Shape-feature approach for locally accurate prediction of CFD simulation results

Dr. D. Steffes-lai, Dr. R. Iza-Teran, C. Gscheidle, Dr. J. Garcke

Fraunhofer SCAI, Schloss Birlinghoven 1, 53757 Sankt Augustin,Germany

Summary:

The CAE product development process can no longer be imagined without numerical simulations. In particular, Computational Fluid Dynamics (CFD) plays a vital role in the prediction of flows around internal or external geometries that can be found in many products from different industrial sectors. The resolution and capability of numerical models have been severely improved, whereas the regulations and the requirements on the product behaviour have drastically increased. This trend leads to increasing simulation times, and very large development trees for a design development. Together with even shorter development times the manual comparison of all these design variants becomes cumbersome and is limited to only few simulations.

To address this problem, we propose a machine learning approach to organize the CFD simulations in a structured way, enabling the interactive exploration and postprocessing of several simulation results. The developed methodology learns from a set of simulations with parameter variations the relationship between input and output quantities. For example, the inputs are different variations of the inflow and viscosity and the output is the velocity or pressure distribution in the domain.
In detail, the approach consists of learning a low dimensional parameterization of the flow fields, so in essence the shapes of the function distributions are learnt as functions of the input parameters providing a completely new way of exploring designs and for postprocessing several design solutions. This low dimensional representation of the simulations allows the organization of the simulations in a structured way, that is finding clusters of simulations that behave similar, which yields to a major advantage in a further prediction step for new sets of parameters. Our proposed method includes a cluster based prediction of new designs by radial-basis function surrogate models which yields to much better forecasting quality in local details compared to baseline proper orthogonal decomposition (POD) approaches.
The approach is demonstrated on an OpenFoam HVAC duct use case.

Keywords:

CFD, data analysis, machine learning, surrogate modeling, clustering, dimension reduction

1 Introduction

In modern production processes and due to the increase of computing power, Computational Fluid Dynamics (CFD) have become an invaluable design tool for industrial applications. The resolution and capability of numerical models have been severely improved, whereas the regulations and the requirements on the product behaviour have drastically increased. This trend leads to increasing simulation times, and very large development trees for a design development. Together with even shorter development times the manual comparison of all these design variants becomes cumbersome and is limited to only few simulations.

Recently [1,2,3,4] methods have been proposed to cope with the large amount of data that is generated during the simulation and to deal with the limitation that classical post-processing starts after the simulation is completed. Most notably, the Laplace-Beltrami (LB) shape-feature approach reduces the dimensionality of the flow fields, such as pressure or velocity, in a data independent way. The method is analogous to a Fourier decomposition of the geometry where the flow is taking place using an orthogonal basis that is dependent only on the geometry. This decomposition has been proven to be adequate for representing functions on the mesh in a compact form. Since the approach is independent of the simulation itself, it can also be used during run-time as soon as a time step is available allowing a significant decrease in post-processing time with a convenient overview over several simulations using a low dimensional representation. The approach has shown to overcome the classical Proper Orthogonal Decomposition (POD) applied to analyse automotive crash data [10].

Still, in product development several input parameters have to be modified and requires the computation of many simulations from which surrogate models can be build using regression. State of the art approaches compute a compact representation of the simulation field, the most prominent of it is the POD decomposition. Once this representation is available, after several runs, surrogate models can be build that approximate complete mesh simulation fields under the assumption that the compact features used for the POD decomposition are correlated with the input parameter variations. Still in this case, it has been demonstrated that local variations of the simulation field are actually not well represented using a POD basis and even more this type of behavior is not detectable using global L2 metrics as it is usually done in practice.

To deal with this challenge, we propose the Laplace-Beltrami (LB) shape-feature approach presented in [2, 4] and show its capabilities in the CFD data analysis application domain (see also [1] for the first application in this area). The methodology consists of mainly three steps: First, dimension reduction, second, clustering in the latent space, and third, surrogate modeling for each cluster found.
These three steps are outlined in Section 2. The approach is demonstrated on the turbulent channel flow through a HVAC duct introduced in Section 3. The results for this use case are summarized in Section 4.

2 LB-Shape Feature Approach

2.1 Dimension reduction

As used in [3, 4, 1] the Laplace-Beltrami operator is computed on a surface mesh embedded in three dimensional space. This operator is actually a Laplace operator, the difference is, that it is evaluated on the surface which is represented in three dimensional space.
The numerical evaluation of the LB operator requires the use of a metric that measures the distance along the surface. This distance is called geodesic distance and it is computed based on the shortest path algorithm for triangular surface meshes as described in [11]. The approach in [2,4] was shown to provide a compact representation for functions defined on the mesh, for example the deformations. In general, for functions defined on a surface, it has been shown that the eigenfunctions of the Laplace-Beltrami operator are optimal for representing smooth functions [6].
The eigenvectors φ_i of the operator can be used to represent a function with n components as follows,

$$f = \sum_{i=1}^{n} \langle f, \varphi_i \rangle \varphi_i .$$

Dimensionality reduction is then achievable using only few coefficients obtained through the scalar product $\langle f, \varphi_i \rangle$ (the spectral coefficients). Then, data analysis can be performed by using only those coefficients.

In the presented use case, the actual flow data is three dimensional, but one can consider sections of the domain, for example the mid plane section or the duct walls. This gives as a result a geometrical surface. The proposed procedure of deriving a reduced representation of the flow data consists of two major steps:
1) Evaluate the discrete Laplace Beltrami operator on each respective geometric surface and calculate the eigenvalue and eigenvectors.
To reduce storage space, only the first ten percent of eigenvectors with the largest eigenvalues are determined. They form a subspace consisting of the spectral components with the largest spatial wavelengths. It is worth noting that the resulting basis solely depends on the geometry of the respective duct part and is independent of the flow solution. This computation is performed only once per geometry part, before the actual numerical simulation.
2) Project all flow variables for each time step and part separately on the respective spectral basis.
If the flow variables are vectors, each component, such as *Ux*, *Uy*, *Uz*, is treated independently. This results in one coefficient per basis vector, flow variable component, time step and geometry part. Considering the independence of the basis from the solution, the coefficients are particular suitable for a comparative analysis of large numbers of different simulations.
In a practical workflow the spectral basis has to be computed only once for each geometry part. This is done prior to the first transient simulation. During the run time of the simulations, the solution of one time step can be transformed to the new representation as soon as it is available. There is no need to store all transient flow data in order to perform an analysis afterwards and only the coefficients for the reduced basis are kept during run time. Besides, the projection onto the new basis can easily be done on subareas of the geometry, e.g. the meshes that result from the parallel execution of the core solver on different processes. In addition, it is possible to do this directly in memory which will reduce IO–operations considerably.

2.2 Clustering

The projection coefficients obtained by the step above can be used as low dimensional representation of the simulations.
In this work we use DBSCAN which stands for "Density-Based Spatial Clustering of Applications with Noise" and represents a density-based spatial clustering method with outlier detection. It is based on the fact that we perceive clusters as denser than the points outside. If points are less dense cluster, we perceive them as outliers, also called noise.
The DBSCAN algorithm can be summarized into the following steps:[12]

- Find the points in the Epsilon neighborhood of every point, and identify the core points with more than *MinPts* neighbors.
- Find the connected components of core points on the neighbor graph, ignoring all non-core points.
- Assign each non-core point to a nearby cluster if the cluster is an *Epsilon* neighbor, otherwise assign it to noise.

A maximum group of density-connected points defines a cluster. That is, all points that are not in one of the clusters found in this way are classified as noise. Thus, the algorithm determines the number of clusters with the help of the density itself. For this, *Epsilon* and *MinPts* are needed as parameters. Since there are no unique optimal values, these must be optimized by algorithms or searches.

2.3 Surrogate Model Computation

A radial basis function (RBF) surrogate model is created for each derived cluster based on the corresponding subset of simulation results to represent the dependency between input parameters and flow variables on the mesh. These surrogate models are used to predict the distribution of the field quantities for a new set of input parameters within the valid parameter design space. The data matrix

$$M = \left(m_{ij}, i = 1, \ldots, N_{data}, j = 1, \ldots, N_{sim}\right)$$

is formed of one simulation per each column (N_{sim} is the total number of simulations considered), where the rows correspond to each data item varied over the simulations, e.g. x, y, z, Ux, p per node in our use case. That is, $N_{data} = N_{nodes} * (3 + N_{vals})$, where $N_{nodes}$is the number of nodes and $N_{vals}$the number of flow variables.
In the LB shape-feature approach, an RBF surrogate model using the multiquadric function is chosen which provides the non-degeneracy of the interpolation matrix for all finite data sets of distinct points and all dimensions [7]. The nonlinear RBF surrogate model is extended with polynomial terms in order

to reconstruct polynomial dependencies, especially linear dependencies, exactly. To accelerate the computation a singular value decomposition (SVD) of the data matrix can be performed and used in the RBF metamodeling. Moreover, the precision of the prediction can be estimated using a cross validation procedure which leads to predicted local model tolerances in each test point.
Details on the chosen RBF surrogate model and its SVD acceleration can be found in [8]. In particular, the chosen surrogate model approach yields to accurate prediction results with only a low number of data points used which is a huge advantage especially over neural network approaches.

3 HVAC Test-Case and Numerical Setup

The turbulent channel flow through a HVAC duct was chosen as a suitable test case. It provides both, an industrial relevant application and complex flow features that allow the demonstration of advanced data analysis. The geometry is set up according to the description in [9] to allow a comparison with their experimental and numerical studies. It consists of a duct with a rectangular cross section and a bend of 90 degrees. In the lower part of the duct, a flap is placed with an angle of 30 degrees towards the wall. Inside the bend and behind the flap, a flow separation is observed. Both a Reynolds Averaged Navier-Stokes (RANS) simulation and a Large-Eddy Simulation (LES) have been performed to investigate the global shape of the flow field, the drag force that acts on the flap, as well as the aeroacoustic noise generated by the flap. However, the results of this study only focus on the RANS computations to demonstrate the principal workflow and methods on less complex input data. In a more general setting, the presented techniques can also be employed to transient LES results if the time component is released by computing statistical quantities, time-averaged values or modal structures, such as by the Proper Orthogonal Decomposition (POD), before running the analysis.

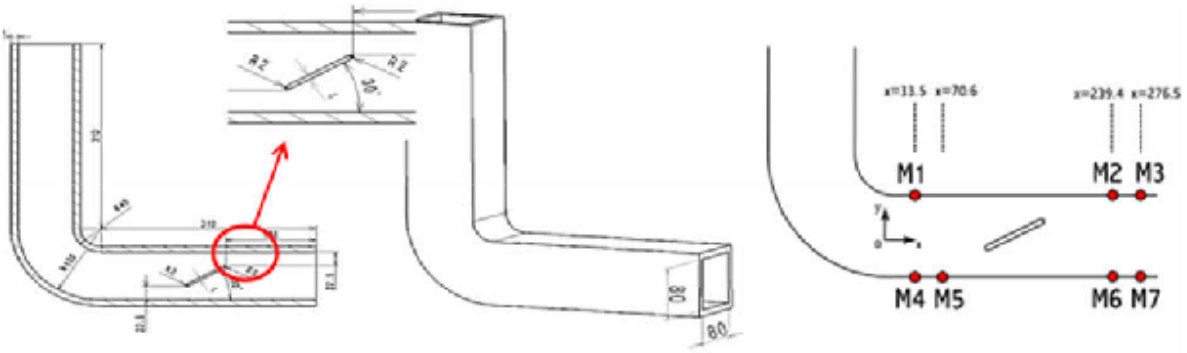

Fig. 1: Geometric dimensions and sensor position of the HVAC duct ([9])

A parameter study of a total of 60 simulations is conducted by varying the volume flow V through the duct as well as the kinematic viscosity v of the fluid both sampled from a symmetric Gaussian distribution with the corresponding mean, -3 sigma and +3 sigma values as shown in Table 1. This leads to a variation of the Reynolds number Re_D in respect to the height of the duct between around 4,400 and 134,000 throughout the study.

	V [m^3/s]	v [m^2/s]	Re_D [-]
+3 sigma	0.128	$1.818 * 10^{-5}$	134,000
mean	0.064	$1.506 * 10^{-5}$	53,120
-3 sigma	0.0064	$1.194 * 10^{-5}$	4,400

Table 1: Configuration of parameter design space used throughout the study and corresponding Reynolds number.

In the following analyses we will investigate two kind of quantities of interest. First, the drag force coefficient C_d, which quantifies the force onto the flap in horizontal direction caused by the surrounding flow. It represents a common scalar quantity of interest that is oftentimes subjects of an optimization task. It is non-dimensionalized by the constant input velocity and density of the flow and the area of the duct‘s cross section. Second, we investigate the global shape of the velocity in x direction (Ux) and pressure field (p) defined on a mesh section in the midplane of the duct. As the flow is mostly constant in the normal direction of the plane, all important features can be observed in this simplified two

dimensional view on the results while reducing computational efforts for the analysis. Here, we are essentially interested in the prediction of the principal shape of the global function and finding regional features such as flow separations in contrast to the analysis of scalar values which is restricted to local effects.

4 Results

The LB-shape feature approach outlined in Section 2 is used to predict the velocity in x direction (*Ux*) and pressure field (*p*) as a nodal grid function for the HVAC duct use case described in Section 3. A dimensionality reduction using the Laplace-Beltrami operator as described yields to the three dimensional embedding of the simulation points as shown in Fig. 2. In this latent space a correlation between the input volume flow V and the embedding coefficients becomes visible. A clustering of the simulation points in the embedding space results in two clusters, Cluster I and Cluster II, and some outlier points which are too far away from the other points and yet not considered in the further data analysis. Simulations within a cluster behave much more similar than outside a cluster. Outliers show different behavior with respect to the distribution of the analyzed field quantities. Table 2 summarizes some features of the two clusters.

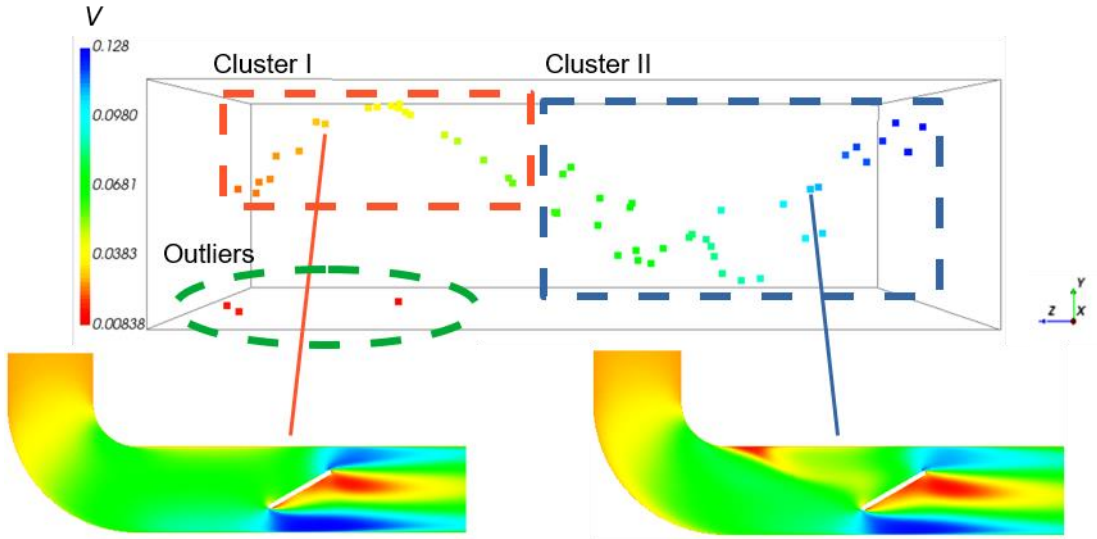

Fig. 2: Embedding of the simulation points in three dimensions and split in two clusters. Color coding shows correlation with input volume flow V.

	Cluster I	**Cluster II**
Count total	20	36
$\boldsymbol{V}_{\mathbf{min}}$	0.02	0.05734
$\boldsymbol{V}_{\mathbf{max}}$	0.05734	0.12787
Count train	14	22
Count test	6	13

Table 2: Features of the two resulting clusters and their split into train and test data sets.

The second cluster with 36 simulation points contains some points with *V* around 0.05 to 0.06 which could not be assigned clearly to that cluster. A further possibility would be to split this cluster up in again two clusters. This might be reasonable since the behavior of these simulation points mark some kind of "transition" between the two clusters regarding the velocity distribution *Ux*. Since a further split of Cluster II in two clusters lead to very few data points per cluster, we decided not to separate them further.

For each of the clusters we perform a train test split of the data points, so that the test data points lay inside the corresponding design space which is visualized in *Fig. 3*.

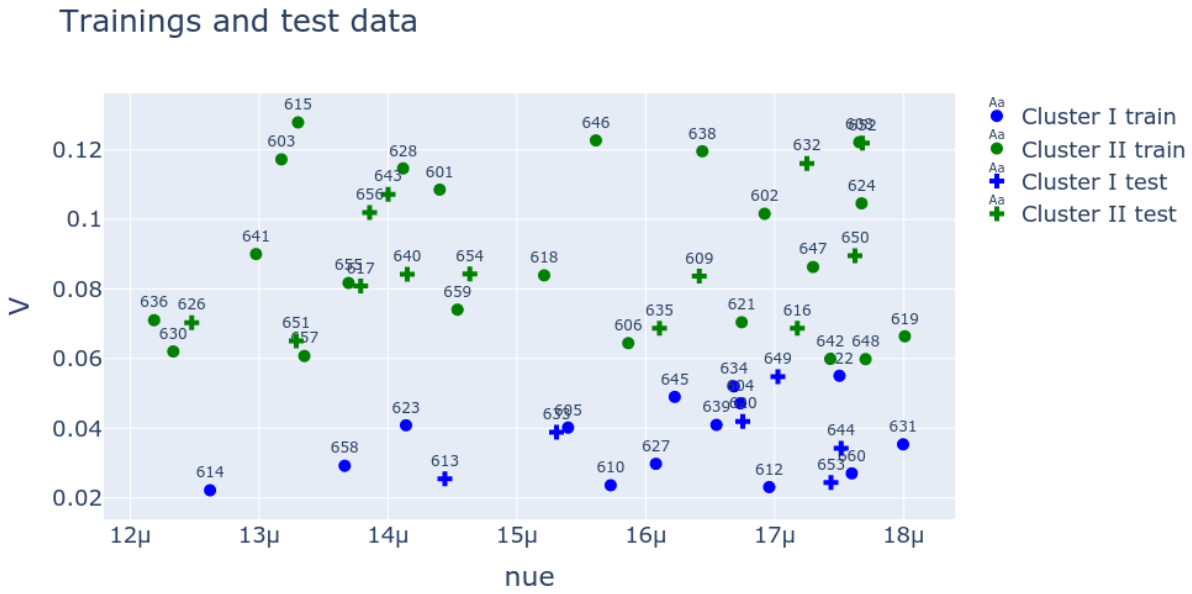

Fig. 3: Overview of the data points in the design space and split into train and test data for Cluster I and Cluster II.

Following the LB-shape feature approach, we create a surrogate model for each cluster using the corresponding training data sets as inputs. We compare the prediction results of the cluster approach against a single surrogate model using a combination of the training data points in Cluster I and Cluster II, called Cluster I+II in the following. As error measure we compute the relative difference with respect to the simulated result, that is

$$diff_{rel} = |\frac{y_{real} - y_{pred}}{y_{real}}|.$$

4.1 Prediction of scalar key values

In a first step, usually some global scalar engineering quantities of interest are investigated. We use the surrogate models per cluster to predict the drag force coefficient C_d. A summary of the results is given in *Table 3*. In all three cases, the relative difference between the simulated and the predicted result is very small, that is for all test points the maximal deviation is < 0.058%, where the mean deviation is < 0.02%. Using one surrogate model per each separate cluster even further improves the overall prediction accuracy. Additionally, the predicted precision of the interpolation is visualized as error bars in each test point in Fig. 4 using Cluster I+II. This shows, that the prediction is highly accurate, only the test point 652 shows some higher tolerance value, which can be explained by the position of the test point at the border of the design space since the used interpolation method is not suited for extrapolation.

		Cluster I (Cluster I+II training)	**Cluster II (Cluster I+II training)**	**Cluster I**	**Cluster II**
	count	6	13	6	13
diff_rel (C_d)	mean	0.000166	0.000209	0.000068	0.000191
	std	0.000183	0.000175	0.000042	0.000171
	max	0.000445	0.000580	0.000104	0.000554

Table 3: Statistics of scalar prediction results for the different sets of training data used.

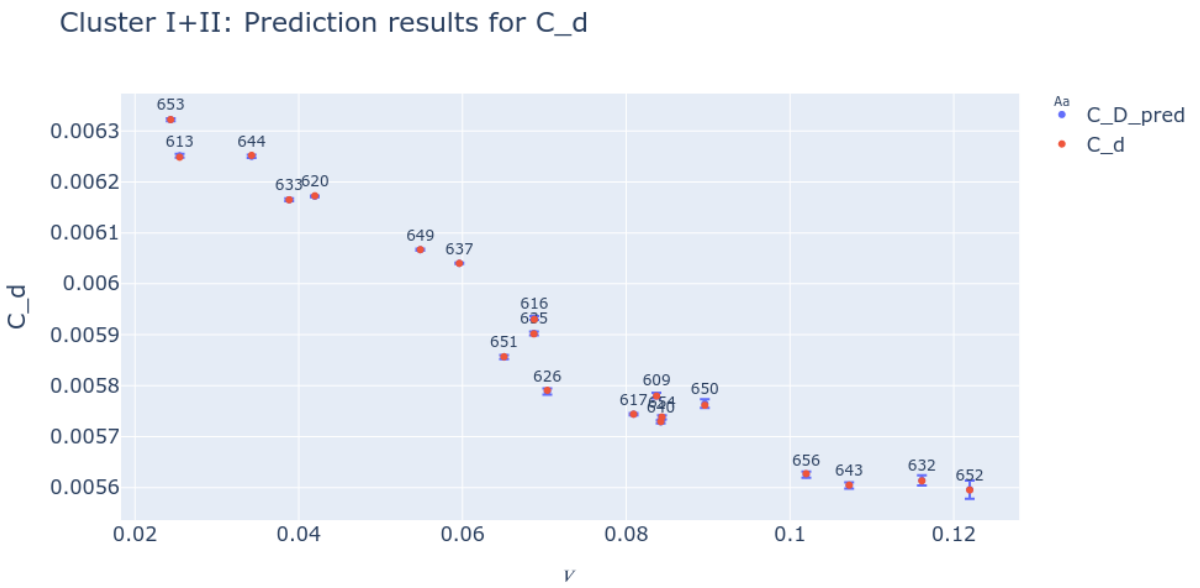

Fig. 4: Prediction results of C_d together with local model tolerances as error bars using Cluster I+II.

4.2 Prediction of field quantities

We have seen that the surrogate models are very well suited for the prediction of the scalar values. Therefore, we use them to predict the field quantities of interest, that is for each node of the mesh the velocity in *x* direction as well as the pressure is predicted. As described in Section 2.3 one metamodel is created for all the nodal values. Again, we compare the prediction quality of using the two separate clusters against the model based on using all data points in the design space.
First, we look at Cluster I corresponding to the points with low input volume flow values. In detail we exemplary investigate the test point 620 within the design space, see *Fig. 3*.

The global shape of the velocity in x direction is very well captured by the prediction, as can be seen in Fig. 5 in both cases, that is using Cluster I as well as with Cluster I+II. Nevertheless, using samples from the entire design space (Cluster I+II) yields to some higher local errors inside the bend and behind the flap, which are exactly the local areas of interest where a flow separation can be observed for higher volume flow values. Thus, the sample points with a high volume flow seem to affect the prediction results especially in these local regions of interest. Thus, the LB-shape feature approach with using only samples from Cluster I locally improves the prediction quality, see Fig. 5, in the regions of interest, as expected.

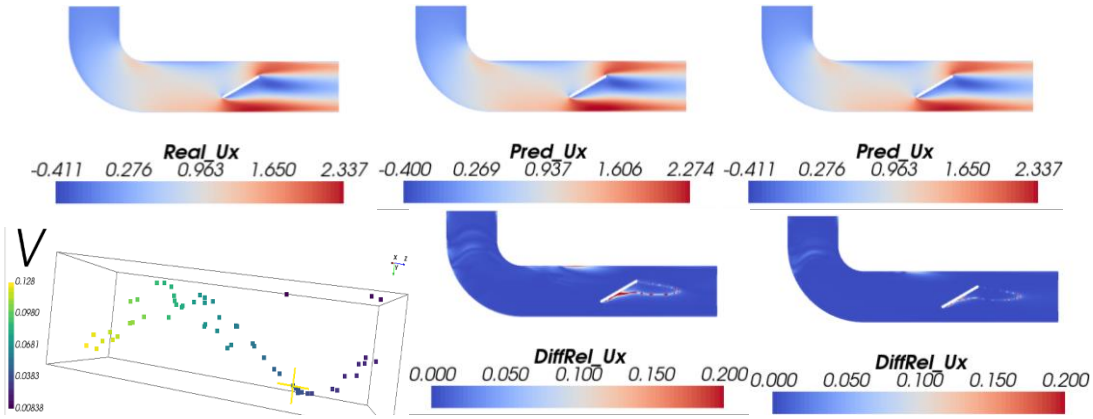

Fig. 5: Comparison of prediction results of grid flow variable Ux exemplary in test point sim620.
1st column: real simulation result and position of test point in latent space,
2nd and 3rd column: predicted result, and relative differences, using simulations in Cluster I+II (left) and Cluster I separately (right) as training data.

Fig. 6 shows the prediction results for the same test point 620 for the pressure field. Again, we observe that using only points from Cluster I improves the overall prediction of the pressure field, and especially in the local areas of interest the accuracy is highly improved.

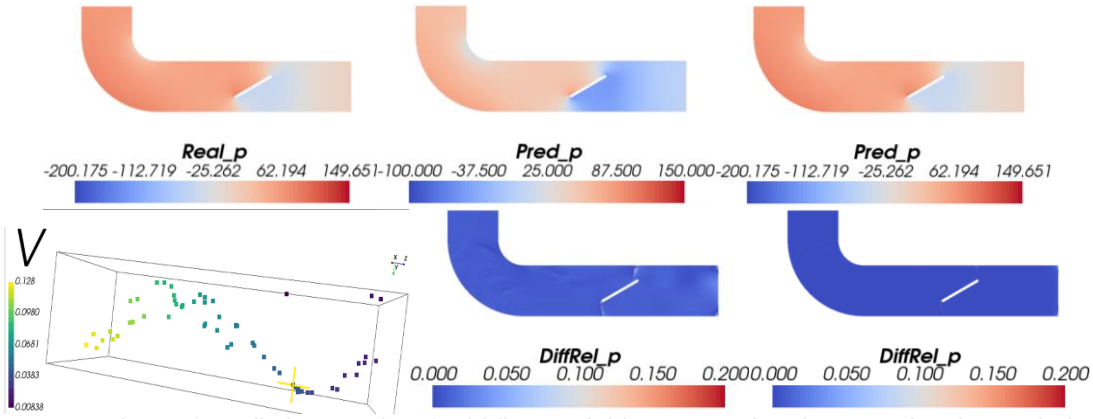

Fig. 6: Comparison of prediction results of grid flow variable p exemplary in test point sim620 belonging to first cluster.
1st column: real simulation result and position of test point in latent space,
2nd and 3rd column: predicted result, and relative differences, using simulations in Cluster I+II (left) and Cluster I separately (right) as training data.

For the other test points in Cluster I very similar prediction results are obtained.
Second, we look at Cluster II corresponding to the points with medium and higher input volume flow values. Exemplary, we investigate the two test points 643 (see Fig. 7) inside the cluster and test point 640 (see Fig. 8) which is one of the socalled transition points. The prediction of the pressure field gives in all cases good and similar results to the points considered in cluster I. Therefore, we focus on the results of the velocity field in *x* direction, especially on the detection of the flow separations. The global shape of the velocity in *x* direction is also well captured by the prediction in both cases, as can be seen in Fig. 7 and Fig. 8. Nevertheless, we observe again higher local errors inside the bend and behind the flap, where a flow separation can be observed looking at the simulation results.

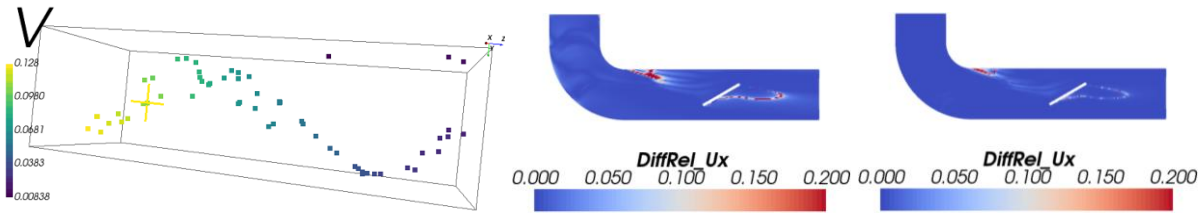

Fig. 7: Comparison of prediction results of grid flow variable Ux exemplary in test point sim643 belonging to second cluster: position of test point in latent space (left), relative differences in Ux using simulations in Cluster I+II (middle) and Cluster II separately (right) as training data.

Using only training points from Cluster II to generate the surrogate model, the local accuracy behind the flap is clearly improved in both the test points considered. On the other hand, the local prediction accuracy is still not adequate inside the bend, especially for the transition point, see Fig. 8. A further split of this cluster could be an adequate way to further improve the prediction quality locally.

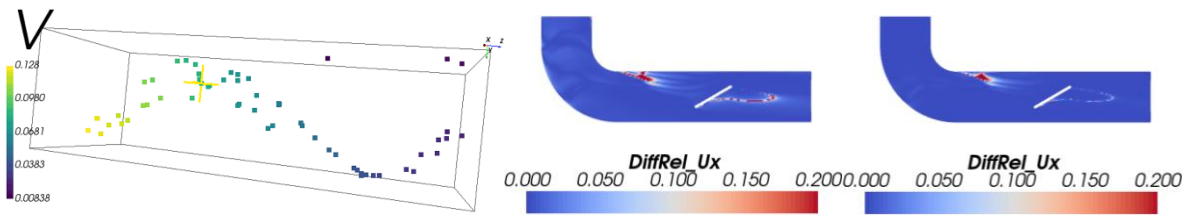

Fig. 8: Comparison of prediction results of grid flow variable Ux exemplary in test point sim640 belonging to second cluster: position of test point in latent space (left), relative differences in Ux using simulations in Cluster I+II (middle) and Cluster II separately (right) as training data.

5 Conclusion

In this study, preliminary results are presented for a new approach for the analysis of turbulent flow data. Thereby, the flow solutions are projected to a spectral basis derived from the Laplace-Beltrami operator on a geometric surface. The resulting coefficients are used to cluster simulations with respect to some field quantities like velocity or pressure. The coefficients are shown to be correlated with the input parameters volume flow and kinematic viscosity. Based on this relationship, a predictor can be constructed using simulations that are a) more similar and b) correlated to the input quantities.
A predictor adapted to the so derived clusters is shown to improve the prediction quality with respect to the case of taking all simulations together. We have shown that with only a few simulations (around 20 per cluster), the prediction of localized velocity changes can be improved with respect to state of the art approaches like the RBF surrogate models which take also non-linearity into account.
Future work will include improvements for clustering and specially the treatment of time dependent cases. An automatic workflow will be developed that, given a set of simulations, automatically detects interesting time steps where clusters start to develop, and next generates predictors for field quantities based on the clusters and the correlation to input parameters.
Furthermore, we research on the mathematical theory to develop a way to justify theoretically to which extension a correlation between spectral coefficients and input parameters can be found.

6 References

[1] Garcke, J., Iza-Teran, R., Gscheidle, C.: “Analysis of Turbulent Flow Data Based on a Spectral Basis Representation”, NAFEMS World Congress 2019, 2019.

[2] Iza Teran, R.: “Enabling the analysis of finite element simulation bundles”, Internat. Jour. for Uncertainty Quantification, 2014, Pages 4(2):95–110.

[3] Iza Teran, R.: “Geometrical Methods for the Analysis of Simulation Bundles”, Dissertation, Universität Bonn, 2016.

[4] Garcke, J., Iza-Teran, R.: “Machine learning approaches for repositories of numerical simulation results”, In 10th European LS-DYNA Conference 2015, 2015.

[5] Iza-Teran, R. and Garcke, J.: “A geometrical method for low-dimensional representations of simulations”. SIAM/ASA Journal on Uncertainty Quantification, 2019.

[6] Aflalo, Y. et.al.: “On the Optimality of Shape and Data Representation in the Spectral Domain”, SIAM J. Imaging. Sciences, Vol. 8, No. 2, 2015, pp. 1141–1160.

[7] Buhmann, M.D.: “Radial Basis Functions: Theory and Implementations”, Cambridge University Press, 2003.

[8] Nikitin, I., Nikitina, L., Clees, T.: “Stochastic Analysis and nonlinear metamodeling of crash test simulation and their application in automotive design”, in Browning, J. E. (Ed.), Computational Engineering: Design, Development and Applications, Nova Science Publishers, New York, 2012, p. 51 – 74.

[9] Jäger, A. et al.: “Numerical and Experimental Investigations of the Noise Generated by a Flap in a Simplified HVAC Duct”, 14th AIAA/CEAS Aeroacoustics Conference (29th AIAA Aeroacoustics Conference), Vancouver, British Columbia Canada, 2008.

[10] Morand L. et. al.: “Learning and analysing the static behaviour of a simulated metal forming process - A comparison between standard approaches and a shape feature approach” (in preparation)

[11] Mitchell, J. S. B. et.al (1987), The discrete geodesic problem. SIAM J. Comput., 16(4):647–668

[12] Ester, Kriegel, H., Sander, J., Xu, X.: “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. In KDD, 1996.