Proceedings of ASME Turbo Expo 2012
GT2012
June 11-15, 2012, Copenhagen, Denmark

GT2012-69706

GRADIENT ENHANCED SURROGATE MODELS BASED ON ADJOINT CFD METHODS FOR THE DESIGN OF A COUNTER ROTATING TURBOFAN

Jan Backhaus, Marcel Aulich, Christian Frey,
Timea Lengyel, Christian Voß
Institute of Propulsion Technology
German Aerospace Center (DLR)
Linder Höhe, 51147 Cologne
Germany
Email: Jan.Backhaus@dlr.de

ABSTRACT

This paper studies the use of adjoint CFD solvers in combination with surrogate modelling in order to reduce the computational cost of the optimization of complex 3D turbomachinery components. The method is applied to a previously optimized counter rotating turbofan, with a shape parameterized by 104 CAD parameters.

Through random changes on the reference design, a small number of design variations are created to serve as training samples for the surrogate models. A steady RANS solver and its discrete adjoint are then used to calculate objective function values and their corresponding sensitivities. Kriging and neural networks are used to build surrogate models from the training data. To study the impact of the additional information provided by the adjoint solver, each model is trained with and without the sensitivity information. The accuracy of the different surrogate model predictions is assessed by comparison against CFD calculations.

The results show a considerable improvement of the fitness function approximation when the sensitivity information is taken into account. Through a gradient based optimization on one of the surrogate models, a design with higher isentropic efficiency at the aerodynamic design point is created. This application demonstrates that the improved surrogate models can be used for design and optimization.

NOMENCLATURE

ADP	Aerodynamic Design Point
CEV	Constant Eddy Viscosity presumption
CRISP	Counter Rotating Integrated Shrouded Propfan
GEK	Gradient Enhanced Kriging
ML	Maximum Likelihood
D	Number of fitness function samples
f	Fitness function value
η_{is}	Isentropic efficiency
M	Mach number
$\dot{m}$	Mass flow
n	Number of design variables
p_{t1}	Total pressure at stage inlet
p_{t2}	Total pressure at stage outlet
$\vec{x}_i$	Fitness function sample points
y	Surrogate model output
y+	Dimensionless wall distance

INTRODUCTION

The availability of cost efficient aerodynamic sensitivity information is regarded as a key technology for the progress of multidisciplinary optimizations in high-dimensional design spaces [1]. Purely aerodynamic optimizations also benefit from sensitivity information. The reason is that *zero order optimiza-*

tion algorithms (as opposed to gradient aware methods) limit the size of the feasible design space through the growth of numerical costs for function evaluations. This is especially true for evolutionary algorithms, which are widely used in CFD based design. In design spaces with in the order of one hundred parameters, surrogate modeling improves on the situation significantly. However, the number of function evaluations necessary to build a surrogate model increases rapidly with the number of design variables. For instance, consider approximating an n-dimensional paraboloid. Since a quadratic polynomial in n dimensions has $(n+2)(n+1)/2$ independent coefficients, the effort to provide the necessary information for a surrogate model to match exactly such a polynomial using only function evaluations increases quadratically with n.

Adjoint CFD methods provide a way to calculate gradients of aerodynamic fitness functions at a cost independent of the number of design variables [2, 3]. Gradient descent methods are a natural way to carry out efficient design optimizations using adjoint solvers [4, 5]. For high-dimensional multistage turbomachinery design these are succesfully employed by Wang and He [6, 7]. However, as reported e.g. by Dwight and Brezillon [8] the robustness of the optimization process may be severely affected by inconsistent or inaccurate function and gradient evaluations. One source of inconsistencies can be circumvented by using the discrete adjoint methodology during the implementation of the adjoint solver [3]. However, discrepancies may still occur if the flow or adjoint solutions do not fully converge. An additional source of inconsistencies is caused by features used in the steady solver which have not been adjoined. Furthermore it cannot be assured that the simulation process delivers a solution for all designs that can be described by the parameterization. This may impact the performance of gradient descent methods.

In contrast to gradient descent methods, experience shows that surrogate modelling approaches exhibit higher robustness in the optimization process. In particular, if the complexity of the model is limited appropriately, surrogate models turn out to be less susceptible to noise in the input data. Furthermore, missing values can be easily handled during model training. Surrogate models have been successfully employed for the 3D shape optimization of turbomachinery components in recent years [9–11]. It is therefore desirable to incorporate adjoint based gradients for the training of surrogate models, when optimizing in a high dimensional design space. The aim of this paper is to demonstrate that gradient enhanced surrogate models trained on sensitivities from an adjoint solver have the potential to significantly reduce computational costs for high dimensional turbomachinery optimization problems.

The numerical methods described, are applied to a counter rotating ducted fan which had been designed and optimized previously at DLR Cologne. The first section of this paper summarizes the preceding optimization and presents the main aerodynamic features of the optimized design. Then the adjoint gradient calculations and the gradient enhanced surrogate models are discussed. A database for the training of the surrogate models is created by randomly varying the design variables at the pre-optimized design. Different surrogate models trained on function values with and without gradient information are compared and validated against parameter studies based on steady CFD simulations.

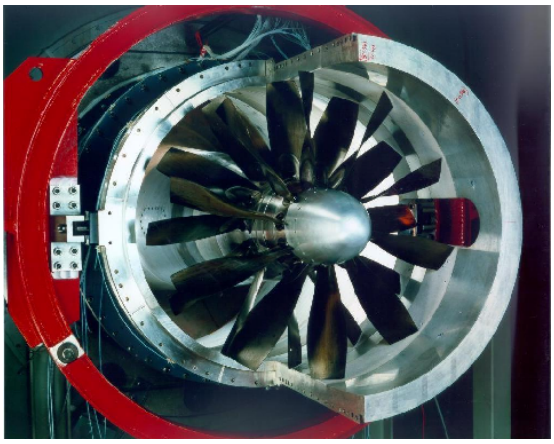

FIGURE 1. CRISP1: 1m-MODEL

In order to assess if the gradient enhanced surrogate models can be applied for turbomachinery optimization, a new design is generated on the basis of the surrogate models for efficiency and mass flow. Finally, steady simulations for the new blade geometry are conducted and the CFD results are compared to the model predictions.

REFERENCE DESIGN

The methods to be presented in the later sections will be demonstrated on a state-of-the-art shrouded counter-rotating fan configuration, called CRISP2 (Counter Rotating Integrated Shrouded Propfan). It is the result of a re-design of an earlier designed configuration called CRISP1.

CRISP1 (Fig. 1) was designed between 1985 and 1991, within a technology project conducted by MTU and DLR. Afterwards it was built as a 1m (inner casing diameter) model and experimentally investigated at DLR until 1999 [12]. The design and optimization of fan- and compressor blades has made significant progress in the last 10 years, based on the development of numerical tools and the growth in available computational power. Therefore a re-design of the blades is conducted. Only aerodynamical aspects of the design are discussed in this paper, mechanical simulations and optimizations for CRISP2 are described in [13].

The targets of the new design are to develop a new engine concept for a high bypass ratio, to achieve very high efficiency and better noise properties in the important operating points. For

the re-design CFD based optimization is used. Evolutionary algorithms and surrogate models are taken from the optimization toolbox AutoOpti [14] and coupled with an evaluation process based on DLR's flow solver for internal flows TRACE. The initial design for this optimization is the CRISP1 geometry. The mesh resolution for the optimization is a trade-off between accuracy and simulation speed and is found by a refinement study; it has about 700.000 structured cells with 35 radial lines and y+ $>$ 30 in order to apply wall functions. The main objective is to increase the isentropic efficiency. Since it is not known in advance, which pressure ratio will deliver the best overall performance, a Pareto-front for isentropic efficiency and total pressure ratio is generated. This is done by defining total pressure ratio as a second objective and varying its target formulation. In the last stage of the optimization, the objective formulation for the total pressure ratio is switched off. During the optimization, constraints on axial Mach number, exit swirl, stall margin and shaft power are enforced.

Each blade geometry is parameterized by means of curvature continuous construction profiles. Four of these profiles are directly changed by the optimization procedure. These are located at 0%, 40%, 75% and 100% of the blade's radius. Another four profiles are placed between these construction profiles. Their parameters are not set directly by the optimization procedure, instead they are interpolated from the neighboring profiles. This allows smooth blade shapes to be generated with an acceptable number of parameters. Each profile is described by the following set of parameters: Stagger angle, leading edge angle, trailing edge angle, leading edge thickness, asymmetry factor of the leading edge, B-Spline control points (De-Boor points), maximum thickness and position of the maximum thickness. The profile's fill factor relates the profile's area to maximum thickness and cord length of the profiles. Free parameters for the x-position of leading and trailing edges allow the axial position and length of the profiles to be varied. Five profiles can be shifted in the circumferential direction by the design parameter θ. A more detailed description of the design system can be found in [10]. The rotational speed of the first rotor (R1) and the speed ratio $N = RPM_{R2}/RPM_{R1}$ are also treated as free variables. Table 1 summarizes the free variables.

Figure 2 shows all converged results calculated during the optimization. The colors mark different stages of the optimization, which result from different formulations of the pressure ratio objective. The optimal solution has an 8% efficiency improvement compared to the initial geometry. Figure 3 shows the radial distribution of the efficiency and total pressure ratio for the optimized geometry and the initial geometry. The radial distribution of the efficiency is uniform and has a very high level in the range from 10% to 90% blade height. The optimization, described here, targeted to improve the area between 10%-90% blade height and this is successfully fulfilled as the figures show. The pressure ratio is increased and evenly distributed for both

Parameter	#
Stagger angle	8
Leading edge angle	8
Trailing edge angle	8
Theta shift	10
Profile spline	46
Maximum thickness	16
Fill factor	8
Sum	104

TABLE 1. SUMMARY OF THE FINAL SET OF DESIGN VARIABLES.

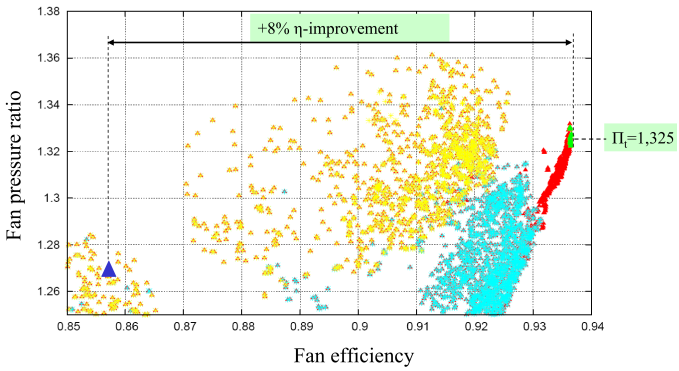

FIGURE 2. PARETO FRONT OF THE CRISP2 OPTIMIZATION

blades from hub to tip. An improvement of the boundary zones is planned for future work.

A pre-shock Mach number of 1.3 – 1.4 (c.f. Fig. 4) causes high losses at a quite small pressure ratio on the initial rotor. By decreasing the pre-shock Mach number to 1.1 – 1.2 and increasing the pressure ratio, the efficiency is raised.

NUMERICAL METHODS

The numerical methods used in this work include a CFD solver, its adjoint and two gradient enhanced surrogate models. Additionally the creation of the training database for the surrogate models is described below.

Flow Solver

The CFD solver used in this paper is the DLR solver for internal flows TRACE [15–17] which has been used in the past for a wide range of aerodynamic optimizations in the context of turbomachinery [9, 10, 18]. Adjoint solvers provide a method to

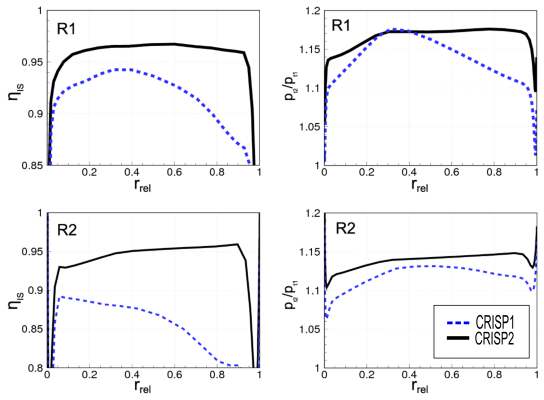

FIGURE 3. COMPARISON OF THE RADIAL DISTRIBUTIONS OF ISENTROPIC EFFICIENCY AND TOTAL PRESSURE RATIO FOR ROTOR 1 AND ROTOR 2 OF CRISP1 AND CRISP2

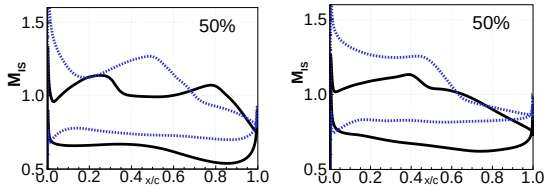

FIGURE 4. MACH NUMBER DISTRIBUTION ON THE 50% PROFILE OF CRISP1 (BLUE LINE) AND CRISP2 (BLACK LINE) FOR R1 (LEFT) AND R2 (RIGHT)

efficiently calculate sensitivities of objective functionals with respect to many design parameters [2, 3]. The flow solver TRACE has therefore been extended by a module for the solution of the adjoint Reynolds-averaged Navier Stokes equations called adjointTRACE [19, 20].

Since the principal aim of the development of the adjoint flow solver is to accelerate optimizations the discrete methodology has been preferred over the continuous one, i.e. the adjoint solver is based on the exact linearization of fluxes, boundary conditions etc. of the steady solver. The discrete methodology ensures that the sensitivities calculated by the adjoint solver are fully consistent with sensitivities calculated by the steady solver. This allows for the validation of the adjoint solver against the established and validated solver TRACE [20]. Possible discrepancies cannot be due to the use of different discretization schemes and thus indicate directly implementation errors in the adjoint solver. The consistency of the adjoint and the steady solver is expected to be important for gradient descent methods; otherwise, at some point during the optimization there could be a mismatch between gradient information and function evaluation so that the optimization may be slowed down or even lead to non-optimal

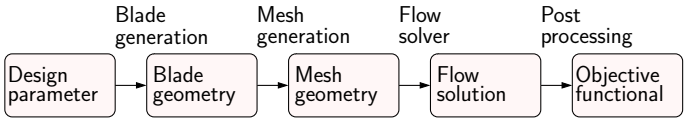

FIGURE 5. PROCESS CHAIN FOR THE EVALUATION OF PERFORMANCE FUNCTIONALS.

solutions [8].

Finite differences are used to calculate the derivatives of the non-linear discrete fluxes. In this way one circumvents the tedious and error prone work of writing differentiated counterpieces for every routine of the higher order upwind spatial scheme of TRACE. Adjoint boundary conditions, however, have been newly developed for the adjoint module but are consistent to the counter part in the steady solver. This applies both to solid walls and the non-reflecting boundary conditions used in turbomachinery configurations at entries, exits and mixing planes [20]. Functionals provided by the adjoint solver include mass flows, pressure ratios, temperature ratios and efficiency, each one being defined between a pair of user specified control surfaces [21].

The adjoint solver does not currently take into account the turbulence model. Instead the so-called *constant-eddy viscosity* approximation (CEV) is made, i.e. the turbulent quantities are frozen during the linearization process. Therefore the sensitivities correspond to derivatives of the functionals under the simplifying assumptions that the eddy-viscosity stays constant when the geometry is slightly deformed [22]. It should be emphasized that the aforementioned possibility to validate the adjoint solver against the steady solver is helpful in this respect since the impact of the CEV approximation on the accuracy of the sensitivities can be assessed beforehand.

Database Creation

To provide function values and corresponding gradients of the fitness function for surrogate modeling, a database with 45 designs is created by varying the previously optimized CRISP2 geometry. A parameter is chosen for variation with a probability of 30%. Then, random, normally distributed changes inside a parameter specific interval are applied to that parameter. For each modified design a steady 3D Navier-Stokes simulation is performed using TRACE. The flow solutions are then postprocessed to calculate values for the performance functionals of efficiency and mass flow (c.f. Fig. 5).

For these functionals the corresponding sensitivities with respect to 104 design parameters (Tab. 1) are calculated using the adjoint solver. All results are stored in a database which will be referred to as the *training database* in the following. The computational cost using all 12 cores of one intel westmere cluster node are as follows: The nonlinear steady solution takes approximately 35 minutes on average and the adjoint solution plus gra-

dient evaluation costs about 40 minutes per functional on average.

Surrogate Models

Surrogate models are used to approximate the objective functions and constraints of a design application. In this way, the time consuming parts of the evaluation process chain (usually CFD calculations) can be replaced by fast evaluations of the approximate model. The fast response time of surrogate models enables a huge amount of member evaluations in a relatively short time. Hence surrogate models are used as a preselection tool to create promising solution candidates for the design problem. For this application, Bayesian neural networks and ordinary Kriging are used. The advantages and disadvantages of Kriging models and neural networks are discussed in [14]. For both models, a direct approach is chosen, where the gradients are part of the model formulation. In the indirect approach, on the contrary, the gradient information would be used to add some artifical samples using first order Taylor approximations, which has disantvantages due to the necessary choice of the step size. Both models also provide uncertainty estimations for their predicted values. These uncertainty estimations are used to restrict the extrapolation behavior when creating new members, thus avoiding predictions in sparsely sampled regions of the parameter space.

Neural Networks The applied neural networks are based on Bayesian feed-forward networks with automatic relevance determination (for detailed information see [23, 24]). The aim of the Bayesian training algorithm is to reduce the database error F_D (mean squared error between real function values $f(\vec{x})$ and network output $y(\vec{x},\vec{w})$ for all D members) and the network complexity F_W (quadratic sum of all W weights) simultaneously. The idea for reducing the network complexity is inspired by "Occam's razor", a principle that prefers simpler solutions to complex ones. The training algorithm is based on the "weight-decay" approach, where the network complexity is regularized by the parameter λ,

$$\begin{aligned}\text{min!} : F &= F_D + \lambda F_W \\ &= \frac{1}{2}\sum_{i=1}^{D}\left(f(\vec{x}_i) - y(\vec{x}_i,\vec{w})\right)^2 + \frac{1}{2}\lambda\sum_{j=1}^{W} w_j^2. \end{aligned} \tag{1}$$

The difficulty is to find a suitable value for the regularization parameter λ. If λ is too big, the neural network becomes too simple and the training database is not reproduced by the neural network approximation. In the other case the complexity of the network increases, which usually results in a huge approximation error, when the neural networks are used for predictions. The training searches for the best compromise between model complexity F_W and database error F_D. In the Bayesian training the regularization parameter λ is estimated by a probabilistic approach. For this approach the basic equation of the "weight-decay" is slightly modified,

$$\begin{aligned}\text{min!} : F &= \beta F_D + \alpha F_W \\ &= \frac{1}{2}\beta\sum_{i=1}^{D}\left(f(\vec{x}_i) - y(\vec{x}_i,\vec{w})\right)^2 + \frac{1}{2}\alpha\sum_{j=1}^{W} w_j^2. \end{aligned} \tag{2}$$

If $\lambda = \alpha/\beta$, the minimum of F is the same for the Bayesian and weight-decay approach. Inside the Bayesian approach a white noise in the function $f(\vec{x})$ is assumed. That noise can be e.g. an error in the CFD (mesh quality, convergence criteria, etc.). Usually, the neural networks estimate more noise than in the real function. This allows simpler networks, which can increase approximation quality. The variance of that noise is $1/\beta$. The parameter α regularizes the variance σ^2 of the weight values, which is $1/\alpha$. The most probable values of α and β are found by using Bayes' theorem. To use gradient information inside the Bayesian approach, a new training algorithm is developed. For this algorithm the basic equation (2) of the Bayesian approach is extended, and reads as

$$\begin{aligned}\text{min!} : F &= \beta F_D + \alpha F_W + \sum_{k=1}^{n}\gamma_k\frac{\partial F}{\partial x_k} \\ &= \frac{1}{2}\beta\sum_{i=1}^{D}\left(f(\vec{x}_i) - y(\vec{x}_i,\vec{w})\right)^2 + \frac{1}{2}\alpha\sum_{j=1}^{W} w_j^2 \\ &\quad + \sum_{k=1}^{n}\frac{1}{2}\gamma_k\sum_{i=1}^{D}\left(\frac{\partial f(\vec{x}_i)}{\partial x_{i,k}} - \frac{\partial y(\vec{x}_i,\vec{w})}{\partial x_{i,k}}\right)^2. \end{aligned} \tag{3}$$

The new summands are the mean squared errors between function gradients $\partial f(\vec{x}_i)/\partial x_{i,k}$ and network gradients $\partial y(\vec{x}_i,\vec{w})/\partial x_{i,k}$. Each γ_k (one for each of the n free parameters) estimates the white noise in the gradients of the real function. Therefore the neural networks may detect errors in the gradient calculations, by estimating it as a function with added white noise. The question, whether these noise estimations may be used for the validation of adjoint gradients over a set of objective function samples, could be a topic for further research.

Gradient Enhanced Kriging In this section, a direct gradient enhanced Kriging (GEK) model approach is proposed. This direct GEK model is used to incorporate the gradient information at the sample points, which is provided by adjoint-TRACE. Here, the gradient enhanced direct Kriging formulation is derived using the conventional ordinary Kriging model for simplicity of the description.

The Kriging method is a statistical prediction of a function $f(\vec{x})$ based on a set of samples $(\vec{x}_i, f(\vec{x}_i)), i \in \{1,...,D\}$ where all the $\vec{x}_i$ are n-dimensional. The vector $\vec{Y}_s$ is defined as $\vec{Y_s}^T = [f(\vec{x}_1),...,f(\vec{x}_D)] = [y_1,...,y_D]$.

Kriging prediction depends on the spatial correlations between the given samples. The correlation is given by a correlation function which only depends on the weighted distance between two points. Gaussian or cubic spline functions are the most common forms of the spatial correlation function. The Gaussian function which is used in this work is defined by

$$\begin{aligned}\mathrm{Corr}(\vec{x}_i,\vec{x}_j) &= \prod_{k=1}^{n} \mathrm{Corr}_k(\theta_k,\vec{x}_i,\vec{x}_j) := \prod_{k=1}^{n} e^{-\theta_k (x_{i,k}-x_{j,k})^2} \\ &= e^{-\sum_{k=1}^{n} \theta_k (x_{i,k}-x_{j,k})^2}.\end{aligned} \tag{4}$$

The so called hyperparameters $\theta_k, k \in \{1,...,n\}$, must be trained to improve the reliability of the model. This training procedure is described below.

The correlations are considered in a correlation matrix, $R_{i,j} = \mathrm{Corr}(\vec{x}_i,\vec{x}_j)$ with the size D^2. With this matrix the ordinary Kriging predictor y and the uncertainty s of the prediction of any point $\vec{x}$ in the design space can be obtained by

$$y(\vec{x}) = \beta + \vec{r}^T(\vec{x}) R^{-1} (\vec{Y}_s - \beta \vec{F}), \tag{5}$$

$$s^2(\vec{x}) = \sigma^2 \left(1 - \vec{r}^T(\vec{x}) R^{-1} \vec{r}(\vec{x}) + \frac{(1 - \vec{r}^T(\vec{x}) R^{-1} \vec{F})^2}{\vec{F}^T R^{-1} \vec{F}} \right), \tag{6}$$

with

$$\begin{aligned}\vec{r}_i(\vec{x}) &= \mathrm{Corr}(\vec{x},\vec{x}_i), \\ \vec{F}^T &= [1,...,1], \\ \beta &= \frac{\vec{F}^T R^{-1} \vec{Y}_s}{\vec{F}^T R^{-1} \vec{F}}, \\ \sigma^2 &= \frac{1}{D} (\vec{Y}_s - \beta \vec{F})^T R^{-1} (\vec{Y}_s - \beta \vec{F}).\end{aligned} \tag{7}$$

In this work the partial derivatives at several sample points are also used for the construction of the GEK model. The partial derivatives are directly used in the modified Kriging formulation by including the correlations between function-derivative and derivative-derivative. These correlations can be obtained by differentiating the correlation function. This leads to an extended correlation matrix:

$$R_{i,j} = R_{j,i} = \begin{cases} \mathrm{Corr}(\vec{x}_i,\vec{x}_j), & i \le D \wedge j \le D \\ \dfrac{\partial \mathrm{Corr}(\vec{x}_i,\vec{x}_l)}{\partial x_l^m}, & i \le D \wedge j > D \\ \dfrac{\partial^2 \mathrm{Corr}(\vec{x}_k,\vec{x}_l)}{\partial x_k^n \partial x_l^m}, & i > D \wedge j > D \end{cases} \tag{8}$$

Here $k,l \in \{1,...,D\}$ and $n,m \in \{1,...,n\}$. If all sample points have full information (i.e. function value and all partial derivatives), the size of the correlation matrix is $(D(1+n))^2$. With this extended correlation matrix and the extended vectors

$$\begin{aligned}\vec{Y_s}^T &= [y_1,...,y_D, \frac{\partial y_1}{\partial x_1},...,\frac{\partial y_D}{\partial x_n}], \\ \vec{F}^T &= [1_1,...,1_D,0,...0], \\ \vec{r}(\vec{x})^T &= [\mathrm{Corr}(\vec{x}_1,\vec{x}),...,\mathrm{Corr}(\vec{x}_D,\vec{x}), \\ &\quad \frac{\partial \mathrm{Corr}(\vec{x}_1,\vec{x})}{\partial x_1^1},...,\frac{\partial \mathrm{Corr}(\vec{x}_D,\vec{x})}{\partial x_D^n}],\end{aligned}$$

the GEK predictor y and the GEK uncertainty s of the prediction of any untried $\vec{x}$ can still be calculated by equation (5) and (6). Obviously, D in equation (7) must be replaced by $M = D(1+n)$ if all sample points have full information. For a more detailed description of GEK see [25, 26].

During the training of the hyperparameters $\theta_k, k \in \{1,...,n\}$ the maximum-likelihood approach is used. This leads to a minimization of

$$ML = |R|^{1/M} \sigma^2. \tag{9}$$

With some straightforward transformations, the partial derivative of (9) with respect to θ_k ($\partial ML/\partial \theta_k = \dot{ML}$) can be calculated as

$$\dot{ML} = \frac{1}{M} |R|^{1/M} \left(tr(R^{-1}\dot{R})\sigma^2 - (\vec{Y}_s - \beta\vec{F})^T R^{-1} \dot{R} R^{-1} (\vec{Y}_s - \beta\vec{F}) \right).$$

An augmented Lagrangian algorithm can be used to minimize (9) while using the condition number of the correlation matrix $c(R) = \lambda_{max}/\lambda_{min}$ (λ is an eigenvalue of R) as a constraint, to ensure that the surrogate model training leaves the condition number sufficiently small. The partial derivative of the condition number is therefore needed, which is given by

$$\dot{c}(R) = \frac{\dot{\lambda}_{max}\lambda_{min} - \lambda_{max}\dot{\lambda}_{min}}{\lambda_{min}^2}, \quad \dot{\lambda} = \frac{\vec{v}^T \dot{R} \vec{v}}{\vec{v}^T \vec{v}}.$$

Here $\vec{v}$ is the corresponding eigenvector of λ. The maximal and minimal eigenvalues and the corresponding eigenvectors are determined by von Mises iteration.

In the presented work the dimension of the correlation matrix R is given by $D(1+n) = 45(1+104) = 4725$. The constrained minimization of (9) using $\dot{c}(R)$ and $\dot{M}L$ becomes a challenging task because R^{-1} has to be calculated via Cholesky decomposition at each iteration. Therefore, the numerically most expensive algorithms are thread-parallelized. The training of the GEK model needs roughly 12 hours on one intel westmere node with 12 cores. To handle substantially larger design spaces in the near future, the training will be accelerated by MPI parallelization and an iterative increase of the input information (sample points and derivatives) during the training.

VALIDATION

Figures 6 and 7 show the validation results for the adjoint solver in terms of the sensitivities of mass flow and isentropic efficiency, respectively. The reference results are created using finite differences of the steady simulations. Note that the underlying perturbation ε for each parameter has to be chosen carefully for the finite differences, to ensure that the perturbed parameters stay in the range where the target functional behaves approximately linear. For nearly all parameters, the CEV and the adjoint sensitivities are in very good agreement. However, for a few parameters, e.g. no. 28 and no. 104, mismatches can be observed (Fig. 6). For the isentropic efficiency, discrepancies between CEV and turbulent simulation can be observed (Fig. 7). One reason for the discrepancies may be inappropriately chosen perturbations for the finite differences. Another reason could be a simplification in the adjoint sensitivity evaluation which assumes that the mesh is kept fixed at the surfaces where the functional is evaluated, i.e. at the entry and exit of the computational domain.

The plots also show a comparison with sensitivities obtained with the steady solver in CEV mode that serve as reference for the adjoint solver. The CEV and standard mode are in very good agreement, showing the validity of the assumption for this functional on this configuration. However, for the isentropic efficiency, the agreement between the adjoint and the CEV results is far better than for the turbulent results. This indicates that the CEV assumption, used in the adjoint flow solver, is violated for the isentropic efficiency of this configuration, at least for some of the parameters. These discrepancies between function point and gradient information have to be treated as noise in the fitness function. As long as there is no adjoint turbulence model in the adjoint solver, a noise tolerant optimization algorithm is necessary.

The prediction accuracy of the gradient enhanced surrogate models is assessed by means of a parameter study. Eight design parameters are separately varied in the study: Stagger angles of two profiles on each rotor, and maximum thickness of two profiles on each rotor. The variations are chosen to cover a broader parameter range than the variations found inside the training database. This results in testing geometries different from the members in the training database. Two different kinds of training are conducted for each surrogate modeling approach, one incorporating the gradient information from the training database, and one using only the function values. The surrogate models prediction for each test member is then compared to the result of a steady CFD calculation. When interpreting the results, keep in mind that the training data is not positioned along the axes for which the parameter study is conducted. Rather, the training points are scattered in the parameter space and may therefore have a huge distance to the test points from the parameter study. Figures 8 – 10 show three exemplary results. Vertical lines in

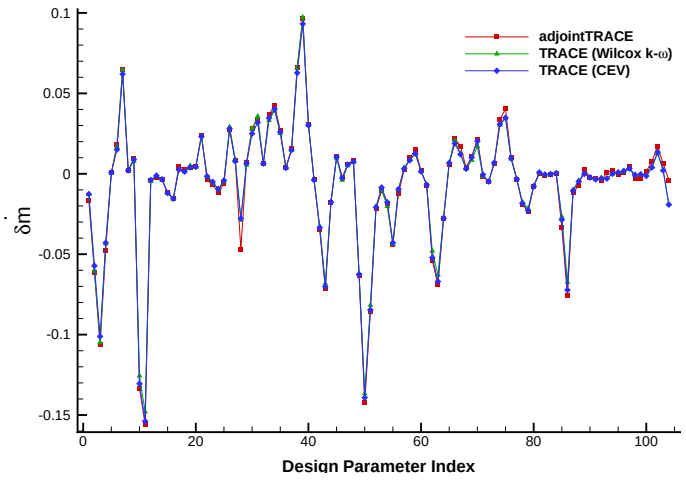

FIGURE 6. SENSITIVITIES OF MASS FLOW FOR ALL DESIGN PARAMETERS AT INITIAL DESIGN.

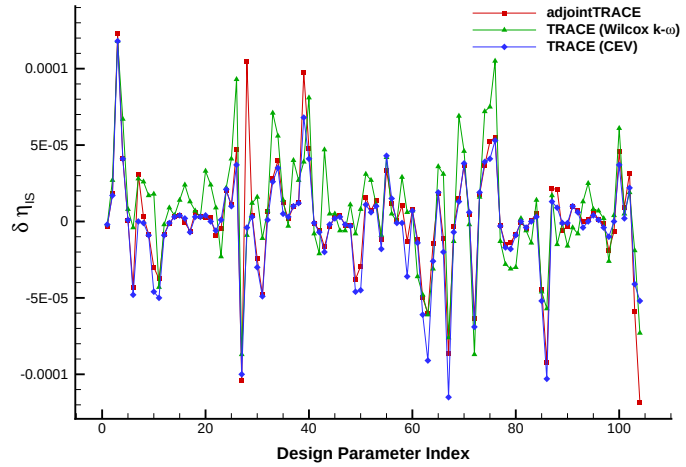

FIGURE 7. SENSITIVITIES OF STAGE ISENTROPIC EFFICIENCY FOR ALL DESIGN PARAMETERS AT INITIAL DESIGN.

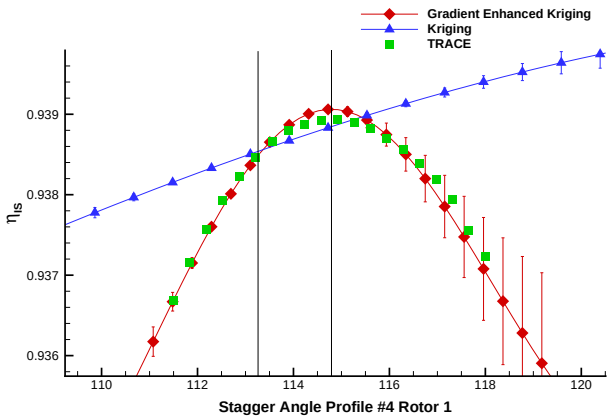

FIGURE 8. PREDICTIONS OF ISENTROPIC EFFICIENCY FROM KRIGING WITH AND WITHOUT GRADIENTS COMPARED AGAINST TRACE CALCULATIONS.

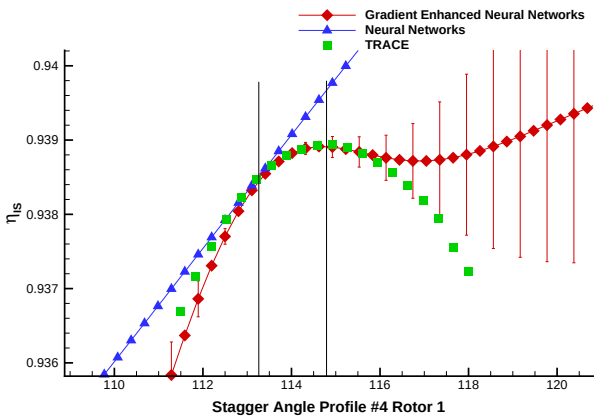

FIGURE 9. PREDICTIONS OF ISENTROPIC EFFICIENCY FROM NEURAL NETWORKS WITH AND WITHOUT GRADIENTS COMPARED AGAINST TRACE CALCULATIONS.

the plots mark the range of training values for this parameter in the training database. The displayed error bars are retrieved from the uncertainty estimations of the surrogate models. This quantity is not to be seen as an absolute error, but as an estimation of the prediction uncertainty. In the diagrams one can observe that a standard surrogate model trained for a 104 dimensional function, provided with 45 function values (CFD simulations) can only capture a global trend for the function. The small number of training points, compared to the dimension, leads to models of too low complexity, which are not able to capture the fitness functions behavior appropriately. By incorporating the gradient information, the approximation quality of the models is improved. The models almost interpolate the function values inside the range of given training data and they also extrapolate the behaviour of the fitness functions outside of the region where training-data was provided. This extrapolation property can be attributed to the gradient information, which allows the models to capture the first order behavior of the fitness function in the vicinity of the training data.

APPLICATION

To show that the previously described methods are suitable for the optimization of turbomachinery components, the gradient enhanced neural network model is used to produce a new design, which is then aerodynamically analyzed. The aim is to find a configuration which has a better isentropic efficiency in the ADP than the pre-optimized geometry.

This is done by performing an optimization using the predictions of function values and gradients from the meta-model for isentropic efficiency, and afterwards, to verify the performance of the resulting geometry by means of a steady CFD simulation. When optimizing using the meta-model it is important to limit

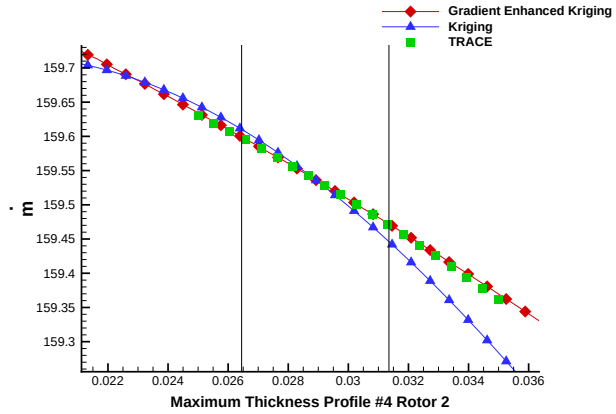

FIGURE 10. PREDICTIONS OF MASS FLOW FROM KRIGING WITH AND WITHOUT GRADIENTS COMPARED AGAINST TRACE CALCULATIONS.

the predicted uncertainty by an upper bound constraint. Otherwise the optimization could converge into regions of sparse training information, resulting in a high prediction uncertainty. In these regions of high uncertainty, the model may erroneously predict high efficiency values which will not be in agreement to CFD simulations. The operating point must also be kept fixed, because the ADP is different from the point of maximum efficiency on the speedline. Without an additional constraint, the optimization would shift the operating point on the speedline in the direction of maximum efficiency by reducing the mass flow of the operating points. This will lead to a shifted speedline, instead of a real design improvement. By constraining the mass flow to the ADP's value, the operating point is approximately kept fixed for the surrogate model based optimization. Consequently two neural networks are built using the training database:

one network for isentropic efficiency and one for the mass flow. The meta-model based optimization is performed using the augmented lagrangian method.

Figure 13 shows the training database with the initial geometry, and the result of the optimization on the meta-model. An increase of 0.13% is reached on this highly pre-optimized datum design through only 45 fitness evaluations + gradient calculations. The plot also confirms the need for a constraint on the mass flow. Most of the efficiency improvements in the randomly created training database are achieved by reducing the mass flow, leading to a different operating point. An unrestricted optimization would follow this descent path and produce geometries with an undesired change of the operating point. Radial distributions of efficiency and total pressure ratio comparing the pre-optimized design with the new design can be seen in Figure 11. Most of the increase in fan efficiency is reached in the region between 60% and 80% of radial height. Comparing the distributions of total pressure ratio for R1 and R2 one observes a re-distribution of aerodynamic load from R1 to R2 and on R1 from higher to lower radii. Figure 12 shows a comparison of the Mach number distribution on the suction sides of the two rotor blades before and after surrogate model optimization. Obviously the pre-shock mach number on R1 is reduced in the upper blade section. This is a direct consequence of aerodynamic load being taken from that region. The increased aerodynamic load on R2 leads to a slightly higher pre-shock mach number there, but in total an increase of efficiency is reached. A comparison of the speedlines for the pre-optimized and the new geometry (Fig. 14) shows, that the operating points isentropic efficiency is increased while keeping the mass flow approximately fixed. The slight difference in the operating points is due to a small prediction error of the meta-model for mass flow at that point. The differences between the speedlines in the off-design regions are a consequence of optimizing only the ADP's efficiency. For future optimizations, one may want to use additional surrogate models to also optimize off-design efficiencies and pressure ratios. This could be achieved by multi-objective optimizations on several meta-models simultaneously.

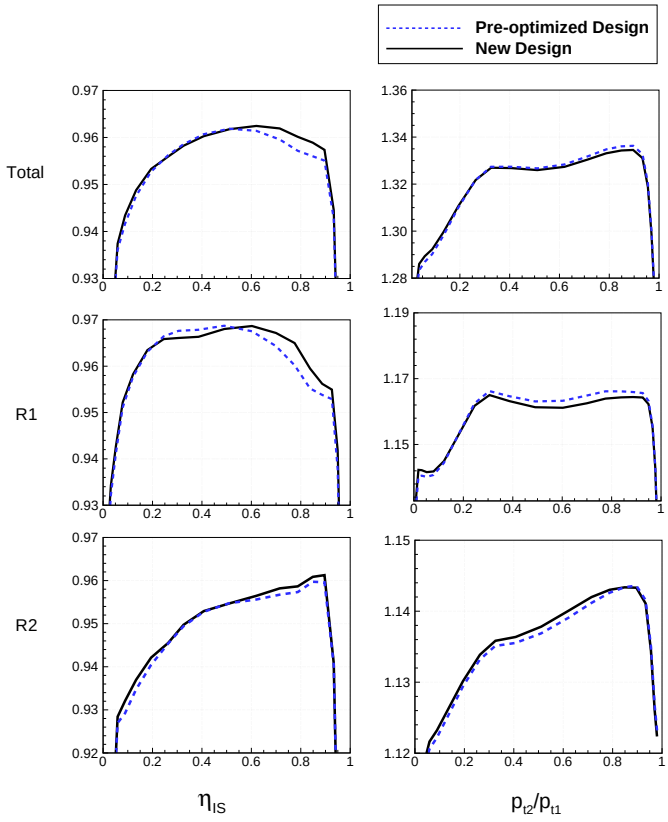

FIGURE 11. RADIAL DISTRIBUTIONS COMPARING EFFICIENCY AND PRESSURE RATIO OF THE PRE-OPTIMIZED DESIGN WITH THE NEW DESIGN FOR THE WHOLE STAGE AND THE TWO ROTORS.

CONCLUSIONS

In this work, an adjoint solver is used and compared against validation calculations for a high dimensional turbomachinery design application. The flow solver and its adjoint are in good agreement for the mass flow. Discrepancies can be observed between the adjoint simulation and the flow simulation with a turbulence model in the case of isentropic efficiency.

In view of the fact that errors in the calculated gradients are possible, noise tolerant optimization strategies seem to be indispensable for employing sensitivities from an adjoint solver during optimizations. Consequently gradient enhanced Kriging and gradient enhanced neural network methods are proposed as surrogate modeling techniques, which are tolerant against noise in the objective functions. Both surrogate models are trained on a small database compared to the number of design variables. When gradients are included, a relatively small number of sample point is sufficient to capture the fitness functions behaviour in the vicinity of the training data as demonstrated by the parameter study.

The challenging design problem of finding a geometry with a higher isentropic efficiency than the extensively optimized CRISP2 geometry with a maximum isentropic efficiency of about 94% is used as an evaluation of the surrogate models prediction qualities. By optimizing on the gradient enhanced neural network, an improved design with a 0.13% higher isentropic efficiency in the ADP is found. A comparison of the aerodynamic behaviour of CRISP2 and the optimized design, shows that aerodynamic load is re-distributed among the blades and among different radial heights. These results show that the surrogate models predict verifiable properties of the 3D aerodynamic design problem with two blade rows and over a hundred of design variables.

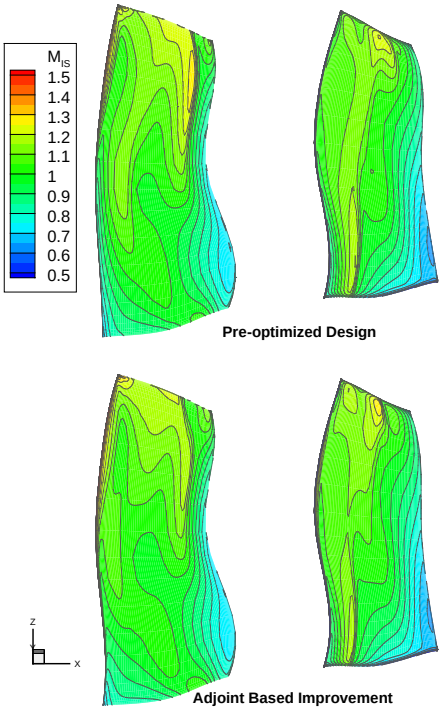

FIGURE 12. COMPARISON OF THE SUCCTION SIDE MACH NUMBER DISTRIBUTION FOR BOTH ROTORS PRE-OPTIMIZED (UPPER) AND NEW DESIGN (LOWER).

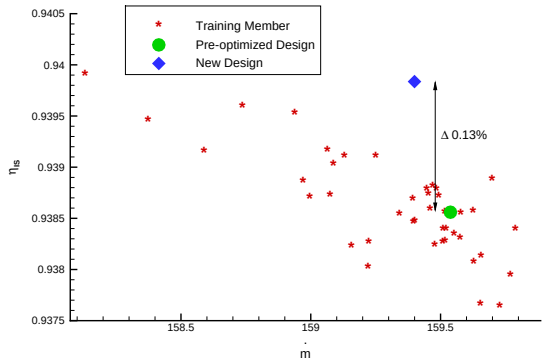

FIGURE 13. SIMULATION RESULT FOR THE NEW DESIGN IN RELATION TO THE TRAINING DATABASE AND THE INITIAL MEMBER.

The application of noise tolerant gradient enhanced surrogate models is proposed for evaluation processes in which function or gradient evaluations may contain errors or fail to deliver a solution in some points of the design space.

A detailed comparison on the effectiveness of the gradient

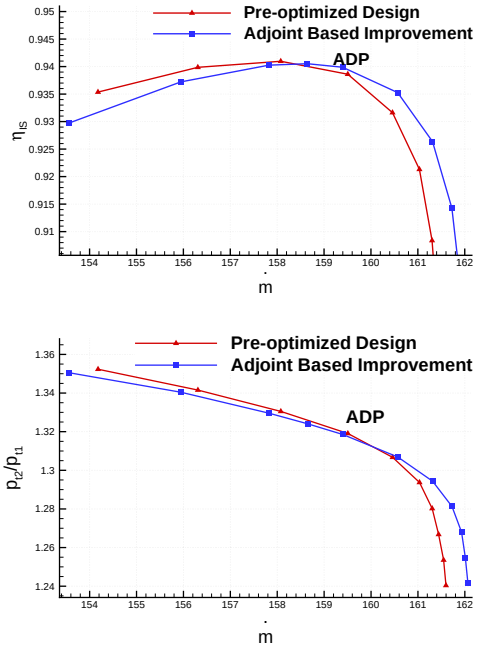

FIGURE 14. SPEEDLINE COMPARING THE PRE-OPTIMIZED AND THE RESULT FROM OPTIMIZING ON THE GRADIENT ENHANCED NEURAL NETWORK.

free versus gradient aware surrogate modelling, considering the influence of noise in the input data, is a topic for further research.

The authors expect an essential reduction of computational costs by the use of gradient enhanced surrogate models, trained on adjoint based sensitivities in the field of multi blade row turbomachinery optimization. This may permit a substantial increase in the manageable design space size.

REFERENCES

[1] Sobieszczanski-Sobieski, J., and Haftka, R., 1997. "Multidisciplinary aerospace design optimization: survey of recent developments". *Structural Optimization,* ***14***, pp. 1–23.

[2] Jameson, A., 1988. "Aerodynamic design via control theory". *Journal of Scientific Computing,* ***3***(3), pp. 233–260.

[3] Giles, M. B., and Pierce, N. A., 2000. "An introduction to the adjoint approach to design". *Flow, Turbulence and Combustion,* ***65***(3–4), pp. 393–415.

[4] Mohammadi, B., and Pironneau, O., 2009. *Applied Shape Optimization for Fluids*. Oxford University Press; second edition, ch. 7, pp. 140–163.

[5] Peter, J. E., and Dwight, R. P., 2010. "Numerical sensitivity analysis for aerodynamic optimization: A survey of approaches". *Computers & Fluids,* ***39***, pp. 373 – 391.

[6] Wang, D. X., and He, L., 2010. "Adjoint aerodynamic design optimization for blades in multistage turbomachines—part i: Methodology and verification". *Journal of Turbomachinery,* ***132***(2), p. 021011.

[7] Wang, D. X., He, L., Li, Y. S., and Wells, R. G., 2010. "Adjoint aerodynamic design optimization for blades in multistage turbomachines—part ii: Validation and application". *Journal of Turbomachinery,* ***132***(2), p. 021012.

[8] Dwight, R., and Brezillon, J., 2006. "Effect of approximations of the discrete adjoint on gradient-based optimization". *AIAA Journal,* ***44***(12), pp. 3022–3071.

[9] Kaplan, B., Nicke, E., and Voss, C., 2006. "Design of a highly efficient low-noise fan for ultra-high bypass engines". In Proceedings of the ASME Turbo Expo 2006, pp. 185–194.

[10] Siller, U., and Aulich, M., 2010. "Multidisciplinary 3D-optimization of a fan stage performance map with consideration of the static and dynamic rotor mechanics". In Proceedings of the ASME Turbo Expo 2010, pp. 317–328.

[11] Lepot, I., Leborgne, M., Schnell, R., Yin, J., Delattre, G., Falissard, F., and Talbotec, J., 2011. "Aero-mechanical optimization of a contra-rotating open rotor and assessment of its aerodynamic and acoustic characteristics". In Proceedings of the 9th European Turbomachinery Conference (ETC).

[12] Schimming, P., Schewe, G., Schmitt, S., Schnell, R., Neise, W., Enghardt, L., Zhang, Y., and Wallscheid, L., 2000. Experimental investigations on a CRISP-1m-model concerning its aeroacoustics, aeroelastics and aerodynamics. Final Project Report. BMBF LFT9601, German Aerospace Center (DLR).

[13] Görke, D., Le Denmat, A.-L., Schmidt, T., Kocian, F., and Nicke, E., 2012. "Aerodynamic and mechanical optimization of CF/PEEK blades of a counter rotating fan". In Proceedings of the ASME Turbo Expo 2012. To be published.

[14] Aulich, M., and Siller, U., 2011. "High-dimensional constrained multiobjective optimization of a fan stage". In Proceedings of the ASME Turbo Expo 2011.

[15] Yang, H., Nürnberger, D., and Kersken, H.-P., 2006. "Towards excellence in turbomachinery computational fluid dynamics". *Journal of Turbomachinery,* ***128***(2), pp. 390–402.

[16] Kügeler, E., Weber, A., Nürnberger, D., and Engel, K., 2008. "Influence of blade fillets on the performance of a 15 stage gas turbine compressor". In Proceedings of the ASME Turbo Expo 2008, pp. 415–424.

[17] Becker, K., Heitkamp, K., and Kügeler, E., 2010. "Recent progress in a hybrid-grid CFD solver for turbomachinery flows". In Proceedings of the 5. European Conference on Computational Fluid Dynamics ECCOMAS CFD 2010.

[18] Kröger, G., Voß, C., and Nicke, E., 2010. "Axisymmetric casing optimization for transonic compressor rotors". In 13th International Symposium on Transport Phenomena and Dynamics of Rotating Machinery (ISROMAC-13).

[19] Frey, C., Nürnberger, D., and Kersken, H., 2009. "Development and application of an adjoint RANS-solver for turbomachinery". In Proceedings of the 8th European Conference on Turbomachinery Conference (ETC).

[20] Frey, C., Nürnberger, D., and Kersken, H., 2009. "The discrete adjoint of a turbomachinery RANS solver". In Proceedings of the ASME Turbo Expo 2009, pp. 345–354.

[21] Frey, C., Ashcroft, G., Backhaus, J., Kügeler, E., and Wellner, J., 2011. "Adjoint-based flow sensitivity analysis using arbitrary control surfaces". In Proceedings of the ASME Turbo Expo 2011.

[22] Kim, C. S., Kim, C., and Rho, O. H., 2003. "Feasibility study of constant eddy-viscosity assumption in gradient-based design optimization". *Journal of Aircraft,* ***40***(6), pp. 1168–1176.

[23] MacKay, D. J. C., 1991. "Bayesian methods for adaptive models". PhD thesis, California Institute of Technology.

[24] MacKay, D. J. C., 1995. "Probable networks and plausible predictions — a review of practical Bayesian methods for supervised neural networks". *Network-Computation in Neural Systems,* ***6***(3), pp. 469–505.

[25] Yamazaki, W., Rumpfkeil, M. P., and Mavriplis, D. J., 2010. "Design optimization utilizing gradient/hessian enhanced surrogate model". In Procedings of the 28th AIAA Applied Aerodynamics Conference.

[26] Han, Z. H., Görtz, S., and Zimmermann, R., 2009. "On improving efficiency and accuracy of variable-fidelity surrogate modeling in aero-data for loads context". In Proceedings of the CEAS 2009 European Air and Space conference.