \times L_y \times L_z = 3\text{H} \times 2\text{H} \times 1\text{H}$ is selected to accommodate both near-wall turbulence and roughness structures. Where H is the channel half height. In order to resolve the finest roughness structures, a resolution of $576 \times 401 \times 192$ voxels is selected, which translates to the wall-parallel grid size $\Delta^+ = 2.6$. With the present

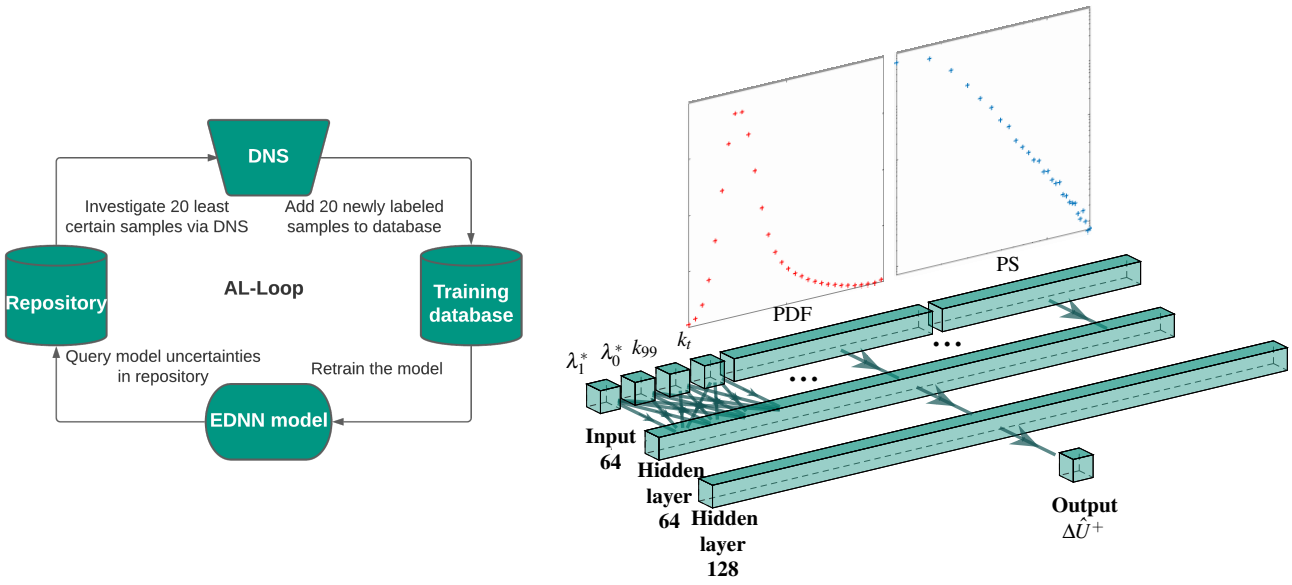

Figure 1: Left: schematic of the AL-framework. Right: illustration of the NN architecture, where input vector contains normalized and discretized roughness PS, PDF and 4 roughness properties. $\lambda^* = \lambda/k_{99}$ indicates normalized wavelength. Note, that layers are fully connected.

mesh size, the smallest roughness structures with wavelength λ_1 are resolved by more than 7 grid points in each wall parallel direction. In wall-normal directions, consine stretching mesh is adopted for the Chebychev discretization. A vertical cell number of 401 is selected. The mesh independence is tested. An exemplary instantaneous velocity field at an arbitrary z-location is shown in Figure 2.

Roughness generation

Artificial irregular rough surfaces are generated through a mathematical roughness generation method where the PS and PDF of the roughness can be prescribed (Pérez-Ràfols & Almqvist, 2019). PDF and PS in the repository are set in a random but controlled form to imitate the realistic distribution of PDF and PS. Weibull, bimodal as well as Gaussian distribution PDF are selected randomly. The Weibull distribution writes:

$$f(k) = K\beta^K k^{(K-1)} e^{-(\beta k)^K} . \tag{3}$$

where the shape parameters K and β can be randomly prescribed to adjust the shape of the PDF profile. k represents the local roughness height. The Bimodal distribution is obtained by generating two Gaussian sequences with random expectation μ and standard deviation σ and take the point-wise minimum. The Bimodal roughness height distribution as a function of wall-parallel coordinates (x,z) writes:

$$k(x,z) = \min\{\Phi_{0,1}(x,z), \Phi_{\mu,\sigma}(x,z)\} . \tag{4}$$

The roughness height is selected to be at the level of $k_t \approx 0.1$H.

Random values of the roll-off length L_r (Jacobs *et al.*, 2017) as well as the power-law decline rate θ_{PS} (Lyashenko *et al.*, 2013) are prescribed for the modelling of roughness PS. Random perturbations are added to the PS. The upper as well as lower bound of the roughness wavelength are set to $\lambda_0 = 0.8$H and $\lambda_1 = 0.04$H, respectively. Eventually, by randomly tuning the shape parameters of the roughness PDF and PS a repository with 400 rough surfaces is generated. The overview of the randomly prescribed PSs and PDFs in the repository along with the realistic PSs and PDFs are shown in Figure 3(a,b). Exemplary surface realizations in the repository are shown in Figure 4.

Sensitivity analysis

In recent years, growing number of attempts have been conducted with the aim to interpret the predictions of neural networks. One of the most frequently used model explanation methods is the sensitivity analysis (SA). The SA approach attempts to quantify the importance of each input feature. The SA is the Jacobian matrix of the NN output $\Delta\hat{U}^+$ in respect to the input elements, which writes:

$$S_j = \frac{1}{50}\sum_{i=1}^{50}\left\langle\left|\frac{\partial\Delta\hat{U}_i^+}{\partial I_j}\right|\right\rangle . \tag{5}$$

Where S_j is the sensitivity of the input element I_j, the subscript $_j$ represents the j-th component of the input vector $\boldsymbol{I}$. The symbol $\langle\cdot\rangle$ is the averaging operation over all samples in the training data set. The derivative $\partial\Delta\hat{U}^+/\partial I_j$ is calculated using an automatic differentiation method.

RESULTS

Evaluation of the AL framework

According to the AL workflow, 20 randomly selected rough surfaces are represented by the red lines in figure 3. The 20 least confident samples are marked with green lines in the figure. After one round of AL, 40 samples in total are included into the database for training the ENN. The relative similarity of the samples in terms of PS and PDF can be observed in the figure 3. It can be seen that the model tends to explore the existing sample parameter space as these newly included samples exhibit significant dissimilarity in terms of the PS and PDF.

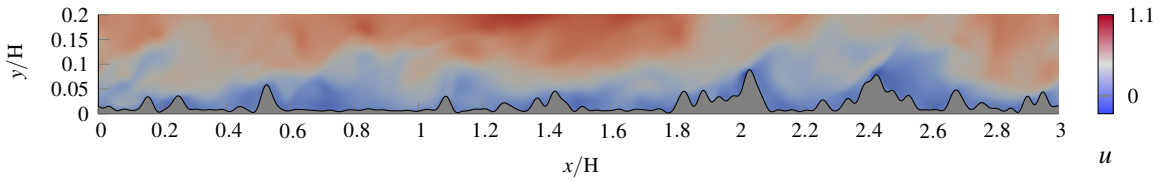

Figure 2: Instantaneous streamwise velocity over simulated rough surface. Contour indicates the mean velocity.

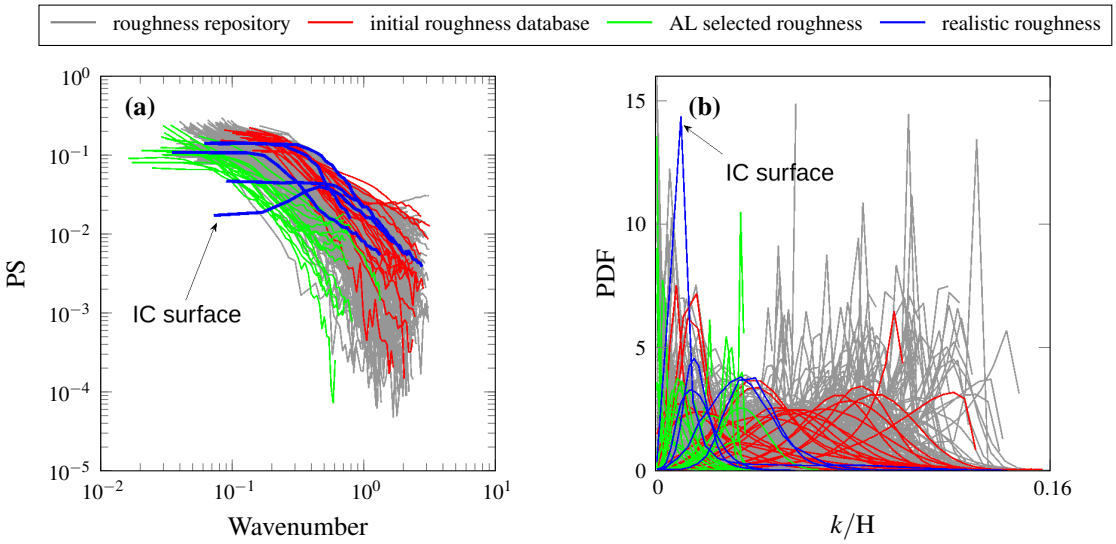

Figure 3: PS (a) and PDF (b) of roughness samples in the roughness pool. Realistic surface scans are colored blue.

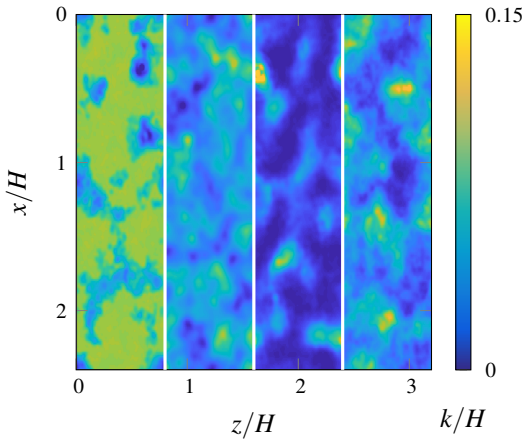

Figure 4: Exemplary artificial rough surface patches

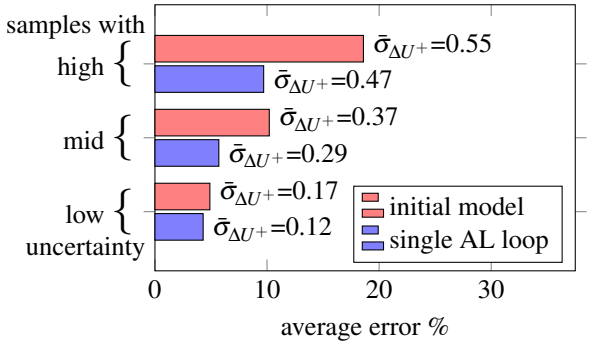

Figure 5: Reduction of average prediction error and uncertainty for three groups of samples after one AL iteration. Here overbar represents averaging process over the samples in each uncertainty level.

In order to evaluate the performance of the ENN, multiple samples from the repository at different uncertainty levels are grouped (see Figure 5). The high uncertainty group contains the 20 samples with highest uncertainty, the low uncertainty group contains samples whose $\sigma_{\Delta U^+} < 0.2$, the mid uncertainty group contains 20 randomly selected samples in the remaining repository. The averaged prediction error $Err = (\Delta U^+_{DNS} - \mu_{\Delta U^+})/\Delta U^+_{DNS}$ as well as the averaged uncertainty $\sigma_{\Delta U^+}$ are compared with the initial model in the figure. A clear correlation between the error and its estimated uncertainty can be observed, i.e. high uncertainty corresponds to high prediction error. For the most confident sample group in the testing set with $\sigma_{\Delta U^+} < 0.2$, a mean error of 4.3% is achieved. Furthermore, the efficiency of the AL is illustrated when comparing both uncertainty drop as well as the error drop from the initial model. We observe a significant reduction in error of 9 percentage points for the high uncertainty group. It can be seen that at higher uncertainty level more significant improvement of prediction accuracy and reduction in uncertainty is achieved. A possible explanation is that a cluster of least confident roughness properties is DNS-annotated and included to the model via AL. This leads to a significant decrease of the uncertainty for the samples that are similar to these newly incorporated roughness. Thus, it is expected that within a few AL iterations, the prediction error level converges to an acceptable level for all considered topographies in the repository.

Realistic surface prediction

The applicability of the model is evaluated with realistic rough surfaces. The realistic surface scans are labeled as Sandpaper, ICE 1∼3 (Velandia & Bansmer, 2019) and IC surface (Forooghi *et al.*, 2018), which originate from sandpaper, ice accretion on simplified aero-engine nacelle and deposit in internal combustion engines respectively. The PS is obtained by applying radial averaging about the origin of the frequency space (Jacobs *et al.*, 2017) i.e. $(q_x, q_z) = (0,0)$, where q_x and q_z are the wavenumber in wall-parallel x and z directions, respectively. The prediction error as well as the uncertainty regarding each realistic surface are listed in the table 1. Overall, a mean error of 3.2% among all realistic surfaces is achieved. While the prediction error as well as the uncertainty for all other roughness are relatively low, the model seems to be less confident with the surface IC. As observed from Figure 3, this uncertainty might be caused by the noncompliant shape of the PS for the IC surface. Therefore, given the satisfactory performance of the AL framework, it is expected that an extended roughness repository will yield an improved predictive model.

Table 1: prediction error and uncertainty for five realistic surfaces

	Error [%]	$\sigma_{\Delta U^+}$ [-]
Sandpaper	1.9	0.51
ICE1	4.2	0.42
ICE2	3.9	0.26
ICE3	4.0	0.49
IC	2.0	0.60
Average	3.2	0.46

Sensitivity analysis

The sensitivity vector of the roughness characteristic statistics $\boldsymbol{S}_{\mathrm{st}}$, PDF $\boldsymbol{S}_{\mathrm{PDF}}$ and PS $\boldsymbol{S}_{\mathrm{PS}}$ are separately shown in Figure 6 (a) to (c), respectively. From figure 6(a) it can be seen that the roughness height statistics, namely k_{99} and k_t show high sensitivity for the prediction. This is due to the fact that these values serve as the indications of the physical scale, reminding that the height range for $\boldsymbol{I}_{\mathrm{PDF}}$ is scaled with the inputs k_t. A larger roughness height translates to higher skin friction due to the increased effective slope and therefore a higher ΔU^+ (Chan *et al.*, 2015). Lower sensitivity is observed for the wavelength inputs λ_0^* and λ_1^*. This may indicate that extreme large and small roughness wavelengths will not dramatically alter the roughness induced drag (Yang *et al.*, 2021).

As can be seen from the figure 6, asymmetric distribution of $\boldsymbol{S}_{\mathrm{PDF}}$ indicates that the PDF input elements that locate around and below the median height show higher sensitivity for the model output. In contrast, the height distribution near the roughness peaks demonstrates lower sensitivity. This indicates that the neural networks consider the PDF values close to the roughness troughs to be more important relative to the ones at the peaks for the prediction of surface drag. The degree of asymmetry of the PDF curve is usually represented by the third central moment of the PDF, namely the skewness $Sk = \int_S (k - k_{\mathrm{md}})^3 \mathrm{d}S / k_{\mathrm{rms}}^3$. The sensitivity analysis of the model in figure 6(b) may support the previous observations regarding roughness skewness, e.g. Jelly & Busse (2018); Flack (2018), that a positively skewed roughness with $Sk > 0$ or peak-dominated roughness is mapped to a greater ΔU^+ value, while a negatively skewed roughness or pit-dominated roughness translates into a slightly lower value of the velocity retardation.

In figure 6 (c), $\boldsymbol{S}_{\mathrm{PS}}$ is plotted as a function of λ^+ with logarithmic scaling. The coherence function between the localized surface force and the roughness structure is investigated by Yang *et al.* (2021). It is reported that a certain range of the roughness wavelengths show considerable correlation to the local streamwise drag force due to the roughness sheltering effect. Beyond this certain wavelength the coherence starts dropping. This critical roughness wavelength, according to the investigation on 12 different types of roughness at $Re_\tau = 500$, locates in the range of $150 < \lambda^+ < 300$. To qualitatively illustrate the link of this specific length scale with the present research on global drag, the range is added with vertical dashed lines to the figure. It can be seen from the plot that the sensitivity of the model reaches the highest value within the area between the dashed lines. This observation may suggest that the model interprets the PS values at this wavelength range as being dominant for the global drag of the surface. At the same time a region with lower sensitivity is observed at smaller wavelengths. This is supported by the decreasing coherence of local surface force towards smaller wavelengths (Yang *et al.*, 2021).

CONCLUSION

A predictive model based on combination of ENN and AL approach for characterizing the hydrodynamic properties of homogeneous irregular roughness is presented in the present work. The uncertainty of the prediction, which is measured by the standard deviation of the ENN outputs, is shown to be a proper quantity for estimation of prediction error level of the model. The results show that with 20 selected roughness samples by the AL framework, a significant decrease of error of 9 percentage points is observed for the group of roughness samples with the highest uncertainty. The performance of the model is examined on artificial roughness and realistic surfaces. In both cases, very good agreement between the prediction of the ENN and DNS results is observed. The maximum prediction error of the model for the 5 realistic surfaces is 4.2%, while the mean error is 3.2%. The result of presented study confirm the potential of the AL approach for prediction of the hydrodynamic roughness properties. The effects of additional AL-loops on the model performance will be investigated in future work and reported in the conference presentation.

ACKNOWLEDGEMENT

Jiasheng Yang and Pourya Forooghi gratefully acknowledge financial support from Friedrich und Elisabeth Boysen-Foundation (BOY-151). This work was performed on the supercomputer HoReKa and the storage facility LSDF funded by the Ministry of Science, Research and the Arts Baden-Württemberg and by the Federal Ministry of Education and Research.

REFERENCES

Aggarwal, C., Kong, X., Gu, Q., Han, J. & Yu, P. 2014 Active learning: A survey. In *Data Classification: Algorithms and Applications* (ed. C. Aggarwal), chap. 22, pp. 572–597. Florida: CRC Press.

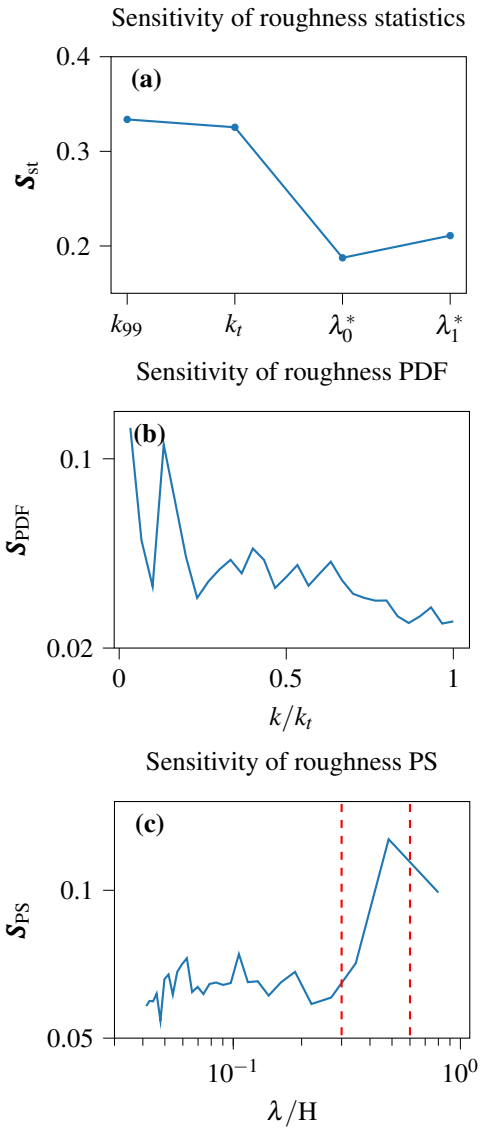

Figure 6: Sensitivity scores of the input vector $\boldsymbol{I}$. (a): $\boldsymbol{S}_{\text{st}}$, (b): $\boldsymbol{S}_{\text{PDF}}$, (c): $\boldsymbol{S}_{\text{PS}}$.

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

Chung, D., Hutchins, N., Schultz, M. P. & Flack, K. A. 2021 Predicting the drag of rough surfaces. *Annu. Rev. Fluid Mech.* **53**, 439–471.

Flack, K.A., Schultz, M.P. & Barros, J.M. 2020 Skin friction measurements of systematically-varied roughness: Probing the role of roughness amplitude and skewness. *Flow, Turbulence and Combustion* **104** (2-3), 317–329.

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

Forooghi, P., Stroh, A., Magagnato, F., Jakirlić, S. & Frohnapfel, B. 2017 Toward a universal roughness correlation. *Journal of Fluids Engineering* **139** (12), 121201.

Forooghi, P., Weidenlener, A., Magagnato, F., Böhm, B., Kubach, H., Koch, T. & Frohnapfel, B. 2018 DNS of momentum and heat transfer over rough surfaces based on realistic combustion chamber deposit geometries. *Int. J. Heat Fluid Flow* **69**, 83–94.

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

Jacobs, Tevis D B, Junge, Till & Pastewka, Lars 2017 Quantitative characterization of surface topography using spectral analysis. *Surface Topography: Metrology and Properties* **5** (1), 013001.

Jelly, T. O. & Busse, A. 2018 Reynolds and dispersive shear stress contributions above highly skewed roughness. *Journal of Fluid Mechanics* **852**, 710–724.

Jouybari, M. A., Yuan, J., Brereton, G. J. & Murillo, M. S. 2021 Data-driven prediction of the equivalent sand-grain height in rough-wall turbulent flows. *Journal of Fluid Mechanics* **912**, A8.

Lee, Sangseung, Yang, Jiasheng, Forooghi, Pourya, Stroh, Alexander & Bagheri, Shervin 2022 Predicting drag on rough surfaces by transfer learning of empirical correlations. *Journal of Fluid Mechanics* **933**, A18.

Lyashenko, I. A., Pastewka, Lars & Persson, Bo N. J. 2013 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.

MacDonald, M., Chung, D., Hutchins, N., Chan, L., Ooi, A. & García-Mayoral, A. 2016 The minimal channel: a fast and direct method for characterising roughness. *J. Phys. Conf. Ser.* **708**, 012010.

Pérez-Ràfols, F. & Almqvist, A. 2019 Generating randomly rough surfaces with given height probability distribution and power spectrum. *Tribol. Int.* **131**, 591 – 604.

Thakkar, M., Busse, A. & Sandham, N. 2017 Surface correlations of hydrodynamic drag for transitionally rough engineering surfaces. *Journal of Turbulence* **18** (2), 138–169.

Velandia, J. & Bansmer, S. 2019 Topographic study of the ice accretion roughness on a generic aero-engine intake. *AIAA Scitech 2019 Forum* .

Yang, J., Stroh, A., Chung, D. & Forooghi, P. 2021 DNS-based characterization of pseudo-random roughness in minimal channels. *arXiv preprint arXiv:2106.08160* .

Yang, X. I. A., Sadique, J., Mittal, R. & Meneveau, C. 2016 Exponential roughness layer and analytical model for turbulent boundary layer flow over rectangular-prism roughness elements. *Journal of Fluid Mechanics* **789**, 127–165.