Improving the Hydrodynamic Performance of Diffuser Vanes *via* Shape Optimization

Tushar Goel[1*], Daniel J Dorney[2**], Raphael T. Haftka[3*], and Wei Shyy[4†]

Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611

**ER42, NASA Marshall Space Flight Center, AL 35812*

†*Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109*

Abstract

The performance of a diffuser in a pump stage depends on its configuration and placement within the stage. The influence of vane shape on the hydrodynamic performance of a diffuser has been studied. The goal of this effort has been to improve the performance of a pump stage by optimizing the shape of the diffuser vanes. The shape of the vanes was defined using Bezier curves and circular arcs. Surrogate model based tools were used to identify regions of the vane that have a strong influence on its performance. Optimization of the vane shape, in the absence of manufacturing and stress constraints, led to a nearly eight percent reduction in the total pressure losses compared to the baseline design by reducing the extent of the base separation.

Keywords: Diffuser, Multiple surrogates, Shape optimization, Hydrodynamics

I.   Introduction

The space shuttle main engine (SSME) is required to operate over a wide range of flow conditions. This requirement imposes numerous challenges on the design of turbomachinery components. One concept that is being explored is the use of an expander cycle for an upper stage engine. A schematic of a representative expander cycle for a conceptual upper stage engine is shown in Figure 1. Oxidizer and fuel pumps are used to feed LOX (liquid oxygen) and LH2 (liquid hydrogen) to the combustion chamber of the main engine. The combustion products are discharged from the nozzle. The pumps are driven by turbines, which use the gasified fuel as the working fluid.

There is a continuing effort to develop subsystems like turbopumps and turbines used in a typical expander cycle based upper stage engine. The requirements on the design of subsystems are influenced by size, weight, efficiency, and manufacturability of the system. Besides, often the requirements for subsystems are coupled, for example, the above-mentioned design constraints require turbopumps to operate at high speeds to have high efficiency with low weight and compact design. This may require seeking alternate designs for the turbopumps and turbines as the current designs may not prove adequate over a wide range of operating conditions. Mack et al. [1, 2] optimized the design of a radial turbine that allows high turbopump speeds, performs comparable to an axial turbine at design conditions, and yields good performance at off-design conditions.

Dorney et al. [3] have been exploring different concepts for turbopump design. A simplified schematic of a pump is shown in Figure 2. Oxidizer or fuel enters from the left. The pressure increases as the fluid passes through the impeller. The fluid emerging from the impeller periphery typically has high tangential velocity, which is partially converted into pressure by passing it over a diffuser. Dorney et al. [3] found that a diffuser with vanes is more efficient than a vaneless diffuser at off-design conditions. Over a range of operating conditions, the performance of a pump is driven by the flow in the diffuser. Generally, the diffuser will stall before the impeller, and the performance of the diffuser will drop off more rapidly at off-design operating conditions. The main objectives of the current effort are to improve the hydrodynamic performance of a diffuser, using advanced optimization techniques, and to study the features of the diffuser vane that influences its performance.

---

[1] Current Affiliation: Livermore Software Technology Corporation, Livermore CA

[2] Aerospace Engineer, Associate Fellow AIAA

[3] Distinguished Professor, Fellow AIAA

[4] Clarence L "Kelly" Johnson Professor, Fellow AIAA

Improvement in performance *via* shape optimization has been successfully achieved in many areas. For example, there are numerous instances of improvement in lift to drag ratio *via* airfoil or wing shape design [4-10], shape of blades is optimized to increase the efficiency of turbines [4, 11] and pumps [12] in past. Shape optimization typically requires a significant number of function evaluations to explore different concepts. If the cost of evaluating a single design is high, as is usually the case, surrogate models of objectives and constraints are frequently used to reduce the computational burden.

Surrogate based optimization approach is widely used in the design of space propulsion systems, such as radial turbines [1, 2], supersonic turbines [4], diffusers [13], rocket injectors [14, 15], and combustion chamber geometry [16]. Detailed reviews on surrogate based optimization are provided by Li and Padula [17], and Queipo et al. [18]. There are many surrogate models, and it is not clear which surrogate model performs the best for any particular problem. In such a scenario, one possible approach to account for uncertainties in predictions is to use an ensemble of surrogates (Goel et al. [19]). This multiple surrogate approach has been demonstrated to work well for several problems including system identification (Goel et al. [20]), compressor blade design (Samad et al. [12]).

Specifically, the objectives of the present study are: i) to improve the hydrodynamic performance of the diffuser *via* shape optimization of vanes, ii) to identify important regions in diffuser vanes that help optimize pressure ratio, and iii) to demonstrate the application of multiple surrogates based strategy for space propulsion systems.

The paper is organized as follows. We define the geometry of the vanes, and numerical tools to evaluate the diffuser vane shapes in Section II. We describe the relevant details in surrogate model based optimization in Section III. The results obtained for the current optimization problem and a discussion of the physics involved in flow over optimal vane is presented in Section IV. Finally, we conclude by recapitulating the major findings of the paper in Section V.

II.   Problem Description

Representative radial locations in the meanline pump flow path (not-to-scale) are shown in Figure 3 (Dorney et al. [3]). The fluid enters from the left, and is guided to the unshrouded impeller *via* inlet guide vanes assembly. The flow from the impeller passes through the diffuser before being collected in the discharge collector. In this study, we focus our efforts on a configuration of 15 inlet guide vanes, seven main and seven splitter blades, and 17 diffuser vanes. The length of the diffuser vanes is also fixed according to the location of the collector. Our goal is to maximize the performance of the diffuser, characterized by the ratio of pressure at the inlet and the outlet, hereafter called as 'pressure ratio'. The performance of the diffuser is governed by the shape of the diffuser vane.

II.A. Vane Shape Definition

The description of geometry is the most important step in the shape optimization. The current shape of the diffuser vanes, referred as baseline design, (shown in Figure 4(A)) was created using meanline and geometry generation codes at NASA [21] and yields a pressure-ratio of 1.074. It is obvious from Figure 4(B) that the existing shape allows the flow to remain attached while passing on the vanes but causes significant flow separation while leaving the vane. The separation induces significant loss of pressure recovery that is the primary goal of using a diffuser. The baseline design was created subject to several constraints, including: i) the number of vanes was set at 17 based on the number of bolts used to attach the two sides of the experimental rig, ii) the vanes had to accommodate a 3/8-inch diameter bolt, with enough excess material to allow manufacturing, and iii) the length of the vanes was set based on the location of the collector. These constraints resulted in an initial optimized design that looked quite similar to the baseline design, and yielded similar performance in both isolated component and full stage simulations. In an effort to more thoroughly explore the design space for diffusers (and allowing for future improvements in materials and manufacturing), the constraints resulting from the bolts (the number and thickness of the vanes) were relaxed. Other factors that must be considered in the following optimization include: i) the designs resulting from the optimization techniques shown in this paper are based on diffuser-alone simulations (different results may be obtained if the optimizations were based on full stage simulations), and ii) no effort was made to determine if the proposed designs would meet stress and/or manufacturing requirements.

To reduce the separation, we represent the geometry of a vane by five sections using a circular arc and Bezier curves as shown in Figure 5. Sections 1, 3, and 4 are Bezier curves, and Section 2 is a circular arc. The shape of the inlet nose (between points B_1 and B_2) is obtained from the existing baseline vane shape. The circular arc in Section 2 is described by fixing the radius r ($r = 0.08$), the location of the center C_I (-3.9, 6.1), the start and end angles [21]. The arc begins at angle $\pi + \theta_r + \theta/2$ (point A_1) and ends at $\pi + \theta_r - \theta/2$ (point A_2). Here, we use $\theta_r = 22.5^o$ and θ is a design variable. We provide the information about the coordinates and slope of the tangents at endpoints such that different Bezier curves (Appendix 1) are generated by varying the length of tangents. The points A_1 and A_2,

2

shown by red dots in Figure 5, and the tangents to the arc, serve as the end points location and tangents used to define the Bezier curves in Sections 1 and 3. Coordinates and the slopes at points B_1 and B_2 (obtained using the data of inlet nose of the baseline design [21]) are used to define Bezier curves in Sections 1 and 4. The Bezier curve in Section 1 (Figure 5) is defined using the coordinates and slope of tangents at B_2 and A_2. The lengths of the tangents t_1 and t_2 control the shape of the curve. While one end point coordinates and slope for Bezier curves in Sections 3 and 4 are known at points A_1 and B_1, the second end point and slope are obtained by defining the location of point P (P_y, P_z). The slope of tangent at point P is taken as *five degrees* more than the slope of the line joining points P and B_1. The additional slope is specified to avoid all the points in Section 4 fall in a straight line. The values of the fixed parameters are decided based on the inputs from designer [21]. The lengths of tangents t_1–t_6 serve as variables to generate Bezier curves in Sections 1, 3 and 4.

The ranges of the design variables, summarized in Table 1, are selected such that we obtain practically feasible vane geometries and the corresponding grids (discussed in next section). We note that present Bezier curve and circular arc based definition of diffuser vane shapes provides significantly different vane geometries than the baseline design, particularly near the outlet region. The main rationale behind this choice of vane shape is the relaxation of manufacturing and stress constraints, ease of parameterization and better control of the curvature of the vane that allows the flow to remain attached.

II.B. Mesh Generation, Boundary Conditions, and Numerical Simulation

Performance of diffuser vane geometries is analyzed using the NASA PHANTOM code [22] developed at Marshall Space Flight Center to analyze turbomachinery flows. This 3D, unsteady, Navier-Stokes code utilizes structured, overset O- and H-grids to discretize and analyze the unsteady flow resulting from the relative motion of rotating components. The code is based on the Generalized Equations Set methodology [23] and utilizes a modified Baldwin-Lomax turbulence model [24]. The inviscid and viscous fluxes are discretized using a third-order spatially accurate Roe's scheme, and second-order central differencing, respectively. The unsteady terms are modeled using a second order accurate scheme.

For this problem, we solve incompressible, unsteady, non-rotating, turbulent, single phase, constant-material-property flow over diffuser vane. The working fluid is water. By taking the advantage of periodicity, only a single vane is analyzed here. A combination of H- and O-grids with 13065 grid points, has been used to analyze diffuser vane shapes. A typical grid is shown in Figure 6. The boundary conditions imposed on the flow domain are as follows. Mass flux, total temperature, and flow angles (circumferential and radial) are specified at the inlet. Mass flux is fixed at the outlet. All solid boundaries are modeled as no slip, adiabatic walls, with zero normal derivative of pressure. Periodic boundary condition is enforced at outer boundaries. With this setup, it takes approximately 15 minutes on a single Intel Xeon processor (2.0 GHz, 1.0 GB RAM) to simulate each design.

III.   Methodology

As discussed earlier, the surrogate model based approach is suitable to reduce the computational cost of optimization. In this section, we discuss different steps and considerations in surrogate modeling and its application to the design problem(s). A stepwise procedure of surrogate based analysis and optimization is explained with the help of Figure 7. Firstly, we identify the objectives, constraints, and design variables. Next, we develop a procedure to evaluate different designs. In context of the current problem, we identified design variables, objectives, and numerical procedure to evaluate different designs in the previous section. To reduce the computational expense involved in optimization, we construct multiple surrogate models of the objectives and constraints. A brief description of popular surrogate models is as follows.

III.A. Surrogate Modeling

There are many surrogate models e.g., polynomial response surface approximations, kriging, radial basis neural network, support vector regression etc. A detailed discussion of different aspects of surrogate modeling was reviewed by Li and Padula [17] and Queipo et al. [18]. We give a brief description of different surrogate models as follows.

III.A.1. Polynomial Response Surface Approximation (PRS, [25])

The observed response $y(\mathbf{x})$ of a function at point $\mathbf{x}$ is represented as a linear combination of basis functions $f_i(\mathbf{x})$ (mostly monomials are selected as basis functions) and coefficients β_i. Error in approximation ε is assumed to be uncorrelated and normally distributed with zero mean and σ^2 variance. That is,

$$y(\mathbf{x}) = \sum_i \beta_i f_i(\mathbf{x}) + \varepsilon; \qquad E(\varepsilon) = 0, V(\varepsilon) = \sigma^2. \tag{1}$$

The polynomial response surface approximation of $y(\mathbf{x})$ is,

$$\hat{y}(\mathbf{x}) = \sum_i b_i f_i(\mathbf{x}), \tag{2}$$

where b_i is the estimated value of the coefficient associated with the ith basis function $f_i(\mathbf{x})$. The coefficient vector $\mathbf{b}$ is obtained by minimizing the error in approximation ($e(\mathbf{x}) = y(\mathbf{x}) - \hat{y}(\mathbf{x})$) at N_s sampled design points in a least square sense as,

$$\mathbf{b} = (X^T X)^{-1} X^T \mathbf{y}, \tag{3}$$

where X is the matrix of basis functions and $\mathbf{y}$ is the vector of responses at N_s design points. The quality of approximation is measured by computing the coefficient of multiple determination R_{adj}^2 defined as,

$$R_{adj}^2 = 1 - \sigma_a^2 \left. (N_s - 1) \middle/ \sum_{i=1}^{N_S} (y_i - \bar{y})^2 \right. , \tag{4}$$

where $\bar{y} = (N_s)^{-1} \sum_{i=1}^{N_s} y_i$ and RMS error at sampling points is given as $\sigma_a = \sqrt{(N_s - N_\beta)^{-1} \sum_{i=1}^{N_S} (y(\mathbf{x}_i) - \hat{y}(\mathbf{x}_i))^2}$. For a good fit, R_{adj}^2 should be close to 1. For more details on polynomial response surface approximation, we refer the reader to text by Myers and Montgomery [25].

III.A.2. Kriging (KRG, [26])

Kriging is named after the pioneering work of D.G. Krige (a South African mining engineer) and was formally developed by Matheron [26]. Kriging estimates the value of an objective function $y(\mathbf{x})$ at design point $\mathbf{x}$ as the sum of a linear polynomial trend model $\sum_{i=1}^{N_v} \beta_i f_i(\mathbf{x})$ and a systematic departure $Z(\mathbf{x})$ representing low (large scale) and high frequency (small scale) variations around the trend model.

$$y(\mathbf{x}) = \hat{y}(\mathbf{x}) = \sum_i \beta_i f_i(\mathbf{x}) + Z(\mathbf{x}). \tag{5}$$

The systematic departure components are assumed to be correlated as a function of distance between the locations under consideration. Gaussian function is the most commonly used correlation function.

$$C(Z(\mathbf{x}), Z(\mathbf{s}), \boldsymbol{\theta}) = \prod_{i=1}^{N_v} \exp\left(-\theta_i (x_i - s_i)^2\right). \tag{6}$$

The parameters β_i, θ_i are obtained by maximizing a likelihood function that is a measure of the probability of the sample data being drawn from it. We use an implementation by Lophaven et al. [27].

III.A.3. Radial Basis Neural Network (RBNN, [28])

The objective function is approximated as a weighted combination of responses from radial basis functions (also known as neurons).

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^{N_{RBF}} w_i a_i(\mathbf{x}), \tag{7}$$

where $a_i(\mathbf{x})$ is the response of the ith radial basis function at design point $\mathbf{x}$ and w_i is the weight associated with $a_i(\mathbf{x})$. Mostly Gaussian function is used for radial basis function $a(\mathbf{x})$ as

$$a = radbas\left(\|\mathbf{s} - \mathbf{x}\| b\right); radbas(n) = e^{-n^2}. \tag{8}$$

The parameter b in the above equation is inversely related to a user defined parameter 'spread constant' that controls the response of the radial basis function. A high 'spread constant' would cause the response of neurons to be very smooth and very low spread constant would result into a highly non-linear response function. Typically, spread constant is selected between zero and one. The number of radial basis functions (neurons) and associated weights are determined by satisfying the user defined error 'goal' on the mean square error in approximation. We set the error goal as the square of *five-percent* of the mean response.

As discussed here, we have many surrogate models and it is unknown a priori, which surrogate would be most suitable for a given problem. Besides, the choice of best surrogate model changes with sampling density and nature of the problem [19]. It has been shown by Goel et al. [19] that in such scenario simultaneously using multiple surrogate models protects us from choosing wrong surrogates. They proposed using a weighted averaged surrogate that is described as follows.

III.A.4. PRESS-based Weighted Average Surrogate Model (PWS, [19])

We develop a weighted average surrogate model as,

$$\hat{y}_{pws}(\mathbf{x}) = \sum_{i}^{N_{SM}} w_i \hat{y}_i(\mathbf{x}), \tag{9}$$

where $\hat{y}_{pws}(\mathbf{x})$ is the predicted response by the weighted average of surrogate models, $\hat{y}_i(\mathbf{x})$ is the predicted response by the ith surrogate model and w_i is the weight associated with the ith surrogate model at design point $\mathbf{x}$. Furthermore, the sum of the weights must be one $\left(\sum_{i=1}^{N_{SM}} w_i = 1 \right)$ so that if all the surrogates agree, $\hat{y}_{pws}(\mathbf{x})$ will also be the same. Weights are determined as follows.

$$w_i^* = \left(E_i \Big/ E_{avg} + \alpha \right)^{\beta}, \qquad w_i = w_i^* \Big/ \sum_i w_i^*,$$

$$E_{avg} = \sum_{i=1}^{N_{SM}} E_i \Big/ N_{SM}; \qquad \beta < 0, \alpha < 1, \tag{10}$$

where E_i is the global data-based error measure for i^{th} surrogate model. In this study, generalized mean square cross-validation error (GMSE) (leave-one-out cross validation or *PRESS* in polynomial response surface approximation terminology), defined in the Appendix 2, is used as global data-based error measure, by replacing E_i by $\sqrt{GMSE_i}$. We use $\alpha = 0.05$ and $\beta = -1$. The above mentioned formulation of weighting schemes is used with polynomial response surface approximation (PRS), kriging (KRG) and radial basis neural networks (RBNN) such that,

$$\hat{y}_{pws} = w_{prs}\hat{y}_{prs} + w_{krg}\hat{y}_{krg} + w_{rbnn}\hat{y}_{rbnn}. \tag{11}$$

For more details about the weighted average surrogate model, we refer the reader to Goel et al. [19].

Next, we use these surrogate models to characterize the importance of different variables and to identify the most and the least important variables for different objectives and constraints. This information is vital to understand the role of the variables. Also, we can fix the least important variables to reduce the dimensionality of the problem. We use global sensitivity analysis method proposed by Sobol [29].

III.B. Global Sensitivity Analysis (GSA, [29])

Sobol [29] presented a variance based, non-parametric approach to perform global sensitivity analysis. In this approach, the response function is decomposed into unique additive functions of variables and their interactions such that the mean of each additive function is zero. This decomposition allows the variance (V) to be computed as a sum of individual partial variance of each variable (V_i) and partial variance of interactions (V_{ij}) of different variables. The sensitivity of the response function with respect to each variable is assessed by comparing the sensitivity indices (S_i, S_{ij}) that is the relative magnitude of partial and total variance of each variable. The details of the method are as follows.

A function $f(\mathbf{x})$ of a square integrable objective as a function of a vector of independent uniformly distributed random input variables, $\mathbf{x}$ in domain [0, 1] is assumed. The function can be decomposed as the sum of functions of increasing dimensionality as

$$f(\mathbf{x}) = f_0 + \sum_i f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) + \cdots + f_{12\ldots N_v}(x_1, x_2, \ldots, x_{N_v}),\tag{12}$$

where $f_0 = \int_{\mathbf{x}=0}^1 f\, d\mathbf{x}$. If the following condition

$$\int_0^1 f_{i_1\ldots i_s}\, dx_k = 0,\tag{13}$$

is imposed for $k = i_1, \ldots, i_s$, then the decomposition described in Equation (12) is unique. In context of global sensitivity analysis, the total variance denoted as $V(f)$ can be shown equal to

$$V(f) = \sum_{i=1}^{N_v} V_i + \sum_{1 \le i,j \le N_v} V_{ij} + \ldots + V_{1\ldots N_v},\tag{14}$$

where $V(f) = E((f - f_0)^2)$, and each of the terms in Equation (14) represents the partial contribution or partial variance of the independent variables (V_i) or set of variables to the total variance and provides an indication of their relative importance. The partial variances can be calculated using the following expressions:

$$\begin{aligned} V_i &= V(E[f \mid x_i]), \\ V_{ij} &= V(E[f \mid x_i, x_j]) - V_i - V_j, \\ V_{ijk} &= V(E[f \mid x_i, x_j, x_j]) - V_{ij} - V_{ik} - V_{jk} - V_i - V_j - V_k, \end{aligned}\tag{15}$$

and so on, where V and E denote variance and expected value respectively. Note that $E[f \mid x_i] = \int_0^1 f_i\, dx_i$, and $V(E[f \mid x_i]) = \int_0^1 f_i^2\, dx_i$. This formulation facilitates the computation of the sensitivity indices corresponding to the independent variables and set of variables. For example, the first and second order sensitivity indices can be computed as

$$S_i = V_i \big/ V(f), \quad S_{ij} = V_{ij} \big/ V(f).\tag{16}$$

Under the independent model inputs assumption, the sum of all the sensitivity indices is equal to one. The first order sensitivity index for a given variable represents the main effect of the variable, but it does not take into account the effect of the interaction of the variables. The total contribution of a variable to the total variance is given as the sum of all the interactions and the main effect of the variable. The total sensitivity index of a variable is then defined as

$$S_i^{total} = \frac{1}{V(f)}\left(V_i + \sum_{j, j \neq i} V_{ij} + \sum_{j, j \neq i}\sum_{k, k \neq i} V_{ijk} + \ldots\right)\tag{17}$$

Note that the above referenced expressions can be easily evaluated using surrogate models of the objective functions. Sobol [29] has proposed a variance-based non-parametric approach to estimate the global sensitivity for any combination of design variables using Monte Carlo methods. To calculate the total sensitivity of any design variable x_i, the design variable set is divided into two complementary subsets of x_i and Z $\left(Z = x_j, \forall j = 1, N_v; j \neq i\right)$. The purpose of using these subsets is to isolate the influence of x_i from the influence of the remaining design variables included in Z. The total sensitivity index for x_i is then defined as

$$S_i^{total} = V_i^{total} \big/ V(f),\tag{18}$$

where

$$V_i^{total} = V_i + V_{i,Z},\tag{19}$$

where V_i is the partial variance of the objective with respect to x_i and $V_{i,Z}$ is the measure of the objective variance that is dependent on interactions between x_i and Z. Similarly, the partial variance for Z can be defined as V_z. Therefore the total objective variability can be written as

$$V = V_i + V_Z + V_{i,Z}. \tag{20}$$

While Sobol [29] had used Monte Carlo simulations to conduct the global sensitivity analysis, we use Gauss-quadrature numerical integration of different partial variance terms in Equation (15) to calculate sensitivity indices.

Surrogate models, which represent objectives and constraints, are used to optimize the performance of the system. If we use multiple surrogates, we will find more than one candidate optimal solutions. The performance of such predicted optimal designs is validated using numerical simulation. If we are satisfied with the performance of the subsystem, we terminate the search procedure. Otherwise, we refine the design space in the region of interest. The design space refinement can be done in multiple ways, i) we can fix the least important variables at the optimal values (as realized by optimization) or mean values to reduce the dimensionality of the problem, ii) we sample more points in the design space, and iii) we identify the region where we expect potential improvement and concentrate on that region. We repeat this optimization procedure till convergence.

IV. Results and Discussion

We present application of the above-mentioned surrogate based optimization framework to optimize the performance of diffuser vanes in this section. We also present a detailed analysis of optimal design.

IV.A. Surrogate Model Construction

First step in surrogate modeling is to sample data in design variable space. Different design of experiment (DOE) techniques are effective in reducing the computational expense of generating high-fidelity surrogate models. The most popular methods DOE methods are Latin hypercube sampling (LHS) and face-centered central composite designs (FCCD). The choice of DOE depends on the nature of problem and the number of samples required.

For this problem, we had nine design variables hence we selected 110 design points to allow adequate data to evaluate 55 coefficients of a quadratic polynomial response surface. Since face-centered central composite design requires unreasonably large number of samples (531 points to approximate 55 coefficients), we used Latin hypercube sampling (LHS) to construct surrogate models. We generated LHS designs using MATLAB routine 'lhsdesign' with 100 iterations for maximize the minimum distance between points. We evaluated each diffuser vane shape using PHANTOM. This dataset was referred as 'Set A'. The range of the data, given in Table 2, indicated potential of improvement in the performance of the diffuser by shape optimization. We constructed four surrogate models, polynomial response surface approximation (PRS), kriging (KRG), radial basis neural network (RBNN), and PRESS-based weighted average surrogate (PWS) of the objective pressure ratio. We used quadratic polynomial for PRS, and linear trend model with Gaussian correlation function for kriging. For RBNN, the spread coefficient was taken as 0.5 and the error goal was the square of *five-percent* of the mean value of the response at data points. The parameters α, and β for PWS model were 0.05 and -1, respectively. The summary of quality indicators for different surrogate models is given in Table 2.

All error indicators are desired to be low compared to the response data, except R^2_{adj}, which is desired to be close to one. The PRESS (Appendix 2) and RMS error (~1.0e-2) were very high compared to the range of data. This indicated that all surrogate models poorly approximated the actual response, and were likely to yield inaccurate results if used for global sensitivity analysis and optimization. To identify the cause of poor surrogate modeling, we conducted a lack-of-fit test (refer to Appendix 3) for PRS. A low p-value (~0.017) indicated that the chosen order of the polynomial was inadequate in the selected design space. Since, the data available at 110 points is insufficient to estimate 220 coefficients in a cubic polynomial, this issue of model inadequacy also reflected lack of data.

IV.B. Design Space Refinement

We addressed the issue of model accuracy or data inadequacy using two parallel approaches. Firstly, we added more data in the design space to improve the quality of fit. We sampled 330 additional points using Latin hypercube sampling such that we had 440 design points to fit a cubic polynomial (220 coefficients). We call this dataset as 'Set B'. Secondly, the low mean value of the response data (1.041) at 110 points compared to the baseline design (1.074) indicated that large portion of the current design space was undesirable due to inferior performance. Hence, it might be appropriate to identify the region where we expect improvements in the performance of the designs, and construct surrogate models by sampling additional design points in that region (reasonable design space approach

[31]). To identify the region of interest, we used the surrogate models, constructed with Set A data (110 design points), to evaluate response at a large number of points in design space. Specifically, we evaluated responses at a grid of four Gaussian points in each direction (total 4^9=262,144 points). We chose Gaussian points instead of usual uniform grids because Gaussian points lie inside the design domain, and are less susceptible to extrapolation errors than corners of uniform grids that might fall outside the convex hull of LHS design points used to construct surrogate models. Any point with a predicted performance (due to any surrogate model) of 1.08 units or better (0.5% improvement over the baseline design) was considered to belong to the potential region of interest. This process identified 29,681 unique points (~11%) in the potential good region. We selected 110 points from this data set using D-optimality criterion in this smaller region. D-optimal designs were generated using MATLAB routine 'candexch' with a maximum of 100 iterations to maximize D-efficiency ([25], pp. 93). This 110 points dataset is called 'Set C'.

As before, we conducted simulations at data points in the Sets B and C using PHANTOM. One point in each set failed to provide an appropriate mesh. The mean, minimum, and maximum values of the pressure ratio for the two datasets are summarized in Table 3. We observed only minor differences in the mean pressure-ratio of the Set B compared to the Set A (Table 2) but the responses in the Set C data clearly indicated high potential of improvement. This demonstrates the effectiveness of reasonable design space approach used to identify the region of interest.

We approximated the data in the Set B and the Set C using four surrogates. We employed a reduced cubic, and a reduced quadratic polynomial for PRS approximation of the Set B and the Set C data, respectively. As can be seen from different error measures in Table 3, the quality of surrogate models fitted to the Set B and the Set C data was better than the surrogate models fitted to the Set A data. This improvement in surrogate approximation was attributed to the increase in the sampling density (Set B) allowing a cubic model and the reduction of the design space (Set C). Both PRS models did not fail the lack-of-fit test (p-value ~0.90+) such that the chosen polynomial order was adequacy in respective design space. The PRESS metric and weights associated with different surrogates suggested that PRS was the best surrogate model for the Set B and the Set C data, unlike kriging for the Set A data.

We used the surrogate models fit to the Set B data for global sensitivity analysis and the surrogate models fitted to the Set C data for optimization. The optimization of the performance using the surrogate models fitted to the Set B data resulted in an inferior design compared to the optimal design obtained using the Set C data based surrogates. Some of the optimal designs from the Set B data based surrogates could not be analyzed. This anomaly arose because large design space was sampled with limited data such that large regions remain unsampled; and hence, susceptible to significant errors, particularly near the corners where optima were found. The same issue restricted the use of surrogate models constructed using Set C data for global sensitivity analysis as there was large extrapolation error outside the region of interest where no point was sampled. Hence, surrogate models constructed using Set B data were more suited for conducting global sensitivity analysis.

IV.C. Global Sensitivity Assessment

We used global sensitivity analysis (GSA) to identify the most and the least important design variables. We used Gauss quadrature numerical integration scheme with four Gauss points along each direction (total 4^9=262,144 points) to evaluate different integrals in GSA. The response at each point was evaluated using surrogate models fit to the Set B data. Corresponding sensitivity indices of main effects of different variables are shown in Figure 8. Though there are differences in the exact magnitude of sensitivity indices from various surrogates, all surrogates indicated that the pressure ratio was most influenced by three variables, P_z, t_2, and P_y. A comparison of sensitivity indices of total and main effect of design variables (using PWS) in Figure 9 suggested that the interactions between variables are small but non-trivial.

IV.D. Validation of Global Sensitivity Analysis Results

To validate the findings of the global sensitivity analysis, we evaluated the variation in the response function (pressure ratio) by varying one variable at a time while keeping remaining variables at the mean values. We specified five equi-spaced levels for each design variable and used trapezoidal rule to compute the actual variance. The responses at design points were evaluated by performing actual numerical simulations. The results of actual variance computations are shown in Figure 10.

The one-dimensional variance computation results also indicated that variables P_z, t_2, and P_y were more important than all other variables. This validated the findings of the global sensitivity analysis. The differences in the results of one-dimensional variance computation and global sensitivity analysis can be explained as follows: i) the number of points used to compute one-dimensional variance is small, ii) one-dimensional variance computation does not account for interactions between variables, and iii) there are approximation errors in using surrogate models for global sensitivity analysis. Nevertheless, the main implication of the result is that the performance of the diffuser vane was pre-dominantly affected by the location of point P and the length of tangent t_2 in Section 1 (Figure 5).

IV.E. Preliminary Optimization of Diffuser Vane Performance

Next, we used surrogate models, fit to the Set C data, to maximize the pressure-ratio by exploring different diffuser vane shapes. To avoid the danger of large extrapolation errors in the unsampled region, we employed the surrogate model (PRS surrogate) from the Set A as a constraint (all points in the feasible region have predicted response greater than a threshold value of 1.08). Our use of PRS from the Set A as the constraint was motivated by the simplicity of the constraint function, and the fact that PRS contributed to the most number of points in the potential region of interest.

We used sequential quadratic programming optimizer to find the optimal shapes. The optimal configuration of blade shapes obtained using different surrogates as function evaluators is shown in Figure 11, and the corresponding optimal design variables are given in Table 4. The optimal designs obtained from all surrogates were close to each other in both function and design space. A few minor differences were observed in relatively insignificant design variables (refer to the results of global sensitivity analysis). Notably all design variables touched the bounds for PRS and were close to the corner for other surrogate models. The small value of θ indicated sharper nose, as was used in the baseline design. Also, most tangents were at their lower bounds that resulted in low curvature sections. Near the point P, the tangents were on their upper limits to facilitate gradual transition in the slope. The optimal vane was thinner in the middle section and was longer compared to the baseline design (Figure 11). The central region of the optimal design was non-convex compared to the convex section for the baseline design.

We simulated all four candidate optimal designs from different surrogate models to evaluate the improvements. The actual and predicted performances from different surrogates are compared in Table 5. We observed that the error in approximation for different surrogates was comparable to their respective PRESS errors. Nevertheless, PRS was the most accurate surrogate model and furnished the best performance shape. RBNN was the worst surrogate model. PRESS-based weighted average surrogate performed significantly better than the worst surrogate. The best predicted diffuser vane yielded significant improvements in the performance (1.117) compared to the baseline design (1.074). We refer to this design as 'intermediate optimal' design.

It is obvious that the design space refinement in the region of interest based on multiple surrogates has pay-offs in the improved performance of surrogates and the identification of optimal design. The high confidence in the optimal predictions was also derived from the similar performance of all surrogate models. The results also showed incentives (protection against the worst design, proper identification of the reasonable design space) of investing a small amount of computational resources in constructing multiple surrogate models (less than the cost of a single simulation) for computationally expensive problems, and then the extra cost of evaluating multiple optima.

We compared instantaneous (at the end of simulation) and time-averaged flow fields from the intermediate optimal design (from PRS) with the baseline design in Figure 12. The intermediate optimal design allowed smoother turning of the flow compared to the baseline design, and reduced the losses due to separation of the flow, which were significantly high in the baseline design. Consequently, the pressure at the outlet was higher for this intermediate optimal design. We also noted the increase in pressure on the vane for the intermediate optimal design.

IV.F. Design Space Refinement via Dimensionality Reduction

We used first design space refinement by identifying the region of interest. This helped in identifying the intermediate optimal design. Since most design variables in the optimal design were at the boundary of the design space (Table 4), further improvements in the performance of the diffuser might be obtained by refining the design space. To reduce the computational expense, we reduce the design space by utilizing the findings of the global sensitivity analysis. We fixed six relatively insignificant design variables at the optimal design (predicted using PRS), and expanded the range of the three most important design variables. The modified ranges of the design variables and the fixed parameters are given in Table 6. We selected 20 design points using a combination of face-centered central composite design (15 points FCCD), and LHS designs (5 points) of experiments. The range of pressure ratio at 20 design points is given in Table 7. Note that all the tested designs in the refined design space perform better than the intermediate optimal design.

As before, we constructed four surrogates in the refined design space. The performance metrics, specified in Table 7, indicated that all surrogates approximated the response function very well. The weights associated with different surrogates, suggested that a reduced quadratic PRS approximation represents the data the best. This result is not unexpected since any smooth function can be represented by a second order polynomial if the domain of application is small enough. As before, PWS model was comparable to the best surrogate.

IV.G. Final Optimization

The design variables for the four optimal designs of the diffuser vane obtained using different surrogate models and corresponding surrogate-predicted and actual (CFD simulation) pressure ratio are listed in Table 8. The error in predictions of surrogates compared well with the quality indicators and all surrogates had only minor differences in the performance. Also, the optimal vane shapes from different surrogates were similar. In this case also, polynomial response surface approximation conceded the smallest errors in prediction. While the performance of the optimal diffuser vane has improved compared to the intermediate optimal design (compare to Table 5), the contribution of the optimization process was insignificant in this phase. One of the data points ($t_2 = 0.60$, $P_y = -2.00$, $P_z = 6.00$) resulted in a better performance (*P-ratio = 1.151*) than the predicted optimal. This result was not surprising because the optimal design existed at a corner that was already sampled leaving little scope for further improvement. Nevertheless, the optimizers correctly concentrated on the best region. As expected, the optimized design is thinner and streamlined to further reduce the losses, and to improve the pressure recovery. We analyzed the optimal diffuser vane shape according to the flow structure and the other considerations as follows.

IV.H. Flow Structure

The instantaneous and time-averaged pressure contours for the optimal vane shape (best data point) are shown in Figure 13. We observed further reduction in separation losses, and smoother turning of the flow compared to the intermediate optimal design obtained before design space refinement (Table 5). Consequently, the pressure rise in the diffuser was higher. The optimal vane shape had a notable curvature in the middle section on the lower side of the vane (near the point P). This curvature decelerated the flow and led to faster increase in pressure (notice the shift of higher pressure region towards the inlet in Figure 12 and Figure 13).

IV.I. Vane Loadings

The shapes and pressure loads on the baseline, intermediate optimal, and final optimal vanes are shown in Figure 14. We noted the increase in the mean pressure on the diffuser vane for the optimal design. The pressure loads near the inlet tip, and the pressure loading on the diffuser vane, given by the area bound by the pressure profile on the two sides, had reduced by optimization. However, the optimized diffuser vane might be susceptible to high stresses as the optimal design was thinner compared to the baseline vane. The intermediate optimal design served as a compromise design with relatively higher pressure ratio (~4%) compared to the baseline design and lower pressure loading on the vanes compared to the optimal design.

In the future, this problem would be studied by accounting for manufacturing and structural considerations like, stresses in the diffuser vanes. One can either specify a constraint to limit stress to be less than the feasible value or alternatively, one can solve a multi-objective optimization problem with two competing objectives, minimization of stress or pressure loading in the vane, and maximization of pressure ratio.

IV.J. Empirical Considerations

Typically, the vane shape design is carried out using empirical considerations on the gaps between adjacent diffuser vanes as shown in Figure 15. The empirical suggestions on the ratio of different gaps [32] and actual values obtained for different vanes are given in Table 9. Contrary to the empirical relations, the ratio of length to width gap (L/W_1), and ratio of width gaps (W_2/W_1) decreased as the pressure ratio increases, though the actual magnitude of length and width of the gaps increase. The discrepancies between the optimal design and the empirical optimal ratios [32] are explained by multiple design considerations used for the empirical optimal. Firstly, we note that the optimization was carried out for a 2D diffuser vane not for the combination of vanes and flow for which empirical ratios are provided. This allowed a variable height of the vane for optimization. However, the empirical ratios are obtained by assuming a constant channel height so that the area ratio of the channel reduces to the ratio of width gaps (W_2/W_1). Nevertheless, this requires further investigation to understand the cause of discrepancies between empirical ratios and that obtained for optimal design.

V.   Summary

We used surrogate model based optimization strategy to maximize the hydrodynamic performance of a diffuser, characterized by the increase in pressure ratio, by modifying the shape of the vanes. The shape of the diffuser vanes was defined by a combination of Bezier curves and a circular arc. Firstly, we defined the shape of the vane using nine design variables and used surrogate models to represent the pressure ratio. We used lack-of-fit test to identify the issues of model inadequacy, and insufficiency of the data to represent the pressure ratio. We addressed these

issues by, i) adding more data points, and ii) identifying the region of interest using the less-accurate surrogate models. More samples were added in the region of interest using the information from multiple surrogate models.

The surrogate models, constructed with increased data and/or in smaller design space, were significantly more accurate than the initial surrogate models. Also, during the course of design space refinement, the best surrogate model changed from kriging (initial data) to polynomial response surface approximation (all subsequent results). Had we followed the conventional approach of identifying the best surrogate model with the first design of experiments, and then used that surrogate model for optimization, we might have not captured the best design. Thus, we can say that the results reflect the improvements in the performance using the design space refinement approach, and using multiple surrogates constructed by incurring a low computational cost.

We conducted a surrogate model based sensitivity analysis to identify the most important design variables in the entire design space. Three design variables controlling the shape of the upper and lower side of the vane were found to be most influential. We used surrogate model in the reduced design space to identify the optimal design in nine variable design space. This intermediate optimal design improved the pressure ratio by more than four percent compared to the baseline design.

Since all the design variables for intermediate optimal design hit the bounds, we further refined the design space by fixing the least important variables on optimal values to reduce the design space, and relaxing the bounds on the most important design variables. The optimal design obtained using the surrogate models in the refined design space further improved the performance of the diffuser by more than eight percent compared to the baseline design. The pressure losses in the flow were reduced, and a more uniform pressure increase on the vane was obtained. However, the optimal vane shape might be susceptible to fail due to high stresses. This behavior was attributed to the absence of stress constraint that allowed using thinner vanes to maximize the performance. In the future, the optimization would be carried out by considering the multi-disciplinary analysis accounting for stress constraint, manufacturability, and pressure increase.

In terms of the vane shapes, thin vanes helped improve the hydrodynamic performance significantly. The interesting aspect was the change in the sign of curvature of the vane on the suction side that allows an initial speeding of the flow followed by a continuous pressure recovery without flow separation.

VI.   Acknowledgements

Appendix 1: Bezier Curve

A typical parametric Bezier curve $f(x)$, shown in Figure 16, is defined with the help of two end points P_0 and P_1, and two control points P_2 and P_3 as follows.

$$f(x) = P_0(1-x)^3 + 3P_1x(1-x)^2 + 3P_2x^2(1-x) + P_3x^3, x \in [0,1]. \tag{21}$$

The co-ordinate of any point on the Bezier curve is obtained by substituting the value of x accordingly. In this study, the location of control points is obtained by using the information about the slope and the length of the tangents at end points (Papila et al. [4]).

Appendix 2: Generalized Mean Square Cross-Validation Error (GMSE or PRESS)

In general, the data is divided into k subsets (k-fold cross-validation) of approximately equal size. A surrogate model is constructed k times, each time leaving out one of the subsets from training, and using the omitted subset to compute the error measure of interest. The generalization error estimate is computed using the k error measures obtained (e.g., average). If k equals the sample size, this approach is called leave-one-out cross-validation (also known as PRESS in the polynomial response surface approximation terminology). Equation (22) represents a leave-one-out calculation when the generalization error is described by the mean square error (GMSE).

$$GMSE = \frac{1}{k}\sum_{i=1}^{k}(y_i - \hat{y}_i^{(-i)})^2, \tag{22}$$

where $\hat{y}_i^{(-i)}$ represents the prediction at $\mathbf{x}^{(i)}$ using the surrogate constructed using all sample points except ($\mathbf{x}^{(i)}, y_i$). Analytical expressions are available for that case for the GMSE without actually performing the repeated

construction of the surrogates for both polynomial response surface approximation (Myers and Montgomery [25], Section 2.7) and kriging (Martin and Simpson [30]); however, here we used Equation (22). The advantage of cross-validation is that it provides nearly unbiased estimate of the generalization error and the corresponding variance is reduced (when compared to split-sample) considering that every point gets to be in a test set once, and in a training set k-1 times (regardless of how the data is divided).

Appendix 3: Lack-of-Fit Test with Non-replicate Data for Polynomial Response Surface Approximation

A standard lack-of-fit test is a statistical tool to determine the influence of bias error (order of polynomial) on the predictions [25]. The test compares the estimated magnitudes of the error variance and the residuals unaccounted for by the fitted model. Lets say, we have M unique locations of the data and at j^{th} location, we repeat the experiment n_j times, such that total number of points used to construct surrogate model is $N_s = \sum_{j=1}^{M} n_j$. The sum of squares due to pure error is given by,

$$ SS_{pe} = \sum_{j=1}^{M} \sum_{k=1}^{n_j} \left(y_{jk} - \overline{y}_j \right)^2 \tag{23} $$

where $\overline{y}_j$ is the mean response at the jth sample location, given as, $\overline{y}_j = \frac{1}{n_j} \sum_{k=1}^{n_j} y_k$. The sum of square of residuals due to lack-of-fit of the polynomial response surface model is,

$$ SS_{lof} = \sum_{j=1}^{M} n_j \left(\hat{y}(\mathbf{x}_j) - \overline{y}_j \right)^2 , \tag{24} $$

where $\hat{y}(\mathbf{x}_j)$ is the predicted response at sampled location $\mathbf{x}_j$. In matrix form, the above expressions are given as,

$$ SS_{pe} = \sum_{j=1}^{M} \mathbf{y}_{n_j}^T \left(I_{n_j} - \frac{1}{n_j} \mathbf{1}_{n_j} \mathbf{1}_{n_j}^T \right) \mathbf{y}_{n_j} , \tag{25} $$

$$ SS_{lof} = \mathbf{y}^T \left(I_{N_s} - X \left(X^T X \right)^{-1} X^T \right) \mathbf{y} - \sum_{j=1}^{M} \mathbf{y}_{n_j}^T \left(I_{n_j} - \frac{1}{n_j} \mathbf{1}_{n_j} \mathbf{1}_{n_j}^T \right) \mathbf{y}_{n_j} , \tag{26} $$

where $\mathbf{1}_{n_j}$ is the $(n_j \times 1)$ vector of ones, I_{n_j} is $(n_j \times n_j)$ identity matrix, I_{N_s} is $(N_s \times N_s)$ identity matrix. We formulate F-ratio using the two residual sum of squares as,

$$ F = \frac{\left(SS_{lof} / d_{lof} \right)}{\left(SS_{pe} / d_{pe} \right)} , \tag{27} $$

where $d_{lof} = N_s - N_\beta$, and $d_{pe} = N_s - M$, are the degrees of freedom associated with SS_{lof} and SS_{pe} , respectively. The lack-of-fit in the surrogate model is detected with α-level of significance, if the value of F in Equation (27), exceeds the tabulated $F_{\alpha, d_{lof}, d_{pe}}$ value, where the latter quantity is the upper 100α percentile of the central F-distribution.

When the data is obtained from the numerical simulations, the replication of simulations does not provide an estimate of noise (SS_{pe}), since all replications return exactly the same value. In such scenario, the variance of noise can be estimated by treating the observations at neighboring designs as "near" replicates ([33], pp. 123). We adopt the method proposed by Neill and Johnson [34], and Papila [35] to estimate the lack-of-fit for non-replicate simulation. In this method, we denote a near-replicate design point $\mathbf{x}_{jk}$ (as the kth replicate of the jth point $\overline{\mathbf{x}}_j$) such that

$$ \mathbf{x}_{jk} = \overline{\mathbf{x}}_j + \boldsymbol{\delta}_{jk} , \tag{28} $$

where $\boldsymbol{\delta}_{jk}$ represents the disturbance vector. Then, the Gramian matrix is written as

$$X = \bar{X} + \Delta, , \tag{29}$$

where $\bar{X}$ matrix is constructed using $\bar{\mathbf{x}}_j$ for near replicate points and matrix $\Delta = X - \bar{X}$. Now the estimated response at the design points (including near-replicates) is given as

$$\hat{\bar{\mathbf{y}}} = \mathbf{y} - \Delta \mathbf{b}, \tag{30}$$

where $\mathbf{b}$ is given in Equation (3). Now, we compute SS_{pe} and SS_{lof} by replacing $\mathbf{y}$, $\mathbf{y}_{n_j}$, and X in Equations (25) and (26) with $\hat{\bar{\mathbf{y}}}$, $\hat{\bar{\mathbf{y}}}_{n_j}$, and $\bar{X}$.

VII. References

1. Mack Y, Goel T, Shyy W, Haftka RT, Surrogate model based optimization framework: A case study in aerospace design. In: Yang S, Ong YS, Jin Y, editors. Evolutionary Computation in Dynamic and Uncertain Environments. Springer Kluwer Academic Press, 2005.

2. Mack Y, Shyy W, Haftka RT, Griffin L, Snellgrove L, Huber F, Radial turbine preliminary aerodynamic design optimization for expander cycle liquid rocket engine. In: 42nd AIAA/ASME/SAE/ASEE joint propulsion conference and exhibit, Sacramento CA, 9-12 July, 2006; AIAA-2006-5046.

3. Dorney DJ, Rothermel J, Griffin LW, Thornton RJ, Forbes JC, Skelley SE, Huber FW, Design and analysis of a turbopump for a conceptual expander cycle upper-stage engine. In: Symposium of advances in numerical modeling of aerodynamics and hydrodynamics in turbomachinery, Miami FL, 17-20 July, 2006; FEDSM 2006-98101.

4. Papila N, Shyy W, Griffin L, Dorney DJ, Shape optimization of supersonic turbines using global approximation methods. Journal of Propulsion and Power. 2002; 18(3): 509-518.

5. Obayashi S, Sasaki D, Takeguchi Y, Hirose N, Multi-objective evolutionary computation for supersonic wing-shape optimization. IEEE Transactions on Evolutionary Computation. 2000; 4(2): 182-187.

6. Sasaki D, Obayashi S, Sawada K, Himeno R, Multi-objective aerodynamic optimization of supersonic wings using Navier-Stokes equations. In: European congress on computational methods in applied sciences and engineering (ECCOMAS), Barcelona Spain, September, 2000.

7. Sasaki D, Morikawa M, Obayashi S, Nakahashi K, Aerodynamic shape optimization of supersonic wings by adaptive range multi-objective genetic algorithms. In: 1st international conference on evolutionary multi-criterion optimization, Zurich, March, 2001; pp 639-652.

8. Huyse L, Padula SL, Lewis RM, Li W, Probabilistic approach to free-form airfoil shape optimization under uncertainty. AIAA Journal, 2002; 40(9):.1764-1772.

9. Giannakoglou KC, Design of optimal aerodynamic shapes using stochastic optimization methods and computational intelligence. Progress in Aerospace Sciences, 2002; 38: 43-76.

10. Emmerich MJ, Giotis A, Ozdemir M, Back T, Giannakoglou K, Metamodel-assisted evolutionary strategies. In: Parallel problem solving from nature VII conference, Granada Spain, September, 2002; pp 361-370.

11. Mengistu T, Ghaly W, Mansour T, Global and local shape aerodynamic optimization of turbine blades. In: 11th multidisciplinary analysis and optimization conference, Portsmouth VA, 6-8 September, 2006; AIAA-2006-6933.

12. Samad A, Kim K-Y, Goel T, Haftka RT, Shyy W, Shape optimization of turbomachinery blade using multiple surrogate models. In: Symposium of advances in numerical modeling of aerodynamics and hydrodynamics in turbomachinery, Miami FL, 17-20 July, 2006; FEDSM 2006-98368.

13. Madsen JI, Shyy W, Haftka RT, Response surface techniques for diffuser shape optimization. AIAA Journal, 2000; 38(9): 1512-1518.

14. Vaidyanathan R, Papila N, Shyy W, Tucker KP, Griffin LW, Haftka RT, Fitz-Coy N, Neural network and response surface methodology for rocket engine component optimization. In: 8th AIAA/USAF/NASA/ISSMO symposium on multidisciplinary analysis and optimization conference, Long Beach CA, 6-8 September, 2000; AIAA-2000-4480.

15. Goel T, Vaidyanathan R, Haftka RT, Shyy W, Queipo NV, Tucker KP, Response surface approximation of Pareto optimal front in multi-objective optimization. Computer Methods in Applied Mechanics and Engineering, doi 10.1016/j.cma.2006.07.010.

16. Mack Y, Goel T, Shyy W, Haftka RT, Queipo NV, Multiple surrogates for shape optimization for bluff-body facilitated mixing. In: 43rd AIAA aerospace sciences meeting and exhibit, Reno NV, 10-13 Jan, 2005; AIAA-2005-0333.

17. Li W, Padula S, Approximation methods for conceptual design of complex systems. In: Chui C, Neaumtu M, Schumaker L, editors. Approximation Theory XI: Gatlinburg, Nashboro Press, Brentwood TN, 2005; pp 241-278.

18. Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Tucker KP, Surrogate-based analysis and optimization. Progress in Aerospace Sciences, 2005; 41: 1-28.

19. Goel T, Haftka RT, Shyy W, Queipo NV, Ensemble of surrogates. Structural and Multi-disciplinary Optimization Journal, doi 10.1007/s00158-006-0051-9 (a different version was presented at 11th AIAA/ISSMO multidisciplinary analysis and optimization conference, Portsmouth VA, 6-8 September, 2006; AIAA-2006-7047).

20. Goel T, Zhao J, Thakur SS, Haftka RT, Shyy W, Surrogate model-based strategy for cryogenic cavitation model validation and sensitivity evaluation. In: 42nd AIAA/ASME/SAE/ASEE joint propulsion conference and exhibit, Sacramento CA, 9-12 July, 2006; AIAA-2006-5047.

21. Dorney DJ, Personal communications. 2006.

22. Dorney DJ, Sondak DL, PHANTOM: Program user's manual. Version 20, April 2006.

23. Sondak DL, Dorney DJ, General equation set solver for compressible and incompressible turbomachinery flows. In: 39th AIAA/ASME/SAE/ASEE joint propulsion conference and exhibit, Huntsville AL, 20-23 July, 2003; AIAA-2003-4420.

24. Baldwin BS, Lomax H, Thin layer approximation and algebraic model for separated flow. In: 16th aerospace sciences meeting, Hunstville AL, 16-18 January, 1978; AIAA-1978-257.

25. Myers RH, Montgomery DC, Response Surface Methodology, John Wiley and Sons, Inc 1995.

26. Matheron G, Principles of geostatistics. Economic Geology, 1963; 58: 1246-1266.

27. Lophaven SN, Nielsen HB, Sondergaard J, DACE: A Matlab kriging toolbox. Version 2.0, Information and mathematical modeling, Technical University of Denmark, 2002.

28. Orr MJL, Introduction to radial basis function networks. Center for cognitive science, Edinburg University, EH 9LW, Scotland UK. 1996, http://www.anc.ed.ac.uk/~mjo/rbf.html.

29. Sobol IM, Sensitivity analysis for nonlinear mathematical models. Mathematical Modeling & Computational Experiment, 1993; 1(4): 407-414.

30. Martin JD, Simpson TW, Use of kriging models to approximate deterministic computer models. AIAA Journal, 2005; 43(4): 853-863.

31. Balabanov VO, Giunta AA, Golovidov O, Grossman B, Mason WH, Watson LT, Haftka RT, Reasonable design space approach for response surface approximation. Journal of Aircraft, 1999; 36(1): 308-315.

32. Stepanoff AJ, Centrifugal and Axial Flow Pumps, 2nd Edition, Krieger Publishing Company, Malabar FL, 1993.

33. Hart JD, Non-parametric Smoothing and Lack of Fit Tests, Springer-Verlag, NewYork, 1997.

34. Neill JW, Johnson DE, Testing linear regression function adequacy without replications. The Annals of Statistics, 1985; 13(4): 1482-1489.

35. Papila M, Accuracy of Response Surface Approximations for Weight Equations Based on Structural Optimization, PhD Thesis, The University of Florida, 2002.
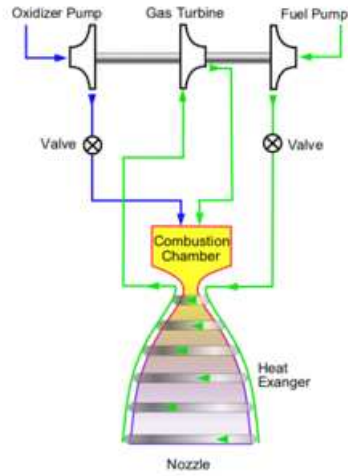
Figures

Figure 1 A representative expander cycle used in the upper stage engine (courtesy wikipedia).
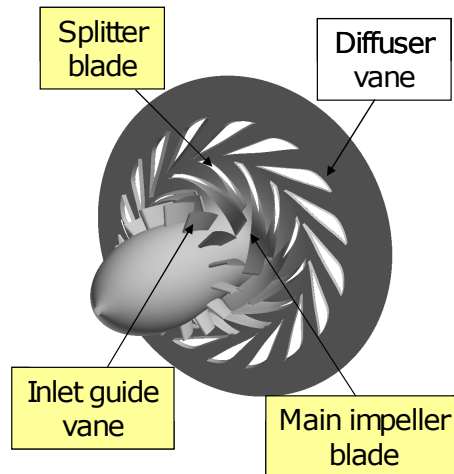

Figure 2 Schematic of a pump.
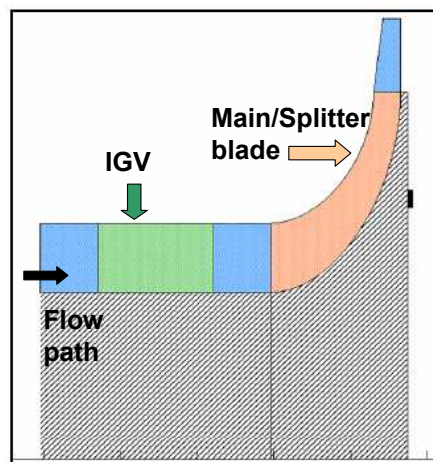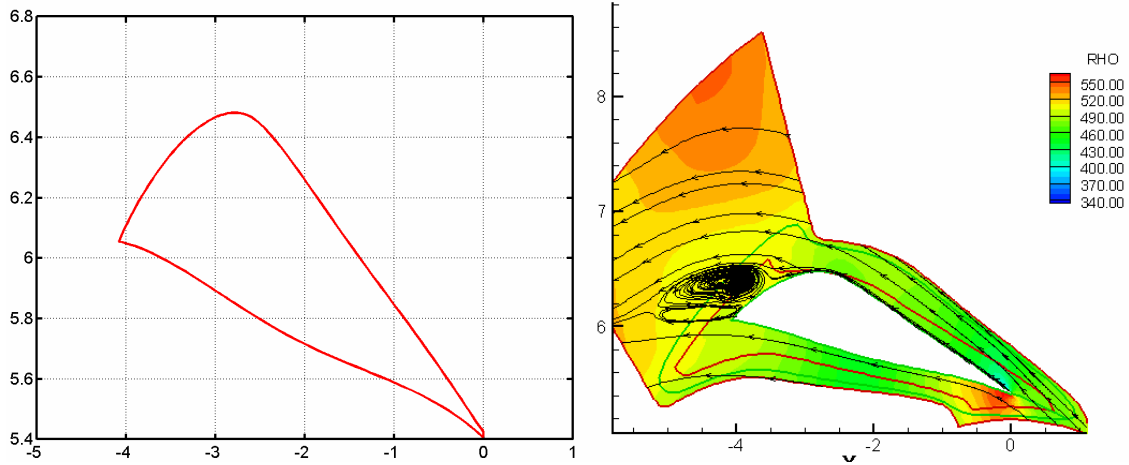

Figure 3 Meanline pump flow path (IGV: Inlet guide vane).

(A) Shape of the diffuser vane   (B) Streamlines and time-averaged pressure

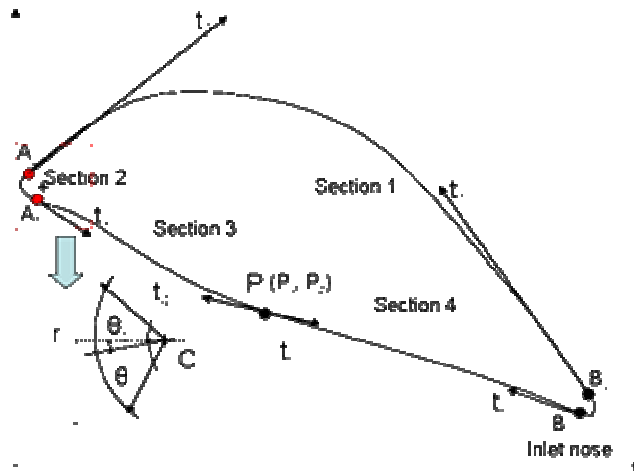**Figure 4 Baseline diffuser vane shape and time-averaged flow.**

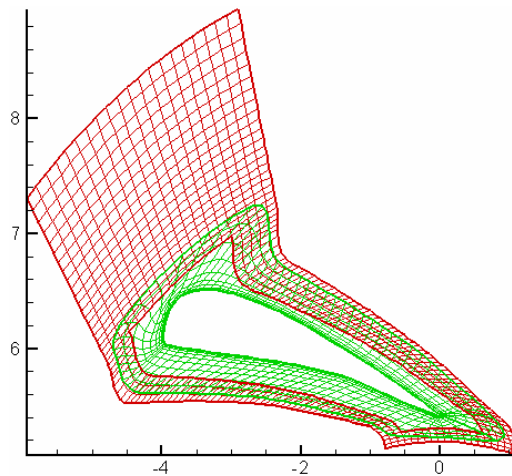**Figure 5 Definition of the geometry of the diffuser vane (refer to Table 1 for variable description).**

Figure 6 A combination of H- and O-grids to analyze diffuser vane. Body-fitted O-grids are shown in green and algebraic H-grid is shown in red.
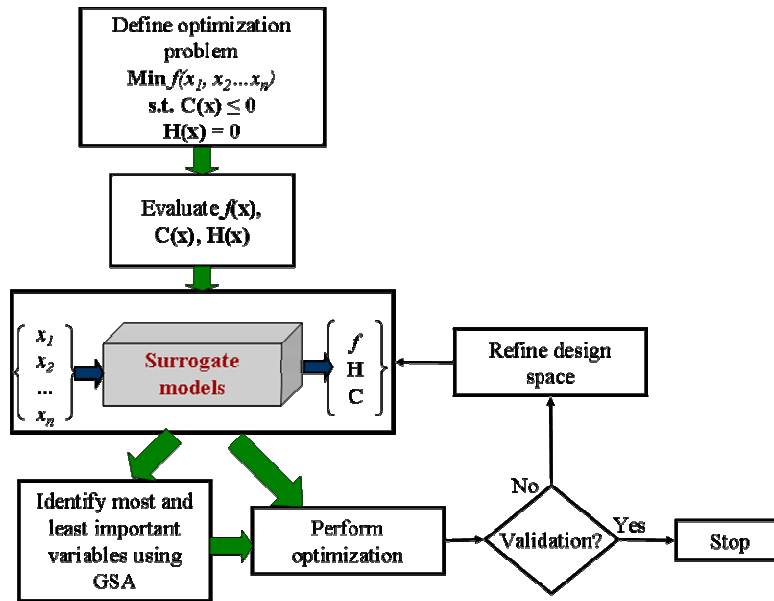
Figure 7 Surrogate based design and optimization procedure.
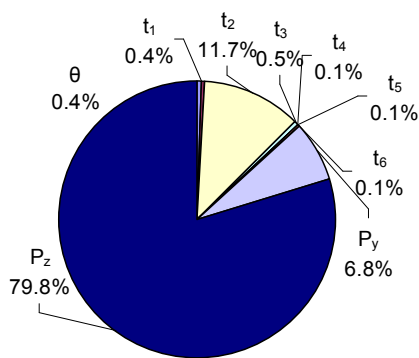
(A) Polynomial response surface
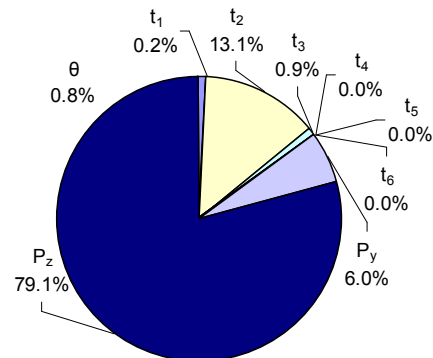
(B) Kriging
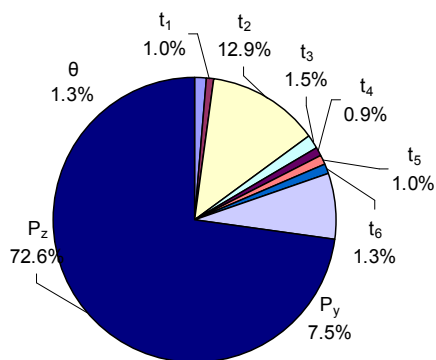
(C) Radial basis neural network
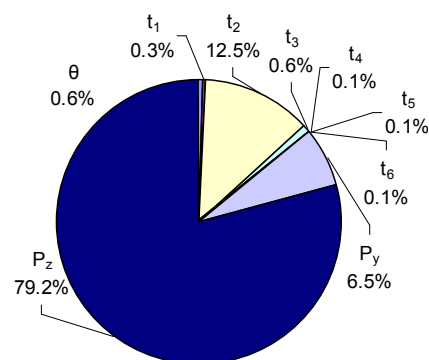
(D) PRESS-based weighted average

Figure 8 Sensitivity indices of main effect using various surrogate models (Set B).
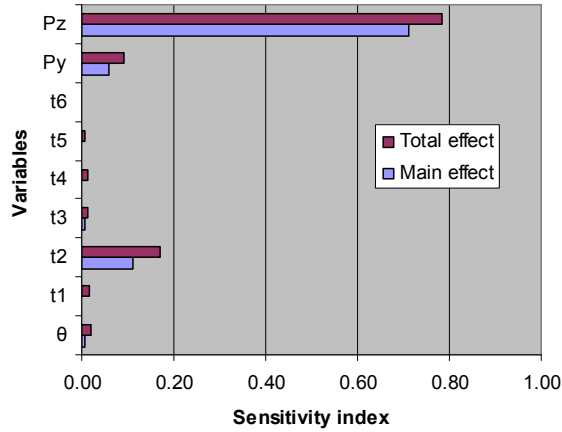
Figure 9 Sensitivity indices of main and total effects of different variables using PWS (Set B).
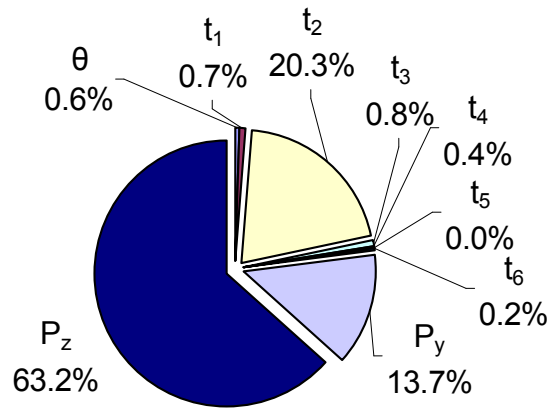
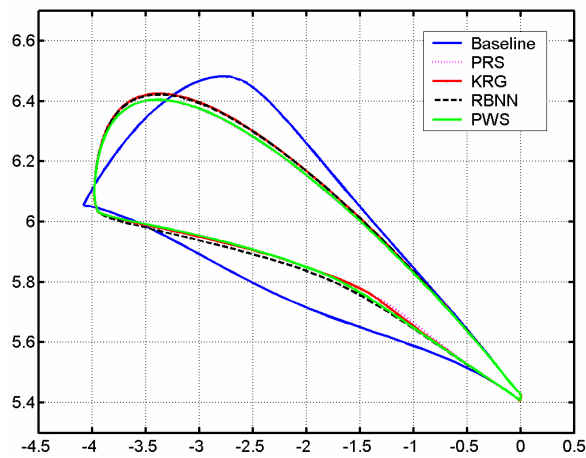**Figure 10 Actual partial variance of different design variables (no interactions are considered).**

Figure 11 Baseline and optimal diffuser vane shape obtained using different surrogate models. PRS indicates function evaluator is polynomial response surface, KRG stands for kriging, RBNN is radial basis neural network, and PWS is PRESS-based weighted surrogate model as function evaluator.
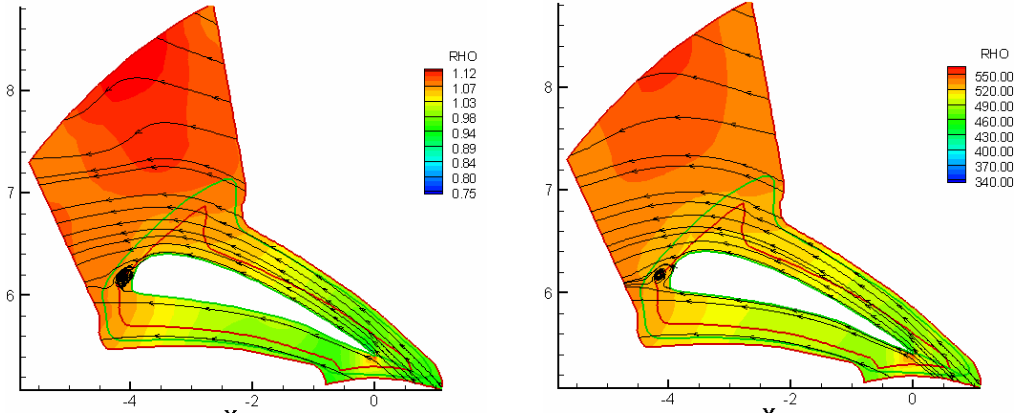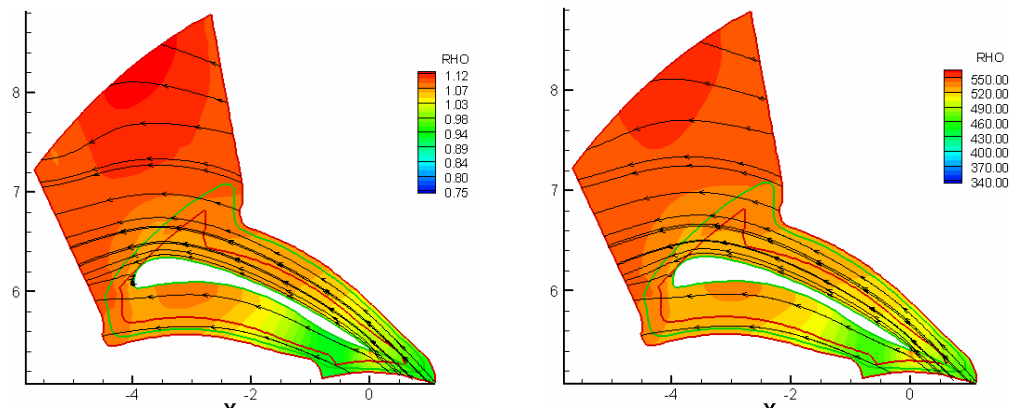
(A) Instantaneous – intermediate optimal design   (B) Time averaged – intermediate optimal design
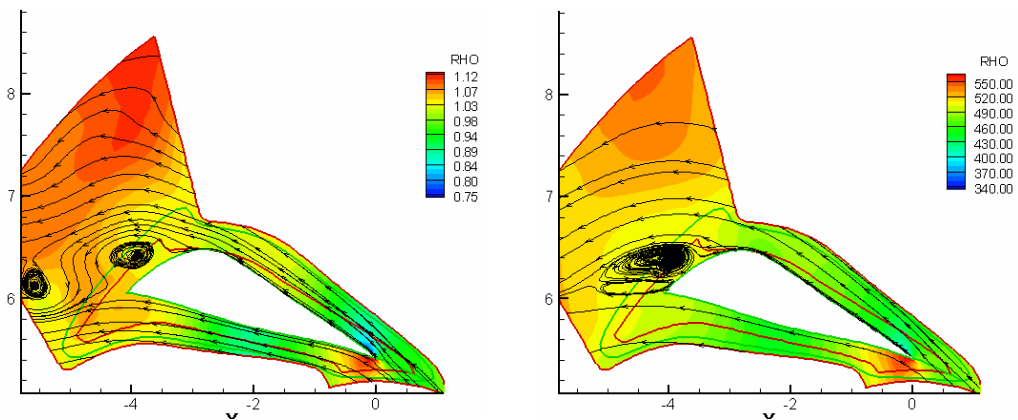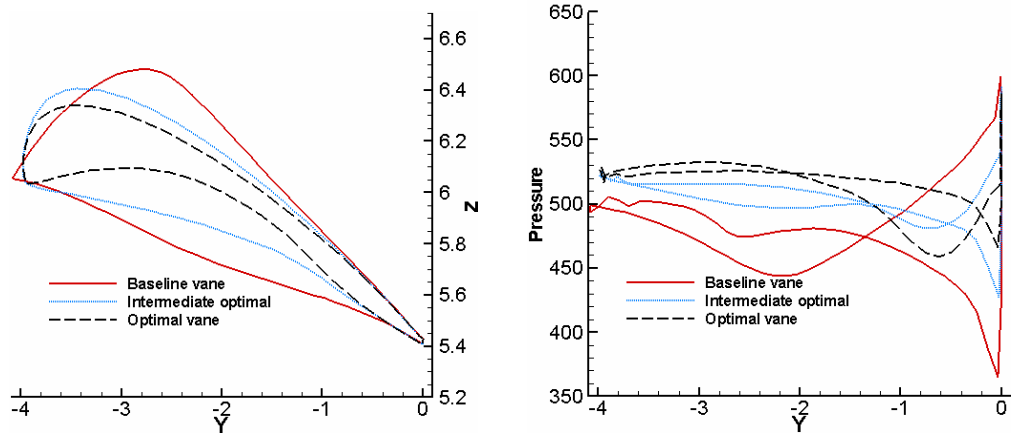

(C) Instantaneous – baseline design   (D) Time averaged – baseline design

Figure 12 Comparison of instantaneous and time-averaged flow fields of intermediate optimal (PRS) and baseline designs.

(A) Instantaneous pressure   (B) Time-averaged pressure

Figure 13 Instantaneous and time-averaged pressure for the final optimal diffuser vane shape.

(A) Different vane shapes   (B) Pressure loadings on the vanes

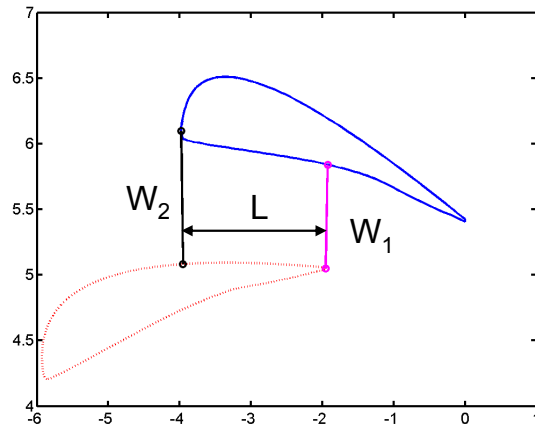**Figure 14 Different vane shapes and corresponding pressure loadings.**
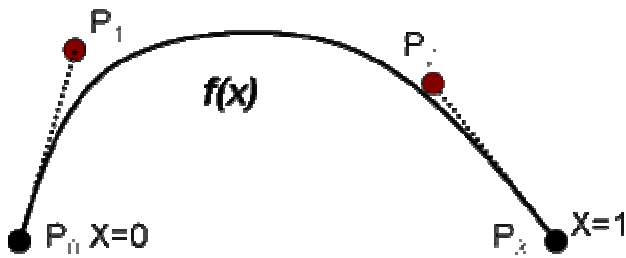

Figure 15 Gap between adjacent diffuser vanes.

Figure 16 Parametric Bezier curve.

Tables

Table 1 Design variables and corresponding ranges. Angle θ controls the size of circular arc, t_1-t_6 define the lengths of the tangents of Bezier curves and determine its curvature, and (Py, Pz) are the coordinates of the point P in the middle of the section. Angle θ is given in degrees and all other dimensions are scaled according to the baseline design.

	Minimum	Maximum
θ	60	110
t_1	1.25	2.50
t_2	0.75	1.50
t_3	0.05	0.30
t_4	0.50	1.00
t_5	0.50	1.00
t_6	1.00	2.00
P_y	-2.50	-2.00
P_z	5.70	5.85

Table 2 Summary of pressure ratio on data points and performance metrics for different surrogate models fitted to Set A. We tabulate the weights associated with the surrogates used to construct PRESS-based weighted surrogate (PWS). PRS: Polynomial response surface, KRG: Kriging, RBNN: Radial basis neural networks, RMSE: Root mean square error, PRESS: Predicted residual sum of squares. Here we give the square root of PRESS so as to facilitate each comparison with RMSE.

Surrogate	Parameter	Value	Weight
Pressure ratio	# of points	110	
	Minimum of data	1.001	
	Mean of data	1.041	
	Maximum of data	1.093	
PRS	R^2_{adj}	0.863	
	RMSE	8.00e-3	
	PRESS	1.33e-2	0.35
	Max error	2.62e-2	
	Mean absolute error	4.40e-3	
KRG	Process variance	1.09e-4	
	PRESS	1.26e-2	0.36
RBNN	PRESS	1.58e-2	
	Max error	2.03e-2	0.29
	Mean absolute error	3.33e-3	
PWS	PRESS	1.12e-2	
	Max error	1.39e-2	
	Mean absolute error	1.95e-3	

Table 3 Range of data, quality indicators for different surrogate models, and weights associated with the components of PWS. PRS: Polynomial response surface, KRG: Kriging, RBNN: Radial basis neural networks, PWS: PRESS-based weighted surrogate, RMSE: Root mean square error, PRESS: Predicted residual sum of squares (in PRS terminology), Here, we give the square root of PRESS so as to facilitate each comparison with RMSE. We used a reduced cubic and a reduced quadratic polynomial to approximate the Set B and Set C data, respectively.

Surrogate	Parameter	Set B Value	Set B Weights	Set C Value	Set C Weights
	# of points	439		109	
	Minimum of data	1.000		1.052	
	Mean of data	1.040		1.075	
	Maximum of data	1.097		1.105	
PRS	R^2_{adj}	0.959		0.978	
	RMSE	4.07e-3		1.65e-3	
	PRESS	4.84e-3	0.50	8.74e-3	0.36
	Max error	1.02e-2		3.84e-3	
	Mean absolute error	2.67e-3		1.05e-3	
KRG	Process variance	1.11e-4	0.27	1.05e-4	0.34
	PRESS	9.14e-3		9.47e-3	
RBNN	PRESS	1.17e-2		1.08e-2	
	Max error	1.53e-2	0.22	2.57e-2	0.30
	Mean absolute error	1.47e-3		2.89e-3	
PWS	PRESS	6.07e-3		7.53e-3	
	Max error	6.85e-3		7.91e-3	
	Mean absolute error	1.30e-3		1.04e-3	

Table 4 Optimal design variables and pressure ratio (P-ratio) obtained using different surrogates constructed using Set C data. PRS is polynomial response surface, KRG is Kriging, RBNN is radial basis neural network, and PWS is PRESS-based weighted surrogate.

Surrogate	θ	t_1	t_2	t_3	t_4	t_5	t_6	P_y	P_z	Predicted P-ratio
PRS	60.00	1.25	0.75	0.05	1.00	1.00	1.00	-2.00	5.85	1.120
KRG	63.58	1.25	0.80	0.05	0.81	0.88	1.43	-2.00	5.85	1.111
RBNN	64.49	1.32	0.78	0.07	0.94	0.54	1.08	-2.04	5.84	1.105
PWS	63.65	1.25	0.75	0.05	1.00	0.53	1.09	-2.00	5.85	1.109

Table 5 Comparison of actual and predicted performance of optimal designs obtained from multiple surrogate models (Set C). Each row shows the results of optimal design using a particular surrogate and different columns show the prediction of different surrogate models at each optimal design. PRS is polynomial response surface, KRG is kriging, RBNN is radial basis neural network, and PWS is PRESS-based weighted surrogate.

Design by	Actual P-ratio	Surrogate-prediction using PRS	KRG	RBNN	PWS
PRS	1.117	1.120	1.103	1.084	1.103
KRG	1.113	1.112	1.111	1.085	1.104
RBNN	1.106	1.106	1.104	1.105	1.105
PWS	1.114	1.116	1.108	1.102	1.109
Average absolute error		0.0014	0.0060	0.0186	0.0072

Table 6 Modified ranges of design variables, and fixed parameters in refined design space.

Variables	Min	Max	Fixed parameters	Value	Fixed parameters	Value
t_2	0.60	0.75	θ	60°	t_4	1.00
P_y	-2.00	-1.50	t_1	1.25	t_5	1.00
P_z	5.85	6.00	t_3	0.05	t_6	1.00

Table 7 Range of data, summary of performance indicators, and weights associated with different surrogate models in the refined design space. PRS: Polynomial response surface, KRG: Kriging, RBNN: Radial basis neural networks, PWS: PRESS-based weighted surrogate, RMSE: Root mean square error, PRESS: Predicted residual sum of squares (in PRS terminology). Here we give the square root of PRESS so as to facilitate each comparison with RMSE.

Surrogate	Parameter	Value	Weights
	# of points	20	
	Minimum of data	1.117	
	Mean of data	1.136	
	Maximum of data	1.151	
PRS	R^2_{adj}	0.956	0.59
	RMSE	1.75e-3	
	PRESS	2.74e-3	
	Max error	2.53e-3	
	Mean absolute error	1.11e-3	
KRG	Process variance	1.21e-4	0.25
	PRESS	7.14e-3	
RBNN	PRESS	1.11e-2	0.16
	Max error	<1.0e-6	
	Mean absolute error	<1.0e-6	
PWS	PRESS	4.21e-3	
	Max error	1.51e-3	
	Mean absolute error	6.62e-4	

Table 8 Design variables and pressure ratio at the optimal designs predicted by different surrogates. PRS is polynomial response surface, KRG is kriging, RBNN is radial basis neural network, and PWS is PRESS-based weighted surrogate.

	t_2	P_y	P_z	Actual P-ratio	Surrogate predictions using			
					PRS	KRG	RBNN	PWS
PRS	0.60	-1.99	6.00	1.150	1.151	1.151	1.155	1.152
KRG	0.60	-1.87	6.00	1.149	1.150	1.152	1.152	1.151
RBNN	0.61	-1.97	5.99	1.150	1.150	1.150	1.155	1.151
PWS	0.60	-1.96	6.00	1.150	1.151	1.152	1.154	1.152
Average absolute error					7.5E-04	1.50E-03	4.25E-03	1.75E-03

Table 9 Actual and empirical ratios of gaps between adjacent diffuser vanes.

	W_1	W_2	L	L/W_1	W_2/W_1	P-ratio
Empirical relations				4.00	1.60	
Baseline vane	0.66	0.94	2.09	3.15	1.42	1.074
Intermediate optimal	0.79	1.07	2.03	2.55	1.35	1.117
Optimal vane	0.94	1.11	2.02	2.15	1.19	1.150