Proceedings of ASME Turbo Expo 2010: Power for Land, Sea and Air
GT2010
June 14-18, 2010, Glasgow, UK

GT2010-22661

MODELLING THE LABYRINTH SEAL DISCHARGE COEFFICIENT USING DATA MINING METHODS

Tim Pychynski*
Klaus Dullenkopf, Hans-Jörg Bauer
Institut für Thermische Strömungsmaschinen (ITS)
Karlsruher Institut für Technologie (KIT)
Kaiserstr. 12, 76131 Karlsruhe, Germany
Email: tim.pychynski@kit.edu

Ralf Mikut
Institut für Angewandte Informatik (IAI)
Karlsruher Institut für Technologie (KIT)
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen,
Germany

ABSTRACT

This paper presents a data-based method to predict the discharge coefficients of labyrinth seals. At first, leakage flow rate data for straight-through and stepped labyrinth seals from various sources was collected and fused in one consistent data base. In total, over 15000 data points have been collected so far covering a 25-dimensional design space. Secondly, this leakage data set was analysed using open-source Data Mining software, which provides several algorithms such as Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN). The suitability of MLR and ANN for modelling labyrinth discharge coefficients and analysing system sensitivity was tested and evaluated. The developed leakage models showed promising prediction qualities within the design space covered by data. Further improvements of model quality may be achieved by continuing data analysis using advanced methods of Data Mining and enlarging the existing data base. The major advantages of the presented method over numerical or analytical models are possible automation of the modelling process, low calculation efforts and high model qualities.

INTRODUCTION

Labyrinth seals still rank among the most widely used sealing technology in turbomachinery. As a non-contacting seal its major advantages are its applicability under extreme operating conditions in the form of high temperatures, high pressure gradients and high rotor circumferential speeds, while offering a simple design and high lifetime at the same time (Denecke [1]).

A labyrinth seal's primary function is the minimisation or control of leakage flow, which makes it a key contributor to overall turbine efficiency and reliability. For example, the control of coolant air flow in a turbine's secondary air system is vital in achieving higher turbine entry temperatures, thus directly influencing efficiency and reliability. Optimisation of leakage flows therefore plays an important role in future turbine design (Steinetz et al. [2]).

Labyrinth seals are complex systems affected by a large number of input parameters. Many different types are possible, such as straight-through or stepped labyrinth seals, which each offer different leakage characteristics. According to today's level of knowledge the exact effects of all **system input parameters** on labyrinth leakage behaviour are not yet completely understood.

The designer has to deal with a high-dimensional design space, in which a global optimal solution needs to be found for a given set of requirements. This requires powerful **design tools** in the form of flow rate prediction models of high quality covering the design space and yielding reliable results at short calculation times and low costs. Conventional design tools such as numerical flow models or existing empirical correlations often do not meet all of these requirements. While numerical models generally offer reliable flow rate predictions over a large range of parameter values (design space), calculation times and model building costs are high. Existing one-dimensional empirical or analytical models, on the other hand, offer short calculation times but are only valid for a limited range of seal configurations and input parameters. Various correlations need to be combined in order to cover a larger

* Address all correspondence to this author.

design space (Zimmermann et al. [3]). Furthermore, a considerable amount of effort would be necessary in order to develop new empirical correlations for future seal geometries based on collected data.

For the optimisation of complex systems with lacking physical understanding of system behaviour a purely data-based approach to model-building ("black-box-modelling") has proven to be very effective (Sjöberg et al. [4] and Juditsky et al. [5]). So called data-driven **surrogate models** reproduce system relationships through mathematical expressions that are trained on the basis of available data, e.g. in the form of Artificial Neural Networks (ANN, Haykin [6]), Multiple Linear Regression models or Decision Trees. Major advantages are short model-building and calculation times at comparably low prediction error.

Numerous examples of successful application of data-based model-building can be found in the literature. Artificial Neural Networks were e.g. used by Klotz et al. to optimise a two-phase flow in a lean prevaporized premixed injection nozzle [7]. Kusiak et al. applied Multiple Linear Regression, Decision Trees and Artificial Neural Networks to optimise a combustion process [8]. A special form of Multiple Linear Regression (Kriging-Model) was suggested by Jeong et al. for multi-objective optimisation of complex systems, which was applied successfully to various engineering problems [9-11].

Data-driven modelling requires a learning data base with a sufficient number of data points. While numerous experimental and numerical studies of labyrinth leakage behaviour have already been carried out both at the ITS and at other research institutions, generally only a small number of up to five parameters have been varied each time. Drawn conclusions on system behaviour as well as models built from the acquired data are therefore only valid for a limited range of parameter values (design space). A comprehensive system analysis by fusion of existing leakage data has so far been prevented by the lack of adequate methods for data analysis. Special computer-based analysis tools such as **Data Mining** methods are required for dealing with large amounts of multidimensional data.

Data Mining is a systematic, computer-based and (semi-) automatic data analysis offering numerous **mathematical tools** (algorithms like Artificial Neural Networks, Clustering methods, Multiple Linear Regression, Nearest Neighbour methods, Analysis of Variance, Decision Trees, etc.) and **strategies** (e.g. cross-validation, automatic parameter evaluation, dealing with missing data). These enable the detection and extraction of information from large and complex data sets and compressing them in the form of a concise mathematical model (Fayyad et al. [12]).

The only attempt at collecting a large amount of leakage data known to the author was undertaken by Tipton et al. who collected leakage flow data on 1839 different labyrinth configurations covering a 14-dimensional design space, while about a fifth of the data was on single knife seals only [13]. Tipton et al. used this data to calibrate an empirical bulk-flow-model for straight-through and stepped labyrinth seals which demonstrated an accuracy of ±5% for the considered labyrinth configurations.

The objective of this paper is twofold. Firstly, existing labyrinth leakage data is collected and fused in one consistent data base in order to cover the high-dimensional design space. Secondly, this data is analysed and one-dimensional leakage prediction models are built using various Data Mining methods such as Artificial Neural Networks. The suitability of the applied approach is tested and evaluated on different data sets of varying complexity. Finally, we also want to show a new perspective to data analysis and model building in turbomachinery.

SYSTEM DEFINITION

Labyrinth seals work on the principle that fluid pressure energy is repeatedly converted into kinetic energy at each tooth tip clearance, part of which is then dissipated into heat by turbulence in the subsequent labyrinth chamber (Trutnovsky et al. [14]). Labyrinth leakage flows are influenced by rotor and stator geometry, flow conditions and rotor rotational speed.

As a first step, system parameters influencing labyrinth leakage have been identified. A general set of 25 input parameters describing numerous configurations of straight-through and stepped labyrinth seals has been defined (Fig. 1). Convergence and divergence of stepped labyrinth seals is indicated by the sign of step height ST_H. The defined parameters are the basis for establishing a consistent data base of leakage flow data and for subsequent model building. Not shown in Fig. 1 are the parameters fin number n, honeycomb wall thickness HC_W, surface roughness of the wall R_{Wmax} and land R_{Lmax} and initial circumferential flow velocity u_{phi}.

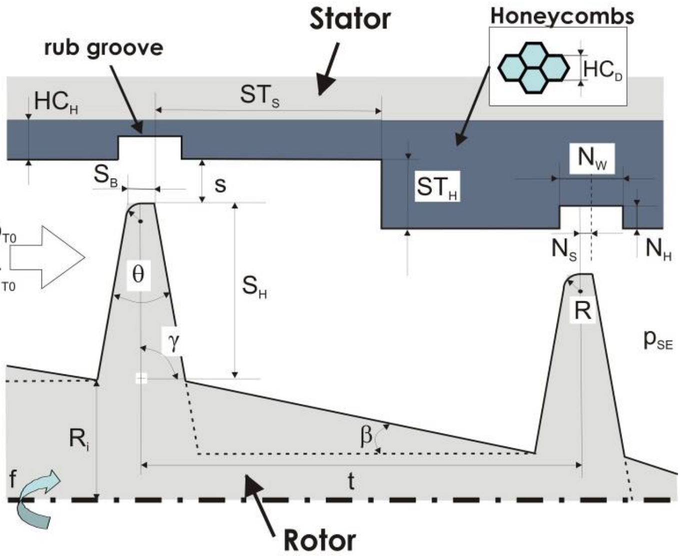

Figure 1: Defined set of input parameters for straight-through and stepped labyrinth seals (20 out of 25 parameters).

In principle, any measure of labyrinth leakage flow can be used for one-dimensional prediction models. In this paper, the non-dimensional coefficient of discharge C_d is used for consistent quantification of the output parameter leakage flow rate:

$$C_d = \frac{\dot{m}}{\dot{m}_{ideal}} \tag{1}$$

The C_d value describes the ratio of real leakage mass flow $\dot{m}$ to the ideal isentropic mass flow $\dot{m}_{ideal}$ through a nozzle of same exit clearance area A_e at the same pressure ratio π (Wittig et al. [15]), where

$$\dot{m}_{ideal} = \frac{Q_{ideal} \cdot p_{T0} \cdot A_e}{\sqrt{T_{T0}}} \tag{2}$$

with

$$Q_{ideal} = \sqrt{\frac{2 \cdot \kappa}{R_{gas} \cdot (\kappa - 1)} \cdot \left[1 - \pi^{\frac{1-\kappa}{\kappa}}\right]} \cdot \pi^{\frac{-1}{\kappa}} \tag{3}$$

for subcritical pressure ratios and

$$Q_{ideal} = Q_{ideal}\left(\pi = \left(\frac{2}{\kappa + 1}\right)^{\frac{\kappa}{\kappa-1}}\right) \tag{4}$$

for supercritical values.

DATA COLLECTION

In order to cover the 25-dimensional design space leakage data for straight-through and stepped labyrinth seals was collected from various sources and fused in one consistent data base. To this end, available data from numerous leakage studies at our institute was supplemented with labyrinth data from other research institutions, which was extracted from published leakage flow diagrams. In order to maximise the total amount of data, a literature review was conducted for identifying publications with relevant leakage flow diagrams or tables as indicated in Fig. 2. A measurement system analysis proved that the error associated with the process of data extraction from diagrams entailed only a negligible increase in data uncertainty compared to typical measurement errors in the region of several percent (mean extraction error: 0,04%, standard deviation: 0,54%).

Both experimental and numerical leakage data from validated CFD-models was included in the data base. Subsequent data analysis did not reveal major discrepancies between CFD and experimental data. Leakage data for fin-on-stator or radial labyrinth seals was not taken into account.

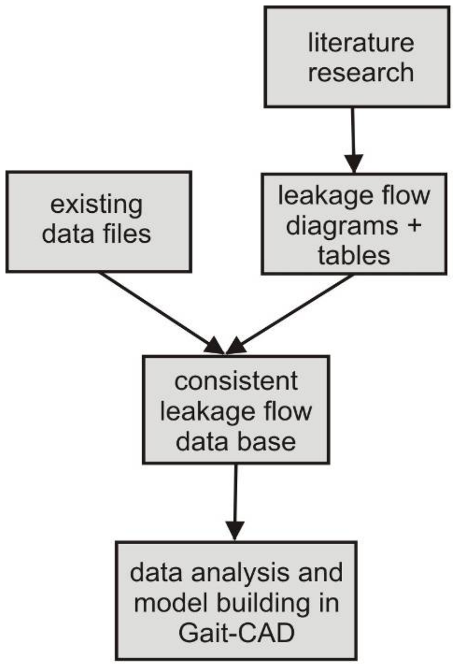

Figure 2: Overview of the data collection, analysis and model building process as performed in this paper.

In many cases geometrical and leakage data values had to be converted to match the set of parameters described in the previous chapter. Parameter values were set to zero if not defined for a certain labyrinth configuration (e.g. $ST_H = 0$ and $\beta = 0$ for straight-through labyrinths or $HC_D = 0$ for seals without honeycomb stator). This was necessary to avoid empty data values while at the same time maintaining a continuous quantitative parameter range as a basis for a global model. In addition to the labyrinth geometric and aerodynamic parameters, the data base also contains information on the respective data source, such as author name or diagram quality. Further classifying parameters such as labyrinth or stator type were introduced to facilitate subsequent data handling and analysis.

Overall 17569 raw data points on labyrinth leakage flows from 16 different authors and 26 publications have so far been recorded in the data base. An overview of labyrinth configurations covered by the collected data as well as the corresponding number of available data points is shown in Table 1. Almost 80 % of the collected data originate from the ITS, which yields the advantages of consistent parameter definitions and only few missing parameter values. Giving exact references to every single source of leakage data would exceed the scope of this paper. The interested reader is encouraged to contact the authors of this paper if he wishes to receive further information on the data sources or covered parameter values.

Labyrinth type	Dimen-sion	Honey-combs	Data points
Straight-through	2D	no	5579
	2D	yes	592
	3D	no	1841
	3D	yes	1108
Stepped convergent	2D	no	1540
	3D	no	3152
	3D	yes	79
Stepped divergent	2D	no	594
	3D	no	2994
	3D	yes	90
		Sum	**17596**

Table 1: Overview of labyrinth configurations covered in the created data base.

During the data collection, special care was taken to avoid recording redundant data, which might lead to distortion of model qualities. It should be pointed out that the data base was structured in a computer-friendly matrix format to enable easy import into the Data Mining software. Information deficits in the form of missing or uncertain data could not be avoided due to lacking published information on parameter values and experimental set-ups. These problems were dealt with in the subsequent data analysis using various techniques of Data Mining. Overall, the process of data collection proved to be the most time-consuming part of the model building process (about 60%).

It should be noted that the data base can also be used as a reference for validating numerical leakage models or identifying labyrinth configurations which still need more studying. So far non- or sparsely populated areas of the design space are for example labyrinth configurations with rub groove and honeycombs, with preswirl as well as stepped labyrinth seals with honeycombs. More experimental and numerical data on labyrinth leakage flow rates is available in the literature. The main difficulty in adding this data will lie in missing published information and fitting the new data to the existing system definition described previously.

DATA ANALYSIS

As a third step, the collected labyrinth data was analysed and pre-processed for subsequent leakage flow modelling. This was achieved using the open source Data Mining Matlab toolbox Gait-CAD (Mikut et al. [16]), which offers the advantages of free adaptability of implemented algorithms and easy automation of the analysis and model building process. Computer-based Data Mining methods are particularly suited for analysing heterogeneous (quantitative and qualitative data, different data sources) and multidimensional data sets with large numbers of data points.

Firstly, correct extraction and recording of raw leakage data was checked by data visualisation and the definition of various verification algorithms. The latter include identification of physically impossible or unusually large/small input and output values as well as logical rules (e.g. "ST_H = 0 for straight-through labyrinth seals" or "ST_H < 0 for convergent stepped labyrinth seals"). A clustering method was employed to support outlier identification.

The collected raw data still contained many empty parameter values, which were not specified in the available publications and data files. For example, the fin tip radius R was not explicitly considered in various leakage studies, leaving the true parameter value open to speculation. Before applying Data Mining methods such as Artificial Neural Networks, these incomplete data points need to be eliminated. This can be achieved either by eliminating whole data points/parameters or by assuming missing data values.

In order to minimise loss of raw data (information), a special elimination strategy was developed to deal with missing data. Data for each parameter was clustered to create additional classification parameters, where *NaN* ("not a number") was included as an additional linguistic term. This conversion yielded two advantages. Firstly, assumptions made on missing data points could later be verified using visualisation techniques (e.g. box- or scatterplots) and Analysis of Variance (ANOVA). Secondly, this enables determining the influence of missing data while evaluating prediction model errors during subsequent model building. In this case, the following verified assumptions were made on selected missing parameter values:

R = 0 mm, p_{SE} = 1 bar, T_{T0} = 300 K, γ = 90°. The following parameters were eliminated due to large amounts of missing data: θ, R_i, $R_{w,max}$ and HC_W. All missing data not covered by these actions was eliminated through deletion of the respective data points. The results of subsequent model building suggested that the assumption of R = 0 mm for missing values is not always correct. However, this error also extends to explicitly defined R values since many authors assumed fin tip radius R to be zero without quantitative verification. The other assumptions, however, did not seem to have a major effect on prediction quality. The mentioned procedures left a reduced pre-processed data base of 21 input parameters and 15297 data points.

The data distribution over the design space was studied using clustering and data visualisation techniques. The histograms of twelve selected input parameters in Fig. 3 give an impression of the covered parameter ranges and data distribution. While the data is not distributed homogeneously over the entire design space, no two data points are identical. It should be pointed out that overall, only little variation of the parameters HC_D, HC_H, N_W, N_H, N_S, R, γ, u_{Phi}, p_{SE} and T_{T0} is covered in the data base. This needs to be taken into account when using a purely data-based approach to modelling leakage flow rates, since physically important parameters may be overshadowed by others and appear insignificant.

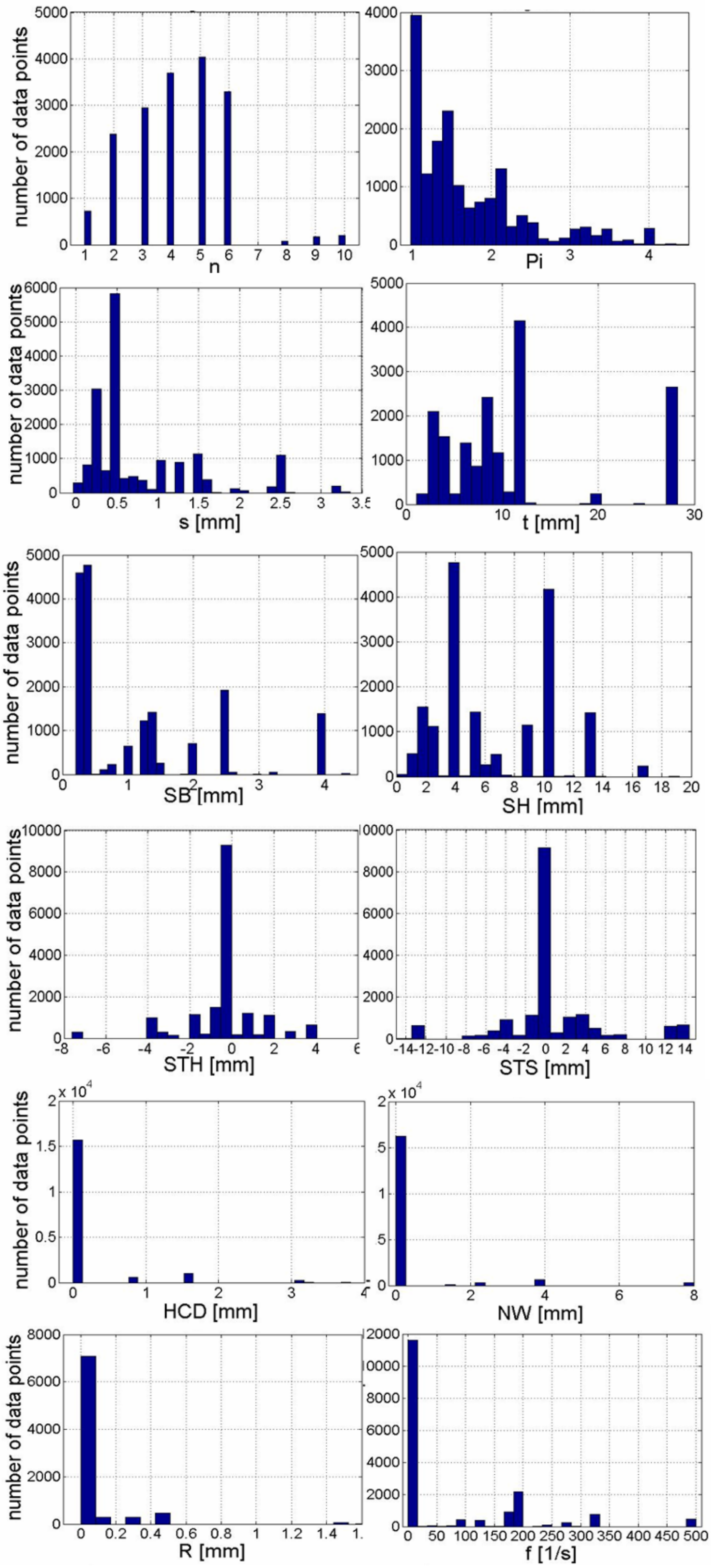

Figure 3: Histograms of the 12 input parameters *n, π, s, t,* S_B, S_H, ST_H, ST_S, HC_d, N_W , *R* and *f* over all data.

MODELLING THE DISCHARGE COEFFICIENT

The pre-processed data was used to create various data-driven one-dimensional prediction models of labyrinth leakage with ANN and Multiple Linear Regression, testing the suitability of this approach. The aim of data-based modelling is to explain the recorded variation of the output parameter C_d in the data with the available information of input parameters (labyrinth geometry and flow conditions). In principle, this is achieved by reproducing the system relationships with a mathematical function $y = \mathrm{f}(\boldsymbol{x}, \boldsymbol{B})$ of any type as indicated in Fig. 4. This function is then fitted to the available multidimensional learning data (training data) by adjusting the function coefficients $\boldsymbol{B}$ accordingly. Systematic tendencies in the data are separated from random variation, which is achieved by applying statistically motivated learning algorithms (e.g. method of least squares for simple polynomial functions or the back-propagation algorithm for ANN). Finally, the model quality is evaluated by analysing model errors on a test data set.

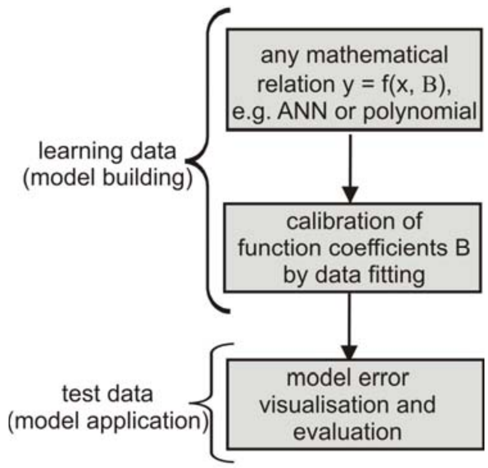

Figure 4: Principle of data-based model building.

System sensitivity:

Univariate and multivariate parameter evaluation was employed first in order to quickly determine the input parameters having the largest effect on leakage flow. This Data Mining strategy represents a type of data-driven analysis of system sensitivity, which is achieved by repeated design of regression models with different combinations of input parameters. This was performed on different data sets of varying complexity, three of which are described in the following:

Data set 1: Only 2D and 3D straight-through labyrinth seals without rotation, honeycombs or rub groove: Eleven input parameters and 5091 data points, 13 data sources.

Data set 2: All labyrinth types covered by the pre-processed data: 21 input parameters and 15297 data points, 16 data sources.

Data set 3: Selected configurations of convergent stepped labyrinth seals without rotation, honeycombs or rub groove: nine varied input parameters (Table 2) and 350 data points, nine data sources.

DATA SET 3			
parameter	**value**	**parameter**	**value**
n	3 / 4 / 5	S_B / t	0.02 to 1.67
s / t	0.04 to 0.17	S_H / t	0.05 to 1.22
π	1.1 to 2	γ	90°
ST_H / t	-0.4 to -0.3	p_{SE}	1 – 14.5 bar
ST_S / t	0.06 to 0.84	T_{T0}	300 – 900 K
R / t	0	$R_{L,max}$	smooth

Table 2: Overview of parameter values for data set 3.

The results of the univariate parameter evaluation for data set 1 are shown in Fig. 5. In this case, the effect of parameter X_i is quantified by the respective quality of a simple leakage prediction model $C_d = f(X_i)$, where the model quality in turn is quantified by the linear correlation coefficient between predicted and true output values. This model quality can be interpreted as a measure for an averaged tendency in the data which is visualised in the equivalent scatter plots in Fig. 4 for the most influential parameters of data set 1. Interactions between input parameters are not taken into account by the univariate approach. The obtained parameter ranking shows that fin number n, relative tip clearance s / t, pressure ratio π, relative fin tip radius R / t as well as relative fin thickness S_B / t and height S_H / t are the most influential parameters for simple straight-through labyrinth flows (in descending order).

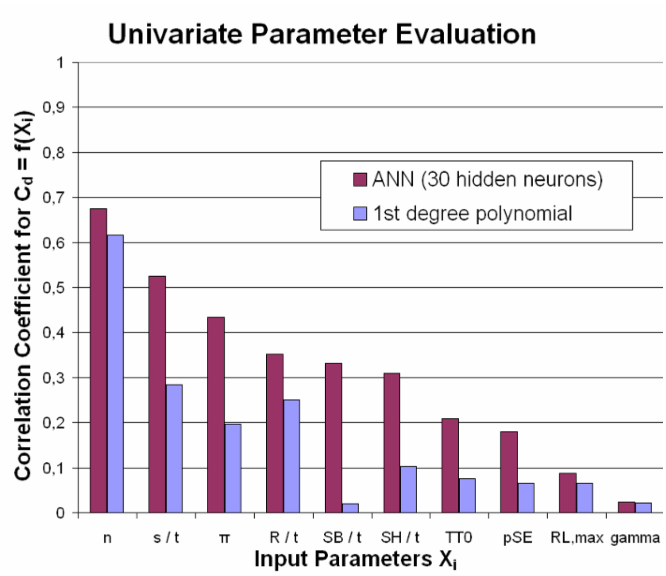

Figure 5: Univariate ranking of ten most influential input parameters with ANN and a linear polynomial function for data set 1.

The linearity of parameter effects was verified by applying both a linear polynomial function and an ANN for parameter evaluation and comparing the respective results. While the univariate approach suggests that fin number n would have a fairly linear effect on leakage flow, all other parameters seem to be rather non-linear. A similar parameter ranking was obtained by the multivariate approach. It should be stressed that all these findings are purely based on the observed data and not on physical evaluation of parameter effects. The results, however, are consistent to what is found in the literature, which proves the potential of the applied data-driven approach for evaluating system sensitivities. Similar results were obtained for data sets 2 and 3.

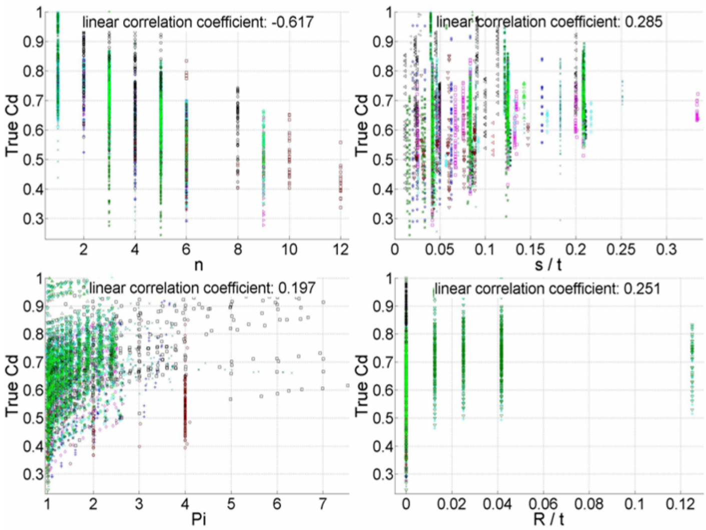

Figure 4: Scatterplots of true C_d-values over the four input parameters n, π, s / t and R / t for entire data set 1. The markers indicate the different data sources.

C_d-model:

Various data-based leakage flow prediction models were built in a short period of time for different data sets of varying complexity. Artificial Neural Networks were chosen for this, which proved to be better suited for reproducing the strong non-linear system relationships than Multiple Linear Regression. The basis for the process of model building were the pre-processed data and parameter ranking described previously. Only non-dimensional parameters were used to guarantee model comparability and transferability to labyrinth seals of different scaling. The fin pitch t was chosen as reference value for geometrical data. Flow conditions are described by the non-dimensional pressure ratio π, axial and tangential Reynolds number Re_{ax} and Re_{tan} as well as the non-dimensional pre-swirl K_0 [17]. The results of three different ANN prediction models for the three data sets are shown in Table 3.

Data set	Linear corr. coeff.	$\bar{\varepsilon}_{rel}$ [%]	σ_{error} [%]	Input parameters	ANN deve-loped in
1	0,958 (0,948)	4,0	6,4	*n, s/t, π,, R/t,* S_B*/t,* S_H*/t*	*15sec*
2	0,956 (0,949)	6,5	10,3	*n, π, s/t,* S_H*/t,* S_B*/t,* ST_S*/t,* ST_H*/t,* HC_d*/t,* N_S*/t, ,* N_W*/t*	*90sec*
3	0.991 (0.952)	1.4	3	*n, π, s/t,* S_H*/t,* S_B*/t,* ST_S*/t,* ST_H*/t*	*5sec*

Table 3: Summary of considered input parameters and model qualities for ANN

All ANN were simple three-layer Multiple Layer Perceptrons (MLP) with 30 hidden neurons, a weighted sum for calculation of neuron state, a Levenberg-Marquardt learning algorithm with 50 learning epochs and a hyperbolic tangent sigmoid transfer (tansig) function (Haykin [6]). Training data parameter values were normalised to mean zero and standard deviation one to improve the efficiency of the learning algorithm. The principle structure of the MLP Artificial Neural Network for data set 2 is shown in Fig. 6. The connections of the input parameters two to nine to the hidden layer are not shown. All obtained ANN were developed within several seconds and are implemented in Matlab code. The interested reader is encouraged to contact the authors of this paper if he wishes to receive further information on exact ANN parameters.

The achieved leakage flow prediction qualities are high with a mean relative model error of 4.0% for simple straight-through labyrinths (data set 1) and 6.5% for all considered straight-through and stepped labyrinth seals (data set 2). The mean relative model error represents the arithmetic mean of positive relative model errors on all data.

$$\bar{\varepsilon}_{rel} = \frac{\sum_{x=1}^{N} \frac{\left| C^*_{d,x} - C_{d,x} \right|}{C_{d,x}}}{N} \tag{5}$$

The local model for convergent stepped labyrinth seals built on data set 3 shows the lowest mean error of 1.4%. Linear correlation coefficients between predicted and true C_d-values are close to one for all three models (0.958, 0.956 and 0.991 respectively), where one would be a perfectly deterministic model.

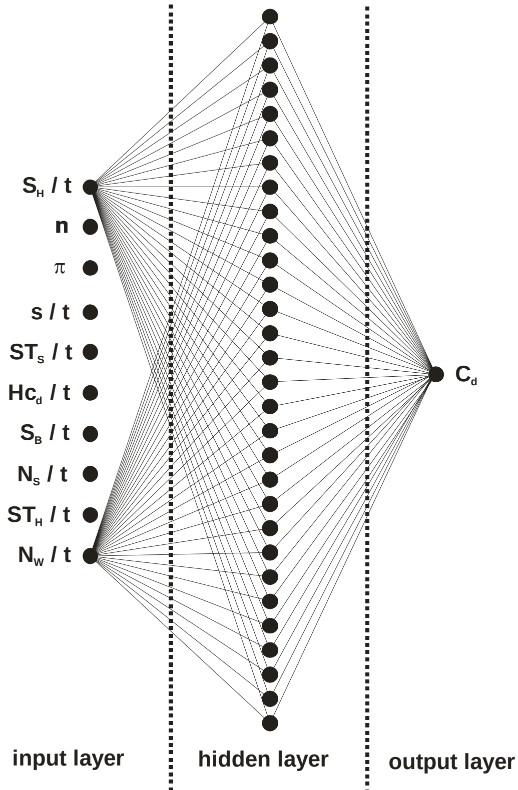

Figure 6: Principle structure of the MLP Artificial Neural Network for data set 2.

All values indicated in Table 3 are averaged from ten different prediction models. Model overfitting and transferability to new application data were monitored using ten-fold cross-validation. For this, the data set is divided randomly into ten subsets, nine of which are used as training data while the remaining tenth subset is only used for testing. The averaged test results over all ten cross validation runs are indicated with parenthesis in Table 3. The very small change in the linear correlation coefficient suggests good prediction results for new labyrinth data within the underlying parameter ranges. The leakage flow model's quality was also evaluated by visualising model prediction over true output and input parameters (Fig. 7 and 8).

In order to demonstrate the performance of the applied approach, the global model (data set 2) and local model (data set 3) are compared for three selected convergent labyrinth configurations (see Table 4 for parameter values). In Fig. 9, predicted and measured discharge coefficients are plotted over the pressure ratio π. While the local model yields very good prediction values with less than 1% error for all three configurations, the global model only shows good predictions for fin number $n = 5$. For $n = 3$, the effect of relative clearance $s\ /\ t$ seems to be represented correctly by the global model. However, the overall prediction is shifted to lower discharge coefficients. This is suspected to be due to the fewer number of data points available for $n = 3$ for convergent stepped labyrinth seals without honeycombs and rub groove as can be seen in Fig. 10. This presumably led to the ANN not being able to learn the correct system relationships in this region of the design space.

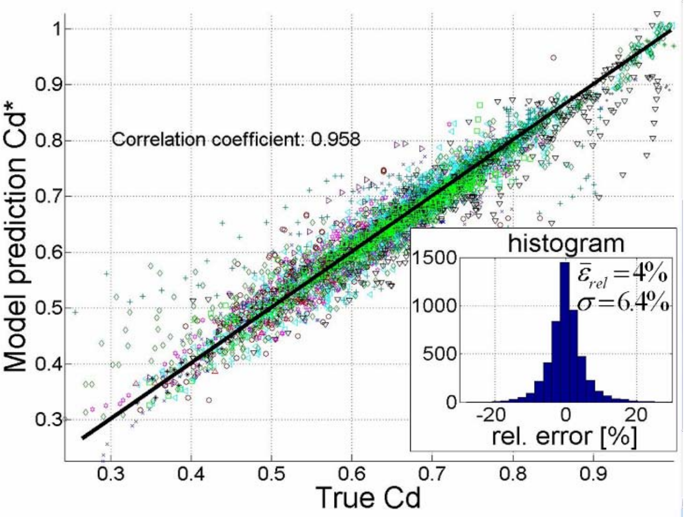

Figure 7: Leakage flow rate prediction C_d* over true C_d-value for data set 1. The markers indicate the different data sources.

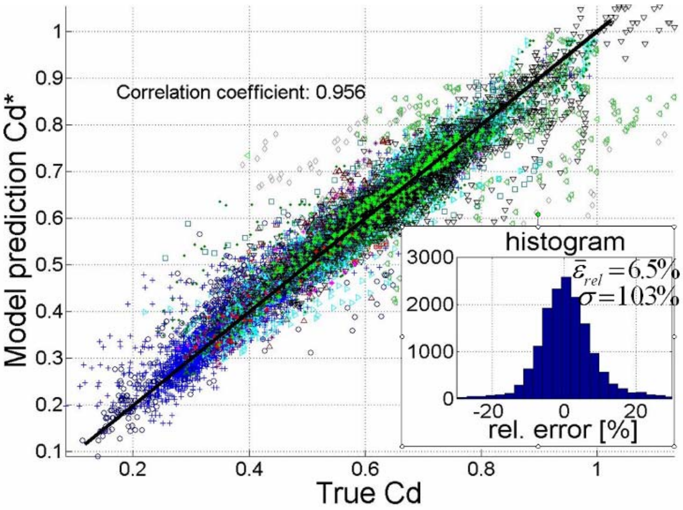

Figure 8: Leakage flow rate prediction C_d* over true C_d-value for data set 2. The markers indicate the different data sources.

DISCUSSION

The high global prediction qualities achieved by the purely data-based models are comparable to more cost-intensive numerical or analytical models. Major advantages of the data-driven models are short model building (~ 1 to 5 minutes) and calculation times (~ 1 second), making them an interesting asset for labyrinth seal design and optimisation. The disadvantages are the lack of physical interpretability in the prediction models as well as limited transferability of the models to areas of the design space not or only sparsely covered by the data.

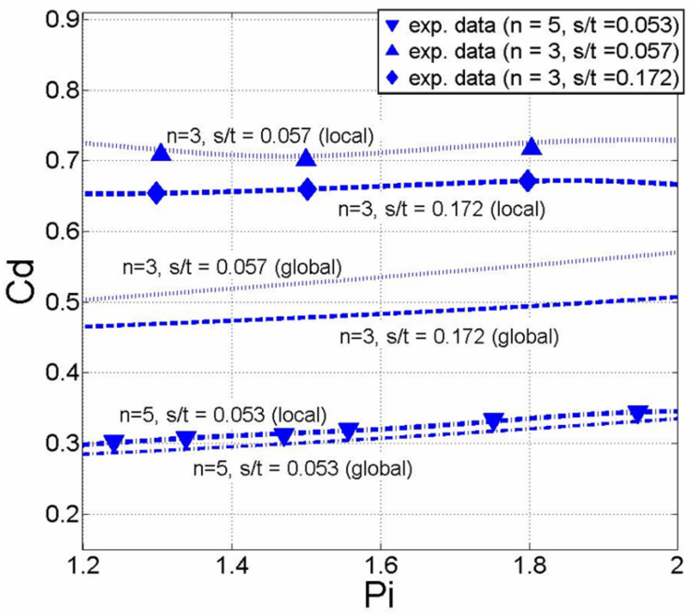

Figure 9: Discharge coefficients predicted with a local model (data set 3) and global model (data set 2) respectively compared to experimental in-house data for three different labyrinth configurations.

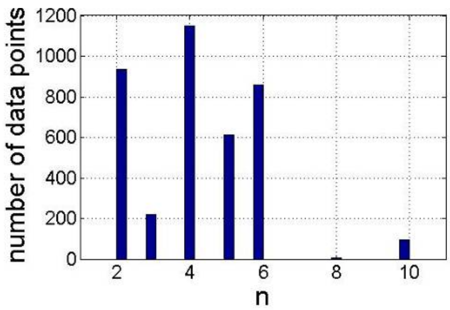

Figure 10: Histogram of fin number n for all convergent stepped labyrinth seals without honeycombs and rub groove (3878 data points in total).

Fig. 9			
parameter	**curve 1** ▼	**curve 2** ▲	**curve 3** ♦
n	5	*3*	3
s / t	0.053	*0.057*	0.172
ST_H / t	-0.268	*-0.143*	
ST_S / t	0.477		
S_B / t	0.036		
S_H / t	0.381		
p_{SE} [bar]	1		
T_{T0} [K]	300		

Table 4: Parameter values for Fig. 9.

The effect of inhomogeneous data distribution on model qualities needs to be evaluated further. Plotting the predicted leakage flow rates showed high model errors of over 15% for single data points within the design space. These may be due to several factors, the effects of which superimpose to create the total observed error:

- Incorrect or inconsistent data
- Insufficient amount of data for model calibration
- Incapability of the mathematical model to reproduce the system relationships in certain areas of the design space
- Inhomogeneous distribution of data in the design space

Further improvement of prediction quality would be possible by eliminating these sources of errors. More effort should therefore be put in verifying data consistency and taking appropriate corrections where necessary. Enlarging the data base would probably yield higher model qualities and increase model transferability to new data. Also, by increasing the ratio of available training data points to calibrated model parameters, more sophisticated ANN could be used (Leakage prediction quality was e.g. shown to improve with increasing number of hidden neurons). Great potential therefore lies in the integration of new labyrinth leakage flow data from additional sources into the data base. Also, missing parameter values could be added to the data base, e.g. by further literature research or by developing a sophisticated estimation strategy.

Other Data Mining methods such as Clustering methods, Fuzzy Logic and Decision Trees, Principal Component or Discriminant Analysis as well as Nearest Neighbour methods might be used to improve the model building process. The negative effects of an inhomogeneous data distribution might be eliminated by systematic weighting of data points in sparsely populated areas of the design space. Judging by the low relative prediction error of 1.4% for the local model on data set 3, another promising strategy to increase global prediction quality might be the combination of various local models for different seal configurations. Also, using a combination of global prediction models for preliminary (rough) optimisation/design and more refined dynamically built local models might lead to improved labyrinth designs.

SUMMARY AND CONCLUSION

Leakage data for straight-through and stepped labyrinth seals was collected from various sources and fused in one large data base, containing 17596 raw data points and covering a 25-dimensional design space. An important basis for this was the definition of a general set of parameters, allowing a consistent definition of all considered labyrinth configurations. The created leakage data base was the foundation for the subsequent data analysis and model building.

The collected data was analysed and pre-processed using Data Mining methods such as ANOVA, Artificial Neural Networks (ANN), various visualisation techniques or automated parameter evaluation by repeated data-based model building. Univariate and multivariate parameter evaluations were performed to quantify parameter effects and interactions.

The results of three one-dimensional leakage rate prediction models with ANN were presented in this paper. A global model built on all available data points showed an average prediction error of 6.5%, which could not yet be achieved using “manually built” empirical correlations. Even better prediction values could be achieved with local models built in a desired region of the entire design space. The models $C_d = f(X_i)$ built with ANN can be used for seal design and optimisation or robustness analyses in the same manner as conventional one-dimensional flow rate correlations, as long as the ANN is implemented in a computer code (e.g. Matlab).

The data-based approach to modelling one-dimensional labyrinth leakage rates presented in this paper has proven high potential for modelling complex systems where sufficient amounts of system data are available. Care needs to be taken, however, in evaluating model overfitting and transferability to new data, which can be done with cross-validation technique. Advantages over conventional leakage models such as bulk-flow- or numerical models are quick and automated model building, short calculation times enabling numerous iterations (e.g. in optimisation algorithms) while offering comparable model qualities. The proposed approach can in principle be transferred to different systems of varying complexity, especially where system relationships are not yet completely understood and large amounts of data are available.

OUTLOOK

Further experimental and numerical studies of labyrinth leakage flows are currently being performed at the ITS. The results will be added to the data base. Efforts at analysing the data base and modelling leakage flows will be continued in the future. We would like to encourage an exchange of available labyrinth leakage data for the purpose of scientific study. A comprehensive data base including all existing leakage data could be analysed using Data Mining methods in order to extract useful new knowledge and increase system understanding.

ACKNOWLEDGMENTS

The authors would like to thank all those involved in the comprehensive and time-consuming process of data collection. Our special thanks go to our colleague Georg Blesinger who initiated this project and played an important role in coordinating the data collection. Great thanks also go to our colleague Tina Weinberger for her counseling on technical issues regarding labyrinth seals and to Jens Denecke who provided a large amount of leakage data.

NOMENCLATURE

A_e	ideal nozzle clearance area	[mm^2]
$\boldsymbol{B}$	vector of function parameters	[-]
C_d	(true) coefficient of discharge	[-]
C_d^*	predicted coeff. of discharge	[-]
f	rotor rotational frequency	[1/s]
$f()$	function	[-]
HC_D	honeycomb cell diameter	[mm]
HC_H	honeycomb cell height	[mm]
HC_W	honeycomb sheet width	[mm]
K_0	dimensionless preswirl	[-]
$\dot{m}$	real seal leakage mass flow	[kg/s]
$\dot{m}_{ideal}$	ideal nozzle leakage mass flow	[kg/s]
n	number of seal fins	[-]
N	total number of data points	[-]
N_H	groove height	[mm]
N_S	groove shift	[mm]
N_W	groove width	[mm]
p_{SE}	static outlet pressure	[bar]
p_{T0}	total input pressure	[bar]
R	fin tip edge radius	[mm]
R_{gas}	ideal gas constant	[J/Kmol)]
Re_{ax}	axial Reynolds number $\frac{\dot{m}}{\mu \cdot \pi \cdot R_i}$	[-]
Re_{tan}	tangential Reynolds number $\frac{\rho \cdot \omega \cdot R_i^2}{\mu}$	[-]
R_i	rotor radius at first fin base	[mm]
R_{Lmax}	land surface roughness	[µm]
R_{Wmax}	fin tip surface roughness	[µm]
S_B	fin width	[mm]
S_H	fin height	[mm]
ST_H	step height	[mm]
ST_S	step shift	[mm]
s	clearance	[mm]
T_{T0}	total seal entry temperature	[K]
t	fin pitch	[mm]
u_{phi}	circumferential fluid inlet velocity	[m/s]
$\boldsymbol{x}$	vector of input parameters	[-]
X_i	Input parameter i	[-]
y	output parameter	[-]

Greek letters

β	rotor step angle	[°]
γ	fin inclination	[°]
$\overline{\varepsilon}_{rel}$	mean relative model error	[-]
κ	specific heat ratio	[-]
μ	dynamic viscosity	[kg/(ms^3)]
π	pressure ratio (p_{T0} / p_{SE})	[-]
ρ	fluid density	[kg/m^3]
σ_{error}	rel. error standard deviation	[-]
θ	fin taper angle	[°]
ω	rotor angular frequency	[rad/s]

REFERENCES

[1] Denecke, J., 2007, “Rotierende Labyrinthdichtungen mit Honigwabenanstreifbelägen – Untersuchung der Wechselwirkung von Durchflussverhalten, Drallverlauf und Totaltemperaturänderung“, PhD thesis, Institut für Thermische Strömungsmaschinen (ITS), Karlsruhe, Germany

[2] Steinetz, B.M., Hendricks, R.C., and Munson, J., 1998, „Advanced Seal Technology Role in Meeting Next Generation Turbine Engine Goals”, AVT-PPS Paper No. 11

[3] Zimmermann, H., Wolff, Q., 1998, “Air System Correlations Part 1: Labyrinth Seals”, ASME-Paper 98-GT-206

[4] Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Deylon, B., Glorennec, P.Y., Hjalmarsson, H., and Juditsky, A., 1995, “Nonlinear Black-Box Modelling in System Identification: a Unified Overview”, Automatica, 31(12): 1725-1750

[5] Juditsky, A., Hjalmarsson, H., Benveniste, A., Deylon, B., Ljung, L., Sjöberg, J., and Zhang, Q., 1995, “Nonlinear Black-Box Modelling in System Identification: Mathematical Foundations”, Automatica, 31(12): 1691-724

[6] Haykin, S., 1998, “Neural Networks: A Comprehensive Foundation”, Prentice Hall, Englewood Cliffs, NJ

[7] Klotz, P., Bartoli, N., Cornez, L., Samuelides, M., and Villedieu, P., 2007, “Database Enrichment for Optimization with Artificial Neural Networks”; ONERA – Toulouse, France

[8] Kusiak, A., and Song, Z., 2006, “Combustion Efficiency Optimization and Virtual Testing: A Data-Mining Approach”, IEEE Transactions on Industrial Informatics, Vol. 2, No.3

[9] Jeong, S., Chiba, K., and Obayashi, S., 2005, “Data Mining for Aerodynamic Design Space”, Journal of Aerospace Computing, Information, and Communication, 2 (11): 452-469

[10] Chiba, K., Jeong, S., Obayashi, S., and Morino, H., 2005, “Data Mining for Multidisciplinary Design Space of Regional-Jet Wing”, IEEE Congress on Evolutionary Computation

[11] Kanazaki, M., Tanaka, K., Jeong, S., and Yamamoto, K., 2006, “Design Exploration of High-Lift Airfoil using Kriging Model and Data Mining Technique”, European Conference on Computational Fluid Dynamics ECCOMAS CFD

[12] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P., 1996, “From Data Mining to Knowledge Discovery in Databases”, AI Magazine, 17:37-54
[13] Tipton, D.L., Scott, T.E., and Vogel, R.E., 1986, “Labyrinth Seal Analysis: Volume III – Analytical and Experimental Development of a Design Model for Labyrinth Seals”, Tech. Rep. AFWAL-TR-85-2103 Vol. III, Allison Gas Turbine Division, General Motors Corporation, Indianapolis, Indiana
[14] Trutnovsky, K., and Komotori, K., 1981, “Berührungsfreie Dichtungen“ (Contactless Seals), Düsseldorf: VDI-Verlag, 4th edition, ISBN 3-18-400490-2
[15] Wittig, S., Dörr, L., and Kim, S., 1983, „Scaling Effects on Leakage Losses in Labyrinth Seals“, ASME Journal of Engineering for Power, 105: 305-309
[16] Mikut, R.; Burmeister, O.; Braun, S., and Reischl, M., 2008, “The Open Source Matlab Toolbox Gait-CAD and its Application to Bioelectric Signal Processing”, Proc., DGBMT-Workshop Biosignal analysis, Potsdam, 109-111
[17] Denecke, J., Färber, J., Dullenkopf, K., and Bauer, H.J., 2005, „Dimensional Analysis and Scaling of Rotating Seals“, ASME-Paper GT2005-68676