JOURNAL OF AEROSPACE COMPUTING, INFORMATION, AND COMMUNICATION
Vol. 2, November 2005

Data Mining for Aerodynamic Design Space

Shinkyu Jeong*
Tohoku University, Sendai, 980-8577, Japan

Kazuhisa Chiba†
Japan Aerospace Exploration Agency, Tokyo, 181-0015, Japan

and
Shigeru Obayashi‡
Tohoku University, Sendai, 980-8577, Japan

Analysis of variance (ANOVA) and self-organizing map (SOM) were applied to data mining for aerodynamic design space. These methods make it possible to identify the effect of each design variable on objective functions. ANOVA shows the information quantitatively, while SOM shows it qualitatively. Furthermore, ANOVA can show the effects of interaction between design variables on objective functions and SOM can visualize the trade-offs among objective functions. This information will be helpful for designers to determine the final design from non-dominated solutions of multi-objective problems. These methods were applied to two design results: a fly-back booster in reusable launch vehicle design, which has 4 objective functions and 71 design variables, and a transonic airfoil design performed with the adaptive search region method.

Nomenclature

A	cross-sectional area of airfoil
c_i	best-matching unit
$E[I(\cdot)]$	expected improvement
$\mathbf{r}'$	vector of correlation values for Kriging model
$\mathbf{R}$	correlation matrix for Kriging model
$s^2(\cdot)$	mean squared error of the predictor
$\mathbf{x}$	vector denoting position in the design space
x	scalar component of $\mathbf{x}$
$\mathbf{y}$	vector of response data
β	constant global model of Kriging model
$\hat{\beta}$	estimated value of constant global model
ϕ	standard normal density
Φ	standard normal distribution

Received 25 April 2005; accepted for publication 5 October 2005.

* Research Associate, Institute of Fluid Science, 2-1-1 Katahira, Sendai, Japan, 980-8577, Member AIAA

† Research Associate, 7-44-1 Jindaiji-Higashi, Chofu, Japan, 182-8522, Member AIAA

‡ Professor, Institute of Fluid Science, 2-1-1 Katahira, Sendai, Japan, 980-8577, Associate Fellow AIAA

$\hat{\mu}_{total}$	total mean of model
$\boldsymbol{\theta}$	vector of correlation parameters for Kriging model
θ_k	k_{th} element of $\boldsymbol{\theta}$
$\hat{\sigma}^2$	estimated variance of Kriging model
$\hat{\sigma}^2_{total}$	total variance of model

I. Introduction

WITH recent advances in computing power and the development of high-fidelity Computational Fluid Dynamics (CFD) codes, design optimization with CFD has become an essential tool in the field of aircraft design. There are several approaches for design optimization using CFD: the Adjoint method, which is a gradient-based method, is very efficient for design problems having large numbers of design variables,[1–2] and Genetic Algorithms (GA), which are population-based methods, are very powerful for multi-objective design problems.[3–5]

However, research in optimization is mainly concerned with finding the optimal solution or the non-dominated (Pareto) solutions of multi-objective problems. In the engineering design field, it is also important to determine the final design from nominated solutions, such as non-dominated solutions of multi-objective problems. Thus, it is preferable for a designer to provide the non-dominated solutions with some useful information for the final design decision. Information about the design space, such as trade-off relations between objective functions and the relations between design variables and objective functions, are among the most useful information for the final design decision. Furthermore, this information will be useful to understand why the final design has good performance and makes it possible to simplify the design problem by eliminating the design variables that do not have a large influence on the objective functions.

The process to find the information from the design results is called ‘data mining.’ In this study, Analysis of Variance (ANOVA)[6–7] and Self-Organizing Map (SOM)[8–9] were used for data mining. The former uses the variance of the objective function due to design variables on response surface models. ANOVA can identify not only the effect of each design variable but also the effect of interactions between design variables on objective functions. ANOVA expresses the information in a quantitative way. On the other hand, SOM employs a nonlinear projection algorithm from high to low dimensions and a clustering technique. This method can visualize not only the relations between design variables and objective functions but also the trade-offs between objective functions. The method expresses the information in a qualitative way.

These techniques were applied to the results of two design problems:

1. The fly-back booster of a reusable launch vehicle design that treats 4 objective functions and 71 design variables; the results showed that the present data-mining techniques are useful to analyze problems with large numbers of design variables.
2. Transonic airfoil design using the adaptive search region method; the results indicated the influence of definition of the design space on the data mining results.

II. Analysis of Variance (ANOVA)

ANOVA uses the variance of the objective functions due to the design variables on the response surface models. The response surface model should be constructed for each objective function to calculate the variance. The response surface model used in the present study is the Kriging model.[7,10,11]

A. Kriging Model

The Kriging model, developed in the field of spatial statistics and geostatistics, predicts the value of the unknown point using stochastic processes. The Kriging predictor is expressed as follows:

$$\hat{y}(\mathbf{x}) = \hat{\beta} + \mathbf{r}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\beta}) \tag{1}$$

where $\mathbf{x} = \{x_1, x_2, \ldots, x_m\}$ is an m-dimensional vector of design variable, $\mathbf{y}$ is an n-dimensional column vector of sampled response data, $\mathbf{1}$ is an n-dimensional unit column vector, and $\mathbf{R}$ is the correlation matrix whose (i, j)

element is:

$$\mathrm{R}(\mathbf{x}^i,\ \mathbf{x}^j) = \exp\left[-\sum_{k=1}^{m}\theta_k\left|x_k^i - x_k^j\right|^2\right] \tag{2}$$

The correlation vector between $\mathbf{x}$ and the n sampled data is expressed as:

$$\mathbf{r}'(\mathbf{x}) = \left[\mathrm{R}(\mathbf{x},\mathbf{x}^1), \mathrm{R}(\mathbf{x},\mathbf{x}^2), \ldots\ldots, \mathrm{R}(\mathbf{x},\mathbf{x}^n)\right] \tag{3}$$

The $\hat{\beta}$ can be calculated using the following equation:

$$\hat{\beta} = \frac{\mathbf{1}'\mathbf{R}^{-1}\mathbf{y}}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}} \tag{4}$$

The unknown parameter, $\boldsymbol{\theta}$, for the Kriging model can be estimated by maximizing the following likelihood function:

$$Ln(\hat{\beta}, \hat{\sigma}^2, \boldsymbol{\theta}) = -\frac{n}{2}\ln(\hat{\sigma}^2) - \frac{1}{2}\ln(|\mathbf{R}|) \tag{5}$$

where $\hat{\sigma}^2$ can be calculated as follows:

$$\hat{\sigma}^2 = \frac{\left(\mathbf{y} - \mathbf{1}\hat{\beta}\right)'\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{1}\hat{\beta}\right)}{n} \tag{6}$$

Maximization of the likelihood function is an m-dimensional unconstrained non-linear optimization problem. A simple genetic algorithm is used to solve this problem.

The accuracy of the predicted value depends largely on the distance from the sample points. Intuitively, the closer point $\mathbf{x}$ is to the sample points, the more accurate the prediction, $\hat{y}(\mathbf{x})$, becomes. This is expressed in the following equation:

$$s^2(\mathbf{x}) = \hat{\sigma}^2\left[1 - \mathbf{r}'\mathbf{R}^{-1}\mathbf{r} + \frac{(1 - \mathbf{1}\mathbf{R}^{-1}\mathbf{r})^2}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}}\right] \tag{7}$$

where $s^2(\mathbf{x})$ is the mean squared error at point $\mathbf{x}$, indicating the uncertainty of the estimated value.

B. Decomposition of Variance

Once the response surface model is made, the effects of design variables on the objective function can be calculated by decomposing the total variance of the model into the variance component due to the design variable. Decomposition is done by integrating variables out of the model $\hat{y}$. The total mean ($\hat{\mu}_{total}$) and the variance ($\hat{\sigma}^2_{total}$) of the model are as follows:

$$\hat{\mu}_{total} \equiv \int \cdots\cdot \int \hat{y}(x_1, \ldots\ldots, x_m)dx_1 \cdots dx_m \tag{8}$$

$$\hat{\sigma}^2_{total} = \int \cdots\cdot \int \left[\hat{y}(x_1, \ldots\ldots, x_m) - \hat{\mu}_{total}\right]^2 dx_1 \cdots\cdot dx_m \tag{9}$$

The main effect of variable x_i and the two-way interaction effect of variables x_i and x_j are given as:

$$\hat{\mu}_i(x_i) \equiv \int \cdots \int \hat{y}(x_1, \ldots, x_m)dx_1 \cdot\cdot dx_{i-1}dx_{i+1} \cdot\cdot dx_m - \hat{\mu}_{total} \tag{10}$$

$$\hat{\mu}_{i,j}(x_i, x_j) \equiv \int \cdots\cdot \int \hat{y}(x_1, \ldots\ldots, x_m)dx_1 \cdot\cdot dx_{i-1}dx_{i+1} \cdot\cdot dx_{j-1}dx_{j+1} \cdot\cdot dx_m - \hat{\mu}_i(x_i) - \hat{\mu}_j(x_j) - \hat{\mu}_{total} \tag{11}$$

$\hat{\mu}_i(x_i)$ and $\hat{\mu}_{i,j}(x_i, x_j)$ quantify the effect of variable x_i and interaction effect of x_i and x_j on the objective function.

The variance due to the design variable x_i is given as:

$$\hat{\sigma}_i^2 = \int \left[\hat{\mu}_i\left(x_i\right)\right]^2 dx_i \tag{12}$$

The proportion of the variance due to design variable x_i to total variance of the model can be calculated by dividing Eq. (12) by Eq. (9):

$$\frac{\hat{\sigma}_i^2}{\hat{\sigma}_{total}^2} = \frac{\int \left[\hat{\mu}_i\left(x_i\right)\right]^2 dx_i}{\int \cdots \int \left[\hat{y}(x_1, \ldots\ldots, x_{\mathrm{m}}) - \hat{\mu}\right]^2 dx_1 \cdots dx_{\mathrm{m}}} \tag{13}$$

This value indicates the effect of design variable x_i on the objective function.

III. Self-Organizing Map (SOM)

A. General SOM algorithm

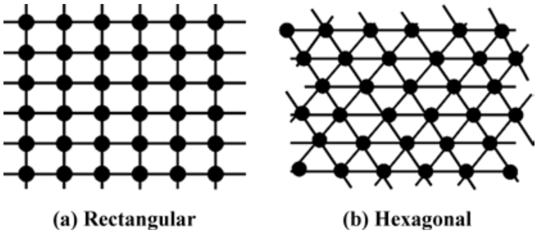
SOM is one of the unsupervised neural networks techniques that classify, organize, and visualize large data sets. SOM is a nonlinear projection algorithm[9] from high- to low-dimensional space. This projection is based on self-organization of a low-dimensional array of neurons. In the projection algorithm, the weights between the input vector and the array of neurons are adjusted to represent features of the high-dimensional data on the low-dimensional map. The closer two patterns are in the original space, the closer the response of two neighboring neurons in the low-dimensional space. Thus, SOM reduces the dimensions of input data while preserving their features.

A neuron used in SOM is associated with weight vector $\mathbf{w}_i = [w_{i1}, w_{i2}, \ldots\ldots, w_{im}](i = 1, \ldots, N)$, where m is the dimension of the input vector and N is the number of neurons. Each neuron is connected to its adjacent neurons by a neighborhood relation and usually forms a two-dimensional rectangular or hexagonal topology as shown in Fig. 1.

The learning algorithm of SOM begins with finding the best-matching unit ($\mathbf{w}_c$), which is closest to the input vector $\mathbf{x}$ as follows:

$$\|\mathbf{x} - \mathbf{w}_c\| = \min \|\mathbf{x} - \mathbf{w}_k\| \quad (k = 1, \ldots\ldots., N) \tag{14}$$

Once the best-matching unit is determined, the weight adjustments are performed not only for the best-matching unit but also for its neighbors to organize the topological mapping. The adjustment depends on the distance (similarity) between the input vector and the neuron. Based on the distance, the best-matching unit and its neighboring neurons become closer to the input vector as shown in Fig. 2. The topology before the adjustment is represented with the solid line and the weight vectors at that stage are represented by black dots. The best-matching unit is the weight vector closest to the input vector **x**. The best-matching unit and its neighbors are adjusted to be closer to the input vector **x**. The adjusted topology is represented by dashed lines and its weight vectors are represented as white dots. With repeated iterations of this learning algorithm, the weight vectors become smooth not only locally but also globally. Thus, the sequence of close vectors in the original space results in a sequence of corresponding neighboring neurons in the two-dimensional map.


(a) Rectangular

(b) Hexagonal

Fig. 1 Topology used in SOM.

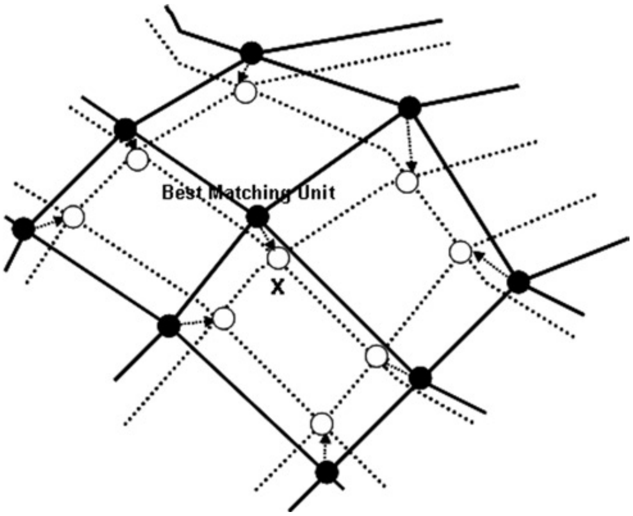


Fig. 2 Adjustment of best-matching unit and its neighbors.

B. Kohonen's Batch-SOM

In this investigation, SOM were generated using commercial software Viscovery® SOMine plus 4.0[12] produced by Eudaptics GmbH. Although SOMine is based on the general SOM concept and algorithm, it employs an advanced variant of unsupervised neural networks, *i.e.* Kohonen's Batch SOM.[8] The algorithm consists of two steps that are iterated until no more significant changes occur: search of the best-matching unit c_i for all input data $\{\mathbf{x}_i\}$ and adjustment of weight vector $\{\mathbf{w}_j\}$ near the best-matching unit. The Batch-SOM algorithm can be formulated as follows:

$$c_i = \arg\min_j \left\| \mathbf{x}_i - \mathbf{w}_j \right\| \tag{15}$$

$$\mathbf{w}_j^* = \frac{\sum_i h_{jc_i}\mathbf{x}_i}{\sum_{i'} h_{jc_i}} \tag{16}$$

where $\mathbf{w}_j^*$ is the adjusted weight vector. The neighborhood relationship between the neuron j and the best-matching unit c_i is defined by the following Gaussian-like function:

$$h_{jc_i} = \exp\left(-\frac{d_{jc_i}^2}{r_t^2}\right) \tag{17}$$

where d_{jc_i} denotes the Euclidean distance between the neuron k and the best-matching unit c_i on the map, and r_t denotes the neighborhood radius, which is decreased with the iteration steps t.

The standard Kohonen algorithm adjusts the weight vector after each record is read and matched. On the other hand, Batch-SOM takes a 'batch' of data (typically all records), and performs a 'collected' adjustment of the weight vectors after all records have been matched. This is much like 'epoch' learning in supervised neural networks. Batch-SOM is a more robust approach, as it mediates over a large number of learning steps. In SOMine, the uniqueness of the map is ensured by adoption of Batch-SOM and the linear initialization for input data. Much like some other SOM, SOMine creates a map in a two-dimensional hexagonal grid. Starting from numerical, multivariate data, the nodes on the grid gradually adapt to the intrinsic shape of the data distribution. As the position on the grid reflects the neighborhood within the data, features of the data distribution can be read from the emerging map on the grid. The trained SOM is systematically converted into visual information.

C. Cluster Analysis

Once the high-dimensional data is projected onto the two-dimensional regular grid, the map can be used for visualization and data mining. It is efficient to group all neurons by similarity to facilitate SOM for qualitative analysis, because the number of neurons on the SOM is large as a whole. This process of grouping is called 'clustering.'

Here, the hierarchical agglomerative algorithm was used for clustering. First, each node itself forms a single cluster and two clusters, which are adjacent in the map, are merged in each step. The distance between two clusters is calculated by using the SOM-ward distance[12]. The number of clusters is determined by the hierarchical sequence of clustering. Relatively small numbers of clusters are used for visualization, while large numbers are used for generation of weight vectors for the respective design variables.

IV. Results

A. Fly-Back Booster of Reusable Launch Vehicle (RLV) Design

The geometry of the fly-back booster[13] used in this design is shown in Fig. 3(a). In this design, the fuselage shape is fixed and only the wing shape design is considered, because the fuselage is filled with the liquid propellant rocket engines, so little change in its size is possible. The design variables used to define wing shape are related to planform, airfoil, wing twist, and position of the wing relative to the fuselage. A wing planform is determined by five design variables as shown in Fig. 3(b). Airfoil shapes are defined at wing root, kink, and tip based on thickness and camber distributions. Both distributions are parameterized using Bezier curves and linearly interpolated in the spanwise direction. Wing twist is refined using a B-spline curve with six control points. The position of the wing root relative to the fuselage is parameterized by *x* and *z* coordinates of the leading edge, angle of attack, and dihedral angle. A total of 71 design variables are used for definition of the wing geometry.

According to trajectory analysis,[14] separation of the booster and orbiter takes place around Mach 3 and the booster turns over, slows down, and cruises at transonic speed and lands at subsonic speed as shown in Fig. 4. To maintain good aerodynamic performance over a wide flight range, the following 4 objective functions are considered in this design. Subsonic, transonic, and supersonic flight conditions of the present design are shown in Table 1.

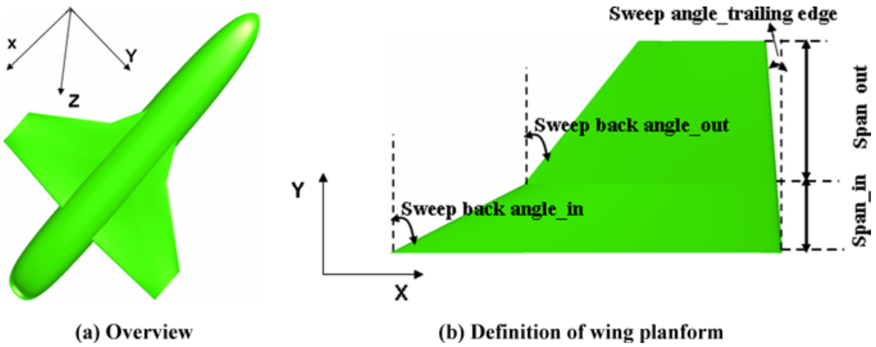
1. Minimization of the difference between supersonic pitching moment and transonic pitching moment.

$$F_1 = \left| C_M^{SUPERSONIC} - C_M^{TRANSONIC} \right| \tag{18}$$

A significant problem related to fly-back booster stability and control may result if the aerodynamic center experiences a large positional shift between supersonic and transonic flight conditions. Thus, it is desirable to design wing shapes with less positional variation in the aerodynamic center, as the vehicle decelerates from supersonic to transonic flight.

2. Minimization of the pitching moment at the transonic flight conditions

$$F_2 = \left| C_M^{TRANSONIC} \right| \tag{19}$$



(a) Overview

(b) Definition of wing planform

Fig. 3 Geometry of fly-back booster.

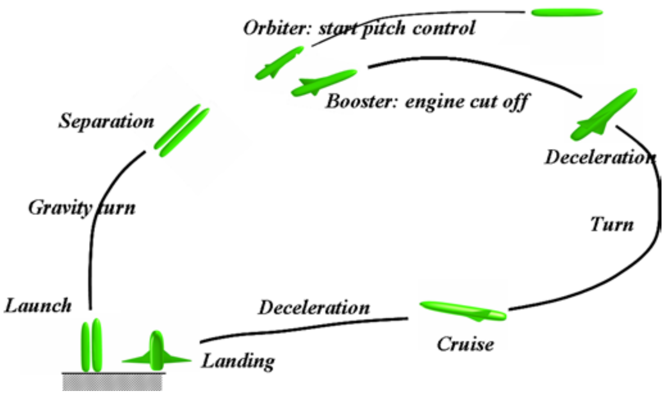


Fig. 4 Typical flight sequence for Two-Stage-to-Orbit fly-back booster.

It is known that the arrow wing ensures high aerodynamic performance, while it also produces a large pitching moment. Thus, it should be minimized under transonic flight conditions to reduce trim drag and improve flight stability.

3. Minimization of drag under transonic flight conditions

$$F_3 = \left|C_D^{TRANSONIC}\right| \tag{20}$$

Trajectory analysis shows that the range of the RLV booster is mostly covered by transonic flight. Thus, the transonic drag should be minimized to increase the flight range.

4. Maximization of lift under subsonic flight conditions

$$F_4 = C_L^{SUBSONIC} \tag{21}$$

To reduce the required runway distance, the lift obtained under subsonic flight conditions should be maximized.

As the optimizer, the Adaptive Range Multi-Objective Genetic Algorithm (ARMOGA)[15] is used. The population size of the present ARMOGA was 8 and 40 generations were performed. From the standpoint of optimization, these values are too small to guarantee the convergence of ARMOGA. However, these values may be sufficient for exploration of the design space. Figure 5 shows three-dimensional plots of 102 non-dominated solutions obtained by ARMOGA. However, it is difficult to understand the features of the design space from Fig. 5. To better understand the design space, ANOVA and SOM were performed with 102 non-dominated solutions.

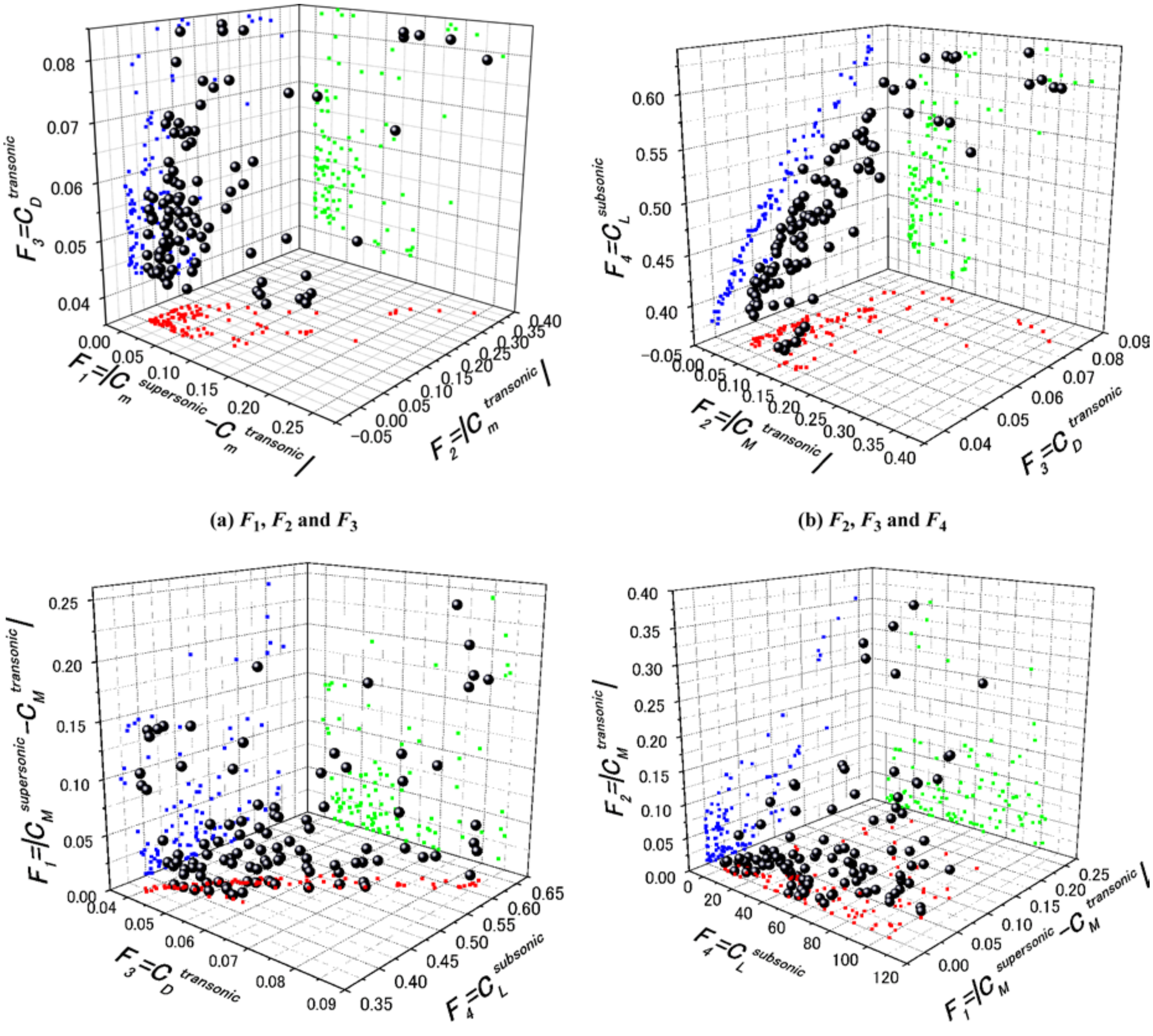
ANOVA

ANOVA was performed for 4 objective functions. Variance of design variables and their interactions, the proportions of which to the total variance were larger than 1.0%, are shown in Fig. 6. In the figure, 'dv' indicates design variables and '-' indicates interactions between two design variables.

According to the results, dv7(x coordinate of relative wing position to fuselage) gives the largest effect on the objective functions F_1 and F_2, and dv18 (rearward camber height at wing tip) gives the largest effect on F_3 and F_4. dv7 and dv18 are illustrated in Fig. 7. With regard to aerodynamics, if the wing position relative to the fuselage is changed, the aerodynamic center is also changed and this varies the pitching moment. In addition, it is well known

Table 1 Three flight conditions used in the design.

	Mach number	Angle of attack (°)	Reynolds number
Subsonic	0.3	0.0°	6×10^7
Transonic	0.8	8.0°	6×10^6
Supersonic	1.2	13.0°	6×10^6



(a) F_1, F_2 and F_3

(b) F_2, F_3 and F_4

(c) F_3, F_4 and F_1

(d) F_4, F_1 and F_2

Fig. 5 Non-dominated solutions projected onto three-dimensional objective function space.

that the shape of the camber line is strongly related to drag and lift performance. Thus, the findings from ANOVA correspond to general knowledge regarding aerodynamics.

SOM

SOM based on four objective functions is generated with 102 non-dominated solutions. Figure 8 shows SOM classified into 10 clusters based on the similarity with regard to four objective functions. The x and y coordinates of SOM have no meaning and clusters were arranged to be included in the rectangular frame. From this figure, we cannot extract any information about the design space. To obtain information about the design space, SOM was colored by each objective function value as shown in Fig. 9. SOM colored by F_1 and F_2 showed similar color patterns. Clusters with large F_1 values had large F_2 values, and clusters with small F_1 values had small F_2 values. Thus, F_1 and F_2 are not in a trade-off relation because both objective functions should be minimized. On the other hand, in the case of SOM colored by F_3 and F_4, although clusters with large F_3 values had large F_4 and those with small F_3 values had small F_4, F_3 and F_4 were in a severe trade-off relation because F_3 should be minimized but F_4 should be maximized.

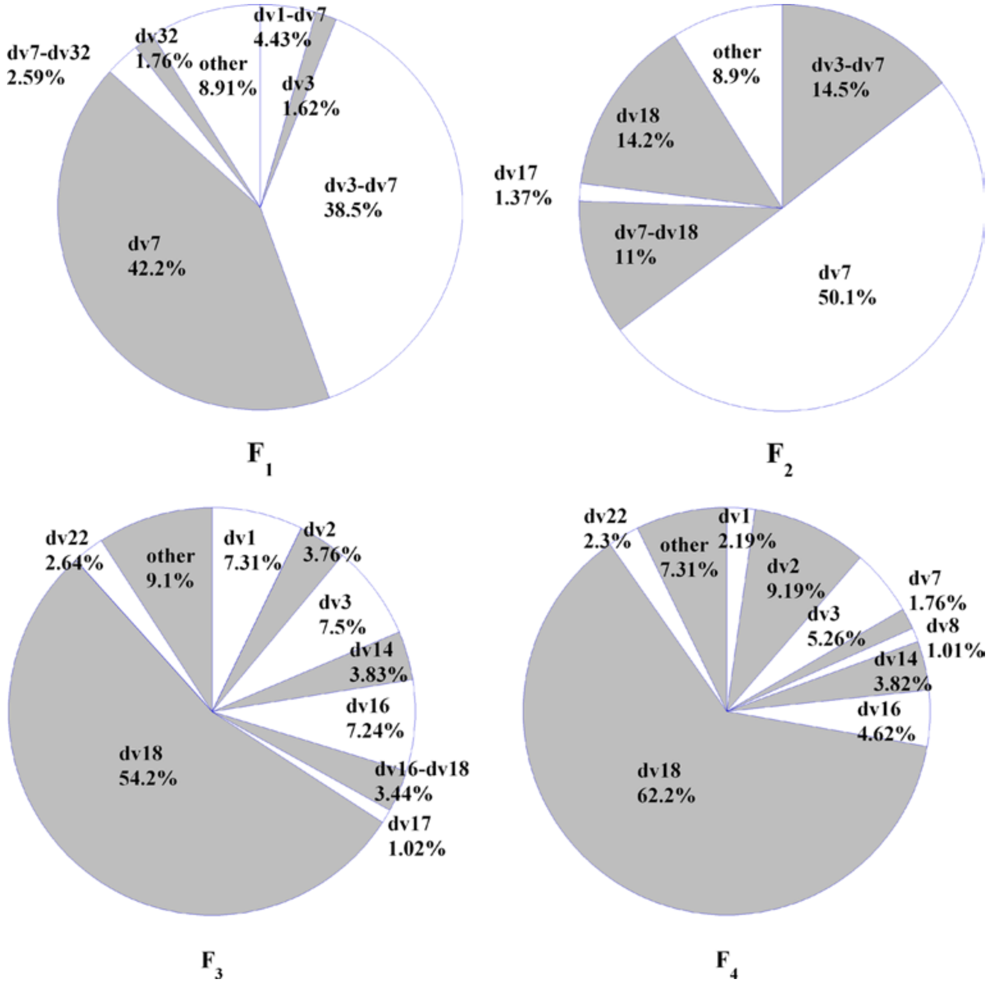
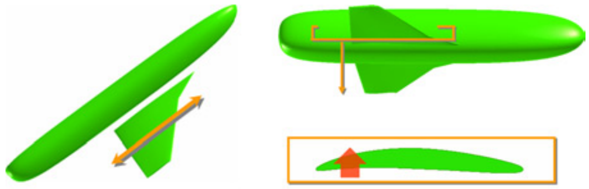


Fig. 6 ANOVA results.

Figure 10 shows SOM colored by three design variables (dv7, dv18, and dv15). In Fig. 10(a), the clusters with large dv7 values are situated in the lower left corner. In Figs. 9(a) and 9(b), these clusters also have large F_1 and F_2 values. Thus, large dv7 values are associated with poor performance of F_1 and F_2. In Fig. 10(b), the clusters with large dv18 values are located in the left-hand side. This color pattern is very similar to those of F_3 and F_4 shown in Figs. 9(c) and 9(d). This indicates that large dv18 values are related to large F_3 and F_4 values. These results indicate



(a) The x coordinate of the wing position relative to the fuselage (dv7)

(b) Rearward camber height at wing tip (dv18)

Fig. 7 Illustations of dv7 and dv18.

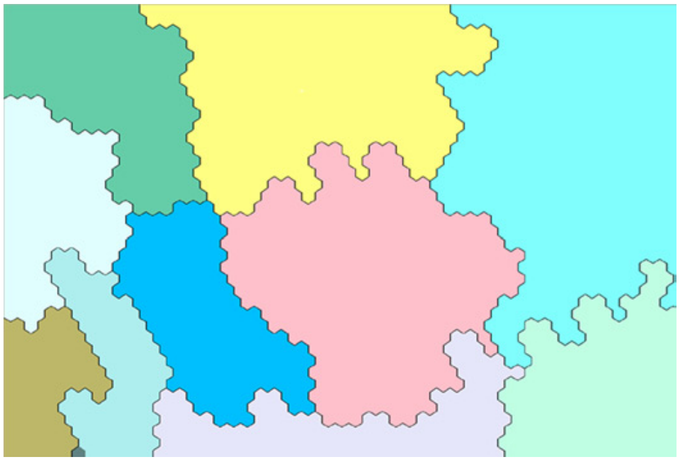


Fig. 8 SOM classified by 10 clusters with 102 non-dominated solution.

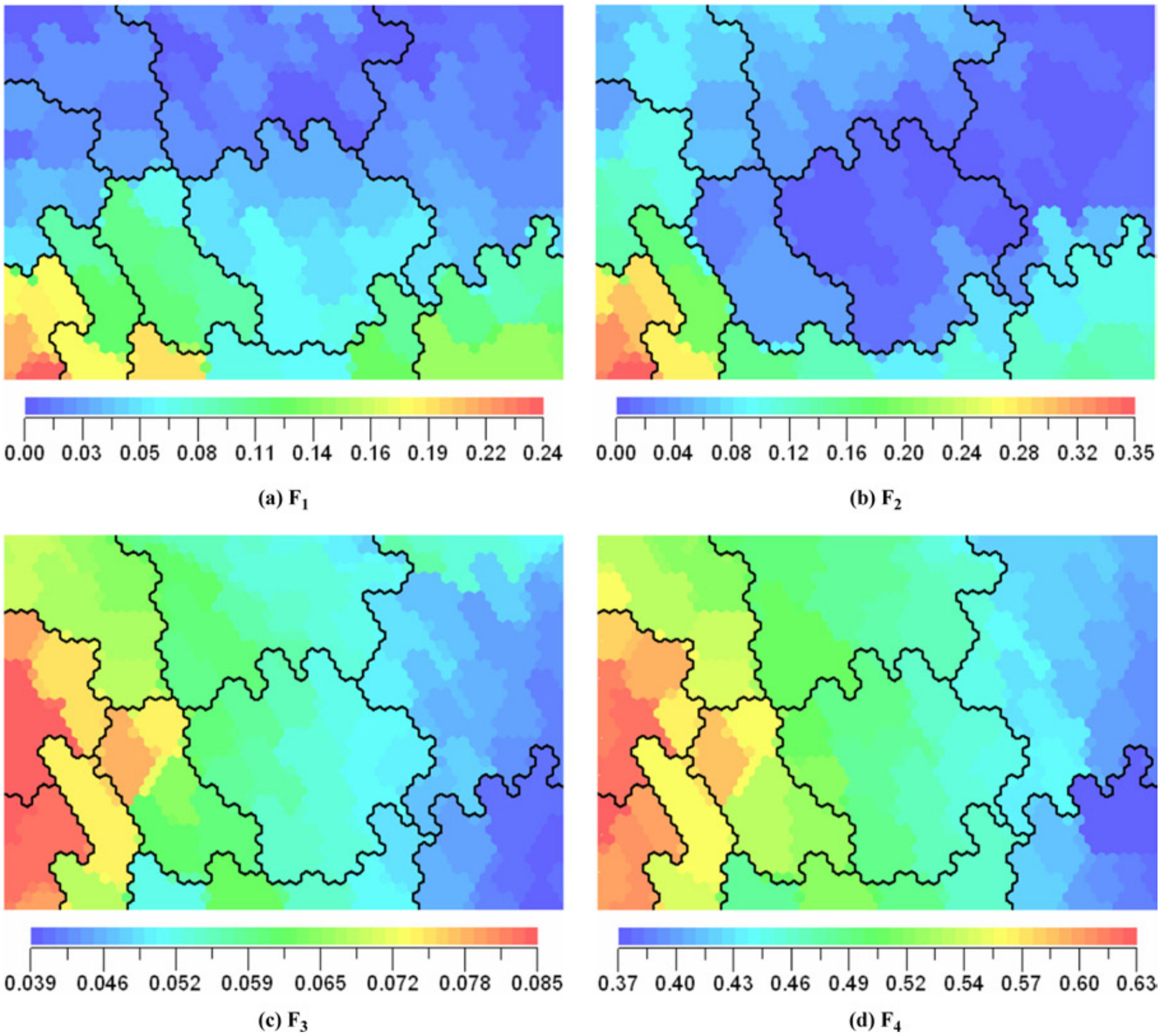
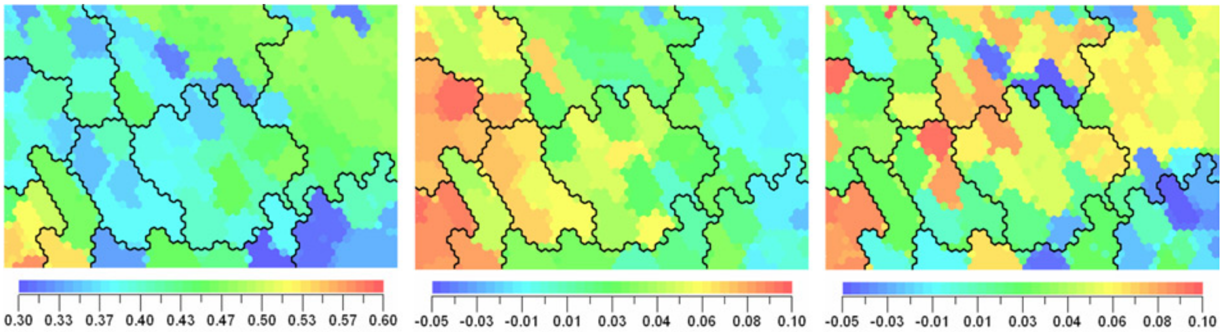


Fig. 9 SOM colored by objective functions.



(a) dv7 **(b) dv18** **(c) dv15**

Fig. 10 SOM colored by design variable.

that dv7 has a large effect on the objective functions F_1 and F_2, and dv18 has a large effect on the objective functions F_3 and F_4. In Fig. 10(c), SOM colored by dv15 (one of the x coordinates near the leading edge at the kink section) shows no noticeable trend of color distribution. This indicates that dv15 has little influence on the objective functions. These results agree with those of ANOVA.

V. Transonic Airfoil Design

In this design, the geometry of the airfoil is defined using PARSEC.[16] Figure 11 shows 11 basic parameters for the PARSEC airfoil. In this design, only a sharp trailing-edge airfoil was considered, and ΔZ_{TE} was set therefore to zero. A total of 10 design variables were used to define the geometry of the airfoil. The design problem is defined as follows:

$$\begin{aligned} &\text{Minimize} && \mathrm{C}_d \\ &\text{subject to} && \text{a) } \mathrm{C}_l \geq \mathrm{C}_{l\_RAE2822} \\ & && b)\ A \cong A_{RAE2822} \end{aligned}$$

under flow conditions of Mach = 0.73 and an angle of attack (AOA) = 2.7°. C_l and C_d are lift and drag coefficients of the designed airfoil, respectively, and $\mathrm{C}_{l\_RAE2822}$ is the lift coefficient of RAE2822 airfoil. A is the cross-sectional area of the airfoil.

In this design, the search region of the optimization problem was successively changed by investigating the probabilistic distribution of the design variables. That is, the adaptive search region method[17] was used. The design method consisted of two stages: i) the optimal search and ii) change of the search region. The optimal search procedure is shown in Fig. 12.

1. Kriging models are constructed for C_l and C_d with N initial sample points. In this study, the number of initial sample points was 50. These points were selected using Latin hypercube sampling[18] to spread the points uniformly in the search region.
2. GA operations
    - Generation of initial population and evaluation.

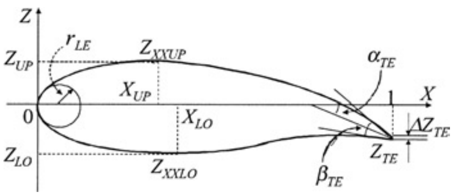


Fig. 11 PARSEC airfoil and its parameters.

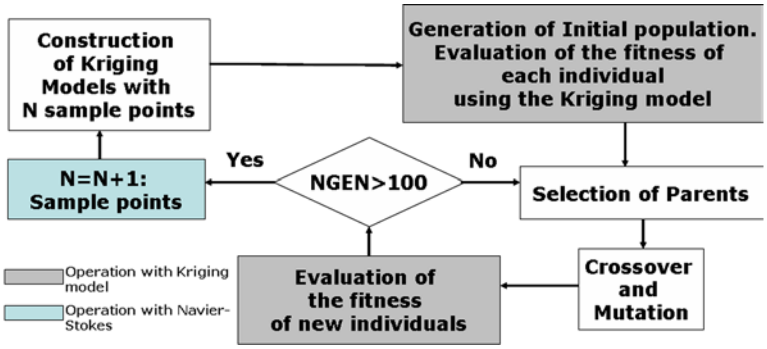

Construction of Kriging Models with N sample points
Generation of Initial population. Evaluation of the fitness of each individual using the Kriging model
N=N+1: Sample points
Yes
NGEN>100
No
Selection of Parents
Crossover and Mutation
Evaluation of the fitness of new individuals
Operation with Kriging model
Operation with Navier-Stokes

Fig. 12 Optimal search procedures.

- Selection of parents
- Crossover and mutation
- Evaluation of new individuals in the Kriging models

When the generation exceeds 100, the point with the largest probability of being superior to the current optimum is selected as an additional sample point. The probability of being superior to the current minimum can be calculated using the 'Expected Improvement'[7] criterion. In the minimization problem, EI is expressed as follows:

$$E\left[I\left(\mathbf{x}\right)\right] = E\left[\max(f_{\min} - \hat{y}, 0)\right] = (f_{\min} - \hat{y})\Phi\left(\frac{f_{\min} - \hat{y}}{s}\right) + s\phi\left(\frac{f_{\min} - \hat{y}}{s}\right) \qquad (22)$$

where Φ and φ are the standard distribution and normal density, respectively.

This routine is iterated until the termination criterion is reached. In the present study, termination the criterion was the maximum number of additional sample points.

Once optimization is over, the validity of the search region is examined. This procedure is shown in Fig. 13. First, the superior population (SP) is generated by GA. In 'SP,' its individuals satisfy all design constraints and the

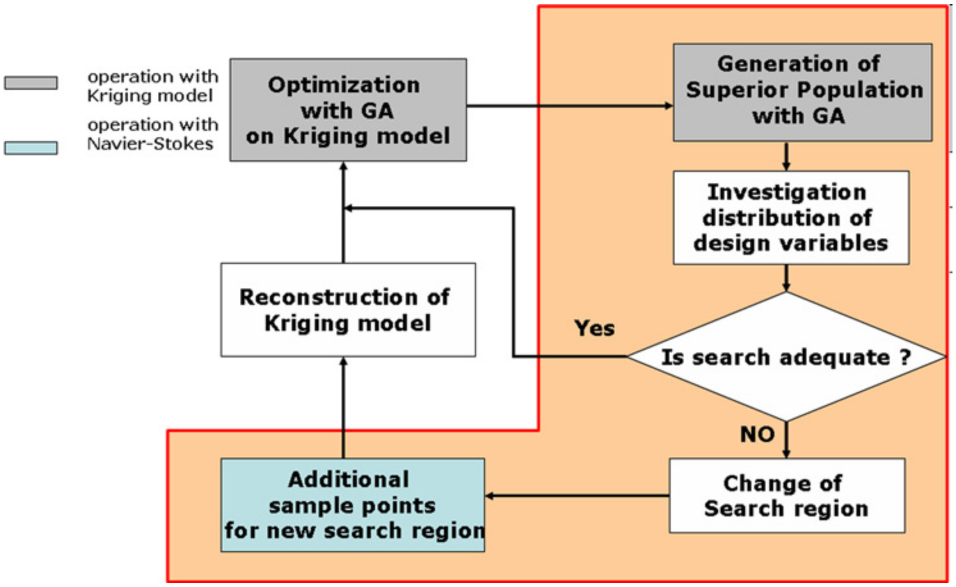

operation with Kriging model
operation with Navier-Stokes
Optimization with GA on Kriging model
Generation of Superior Population with GA
Investigation distribution of design variables
Reconstruction of Kriging model
Yes
Is search adequate ?
NO
Additional sample points for new search region
Change of Search region

Fig. 13 Change of search region.

objective function values of its individuals are larger than certain values. The distribution of design variables in the SP is investigated and the validity of the search region is checked. If the search region is invalid, the search region is redefined using the probabilistic method. A few additional sample points are required for the extended region of the redefined search region to ensure the accuracy of the Kriging models. This routine is iterated until no search region modification occurs.

The initial and final search regions are shown in Fig. 14. The final search region of all design variables, except Z_{TE} and Z_{XXUP}, expanded outside of the initial search region. This final search region was obtained after 3 search region redefinitions. For this design, data mining techniques are performed in both the initial and final search regions to examine the dependency on the search region.

ANOVA

First, ANOVA is performed in the initial search region with 56 sample data. First, 50 points are selected by Latin hypercube sampling and the last 6 data are selected by the optimal search algorithm shown in Fig. 12. The variance of design variables and the interactions in which the proportion to the total variance is larger than 2.0% are shown in Fig. 15. According to Fig. 15, Z_{UP} and Z_{LO} have comparatively large effects on C_l, and Z_{UP} has the largest effect on

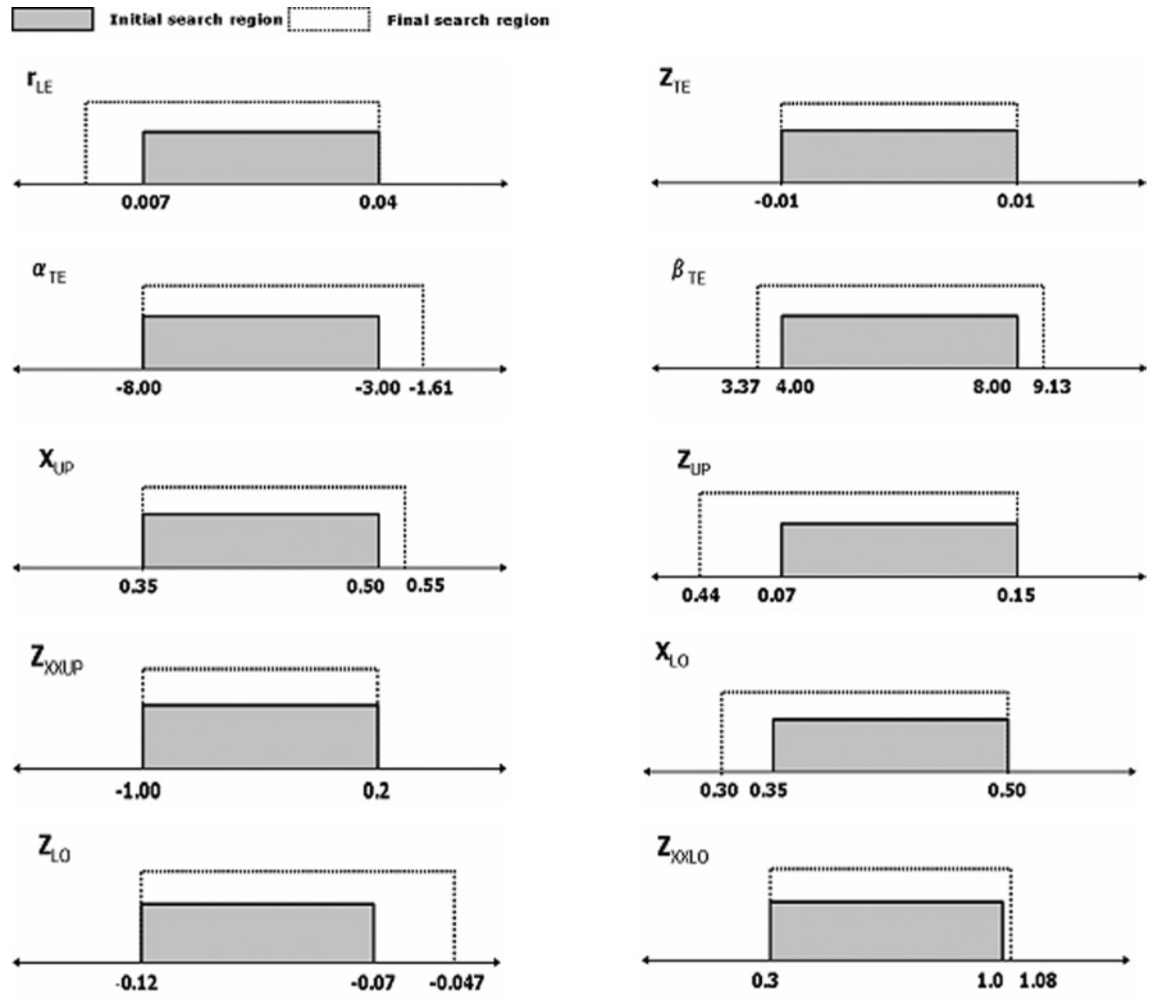


Fig. 14 Comparison of initial search region and final search region.

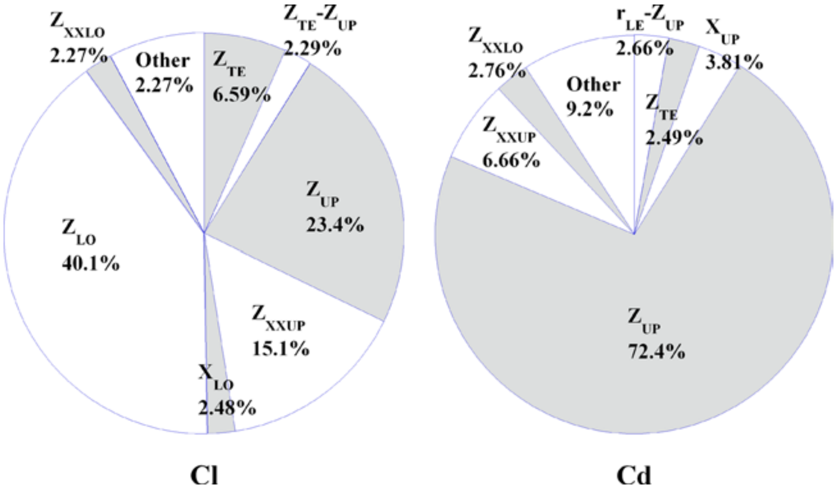


Fig. 15 ANOVA results in initial search region.

C_d. These parameters are related to the definition of airfoil thickness. These findings coincide with the aerodynamic knowledge.

ANOVA is also performed in the final search region with 110 sample data. The first 56 points are the same as those used in the initial search region and the rest of the points are selected during the optimization process using the adaptive search region method. According to Fig. 16, Z_{UP} and Z_{LO} have large effects on C_l, and Z_{UP} and Z_{UP}-Z_{LO} have comparatively large effects on C_d. The proportion of r_{LE}, which is small in the initial search region, is large in the final search region. This finding corresponds to the aerodynamic knowledge that the leading-edge radius is important for drag performance. This means that the effects of design variables on objective functions may vary according to the definition of the search region. Thus, elimination of a design variable, which initially has little influence on any of the objective functions before a validity check of the search region is performed, may lead to an undesirable design result.

SOM

SOM was also applied to the results of the transonic airfoil design. Figure 17 shows the SOM colored by C_l and C_d in the initial search region. The upper left corner and the lower right corner in Figs. 17(a) and 17(b) show similar color patterns. On the other hand, the upper right corner and the lower left corner show opposite color patterns. From

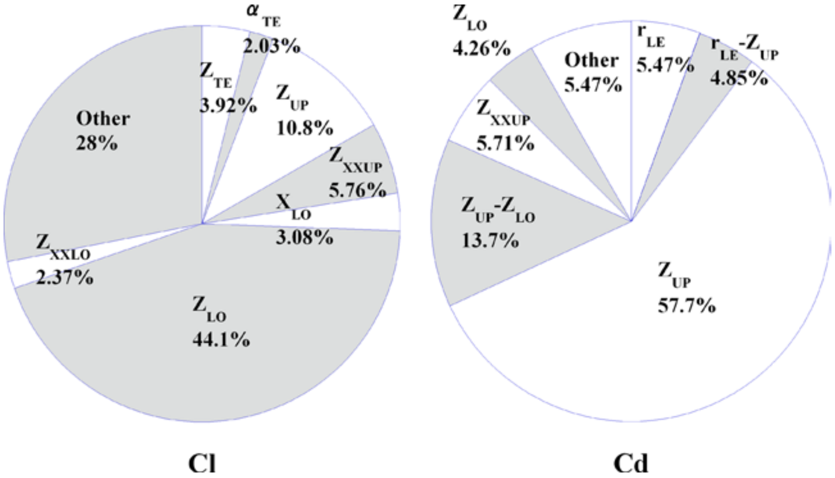
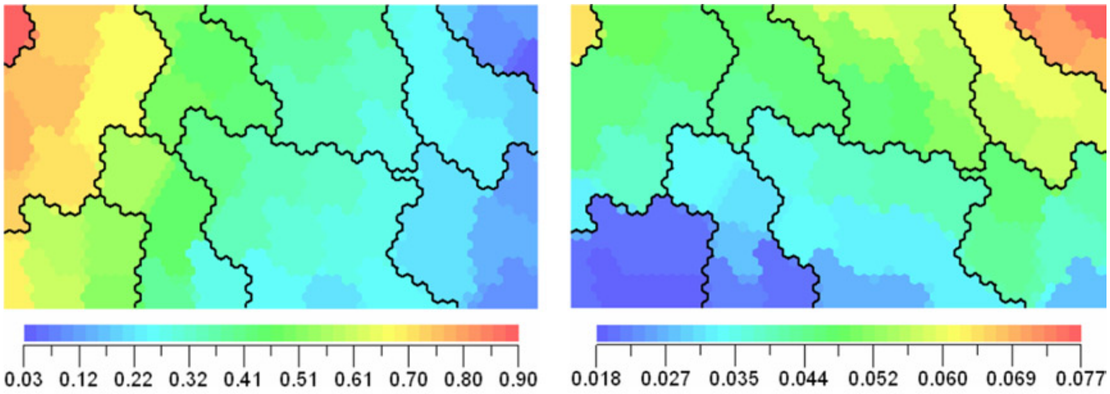


Fig. 16 ANOVA results in final search region.



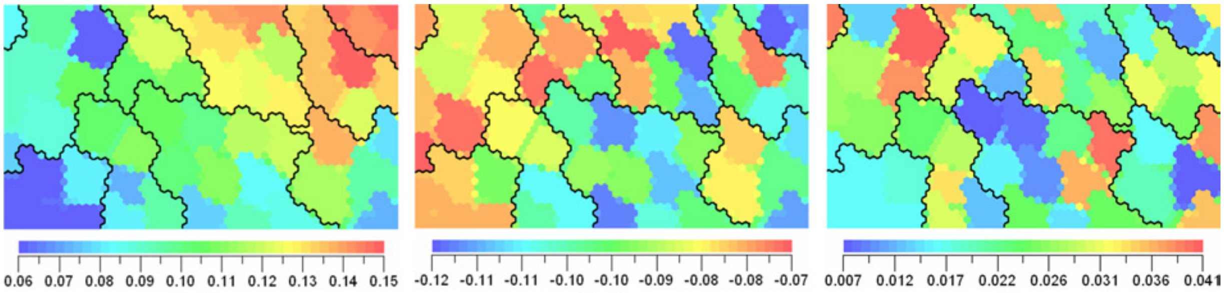
(a) Cl

(b) Cd

Fig. 17 SOM colored by objective functions in initial search region.

these observations, we can classify 56 data into 4 groups: i) both C_l and C_d are large (upper left corner), ii) both C_l and C_d are small (lower right corner), iii) C_l is large and C_d is small (lower left corner), and iv) C_l is small and C_d is large (upper right corner). Thus, Latin hypercube sampling used for the initial sample point selection successfully generated a wide variety of airfoils with various C_l and C_d performances.

Figure 18 shows the SOM colored by three design variables. In Fig. 18(a), the clusters with large Z_{UP} values are located in the upper right side and the clusters with small Z_{UP} values are located in the lower left side. The distribution

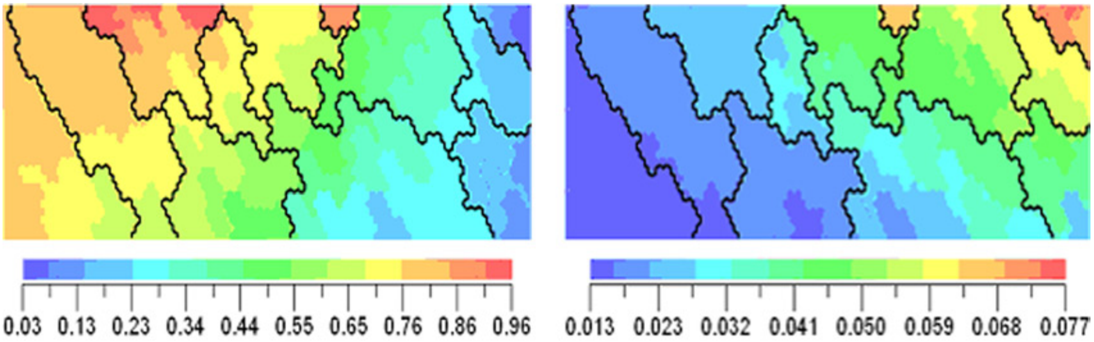


(a) Z_{UP}

(b) Z_{LO}

(c) r_{LE}

Fig. 18 SOM colored by design variables in the initial search region.



(a) C_l

(b) C_d

Fig. 19 SOM colored by objective function in final search region.

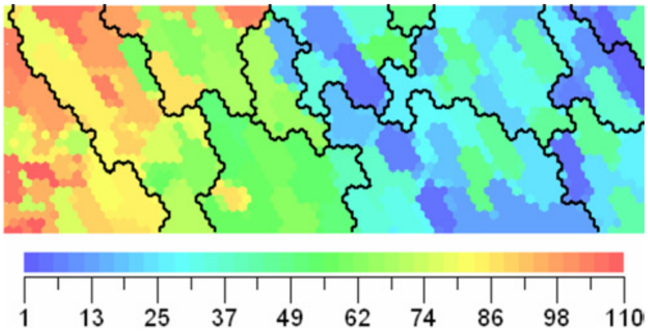
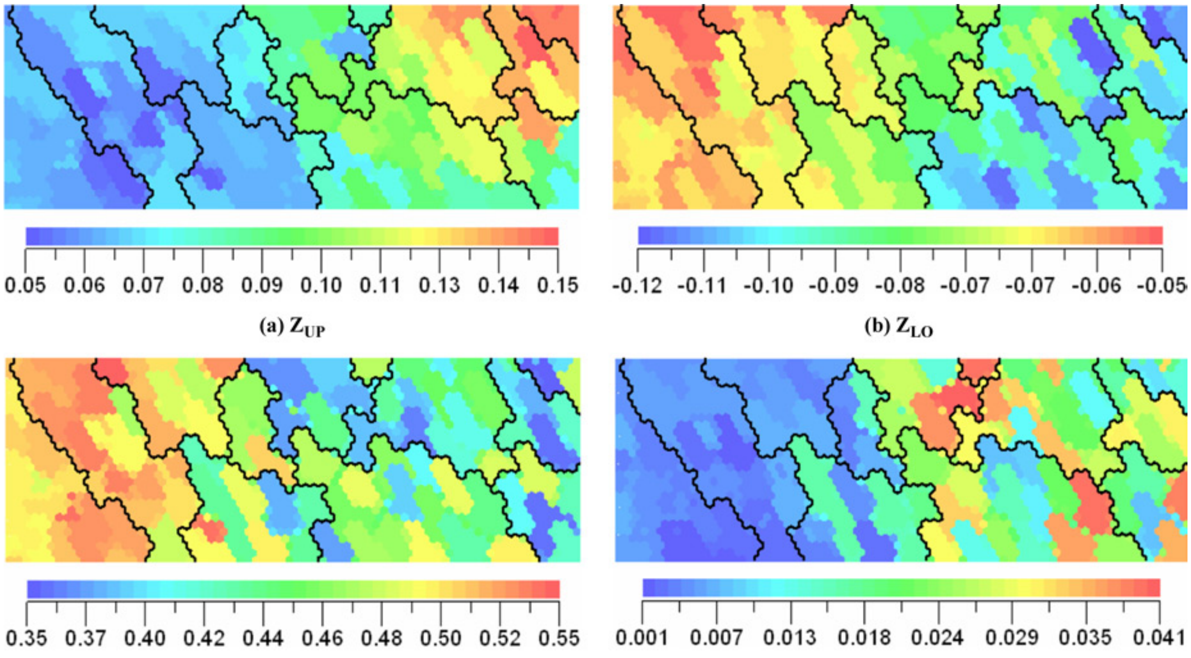


Fig. 20 SOM colored by generated order.

of color is similar to that of SOM colored by C_d. In Fig. 18(b), although the distribution of color is not as clear, it is similar to that of SOM colored by C_l. In Fig. 18(c), colored by r_{LE}, there is no noticeable color pattern. These results indicate that Z_{UP} and Z_{LO} have some effect on C_d and C_l, respectively, and r_{LE} has little effect on C_l and C_d. These results coincide with those of ANOVA applied for the same search region.

SOM was also applied to the final design space with 110 data. Figure 19 shows the SOM colored by C_l and C_d in the final search region. In Fig. 19(a), the clusters with large C_l values are located in the left-hand side and the clusters with small C_l values are located in the right-hand side. On the other hand, in Fig. 19(b), the clusters with large C_d values are located on the right side and those with small C_d values are located on the left side. From these SOM, we can classify all data into two groups: i) both C_l and C_d performances are good, and ii) both C_l and C_d performances are poor. Figure 20 shows the SOM colored by the generated order. The data, with generated order from 1st to 50th were generated by the Latin hypercube sampling, and those with generated order later than 50th were generated through the optimization process. The color distribution of this map is similar with that shown in Fig. 19(a). The



(a) Z_{UP}

(b) Z_{LO}

(c) X_{UP}

(d) r_{LE}

Fig. 21 SOM colored by design variables in final search region.

clusters with high generated order showed relatively good C_l and C_d performances. These observations indicated that the optimization method used in this investigation worked well.

Figure 21 shows the SOM colored by four design variables. The color pattern of Fig. 21(a) is similar to that of Fig. 19(b), indicating that small Z_{UP} values are related to good C_d performance of the airfoil. The color distributions of Figs. 21(b) and 21(c) are similar to those of Fig. 19(a), indicating that high values of Z_{LO} and X_{UP} are associated with good C_l performance. In Fig. 21(d), clusters with small r_{LE} values are located on the left-hand side, similar to Fig. 19(b). A small r_{LE} value is related to good C_d performance. These results coincide with those of ANOVA in the final search region.

VI. Conclusions

In the present study, two data mining techniques—analysis of variance (ANOVA) and self-organizing map (SOM)—were applied to optimization design results. ANOVA shows the effect of each design variable on objective functions quantitatively and SOM shows the information qualitatively. Furthermore, ANOVA can show the effect of interactions of design variables on objective functions. On the other hand, SOM can show the trade-offs between objective functions. The acquired information helps the designer to determine the final design from the non-domination solutions of multi-objective problems and can also be used to identify why the obtained optimum solution has good performance. Furthermore, the information will make it possible to simplify the design space by eliminating design variables that have little effect on the design problem.

However, as in the transonic airfoil design case using the adaptive search region method, the acquired information was dependent on the definition of the search region. To obtain the correct information, the data for data mining should be selected carefully.

Further studies are required to investigate methods of selecting the data for data mining. The results of most data mining techniques are largely dependent on the data used. Robust data selection methods are necessary to ensure the consistency of information obtained from data mining.

References

[1] Reuther, J., and Jameson, A. ,“Aerodynamic Shape Optimization of Wing and Wing-Body Configuration Using Control Theory,” AIAA Paper 95-0123.

[2] Kim, H., Sasaki, D., Obayashi, S., and Nakahashi, K, “Aerodynamic Optimization of Supersonic Transport Wing Using Unstructured Adjoint Method,” *AIAA Journal*, Vol. 39, No. 6, June 1991, pp. 1011–1020.

[3] Coello, C. A. C., and Lamont, G. B., Application of Multi-Objective Evolutionary Algorithms–Advance in Natural Computation, Vol. 1, World Scientific, 2004.

[4] Deb, K., Multi-Objective Optimization Using Evolutionary Algorithms, Wiley, 2001.

[5] Oyama, A., Obayashi, S., and Nakahashi, K. “Euler/Navier-Stokes Optimization of Supersonic Wing Design Based on Evolutionary Algorithms,” *AIAA Journal*, Vol. 37, No. 10, Aug. 1999, pp. 1327–1328.

[6] Efron, B., and Stein, C., “The jackknife estimate of variance,” The annals of Statistics, Vol. 9, No. 3, 1981, pp. 586–596.

[7] Donald, R. J., Matthias, S., and William J. W., “Efficient Global Optimization of Expensive Black-Box Function,” *Journal of global optimization*, Vol. 13, 1998, pp. 455–492.

[8] Kohonen, T., *Self-Organizing Maps*, Springer, Berlin, Heidelberg, 1995.

[9] Krzysztof, J. C., Witold, P., and Roman, W. S., Data Mining Methods for Knowledge Discovery, Kluwer Academic Publisher, 1998.

[10] Sack, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., “Design and analysis of computer experiments (with discussion),” Statistical Science 4, 1989, pp. 409–435.

[11] Jeong, S., Murayama, M., and Yamamoto, K., “Efficient Optimization Design Method Using Kriging Model,” *Journal of Aircraft*, Vol. 42, 2005, pp. 413–420.

[12] Eudaptics software GmbH, http://www.eudaptics.com/somine/index.php?sprache=en, last access on April 14, 2005.

[13] Chiba, K., Obayashi, S., and Nakahashi K., “Tradeoff Analysis of Aerodynamic Wing Design for RLV,” Proceedings of International Conference Parallel Computational Fluid Dynamics 2004, Spain, 2004. (To be published)

[14] Iwata, T., Sawada, K., and Kamijo, K., “Conceptual Study of Rocket Powered TSTO with Fly-back Booster,” AIAA Paper 2003-4813.

[15] Sasaki, D., and Obayashi, S., “Efficient Search for Trade-Offs by Adaptive Range Multi-Objective Genetic Algorithms,” Journal of Aerospace Computing, Information, In addition, Communication, Vol. 2, 2005, pp. 44–64.

[16]Sobieczky, H., “Parametric Airfoils and Wings,” *Recent Development of Aerodynamic Design Methodologies -Inverse Design and Optimization-*, edited by Fuji, K. and Dulikaravich, G. S., Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig/Wiesbaden, 1999, pp. 71–87.

[17]Jeong, S., Yamamoto, K., and Obayashi S., “Kriging-Based Probabilistic Method for Constrained Multi-Objective Optimization Problem,” AIAA Paper 2004-6437.

[18]Mckay, M. D., Beckman, R. J., and Conover, W. J., “A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code,” *Technometric*, Vol. 21, No. 2, 1979, pp. 239–245.