Towards interpretable data-driven turbulence models

H. Mandler* and Bernhard Weigand*

Recently, numerous data-driven turbulence models have been proposed, most of which show the promising potential to outperform traditional models in complex applications. However, a considerable subset of these closures includes black-box constituents like neural networks (NNs), which are lacking interpretability and whose predictions cannot be readily explained by a human. Both interpretability and explainability are crucial properties of a closure model, though, in order to ensure its reliability, to understand and gain trust in its predictions, and possibly extract hidden domain knowledge.

Hence, whitebox models need to be developed, which are explainable by construction, using, for instance, nonlinear or symbolic regression. As the human intuition for a suitable ansatz function may fail in high-dimensional problems, there are nevertheless use cases for blackbox models. This talk is, therefore, concerned with methods which provide insight into NNs and their prediction process. These methods find answers to questions like: Do all inputs of the model contribute equally to the output or are there particularly relevant or irrelevant inputs? How does each of the inputs contribute to a certain prediction? And how sensitive does the model output react to noisy inputs?

In order to measure the global influence of a particular input on the NNs output, permutation feature importance will be introduced as a simple and efficient sensitivity metric. Furthermore, it will be demonstrated how each input's contribution to individual predictions can be assessed by virtue of Shapley values. An example result for the local importance of a model input is shown in Fig. 1. Finally, an efficient way to quantify and even prescribe a NN's robustness w.r.t. noise will be presented.

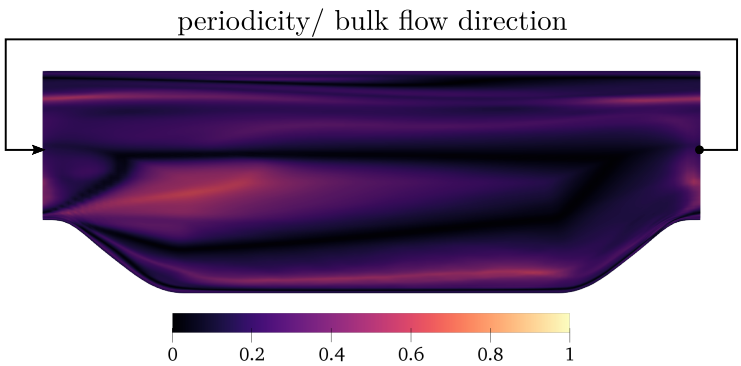

Figure 1: Normalized contribution of a particular input to the model output illustrated for a 2D flow over periodic hills (0 indicates no contribution of this input, whereas a value of 1 indicates vanishing contributions of all other inputs). Apparently, this input is the main driver for the predictions downstream of the hill crest.

*Institute of Aerospace Thermodynamics, University of Stuttgart, Pfaffenwaldring 31, 70569 Stuttgart, Germany