14th International Symposium on Turbulence and Shear Flow Phenomena (TSFP14)
Heidelberg, Germany, July 28–31, 2026

MULTI-TASK GAUSSIAN PROCESS REGRESSION FOR ACTIVE DRAG REDUCTION IN TURBULENT BOUNDARY LAYER FLOWS

Fabian Hübenthal
Institute of Aerodynamics (AIA)
RWTH Aachen University
Wüllnerstraße 5a, 52062 Aachen, Germany
f.huebenthal@aia.rwth-aachen.de

Matthias Meinke
Institute of Aerodynamics (AIA)
RWTH Aachen University
Wüllnerstraße 5a, 52062 Aachen, Germany
m.meinke@aia.rwth-aachen.de

Wolfgang Schröder
Institute of Aerodynamics (AIA)
RWTH Aachen University
Wüllnerstraße 5a, 52062 Aachen, Germany
office@aia.rwth-aachen.de

Dominik Krug
Institute of Aerodynamics (AIA)
RWTH Aachen University
Wüllnerstraße 5a, 52062 Aachen, Germany
d.krug@aia.rwth-aachen.de

ABSTRACT

As environmental goals and increasing energy costs pose technological and economic challenges to air travel and other fields, improvements in aerodynamics are required to reduce energy demand, costs, and environmental impact (Ricco *et al.*, 2021). A promising technique to actively reduce the aerodynamic viscous drag are spanwise traveling transversal surface waves, as analyzed in Albers *et al.* (2019). Given the flow conditions, the goal is to identify the actuation parameters that optimize net power savings, while ensuring that other critical aerodynamic properties, including the lift-to-drag ratio for airfoils, are either neutral or positively impacted.

The objectives of drag reduction and net power savings are evaluated by wall-resolved large-eddy simulations (LESs) with the structured finite volume solver of the in-house simulation framework m-AIA by the AIA (2024). In view of the considerable computational cost of the LESs and the vast design space of actuation parameters, surrogate modeling is key for surrogate-based optimization to efficiently achieve a satisfactory optimum, insightful data, and interpretative knowledge.

In this study, multi-task Gaussian process regression surrogate models with different levels of architectural complexity are analyzed, extending previous work published in Hübenthal *et al.* (2024). The efficacy of these models in learning and leveraging correlations of relative drag reduction and relative net power savings is compared in terms of prediction quality.

METHODOLOGY

The two key constituents of computational fluid dynamics (CFD)-driven surrogate-based optimization are the CFD simulation model and the machine learning (ML) surrogate model. Both models and key metrics for evaluating the surrogate models are introduced briefly below.

CFD simulation model The objective functions relative drag reduction Δc_d and relative net power savings ΔP_net, as defined in Hübenthal *et al.* (2024), are evaluated using wall-resolved large-eddy simulations (LESs) based on the body-fitted deformable structured grid finite volume (FV) solver of the in-house simulation framework m-AIA[1].

The simulation data[2] published in Albers *et al.* (2023) for training the surrogate models are based on a zero-pressure gradient (ZPG) flat plate approximation of the active drag reduction technique using spanwise traveling transversal surface waves, which is validated in Albers *et al.* (2020) and is depicted in Figure 1.

The actuation using spanwise traveling surface waves are parametrized by the wavelength λ, time period T and amplitude A. A fully developed turbulent boundary layer (TBL) is established using the reformulated synthetic turbulence generation (RSTG) at the inflow. The onset of the surface actuation is located at the streamwise position x_0. The boundary conditions (BC) are indicated at the respective boundaries. In the following, the geometric and actuation parameters are nondimensionalized using the inner scaling denoted by $(\cdot)^+$ based on the friction velocity u_τ and the kinematic viscosity ν both at x_0 averaged in the spanwise direction z. The flow conditions are predefined by the momentum thickness based Reynolds number $\mathrm{Re}_\theta = \frac{u_\infty \theta}{\nu} = 1000$ at x_0 and the Mach number $\mathrm{M} = \frac{u_\infty}{a} = 0.1$ using the freestream velocity u_∞, the momentum thickness θ, the kinematic viscosity ν, and the isothermal ideal gas speed of sound a of air.

The CFD model provides the dataset $\mathcal{D}$ consisting of 1 reference/unactuated and 79 actuated LESs in total (Albers *et al.*, 2020) to analyze the dependence of the relative drag reduction $\Delta c_\mathrm{d}(\lambda^+, T^+, A^+)$ and the relative net power savings $\Delta P_\mathrm{net}(\lambda^+, T^+, A^+)$ as the primary objective on the actuation parameters λ^+, T^+, A^+ with $[0,0,0] \leq [\lambda^+, T^+, A^+] \leq [3000, 120, 78]$. The maximal amplitude velocity $v_A^+ = A^+/T^+$ and the area ratio $a = A^+_\mathrm{act}/A^+_\mathrm{ref}$ of the actuated versus the reference/unactuated case are derived based on the explicitly defined formula of the spanwise surface wave presented in Hübenthal *et al.* (2024).

ML surrogate models Due to the computational expense associated with wall-resolved LESs, the relatively inexpensive yet flexible and interpretable surrogate model of

[1] https://git.rwth-aachen.de/aia/m-AIA/m-AIA
[2] https://b2share.fz-juelich.de/

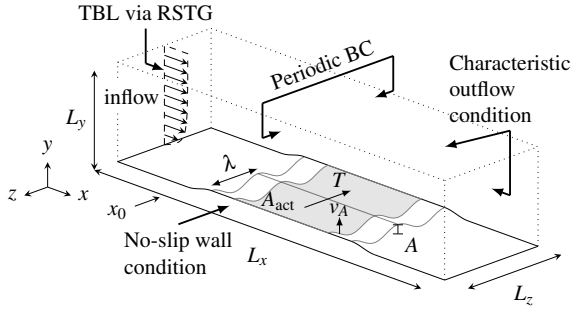


Figure 1: Turbulent flat plate CFD model with spanwise traveling surface waves: The actuated area A_{act} used to calculate the objectives Δc_{d} and ΔP_{net} is highlighted in gray.

Gaussian process regression (GPR) was selected. GPR is data-driven and allows for the incorporation of prior physical knowledge in a Bayesian manner. Furthermore, Gaussian process (GP) models are stochastic by nature. This means they provide a mean prediction along with an estimate of its uncertainty. As outlined in Hübenthal *et al.* (2023), both stochastic estimates can be used jointly in closed-loop Bayesian optimization (BO) workflows to trade off exploitation and exploration to achieve global optima. More details on the fundamentals of GPs can be found in Williams & Rasmussen (2006) and for more information on BO see Shahriari *et al.* (2015).

As the initial hypothesis, positive drag reduction $\Delta c_{\text{d}} > 0$ as task $t = 1$ is necessary for a positive net power saving $\Delta P_{\text{net}} > 0$ as task $t = 2$. Therefore, it is assumed that the two objective functions are causally correlated to some degree, a fact that can be learned and leveraged by surrogate models. In this study, the correlations are learned using multi-task Gaussian process regression (mGPR) with different levels of architectural complexity. These levels range from vanilla single-task GPR (sGPR) (Williams & Rasmussen, 2006) as the baseline for comparison, which cannot directly learn correlations, to the intrinsic model of coregionalization (ICM) (Bonilla *et al.*, 2007) and the linear model of coregionalization (LMC) (Alvarez & Lawrence, 2011) with up to five distinct latent Gaussian processes, and their deep two layer versions DICM and DLMC, to the hierarchical generalization of deep Gaussian processes (DGPs) (Damianou & Lawrence, 2013) with up to three GPs deep and wide, and the highest level of complexity, extending the previous work in Hübenthal *et al.* (2024).

Rather than providing an extensive list of the formulas that define the various models previously referenced, this paper proposes a compact graphical representation. This representation is intended to facilitate rapid and intuitive identification of the differences in the architecture and complexity of each model analyzed. In Figure 2 the graph representation of the sGPR as the simplest model is given. The half elliptic node represents the prior of a Gaussian process with prior mean m, prior kernel k and likelihood function l. The hyperparameters of the prior to be tuned are indicated as $\boldsymbol{\theta}$. The circle with $|\mathcal{D}$ stands for Bayesian inference to condition the prior to the data $\mathcal{D}$ to determine the posterior. In the final node, the expectation operator E is applied to obtain the posterior mean prediction $m_t|\mathcal{D}$ for each task t, here $m_1|\mathcal{D}(\boldsymbol{x}) = \Delta c_{\text{d}}(\boldsymbol{x})$ and $m_2|\mathcal{D}(\boldsymbol{x}) = \Delta P_{\text{net}}(\boldsymbol{x})$ such that $\boldsymbol{m}|\mathcal{D} = [\Delta c_{\text{d}}, P_{\text{net}}]^\top$, at input location $\boldsymbol{x} = [\lambda^+, T^+, A^+, v_A^+, a]^\top$. The set of actuation parameters is extended by the derived physical properties maximal amplitude velocity v_A^+ and the area ratio a. Alternatively, the standard deviation operator S can be used to quantify the uncertainty of the model's prediction as the standard deviation of the posterior $s_t|\mathcal{D}(\boldsymbol{x})$. As an example of a graph representation of a multi-task model, Figure 6 shows the LMC model, which is the best performing multi-task model with relatively moderate complexity. The locations in the architecture where the input and output correlations are explicitly learned are indicated with curly brackets.

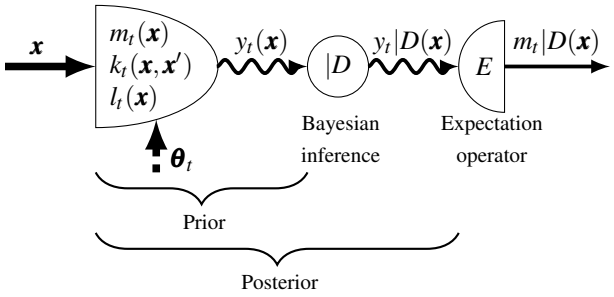


Figure 2: Graph representation of a standard sGPR trained for each task t separately

The models are implemented and analyzed using the GPyTorch[3] library referred to in Gardner *et al.* (2018). All models use a constant prior mean and an automatic relevance determination (ARD) kernel, such as a Matérn or radial basis function (RBF), to model different degrees of differentiability of the outputs, for example. GPyTorch's `IndexKernel` is used for ICM, LMC, DICM, and DLMC. For the sGPR, ICM, and LMC models the model hyperparameters are tuned with exact inference (Williams & Rasmussen, 2006; Bonilla *et al.*, 2007) using maximum likelihood estimation (MLE) as the loss function or maximum a posteriori (MAP) estimation when having hyperpriors $p(\boldsymbol{\theta})$ to best fit the models to the data. The more complex models, such as DICM, DLMC, and DGP, for which exact inference is no longer analytically tractable, doubly stochastic variational inference (Salimbeni & Deisenroth, 2017) is applied to approximate the posterior and to tune the corresponding model and variational hyperparameters. In this approach, the evidence lower bound (ELBO) is used as the loss function. In both cases the ADAM optimizer with a learning rate of 0.1 is employed to solve the resulting hyperparameter optimization problem iteratively, with 1000 training iterations to ensure convergence of the hyperparameters. Additionally, a total of 25 multistarts are performed to circumvent suboptimal local minima during the training process, which is particularly important for multi-layer models with a multitude of variational hyperparameters. In- and outputs are scaled to zero mean and unit variance to enhance numerical stability due to the potentially ill-conditioned covariance matrix inverted for sGPR, ICM, and LMC.

Metrics The performance of the ML surrogate models is assessed using accuracy-based metrics, such as the coefficient of determination and the mean standardized log loss, and correlation-based metrics, such as the local Pearson correlation coefficient.

The **coefficient of determination** R_t^2 of task t is a deterministic measure of the proportion of variance in the output data that is explained by the model. It is defined as:

$$R_t^2 \overset{\text{GPR}}{=} 1 - \frac{\sum_{i=1}^{N} \left(y_{i,t}^* - m_t|\mathcal{D}(\boldsymbol{x}_i^*)\right)^2}{\sum_{i=1}^{N} \left(y_{i,t}^* - \text{mean}(\boldsymbol{y}_t^*)\right)^2} \tag{1}$$

Here $m_t|\mathcal{D}(\boldsymbol{x}_{*,i})$ is the prediction of the model at the test

[3] https://gpytorch.ai/

point $\boldsymbol{x}_i^*$, which is the predictive posterior mean as the most probable prediction for a GPR model, while mean$(\boldsymbol{y}_t^*)$ is the mean of the output data of task t. Thus, the fraction represents the proportion of the mean squared error MSE, quantifying the average squared deviation of the predictive mean from the true outputs, to the total variance in the test data. An R^2 value of 1 implies perfect predictive accuracy, while a value of 0 indicates that the model explains no more variance than the mean of the data, and negative values indicate that the model performs worse than a trivial mean predictor.

The **mean standardized log loss** MSLL_t of task t is a stochastic measure that not only accounts for the accuracy, but also factors in the uncertainty of the model's predictions. The MSLL standardizes the posterior prediction probability of the model p against a trivial baseline predictor p_{ref}, defined as a Gaussian with mean and variance equal to the empirical mean and variance of the training dataset $\mathcal{D}$. In this sense, the MSLL measures how much a model improves over a mean predictor that ignores the inputs entirely and the standardization ensures comparability across different models and datasets. The metric is defined in Equation (2), with the decomposed expression for a GPR model given by Equation (3).

Here $p(y_{i,t}^*|\mathcal{D}(\boldsymbol{x}_{*,i}))$ is the predictive distribution of the model for the input $\boldsymbol{x}_{i,t}^*$, and $p_{\mathrm{ref}}(y_{i,t}^*|\mathcal{D})$ is the baseline Gaussian defined by the training data mean and variance. A negative MSLL indicates that the model outperforms the baseline, values close to zero imply comparable performance, and positive values indicate that the model is less informative than the trivial predictor.

$$\mathrm{MSLL}_t = \underbrace{\frac{1}{N}\sum_{i=1}^{N}}_{\text{Mean}} \underbrace{-\ln\left(\frac{p\left(y_{i,t}^*|\mathcal{D}(\boldsymbol{x}_{i,t}^*)\right)}{p_{\mathrm{ref}}(y_{i,t}^*|\mathcal{D})}\right)}_{\text{Standardized log loss}} \tag{2}$$

$$\overset{\text{GPR}}{=} \frac{1}{N}\sum_{i=1}^{N}\Bigg[\underbrace{\overbrace{\frac{\left(y_{i,t}^* - m|\mathcal{D}(\boldsymbol{x}_{i,t}^*)\right)^2}{2s_t^2|\mathcal{D}(\boldsymbol{x}_{i,t}^*)}}^{\text{Model prediction error}} - \overbrace{\frac{\left(y_{i,t}^* - \mathrm{mean}(\boldsymbol{y}_t^*)\right)^2}{2\,\mathrm{std}^2(\boldsymbol{y}_t^*)}}^{\text{Reference prediction error}}}_{\text{Standardized prediction error}}$$

$$+ \underbrace{\overbrace{\ln\left(s|\mathcal{D}(\boldsymbol{x}_{i,t}^*)\right)}^{\text{Model log uncertainty}} - \overbrace{\ln\left(\mathrm{std}(\boldsymbol{y}_t^*)\right)}^{\text{Reference log uncertainty}}}_{\text{Standardized log prediction uncertainty}}\Bigg] \tag{3}$$

The **local correlation coefficient** $r_{t,t'}(\boldsymbol{x})$ between task t and task t' at input location $\boldsymbol{x}$ for GPR is inspired by the Pearson correlation coefficient. This measure describes the local dependence structure between predictive outputs, i.e., between $f_t(\boldsymbol{x})$, here Δc_{d} as task $t=1$, and $f_{t'}(\boldsymbol{x})$, here ΔP_{net} as task $t'=2$, learned and leveraged by a mGPR model. For real-valued functions $f_t(\boldsymbol{x})$ and $f_{t'}(\boldsymbol{x})$, the one-point location correlation of both outcomes are evaluated at the same input location $\boldsymbol{x}'=\boldsymbol{x}$, is defined by Equation (4).

$$r_{t,t'}(\boldsymbol{x}) = \frac{\mathrm{cov}_f[f_t(\boldsymbol{x}), f_{t'}(\boldsymbol{x})]}{\sqrt{\mathrm{cov}_f[f_t(\boldsymbol{x}), f_t(\boldsymbol{x})]\,\mathrm{cov}_f[f_{t'}(\boldsymbol{x}), f_{t'}(\boldsymbol{x})]}} \tag{4}$$

$$\overset{\text{GPR}}{=} \frac{k(t,\boldsymbol{x},t',\boldsymbol{x})}{\sqrt{k(t,\boldsymbol{x},t,\boldsymbol{x})\,k(t',\boldsymbol{x},t',\boldsymbol{x})}} \tag{5}$$

The coefficient ranges from -1 representing perfect locally linear negative correlation to 1 showing perfect locally linear positive correlation, with 0 representing no locally linear correlation. Within the mGPR framework, these covariances can be expressed directly in terms of the prior kernel function $k(t,\boldsymbol{x},t',\boldsymbol{x}'=\boldsymbol{x})$ for the single-layer ICM and LMC according to the function-space interpretation (Williams & Rasmussen, 2006), which can be written as Equation (5). ICM models with a stationary kernel are a special case, since their correlation is constant for the entire input space. For posterior correlations $r_{t,t'}(\boldsymbol{x})|\mathcal{D}$, the prior covariance k is replaced by the posterior kernel $k|\mathcal{D}$. In general, for deep models like DICM, DLMC, and DGP the local correlation coefficient for the prior and posterior is no longer analytically tractable and requires numerical approximation.

RESULTS AND DISCUSSION

Synthetic use-case-related simulation data with 5-fold cross validation are applied to determine the prior choices of each model that perform best in terms of the performance metrics introduced before. Consequently, the constant prior mean $m(\boldsymbol{x})=C$ and the Matérn prior kernel $k(\boldsymbol{x},\boldsymbol{x}') = k_{\text{Matérn},\nu=2.5}(\boldsymbol{x},\boldsymbol{x}')$ for the actuation parameters and Gaussian likelihood for each GP (sub)model are chosen.

The results of the analysis based on the performance metrics coefficient of determination R^2 and mean standardized log loss MSLL are summarized in Table 1 providing the median and the halved interquartile range IQR/2 in the form of Median $\pm$ IQR/2. The model performance metrics are evaluated on the LES data using 100 randomly sampled data splits and k-fold cross validation with $k=5$. The models are listed in ascending order of complexity, which can be measured in the number of model and variational hyperparameters. The sGPR represents the single-task baseline with the unextended input $\boldsymbol{x}=[\lambda^+,T^+,A^+]$. Except from the sGPR model, all models utilize the extended input $\boldsymbol{x}=[\lambda^+,T^+,A^+,a,v_A^+]$. The asterisk * indicates deep models that include skip connections to forward the input $\boldsymbol{x}$ to subsequent latent or output GP nodes, which are often used for very deep neural networks to combat the degradation problem. The two models that performed best according to the corresponding metric are highlighted in bold. In general, the closer the median of R^2 is to 1 and the smaller the median of MSLL is, the better the model's fit on the test data is. At the same time, the halved interquartile range IQR/2 measures the variability of the performance. The lower the halved interquartile range IQR/2, the more reliable the model achieves the median performance.

Based solely on the performance metrics presented in Table 1, the sGPR+ with the extended input space and the LMC demonstrate the most accurate predictions of both objective functions Δc_{d} and ΔP_{net}. In general, the extension of the input space from $[\lambda^+,T^+,A^+]$ to $[\lambda^+,T^+,A^+,a,v_A^+]$ is more effective in improving the model accuracy than increasing the model complexity alone by moving from single-task to multi-task models. The hypothesis that the area ratio a supports primarily the Δc_{d} prediction and the maximal amplitude velocity v_A^+ helps primarily the ΔP_{net} prediction is verified by investigating the Sobol' indices obtained by variance-based sensitivity analysis (Sobol, 2001) of both models sGPR and sGPR+ using SciPy[4]. Despite the fact that the area ratio a and the maximal amplitude velocity v_A^+ are derived from the actuation parameter set itself and thus are redundant, these features aid in the transportation of physical information, thereby significantly improving the prediction accuracy and reliability, as measured by low halved interquartile ranges. As an interpretation, the explicit addition of these derived features provides a clear support for the models, thereby eliminating the necessity for them to implicitly learn their physical importance. Furthermore, the input space extension enhances the data efficiency

[4]SciPy's implementation of the Sobol' analysis

of the sGPR+ model, thereby enabling it to attain near-optimal R^2 values at approximately half the number of data points required by the sGPR model with actuation parameters only.

Table 1: The performance metrics of the models are listed in ascending order of model complexity and are presented in the form of Median $\pm$ IQR/2: * = With skip connections

Model	R^2_t		MSLL$_t$	
	Δc_d	ΔP_net	Δc_d	ΔP_net
sGPR	0.912 ± 0.051	0.968 ± 0.101	-1.13 ± 0.85	-1.80 ± 0.27
sGPR+	$\mathbf{0.957 \pm 0.031}$	$\mathbf{0.992 \pm 0.007}$	$\mathbf{-2.00} \pm 0.21$	$\mathbf{-2.65} \pm 0.26$
ICM	0.840 ± 0.064	0.968 ± 0.061	-0.82 ± 0.32	-1.74 ± 0.36
LMC	$\mathbf{0.945} \pm 0.044$	$\mathbf{0.992 \pm 0.005}$	$\mathbf{-1.71} \pm 0.33$	$\mathbf{-2.66} \pm 0.28$
DICM	0.931 ± 0.036	0.948 ± 0.030	-1.15 ± 0.55	-1.18 ± 0.49
DICM*	0.924 ± 0.035	0.960 ± 0.035	-0.76 ± 1.13	-1.16 ± 0.96
DLMC	$0.934 \pm \mathbf{0.032}$	0.955 ± 0.028	-1.20 ± 0.65	-1.38 ± 0.55
DLMC*	0.937 ± 0.041	0.963 ± 0.033	-1.00 ± 1.09	-1.28 ± 0.67
DGP	0.313 ± 0.142	0.122 ± 0.163	$+0.06 \pm \mathbf{0.16}$	$+0.08 \pm \mathbf{0.12}$
DGP*	0.828 ± 0.068	0.836 ± 0.083	$+0.09 \pm \mathbf{0.19}$	$+0.17 \pm \mathbf{0.20}$

Among the multi-task models, those explicitly learning correlations, such as ICM, LMC, DICM, and DLMC, demonstrate superior performance in comparison to those implicitly learning correlations, including the best performing model among DGP models. According to the results in Table 1, the LMC with a low to medium complexity is the best performing multi-task model with respect to both performance metrics. The LMC model even slightly outperforms the sGPR model in terms of the net power savings ΔP_net. Both GPR+ and LMC clearly improve the R^2 score in comparison to the baseline GPR model for both metrics. The MSLL score is significantly enhanced, which hints at advantageous uncertainty calibration of the model in the proximity of the available data points.

Unlike the single-task models, multi-task models such as LMC have the capacity to explicitly or implicitly learn and leverage correlations between the targets Δc_d and ΔP_net. This capacity can hypothetically result in further improved prediction accuracy and data efficiency to achieve satisfactory prediction quality. Figure 3 provides an isosurface plot of the explicitly learned correlations $r_{\Delta c_\mathrm{d},\Delta P_\mathrm{net}}(\lambda^+,T^+,A^+)$ between Δc_d and ΔP_net in the actuation parameter space spanned by λ^+, T^+ and A^+. Each dot represents an LES as part of the training dataset $\mathcal{D}$. Red areas indicate regions of positive and blue isosurfaces show regions of negative local linear correlations. The objectives Δc_d and ΔP_net are clearly partially harmonizing, i.e., regions of positive local correlations $r_{\Delta c_\mathrm{d},\Delta P_\mathrm{net}} > 0$, partially uncorrelated, i.e., regions of near-zero local correlations $r_{\Delta c_\mathrm{d},\Delta P_\mathrm{net}} \approx 0$, and partially conflicting, i.e., regions of negative local correlations $r_{\Delta c_\mathrm{d},\Delta P_\mathrm{net}} < 0$. Figure 3 reveals the complex correlations uncovered by the LMC model. These complex correlations cannot be captured by simpler models, such as the ICM model, which is only capable of learning constant correlations for the entire input space, explaining the lower R^2 and lower MSLL scores than the LMC model.

As a hypothesis and in general, deep Gaussian process models, such as DICM, DLMC, DGP can learn highly complex and nonlinear correlations and latent features, but also introduce more ($\mathcal{O}(100)$) model hyperparameters of the prior mean, prior kernel and the likelihood of each sub-GP-model to be tuned. Since exact inference is no longer tractable for deep models variational inference needs to be applied to approximate the posterior introducing orders of magnitude more ($\mathcal{O}(100,000)$) variational hyperparameters alongside with the model hyperparameters. The performance metrics analysis reveals limitations of deep GP models when trained on small simulation datasets in terms of the number of LESs, where insufficient data density hinders the convergence of the hyperparameter tuning and weakens correlation learning. Therefore, the quality of prediction by DICM, DLMC, and DGP is inferior to that of the less complex sGPR and LMC model. Especially, the DGP models suffer from poor calibration of the uncertainty in the model resulting in even slightly positive MSLL scores. Therefore, the posterior standard deviation $s|\mathcal{D}(\boldsymbol{x})$ becomes narrow and spatially near-uniform and in some cases noisy hinting at a susceptibility to overfitting. As an exception, the skip connections of DGP* significantly helped to improve the prediction quality by transporting the physically motivated inputs to the latent and output GP layers. The spread of performances measured as the interquartile range of the MSLL is significantly lower for the DGP than for the other models hinting at a relatively flat ELBO landscape. This means that different hyperparameter settings can explain the data equally well but lead to different predictive behaviors resulting in a relatively large spread in the R^2 metric.

In addition to the accuracy- and correlation-based metrics, the physical plausibility of the input and output spaces is checked, which is visualized exemplarily in the isosurface plots of the posterior mean predictions over the input space spanned by $[\lambda^+,T^+,A^+]^\top$ of the relative drag reduction Δc_d in Figure 4 and of the relative net power savings ΔP_net in Figure 5 for the sGPR baseline model with the unextended input space and the LMC as the best-performing multi-task model. In general, the sGPR+ and the LMC not only show similar performances in terms of the R^2 and MSLL scores, but also delivered qualitatively the same isosurface plots for both objectives with slight difference distant from the data points. For this reason, Figure 4 and Figure 5 only show the isosurface plots of the LMC, which are also representative of the sGPR+. Generally, the sGPR is more optimistic at predicting large relative drag reduction and net power savings for $A^+ \to 0$ and $T^+ \to 0$ than the LMC. The isosurfaces of LMC for $A^+ \to 0$ and $T^+ \to 0$ approach zero or negative values of Δc_d, whereas the isosurfaces for ΔP_net tend to zero distinctly faster. This observation is in accordance with the region of learned negative local correlations for small periods T^+ and medium to small amplitudes A^+ shown in Figure 3. The maximal drag reduction with $\Delta c_{\mathrm{d,\,max}} \approx 30\%$ for both models is of similar order, but slightly shifted toward larger T^+ values for the LMC model, see Figure 4. Similarly, both models predict maximal net power savings around $\Delta P_{\mathrm{net,\,max}} \approx 10\%$ for sGPR and $\Delta P_{\mathrm{net,\,max}} \approx 13\%$ for LMC within the LES data, see Figure 5 However, distant from the LES data, the sGPR model overpredicts net power savings of $\Delta P_{\mathrm{net,\,max}} \geq 40\%$ for small amplitudes A^+ and periods of T^+. This overprediction is identified as physically implausible in the context of the locally achieved drag reduction in Figure 4. It is also evident when considering that low amplitudes correspond to minimal actuation, such that $\lim_{A^+\to 0}[\Delta c_\mathrm{d},\Delta P_\mathrm{net}] = [0,0]$ needs to hold. Regions identified as physically implausible are highlighted in Figure 4 and 5 with red ellipses. Further topological differences are indicated by black circles, such as the noise-like Δc_d predictions of LMC for $A^+, T^+ \to 0$ and the local dips for ΔP_net predictions of sGPR in the center of the input domain.

CONCLUSIONS AND OUTLOOK

In summary, sGPR+ with the extended input space produces the best quality in terms of the relative drag reduction in front of the LMC. The LMC slightly outperformed the sGPR

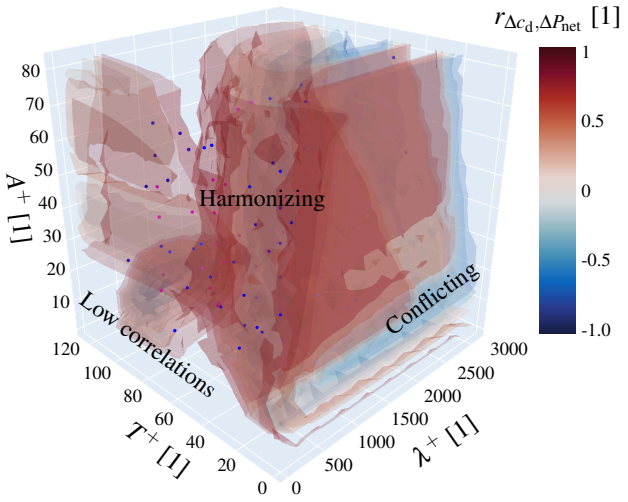


Figure 3: Local posterior correlation coefficient $r_{\Delta c_d,\Delta P_{net}}(\lambda^+,T^+,A^+)$ of the LMC model with median performance during the 5-fold cross validation: Each dot represents an LES.

with respect to relative net power savings. Both yield higher predictive confidence and reliability than the sGPR baseline, which suffers from areas of clearly physically implausible predictions. Extending the input space with physically derived features, such as the area ratio and maximal actuation velocity, proves to be more effective than increasing the complexity and therefore flexibility and expressiveness of the models to learn and leverage inter-objective correlations by using deep multi-task GP models. Furthermore, the LMC model delivers a clear picture of the complex inter-objective correlations that are explicitly learned, and which are shaped by the partially harmonizing, uncorrelated and conflicting regions. As a hypothesis, this motivates the use of more complex models, such as deep GP variants to learn and leverage these complex correlations. However, the analysis also reveals limitations of deep GP models when trained on small simulation datasets in terms of the number of LESs, where insufficient data density hinders hyperparameter convergence and weakens correlation learning. A multi-model approach within a closed-loop, surrogate-based optimization may address this issue, since the GP models with moderate complexity achieve acceptable performance more data efficiently. This involves starting with a simple model and progressing to increasingly more complex models as more data becomes available.

Future work is dedicated toward structurally physics-informed DGP models and the subset of cascaded GPs with physical features assigned to input, latent, and output nodes. This means, the DGP model does not have to implicitly learn these features and correlations. Instead, it can leverage these insights to encode prior knowledge into the architecture, like the extended input space achieved for sGPR+. This strategy could reduce the number of hyperparameters and increase the amount of physically meaningful data fed into the models per LES. Therefore, it could reliably guide the hyperparameter tuning process toward models with enhanced interpolation and extrapolation capabilities. Furthermore, the research is planned to be extended toward a larger input space and more practically relevant Mach numbers, such as $M = 0.2$ to $M = 0.7$. For that, it is hypothesized that the derived physical property of actuation wave speed $v_\lambda^+ = \lambda^+/T^+$ is a key feature because compressibility effects, such as local shocks, may occur at increased Mach numbers as shown in (Shao *et al.*, 2025).

REFERENCES

AIA, RWTH Aachen 2024 m-AIA: A multi-physics PDE solver framework with a focus on fluid dynamics equations. *Zenodo* .

Albers, Marian, Meysonnat, Pascal S, Fernex, Daniel, Semaan, Richard, Noack, Bernd R & Schröder, Wolfgang 2020 Drag Reduction and Energy Saving by Spanwise Traveling Transversal Surface Waves for Flat Plate Flow. *Flow, Turbulence and Combustion* **105** (1), 125–157.

Albers, Marian, Meysonnat, Pascal S & Schröder, Wolfgang 2019 Actively reduced airfoil drag by transversal surface waves. *Flow, Turbulence and Combustion* **102** (4), 865–886.

Albers, Marian, Semaan, Richard, Fernex, Daniel, Noack, Bernd R, Meysonnat, Pascal S, Schröder, Wolfgang & Lintermann, Andreas 2023 Actuated turbulent boundary layer flows dataset. *Tech. Rep.*. Jülich Supercomputing Center.

Alvarez, Mauricio A & Lawrence, Neil D 2011 Computationally efficient convolved multiple output gaussian processes. *The Journal of Machine Learning Research* **12**, 1459–1500.

Bonilla, Edwin V, Chai, Kian & Williams, Christopher 2007 Multi-task gaussian process prediction. *Advances in neural information processing systems* **20**.

Damianou, Andreas & Lawrence, Neil D 2013 Deep gaussian processes. In *Artificial intelligence and statistics*, pp. 207–215. PMLR.

Gardner, Jacob, Pleiss, Geoff, Weinberger, Kilian Q, Bindel, David & Wilson, Andrew G 2018 Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. *Advances in neural information processing systems* **31**.

Hübenthal, Fabian, Albers, Marian, Meinke, Matthias & Schröder, Wolfgang 2023 Surrogate-based optimization for active drag reduction of turbulent boundary layer flows. *PAMM* **23** (4), e202300190.

Hübenthal, Fabian, Meinke, Matthias & Schröder, Wolfgang 2024 Surrogate modeling for active drag reduction in turbulent boundary layer flows using multitask gaussian process regression. *PAMM* **24** (4), e202400186.

Ricco, P., Skote, M. & Leschziner, M. A. 2021 A review of turbulent skin-friction drag reduction by near-wall transverse forcing. *Progress in Aerospace Sciences* **123**.

Salimbeni, Hugh & Deisenroth, Marc 2017 Doubly stochastic variational inference for deep gaussian processes. *Advances in neural information processing systems* **30**.

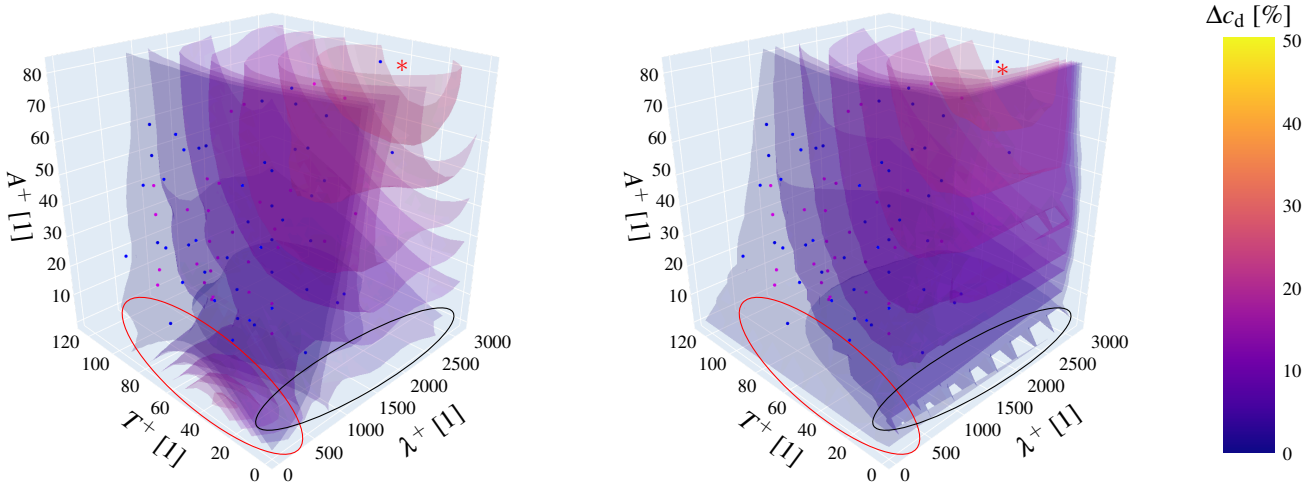
Shahriari, Bobak, Swersky, Kevin, Wang, Ziyu, Adams, Ryan P & De Freitas, Nando 2015 Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE* **104** (1), 148–175.

Shao, Xiao, Albers, Marian, Meinke, Matthias & Schröder, Wolfgang 2025 Compressibility effects on drag reduction by spanwise traveling transversal surface waves in turbulent boundary layers. *Physical Review Fluids* **10** (9), 094603.

Sobol, Ilya M 2001 Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation* **55** (1-3), 271–280.

Williams, Christopher KI & Rasmussen, Carl Edward 2006 *Gaussian processes for machine learning*, , vol. 2. MIT press Cambridge, MA.

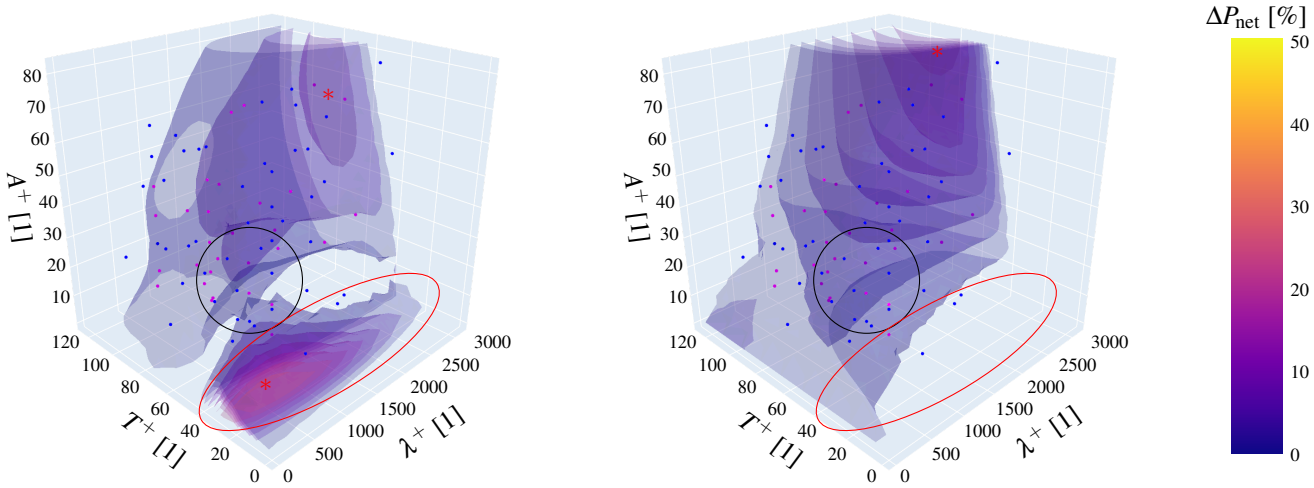
Acknowledgment: The authors gratefully acknowledge the CoE RAISE project for supporting parts of this research in the context of use case 3.1, which receives funding from the European Union's Horizon 2020 – Research and Innovation Framework Programme 2020-INFRAEDI-2019-1 – under Grant Agreement No. 951733. https://www.coe-raise.eu/.



(a) Δc_d for sGPR with $\boldsymbol{x} = [\lambda^+, T^+, A^+]^\top$

(b) Δc_d for LMC with $\boldsymbol{x} = [\lambda^+, T^+, A^+, a, v_A^+]^\top$

Figure 4: Posterior mean predictions $m|\mathcal{D}(\boldsymbol{x})$ for Δc_d with median performance during the 5-fold cross validation: Each dot represents an LES with magenta dots for LESs with $\Delta c_d > 0$ and $\Delta P_{net} > 0$. The red asterisk indicates the maximal posterior prediction $\Delta c_{d,\,max}$.



(a) ΔP_{net} for sGPR with $\boldsymbol{x} = [\lambda^+, T^+, A^+]^\top$

(b) ΔP_{net} for LMC with $\boldsymbol{x} = [\lambda^+, T^+, A^+, a, v_A^+]^\top$

Figure 5: Posterior mean predictions $m|\mathcal{D}(\boldsymbol{x})$ for ΔP_{net} with median performance during the 5-fold cross validation: Each dot represents an LES with magenta dots for LESs with $\Delta c_d > 0$ and $\Delta P_{net} > 0$. The red asterisk indicates the maximal posterior prediction $\Delta P_{net,\,max}$.

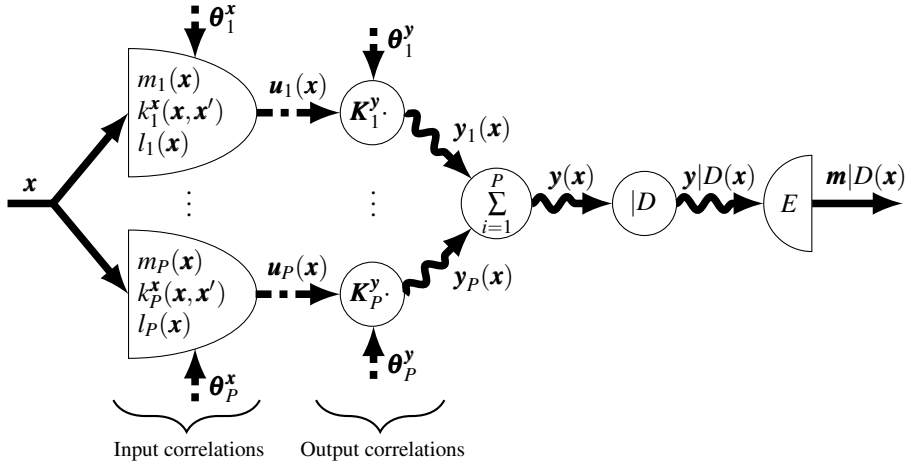


Figure 6: Graph representation of LMC with P, in this study $P \leq 5$, distinct latent Gaussian processes: The structural locations where the input and output correlations are explicitly learned are highlighted.