DERIVATION OF PRESSURE LOSS MODELS FOR TURBINE CENTER FRAMES VIA AN L1-REGULARIZED REGRESSION

Marian Stagg*, Wolfgang Sanz*, Peter Leitl**, Maximilian Kurzthaler**, Paul Pieringer***

*Institute for Thermal Turbomachinery and Machine Dynamics, Graz Univ. of Technology, AT; **Bionic Surface Technologies GmbH, Graz, AT; ***ABES Pircher & Partner GmbH, Graz, AT

ABSTRACT
The turbine center frame (TCF) is a key component of modern aero engines, guiding the airflow from the high-pressure to the low-pressure turbine. Due to an increasing flow area, it has a diffusing effect, making it prone to flow separations. Open literature offers various layout guidelines based on performance correlations, but in order to generate parsimonious models, many authors restrict their investigations to a small parameter space while neglecting other important factors. On the other hand, a correlation based on a large number of variables might be more accurate yet very impractical. An L1-regularized least squares fit (LASSO) yields a possibility to balance the contradictory demands for a simple and accurate model. In the current work, the LASSO method is used to correlate the total pressure loss of a straight strutted TCF with various geometrical and inlet-flow parameters. The strength of the regularization allows setting the tradeoff between the model's complexity and accuracy. In this way crucial parameters for the performance of the TCF could be identified and a simple polynomial pressure loss model for the design optimization could be found.

KEYWORDS
Turbine center frame, pressure loss correlations, least squares regression, LASSO algorithm

NOMENCLATURE
TCF turbine center frame
HPT high-pressure turbine
LPT low-pressure turbine
OLS ordinary least squares
CV cross validation
$p_i$ correlation parameter
$\lambda$ penalty weight
LASSO least abs. shrinkage and selection operator
$p_{t,in}$ mass flow avg. inlet total pressure
$p_{t,out}$ mass flow avg. outlet total pressure
$p_{t,loss}$ total pressure loss $= \frac{p_{t,in} - p_{t,out}}{p_{t,in}}$
$p_{t,mean}$ mean total pressure loss
$\rho_i$ correlation coefficient
INTRODUCTION

In recent years, fuel consumption has become an important performance measure for modern aero engines, and increasing the bypass ratio has proven to be a very efficient way to reduce overall fuel consumption. The increasing bypass ratio leads to ever-increasing fan diameters, and in order to avoid very high velocities at the fan tips, the rotational speed has to be lowered. For widely used engine configurations with two spools, this directly impacts the design of the low-pressure turbine as it is located on the same shaft. To keep the power output and stage number at the same level while reducing the rotational speed, the diameter of the low-pressure turbine also needs to grow. The larger diameter of the low-pressure turbine raises the importance of the flow channel between the low-pressure and high-pressure turbine. The larger radial offset and a diffusing effect of the component, in combination with strong curvatures, lead to a very complex and sensitive flow behavior.

There are various design guidelines for diffusers in general. A fundamental work was published by Sovran and Klomp (1967), where they investigated various families of straight-walled annular diffusers. An essential finding of their work was the outstanding importance of the area ratio and the normalized diffuser length in predicting the diffuser's performance. Despite their greatly simplified geometries, their correlations still provide valuable predictions and the TCFs of many modern aero engine architectures cluster in a parametric region, which was identified as optimal by their work (Göttlich 2011; Gräsel et al. 2006). Notwithstanding, other less crucial geometrical parameters still influence the performance of TCFs (Gräsel et al. 2006; Wallin and Eriksson 2008; Mimic et al. 2018).

Furthermore, also the inlet flow conditions severely impact the airflow through the channel. The Reynolds and Mach numbers are some of the most important influencing parameters. Both result from a nondimensionalization of the governing Navier-Stokes system, underlining their deep physical meaning. For an engine, especially the Mach number significantly impacts loss mechanisms in TCFs (Göttlich 2011; Singh and Arora 2019), whereas the influence of the Reynolds number decreases for higher values (Singh and Arora 2019). The swirl strength at the inlet is another critical influential factor as it affects the radial pressure gradients and, hence, the stability of the flow (Göttlich 2011; Pramstrahler et al. 2022). Many additional variables mainly concern flow non-uniformities where wake flow or boundary layers might be examples. They often appear as low-momentum fluid in a faster moving main flow, and the low momentum fluid is, in a relative sense, more affected by the pressure gradient caused by the main flow, which subsequently triggers the formation of vortical structures (Göttlich 2011; Dominy and Kirkham 1994). On the other hand, flow non-uniformities also increase the blockage and thus reduce the effective area and thus the pressure recovery, which is why their effects are sometimes also lumped together in a single factor (Sovran and Klomp 1967; Ackeret 1967).

In an early design phase, various correlations based on the above parameters represent essential tools for the layout process. However, many correlations restrict themselves to very parsimonious models based on a few parameters to retain interpretability while sacrificing accuracy. In order to increase the accuracy, more parameters have to be included in the performance prediction models, but as a consequence, the resulting model will be much more complex and unhandy.

But there are methods to balance those contradictory demands for an accurate yet simple model and to rank and select the most important influencing parameters (Ho 1995; Ding et al. 2018), given that a sufficient dataset can be provided. A remarkable feature selection method is the "Least Absolute Shrinkage and Selection Operator" (LASSO) (Tibshirani 1996; Brunton and Kutz 2019), built upon the idea of L1 minimization. The algorithm extends the classical formulation of the least squares regression by adding a norm-based penalty term. The introduction of the additional term has two desirable effects: On the one hand, it helps to prevent overfitting (Dellacasagrande et al. 2022;
Goodfellow et al. 2016; Brunton and Kutz 2019), and secondly, it facilitates the discovery of parsimonious models. Concerning the problem described above, this means that in contrast to a model defined in the ordinary least squares (OLS) sense which will use all available parameters to predict the pressure loss, the LASSO operator will blank out many of them by setting their coefficients to zero. The strength of the penalty term can control the actual number of active parameters, whereby less critical parameters will drop out first. The limit case of a vanishing penalty term will approach the OLS solution, while all coefficients will shrink to zero, and the predictions will approach the mean value for an infinitely strong term.

In this way, the penalty term allows manually setting the tradeoff between the accuracy and complexity of the model, depending on whether a conceptual or a more detailed TCF design model is needed. As the derived models are available in a closed-form equation, they can easily be derived along the various parameters, thus allowing the evaluation of the influence of different parameters on the pressure loss.

**GENERATION OF THE DATASET**

In order to evaluate the impact of various influential factors on the airflow through a TCF, a comprehensive dataset is necessary. A broad parametric space was defined to generate such a dataset, including geometrical as well as inflow-condition variations. The parameters which are varied to create the dataset are listed in Table 1 and are described in further detail in the following sections. The defined parametric space was sampled randomly, and an automated workflow was set up to convert a set of parameters into a computational mesh and an additional file with the boundary conditions for CFD simulations.

**Table 1: Summary of varied parameters**

<table>
<thead>
<tr>
<th>Geometrical parameters</th>
<th>Inlet parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>vel_mean</td>
</tr>
<tr>
<td>phi_m</td>
<td>temp_mean</td>
</tr>
<tr>
<td>d_h</td>
<td>cax_lean</td>
</tr>
<tr>
<td>L_</td>
<td>ccirc_mean</td>
</tr>
<tr>
<td>duct_pc0, duct_pc1, ...</td>
<td>d_hub</td>
</tr>
<tr>
<td>blade_xpos</td>
<td>blade_th</td>
</tr>
<tr>
<td>blade_len</td>
<td>blade_pc0, blade_pc1, ...</td>
</tr>
<tr>
<td>blade_th</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Hub and shroud perturbation factors</td>
<td>Thickness of the shroud boundary layer</td>
</tr>
<tr>
<td>Axial position of the strut</td>
<td>Mean inlet axial velocity</td>
</tr>
<tr>
<td>Length of the strut</td>
<td>Mean inlet static temperature</td>
</tr>
<tr>
<td>Thickness of the strut</td>
<td>Axial velocity profile inclination (hub to shroud)</td>
</tr>
<tr>
<td>Strut perturbation factors</td>
<td>Mean inlet tangential velocity</td>
</tr>
<tr>
<td></td>
<td>Thickness of the hub boundary layer</td>
</tr>
</tbody>
</table>
Geometrical parameters

Starting point of the geometry variation were four state-of-the-art TCF geometries with straight strut blades, which have been tested at the institute in cooperation with aeroengine manufacturers. They acted as cornerstones to define and constrain the investigated parametric space. The parameterization was found, following a two-level approach. In the first step, a few rather fundamental parameters were defined. They allow a rough geometry approximation in terms of a simple straight-walled annular diffuser, and they embed the available TCF geometries in a well-known parametric space used by many researchers (Sovran and Klomp 1967; Gräsel et al. 2006; Göttlich 2011; Singh and Arora 2019).

Figure 1: Geometrical parametrization

This preliminary geometry is shown in Figure 1 as a dashed line, and it is specified by the area ratio (AR), the mean slope (\(\phi_m\)), the inlet channel height (\(d_h\)), and the normalized channel length (\(L_\))_. However, there are still large differences between the contours of the real duct and the annular approximation, indicated by the red arrows. In order to refine the approximation and to better capture the actual contour lines, perturbation functions are introduced. Four perturbation functions for the hub and shroud contours are constructed by subtracting the annular approximation from each of the four available geometries. The perturbation functions are orthogonalized and sorted according to their importance, using a singular value decomposition. Finally, this leads to four tailored perturbation functions that accurately describe the hub and shroud contours of the four available TCF geometries, embedding them in the same parametric space and allowing for smooth transitions between them. Each of the four perturbation functions is associated with a weighting factor \(\text{duct}_pc0\), \(\text{duct}_pc1\), \(\text{duct}_pc2\), and \(\text{duct}_pc3\), which provide a low-dimensional parameter set. A similar approach was taken by Wallin et al. (2006), but instead of "empirical" functions, they used a more generic way to construct the perturbations.

The load-bearing structures are described similarly, where the first rough approximation is made by shifting and stretching a default strut geometry. This approximation introduces three additional parameters: the axial position of the strut (\(\text{blade}_x\)), the thickness of the strut (\(\text{blade}_t\)), and the length of the strut (\(\text{blade}_l\)). The deviations to the real strut geometry are again described by orthogonalized perturbation functions associated with the factors \(\text{blade}_pc0\), \(\text{blade}_pc1\), \(\text{blade}_pc2\), and \(\text{blade}_pc3\). Thus, a complete TCF geometry with hub, shroud, and multiple strut contour lines can be described by the 15 parameters in Table 1 or created if they are specified.
Despite this parsimonious description, the resulting parametric space is still very large and cannot be covered completely with CFD simulations. In order to restrict the investigations to a subspace, only the vicinity of the provided TCF geometries was investigated, assuming that the more relevant geometries characterized by a low pressure loss and high pressure recovery will be "similar" to existing ones. The four available TCF geometries serve as cornerstones, and the convex hull, enclosing them in parameter space, represents the limits for the region of interest. Within these limits, new parametric combinations are chosen randomly, translated into contour lines, and handed to the mesh generator. Due to the high dimensionality, it is difficult to specify limit geometries. However, to give a feeling of how strong the geometrical variations within the data set are, Figure 2 shows three different geometries which are relatively far apart regarding their mean inclination, area ratio, and nondimensional length.

**Inlet-flow parameters**

A similar approach was taken to parametrize the inlet boundary conditions. The mean axial velocity (vel_mean) and the mean temperature (temp_mean) were defined as high-level parameters, allowing the independent control of the Mach and Reynolds number. To simplify the problem and to keep the parametric space as small as possible at the beginning, only a radial distribution of the flow variables at the inlet was prescribed. This assumption neglects flow non-uniformities in circumferential direction, and thus the actual pressure loss in a real engine will be likely higher due to the mixing losses of incoming wake-like structures (Göttlich 2011). Even though this simplification does not interfere with general statements, it is worth keeping it in mind regarding further extensions of the parametric space.

Figure 3: Base functions for the parametrization of the inlet boundary values
The radial distribution of the different flow variables is assembled from a linear superposition of base functions shown in Figure 3, in which every function introduces one additional parameter \((c_{\text{x,lean}}, c_{\text{circ,mean}})\). Finally, the velocity deficit due to the boundary layer is approximated with a wall function (Krug et al. 2017), leading to two additional parameters to control its thickness on the hub and shroud wall \((d_{\text{hub}}, d_{\text{shr}})\). Finally, the velocity and temperature profile at the inlet is scaled to fulfill the initially prescribed parameters \(v_{\text{el,mean}}\) and \(t_{\text{emp,mean}}\). This method offers a flexible framework to incorporate new flow features by introducing new base functions. The sampling of the defined parametric space was carried out randomly, whereby the different parameters were constrained by values taken from previous computational and experimental studies to restrict the investigations to a relevant domain.

**Numerical setup**

For the CFD calculations, ANSYS Fluent was used. A preliminary grid independence study was carried out on two different geometries, an aggressive and a less aggressive one. Four different block-structured meshes are used for each geometry, ranging from approximately 300,000 cells to more than two million cells. The variation of the density, Mach number, static temperature, total temperature, and total pressure at the inlet and outlet between the coarsest and the finest mesh was less than 0.12% and also the wall shear along the center of the shroud surface did only vary slightly during this grid convergence study. For the final calculations, the second coarsest mesh with 548,160 cells was chosen to guarantee \(y^+\) values close to 1 for all cases investigated. The two-equation k-omega SST model (Menter 1993), widely used in turbomachinery applications, was chosen for turbulence prediction.

![Figure 4: Mean y+ values (left) and maximum y+ values (right) for the 300 samples of the dataset](image)

The mean and maximum \(y^+\) values are monitored to ensure the quality of the automated CFD calculations, and the results are shown in Figure 4. The mean \(y^+\) values are shown on the left side of Figure 4, and most of the calculations yield mean \(y^+\) values of around 1.4, while one calculation exceeds a mean value of 2. Regarding the maximum \(y^+\) values, most of the calculations yield values of around 3.5. In four cases, the maximum \(y^+\) value slightly exceeds 5 (Figure 4 right).
MODEL DEFINITION

Each CFD simulation represents a unique data sample that associates a total pressure loss with a combination of the different geometrical and inlet-flow parameters $p_i$. Along with a least squares approach, these combinations of target values and parameters can be used to calculate the coefficients $c_i$ of a prediction function (see Eq. (1)). This prediction function will be referred to as a model in the further course of this work.

$$p_{t \text{ loss}} = p_{t \text{ mean}} + c_1 \cdot p_1 + c_2 \cdot p_2 + \cdots$$ (1)

The model allows the prediction of the total pressure loss for a new parametric combination, and the influence of the individual parameters on the total pressure loss. In order to estimate the importance of the different parameters, the initially stated OLS will be extended with a penalty term leading to the LASSO formulation.

Ordinary Least Squares Method (OLS)

The OLS framework offers a simple and closed-form method to get the unknown model coefficients $c_i$ of Eq. (1). The different combinations of total pressure loss and parametrizations given by the available dataset can be inserted into the model Eq. (1). Each data sample provides one additional equation, and the resulting system can be expressed by Eq. (2). The vector $\tilde{p}_{t \text{ loss}}$ contains the total pressure loss values of the different data samples, the matrix $\tilde{P}$ contains the parameter values whereby each row can be associated with the corresponding data sample. Finally, the vector $\tilde{c}$ contains the unknown coefficients.

$$\tilde{p}_{t \text{ loss}} = \tilde{P} \cdot \tilde{c}$$ (2)

If the number of data samples is larger than the number of available coefficients, this leads to an over-determined system that almost always has no exact solution. Such systems can be solved in a least-squares sense, whereby the loss function $E$ is defined as the sum of squared deviations shown by Eq. (3) and is then minimized. Due to its structure, the problem leads to a closed-form solution containing the well-known Moore-Penrose pseudo inverse (Eq. (4)).

$$E = (\tilde{p}_{t \text{ loss}} - \tilde{P} \cdot \tilde{c})^T \cdot (\tilde{p}_{t \text{ loss}} - \tilde{P} \cdot \tilde{c})$$ (3)

$$\tilde{c} = (\tilde{P}^T \cdot \tilde{P})^{-1} \cdot \tilde{P} \cdot \tilde{p}_{t \text{ loss}}$$ (4)

In the case of linear parameters, as stated in Eq. (1), this procedure is equivalent to a multidimensional linear least-squares regression. If the method should be extended to non-linear functions, the model assumption of Eq. (1) can be extended with non-linear kernel functions applied to the parameters. Polynomial kernels of arbitrary degree, as shown in Eq. (5), often represent a convenient choice. In this case, the parameters of various degrees are generally referred to as features rather than parameters.

$$p_{t \text{ loss}} = p_{t \text{ mean}} + c_1 \cdot p_1 + c_2 \cdot p_2 + \cdots + c_{n+1} \cdot p_2^n + \cdots$$ (5)
Regarding the approximation capabilities of a Taylor series, a high polynomial degree seems preferable to give a good prediction of the total pressure loss. However, especially for many parameters, the number of features and thus the number of free coefficients will grow extremely fast with an increasing degree, leading to an under-determined problem if the number of features exceeds the number of data samples. This problem has no unique but infinitely many solutions, and the least squares procedure will yield one of them but provides no additional possibilities to control the outcome. Furthermore, too expressive features will likely result in spurious oscillations of the response surface (overfitting). Additional constraints must be imposed to recover the uniqueness of the problem and avoid those oscillations. The various norms of $\bar{c}$ are often used to define these constraints, which are imposed as penalty terms. An L2-based constraint, for example, leads to the so-called "ridge regression" (Hoerl and Kennard 2000), and an L0 constraint results in the "best subset selection" (Bertsimas et al. 2015).

THE LASSO METHOD

The LASSO algorithm represents a specific method to incorporate an L1-based penalty term into the OLS. The initial error function of Eq. (3) is extended with a penalty term proportional to the L1 norm of the coefficient vector $\bar{c}$ weighted with a parameter $\lambda$ (Eq. (6)) where the subindices indicate the norm type.

$$E = |\bar{p}_{t \text{ loss}} - \bar{P}|_2^2 + \lambda \cdot |\bar{c}|_1$$

(6)

The modification restores the uniqueness of the solution in almost every case (Tibshirani and Wasserman 2017). Unlike the OLS, the LASSO does not have a closed-form solution; nonetheless, it leads to a convex optimization problem that can be solved efficiently by an iterative procedure (Tibshirani 1996). Furthermore, the LASSO circumvents problems arising from overfitting, and it has the remarkable tendency to generate parsimonious models, making it a valuable tool for feature selection (Brunton and Kutz 2019).

Avoidance of Overfitting

Overfitting is a common problem arising in machine learning. If the available feature library is too expressive, the model tends to specialize in the given train dataset rather than providing general statements (Goodfellow et al. 2016; Brunton and Kutz 2019). Even though yielding a very small train error, the model will likely perform poorly on new data samples, and the true prediction error will be much higher than the train error would suppose (Arlot and Celisse 2010). For example, Figure 5 shows how a polynomial of 10th degree is fitted to a noisy dataset via an OLS (left) and a LASSO (right) procedure. A cross-validation procedure was used to identify the $\lambda$ value for the penalty term. The reported mean squared error of the OLS was 0.0068, thus lower than the error of the LASSO. At first glance, this suggests that the OLS should be preferred over the LASSO method. From a heuristic point of view, however, it seems doubtful that the OLS model would perform well on new and unseen data samples, especially when considering extrapolations. The LASSO shows a slightly higher error, but it removes spurious oscillations visible in the OLS fit (left). Furthermore, the LASSO is capable of identifying the known ground truth model (second-order parabola) used to generate this exemplary dataset.
Sparsity promotion

In contrast to the ridge regression (Hoerl and Kennard 2000), which could also be used to regularize the problem and avoid the difficulties shown in Figure 5, the LASSO additionally has a sparsity-promoting property. For an increasing strength of the penalty term in Eq. (6), more and more coefficients of the initial model of Eq. (1) are shrunk to zero and thus drop out. Tibshirani (1996) provides an intuitive and illustrative explanation of the reason for this behavior. The formulation of Eq. (6) forces solutions to lie on the intersection of L1 isolines and isolines of the error function Eq. (3). Due to the "spikey" nature of the L1 norm and the quadratic form of the error function Eq. (3), chances are high that those intersections lie on a coordinate axis. This, however, implies that some of the coefficients \( \hat{c} \) will be exactly zero.

The increasing penalization in Eq. (6) causes many model coefficients of the original model Eq. (1) to drop to zero, as shown in Figure 6. As this happens, the contribution of the according parameter or feature is removed, simplifying the resulting model. Higher \( \lambda \) values will thus lead to much simpler but also less accurate models. An advantage of all the models defined in closed form is that they can be derived along the different variables.

\[
\begin{align*}
\lambda - & \quad p_t \text{ loss} = p_t \text{ mean} + c_1 p_1 + c_2 p_2 + \ldots \\
\downarrow & \quad p_t \text{ loss} = p_t \text{ mean} + 0 \cdot p_1 + c_2 p_2 + \ldots \\
\lambda + & \quad p_t \text{ loss} = p_t \text{ mean} + 0 \cdot p_1 + 0 \cdot p_2 + \ldots
\end{align*}
\]

Figure 6: Drop out of parameters with increasing \( \lambda \)

Model selection

In order to provide a reasonable range of \( \lambda \) values, the bounds were defined based on the following considerations: For a penalty term strength growing to infinity, all coefficients will shrink to zero, and due to its definition Eq. (1), the model prediction will approach the mean total pressure loss. Given that no additional information regarding the inlet conditions or geometrical properties is available, the mean total pressure loss represents the best prediction. In this sense, this limit case provides the simplest prediction model for the total pressure loss.

For lower \( \lambda \) values, however, the model approaches the OLS solution, which can be problematic as it tends to overfit if the available features are too expressive (Figure 5). Several methods are available to set \( \lambda \) to a value where an overfitting is avoided, and the model still provides very accurate predictions. Cross Validation (CV) is probably one of the most popular methods due to its universality.
and its few primary assumptions (Arlot and Celisse 2010; Homrighausen and McDonald 2012; Ding et al. 2018; Goodfellow et al. 2016). The dataset is split into a train set and a test set. The former is used to fit the model by calculating its coefficients $\hat{c}$, while the latter is used to estimate the prediction error. The procedure is repeated for multiple $\lambda$ values, and finally, the one with the smallest prediction error is chosen. This $\lambda$ value found by cross validation defines the lower bound for the domain of reasonable models.

Still, one thing has to be kept in mind: an increasing $\lambda$ value induces a small amount of bias to reduce the variance of the model (Tibshirani 1996), and the bias grows with $\lambda$. For the most-conceptual models, this implies that the predictions are closer to the mean value than they should be. As a result, the derivatives will be slightly underestimated, even though the location of the optimum will not be affected.

RESULTS

The final dataset used to train the LASSO models contains 300 data samples. Fifteen geometrical and four inlet flow parameters were used to define the different cases. All the prediction models are based on normalized parameters, where the mean is removed and the values are scaled to a variance of one. The total pressure loss, as well as the pressure recovery factor, were defined as the target values, which had to be predicted, based on multivariate polynomial features of the different parameters, as shown in Eq. (5). Combinations up to the second degree (products and squared parameters) were used to create the features from the available parameters. For the 19 available parameters, this led to a total number of 209 features (19 linear parameters, 19 squared parameters and 171 products of the individual parameters).

In order to train the models, the LassoCV object was used, which is part of the Python package Scikit-Learn. A "leave one out cross-validation" (LOOCV) procedure was applied to identify the most accurate model. Alternatively, a model selection was also conducted based on the Akaike Information Criterion (AIC) (Arlot and Celisse 2010; Brunton and Kutz 2019). Both of these methods yielded the same result.

LASSO path and model selection

The strength of the regularization term in Eq. (6) was found to $\lambda = 3.68e^{-5}$ for the most accurate model. This model used 33 features and showed a mean train error of 4.56 % of the mean total pressure loss. An additional dataset containing 50 data samples was used to estimate the true prediction error. The average prediction error on these new samples was 4.43 % of the mean total pressure loss. The test error reaches a maximum value of 19.31 % of the mean pressure loss for an increasing $\lambda$ which represents the maximum deviation from the mean value within the data set. This indicates that the total pressure loss of all data samples varies by this value.

Figure 7 (left) shows the values of the different coefficients in equation Eq. (1) as functions of the penalty parameter $\lambda$. For the model development, the nondimensional Reynolds and Mach number were used instead of the directly related parameters mean velocity (vel_mean) and mean temperature (temp_mean). For the sake of clarity only the coefficients of the five most important features (Mach number, shape factor for the linear distribution of the axial inlet velocity cax_lean and its square, Reynolds number Re and mean circumferential inlet velocity ccirc_mean) are shown in color while the rest is greyed out. The dashed vertical line shows the most accurate model, and thus its corresponding $\lambda$ value represents the lower bound of reasonable models. Below this value, overfitting would occur.
It can be observed that for higher values more and more coefficients shrink to zero until finally all the terms drop out of the equation, and only the mean total pressure loss remains active. Figure 7 (right) shows the number of active features (non-zero coefficients), the train error and the test error as functions of $\lambda$. Again, the dashed line represents the most accurate model. For smaller $\lambda$ values, the train error is reduced further but the test error rises again. For higher $\lambda$ values, the number of non-zero coefficients and thus the model complexity decreases while the train and the test error increase.

**Feature importance**

The result of the LASSO path also yields a ranking for the used features according to their importance in predicting the total pressure loss. Figure 8 shows train and test errors of the models developed for predicting the total pressure loss and the pressure recovery differing by the number of used features. On the abscissa, the different features are sorted towards the right side according to descending importance. Each bar shows the mean train and test error of the different models, which
include the corresponding feature and all the previous ones. The values on the ordinate are normalized by the mean total pressure loss and the mean pressure recovery of the whole data set. For clarity, only the first ten most essential features are shown. It can be observed that the accuracy of the models can be increased when more features are considered. If only the mean values are used for prediction, the total pressure loss prediction yields a mean test error of 19.31\% of the mean total pressure loss, whereas the mean test error of the pressure recovery prediction is 21.03\% of the mean pressure recovery. Regarding the total pressure loss, the most influential features refer mainly to the inlet conditions of the duct. The Mach number is identified as the most crucial parameter for the current dataset, and the axial velocity shape factor \( c_{\text{ax, lean}} \), describing the linear component of the inlet axial velocity, is ranked after. The next features to follow concern the Reynolds number and the swirl strength controlled by \( c_{\text{circ, mean}} \).

The first geometrical feature, blade_len, defining the length of the load-bearing strut, appears in sixth place. The first mixed feature appears in seventh place as the product of the mean slope of the channel \( \phi_m \) and the axial velocity shape factor \( c_{\text{ax, lean}} \). The next ones to follow are a shaping factor for the duct geometry \( \text{duct_pe0} \) characterizing the strength of the curvature in combination with \( c_{\text{ax, lean}} \) and, finally, the Mach number in combination with \( c_{\text{ax, lean}} \). For predicting the pressure recovery, geometrical features play a more critical role as they appear more frequently within the high-ranked features. Especially the area ratio (AR) and the normalized duct length (\( L_\) ) significantly impact the pressure recovery factor. The other features shown in Figure 8 are not explained in detail but are of similar importance. Their ranking can change in case of a modified dataset.

**Pressure loss derivatives**

Equation (7) shows the 4\textsuperscript{th} model of Figure 8 using three features and the mean value. The mean error of the chosen model is reported to be slightly higher than 9.2\% of the mean total pressure loss, implying that the chosen model is still on an abstract rather than a detailed level. It uses the mean total pressure loss, the Mach number, and the squared linear inlet axial velocity profile shape-factor (Figure 3) to predict the loss.

\[
p_{t,loss} = p_{t,mean} + c_1 \cdot Mach + c_2 \cdot c_{\text{ax, lean}} + c_3 \cdot c_{\text{ax, lean}}^2 \quad (7)
\]

\[
\frac{\partial p_{t,loss}}{\partial c_{\text{ax, lean}}} = c_2 + 2 \cdot c_3 \cdot c_{\text{ax, lean}} \quad (8)
\]

Despite its simplicity, the model shown in Eq. (7) can reduce the prediction error from 19.31\% to 9.2\%. The model predicts a positive linear correlation between pressure loss and Mach number, implying that a higher Mach number leads to higher losses. This correlation might be explained by increasing losses in the boundary layer due to the higher velocities within the duct geometry. Further, the identified model uses the parameter \( c_{\text{ax, lean}} \), which controls the inclination of the axial velocity profile at the duct inlet. Here, a quadratic dependency is observed, where the quadratic relation might mainly stem from mixing-out losses being realized within the channel. Velocity differences between hub and shroud equalizes and the mass flow averaged total pressure is smaller for the flat profile than for the inclined one. For the simplified example of an incompressible flow through a straight channel without boundary layer (linear velocity profiles at inlet and outlet), the total pressure loss prescribes
an almost quadratic path if plotted over the inclination of the inlet profile. Eq. (9) shows the total pressure loss for the simplified case as a function of $c_{ax_{\text{lean}}}$, assuming that the inclination of the velocity profile at the outlet is zero. If the representative values for the mean velocity, the static pressure, the channel height and an inclination similar to the upper limit shown in Figure 9 are inserted, Eq. (9) predicts a total pressure loss of around 2%, which is in the same order of magnitude as observed in the data set. This result supports the assumption that those mixing-out losses are the main cause of the identified quadratic relation.

$$p_{t\text{loss}} = 1 - \frac{u_m^2 \cdot \frac{P}{2} + p_{stat}}{\left(\frac{1}{4} \cdot h^2 \cdot c_{ax_{\text{lean}}}^2 + u_m^2\right) \cdot \frac{P}{2} + p_{stat}}$$

Nonetheless, the smallest losses, regarding the parameter $c_{ax_{\text{lean}}}$, are not observed for a flat profile at the inlet, as Eq. (9) would suggest, but for an inclined profile. The shift of the optimum is due to a negative linear correlation between parameter $c_{ax_{\text{lean}}}$ and the total pressure loss, which is superimposed on the quadratic relation. As this behavior cannot be observed in the simplified example, it might stem from the geometry of the channel, even though the geometrical variations were too small to show this relation. Regarding the flow properties, it was observed that higher velocities near the hub surface were beneficial to suppress large backflow regions and a strong thickening of the boundary layer.

Figure 9: Radial velocity distributions at the duct inlet for lowest pressure loss in comparison with the most inclined and the least inclined profile

Figure 9 shows the identified best inclination of the velocity profile at the inlet compared to the limit cases concerning this crucial parameter. The optimum inlet velocity profile is inclined such that the velocity near the shroud surface is higher than that near the hub surface. Interestingly, the variation of the inlet condition within the bounds visible in Figure 9 outweighs the influence of the geometrical variations, as shown in Figure 2, by far. In order to have an effect of similar magnitude on the total pressure loss, the geometrical variations would need to be much larger.
CONCLUSIONS

This work aimed to show the capability of the LASSO method for identifying simple and derivable prediction models for the total pressure loss and pressure recovery of turbine center frames. The method was applied to an extensive dataset, covering a wide range of geometrical and inlet-flow related parameters. For the pressure recovery, a shape factor for the linear radial distribution of the axial velocity at the inlet, the area ratio and the normalized length were identified as the most crucial parameters, which agrees very well with the work of Sovran and Klomp (1967) and confirms the validity of the used dataset as well as the capability of the LASSO method to identify crucial parameters.

Concerning the total pressure loss, the Mach number was identified as the most critical parameter, with the same inlet-flow shape factor $c_{a\text{,}\text{ax\_lean}}$ and the Reynolds number ranked after. Interestingly, geometrical parameters were mostly ranked behind, indicating the high importance of the inlet flow conditions in general. Nevertheless, it has to be pointed out that this is true for the chosen limits in parametric space, visible in Figure 2 and Figure 9.

Due to the selected definition, the total pressure loss models discovered by the LASSO method are of polynomial structure, and thus derivatives are easily found in a closed form. This allows preliminary optimizations simply by setting the derivatives to zero. Depending on the current design stage, the penalty weight factor of the LASSO method can be used to decide between a simple model at a rather abstract level, leading to a formulation as shown in Eq. (8), and a more detailed model, where the complexity and accuracy of the model may be increased, including more of the available parameters and features.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from Take Off program within the project ARIADNE. Take Off is a Research, Technology and Innovation Funding Program of the Republic of Austria, Ministry of Climate Action. The Austrian Research Promotion Agency (FFG) has been authorized for the program management. Furthermore, the authors are grateful to GE Aviation for providing the geometries used to create the data set.

REFERENCES


