Data Mining for Multidisciplinary Design Space of Regional-Jet Wing

Kazuhisa Chiba
Institute of Space Technology and Aeronautics
Japan Aerospace Exploration Agency
E-mail: chiba@chofu.jaxa.jp

Shinkyu Jeong, Shigeru Obayashi
Institute of Fluid Science
Tohoku University
E-mail: jeong@edge.ifs.tohoku.ac.jp, obayashi@ieee.org

Hiroyuki Morino
Nagoya Aerospace Systems
Mitsubishi Heavy Industries, Ltd.

***Abstract*— The Data Mining technique is an important facet of solving multi-objective optimization problem. Because it is one of the effective manner to discover the design knowledge in the multi-objective optimization problem which obtains large data. In the present study, two Data Mining techniques have been performed for a large-scale, real-world Multidisciplinary Design Optimization (MDO) to provide knowledge regarding the design space. The MDO among aerodynamics, structures, and aeroelasticity of the regional-jet wing was carried out using high-fidelity evaluation models on Adaptive Range Multi-Objective Genetic Algorithm. As a result, nine non-dominated solutions were generated and used for tradeoff analysis among three objectives. All solutions evaluated during the evolution were analyzed for the influence of design variables using a Self-Organizing Map (SOM) and a functional Analysis of Variance (ANOVA) to extract key features of the design space. SOM and ANOVA compensated with the respective disadvantages, then the design knowledge could be obtained more clearly by the combination between them. Although the MDO results showed the inverted gull-wings as non-dominated solutions, one of the key features found by Data Mining was the non-gull wing geometry. When this knowledge was applied to one optimum solution, the resulting design was found to have better performance compared with the original geometry designed in the conventional manner.**

I. Introduction

Recently, the design optimization using high-fidelity evaluation models becomes one of the essential tools for aircraft design. Optimization problems are concentrated only on finding the optimal solution. Multi-objective optimization obtains only non-dominated solutions. However, it is essential for designers to find the information regarding the design space, such as relations between design variables and objective functions. The design information directly helps the designer to determine the next geometry. The process to find design information from huge database, for example optimization results, is called Data Mining. This technique is an important facet of solving optimization problem and it has a role of post-process for optimization problem[1].

SOM suggested by Kohonen[2] is one of neural network models. SOM can serve as a cluster analyzing tool for high-dimensional data. The cluster analysis of the objective function values will help to identify design tradeoffs and influence of design variables. Furthermore, ANOVA[3], which is one of the approximation models, presents the correlations between objective functions and design variables. Effective design variables can be identified quantitatively for objective functions and other characteristic functions. In this study, the above two Data Mining techniques are applied to a large-scale, real-world MDO problem for regional jet aircraft[4] in the field of aeronautics and space engineering, and then knowledge in the multidisciplinary design space is acquired. SOM and ANOVA have the disadvantages, respectively. It redeems each disadvantage to employ these two techniques simultaneously, and then Data Mining is carried out effectively.

II. MDO Problem

A. Objective Functions

In this system, minimization of the block fuel at a required target range derived from aerodynamics and structures was selected as an objective function. In addition, two more objective functions were considered — minimization of the maximum takeoff weight and minimization of the difference in the drag coefficient between two Mach numbers, which are cruise Mach and target Maximum Operating Mach number (MMO), to prevent decrease MMO.

B. Geometry Definition

First, the planform was given by Mitsubishi Heavy Industries, Ltd. The front and rear spar positions were fixed in the structural shape based on the initial aerodynamic geometry. The wing structural model was substituted with shell elements.

The design variables were related to airfoil, twist, and wing dihedral. The airfoil was defined at three spanwise cross-sections using the modified PARSEC[5] with nine design variables (x_{up}, z_{up}, $z_{xx_{up}}$, x_{lo}, z_{lo}, $z_{xx_{lo}}$, α_{TE}, β_{TE}, and $r_{LE_{lo}}/r_{LE_{up}}$) for each cross-section as shown in Fig. 1. The twists were defined at six spanwise locations, and then wing dihedrals were defined at kink and tip locations. The twist center was set on the trailing edge in the present study. The entire wing shape was thus defined using 35 design variables. The detail of design variables is summarized in Table I. In the present study, the geometry of each individual was generated by the unstructured dynamic mesh method[6], [7] using displacement from the initial geometry.

C. Optimizer

ARMOGA[8] is an efficient Multi-Objective Evolutionary Algorithm (MOEA) designed for MDO problems including aerodynamic evaluation with large computation time. ARMOGA can be used to obtain the non-dominated solutions efficiently because of the concentrated search of the probable design space, while keeping diversity.

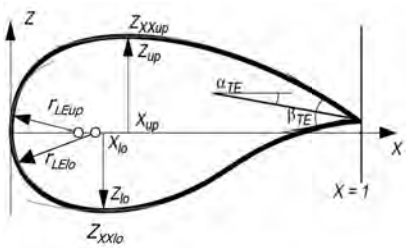


Fig. 1. Illustration of the modified PARSEC airfoil shape defined by nine design variables.

TABLE I

DETAIL OF DESIGN VARIABLES.

serial number		*correspondent design variable*
1 to 9	PARSEC airfoil	35.0% semispan location $(x_{up}, z_{up}, z_{xx_{up}}, x_{lo}, z_{lo}, z_{xx_{lo}}, \alpha_{TE}, \beta_{TE}, r_{LE_{lo}}/r_{LE_{up}})$
10 to 18	PARSEC airfoil	55.5% semispan location
19 to 27	PARSEC airfoil	77.5% semispan location
28 to 33	Twist angle	19.3%, 27.2%, 35.0%, 55.5%, 77.5%, 96.0%
34, 35	Dihedral	35.0%, 96.0%

D. Optimization Results

The population size was set to eight, and then roughly 70 Euler and 90 Reynolds-averaged Navier-Stokes (N-S) computations were performed in one generation for Computational Fluid Dynamics (CFD) evaluation. It took roughly one hour of CPU time for single Euler computation, and it also took roughly nine hours for single N-S computation on NEC SX-5 and SX-7 vector machines per PE. The population was re-initialized every five generations for the range adaptation. First, evolutionary computation was performed for 17 generations. Then, the evolutionary operation was restarted using eight non-dominated solutions extracted from all solution of 17 generations, and two more generations were computed. A total evolutionary computation of 19 generations was carried out. The total of all solutions was 130 individuals and nine non-dominated solutions were generated. The evolution may not converge yet. However, the results were satisfactory because several non-dominated solutions achieved significant improvements over the initial design. Furthermore, a sufficient number of solutions were searched such that the sensitivity of the design space around the initial design could be analyzed. This will provide useful information for designers.

III. DATA MINING

When the optimization problem has only two objectives, tradeoffs can be visualized easily. However, if there are more than two objectives, the technique to visualize the computed all evaluated and non-dominated solutions is needed. In the present study, SOM and ANOVA were employed. Data Mining and knowledge discovery is the new field to extract the knowledge from database including the data which statistical analysis cannot treat. It has the sense to transform analysis results into the concrete proposal.

A. Self-Organizing Map

SOM is not only a technique for visualization but also a tool for the intelligent compression of information. That is, SOM can be applied for data mining to acquire knowledge regarding the design space. In the present study, Viscovery® SOMine[9] (Eudaptics GmbH, Austria) was employed.

1) Viscovery SOMine: Although SOMine is based on the SOM concept and algorithm, it employs an advanced variant of unsupervised neural networks, *i.e.* Kohonen's Batch-SOM.

The algorithm consists of two steps that are iteratively repeated until no more significant changes occur. First the distances between all data items $\{\boldsymbol{x}_i\}$ and the model vectors $\{\boldsymbol{m}_j\}$ are computed and each data item $\boldsymbol{x}_i$ is assigned to the unit c_i that represents it best.

In the second step, each model vector is adapted to better fit the data it represents. To ensure that each unit j represents similar data items as its neighbors, the model vector $\boldsymbol{m}_j$ is adapted not only according to the assigned data items but also with regard to those assigned to the units in the neighborhood. The neighborhood relationship between two units j and k is usually defined by a Gaussian-like function

$$h_{jk} = \exp\left(-\frac{d_{jk}^2}{r_t^2}\right) \tag{1}$$

where d_{jk} denotes the distance between the units j and k on the map, and r_t denotes the neighborhood radius which is set to decrease with each iteration t.

Assuming a Euclidean vector space, the two steps of the Batch-SOM algorithm can be formulated as

$$c_i = \arg\min \|\boldsymbol{x}_i - \boldsymbol{m}_j\| \tag{2a}$$

$$\boldsymbol{m}_j^* = \frac{\sum_i h_{jc_i} \boldsymbol{x}_i}{\sum_i h_{jc_i}} \tag{2b}$$

where $\boldsymbol{m}_j^*$ is the updated model vector.

In contrast to the standard Kohonen algorithm, which makes a learning update of the neuron weights after each record being read and matched, the Batch-SOM takes a 'batch' of data, typically all records, and performs a 'collected' update of the neuron weights after all records have been matched. This is much like 'epoch' learning in supervised neural networks. The Batch-SOM is a more robust approach, since it mediates over a large number of learning steps. Most important, no learning rate is required. The SOMine implementation combines four enhancements to the plain Batch-SOM algorithm(See Ref. [10] for more details). In SOMine, the uniqueness of the map is ensured by the adoption of the Batch-SOM and the linear initialization for input data.

Much like some other SOMs[11], 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. Since the order on the grid reflects the neighborhood within the data, features of the data distribution can be read off from the emerging map on the grid.

In SOMine, the trained SOM is systematically converted into visual information. The tool provides an extensive built-in capability for both pre-processing and post-processing as well as for the automatic colorcoding of the map and its components. SOMine is particularly useful in the determination of dependencies between variables as well as in the analysis of high-dimensional cluster distributions.

2) Cluster Analysis: Once SOM projects input space on a low-dimensional regular grid, the map can be utilized to visualize and explore properties of the data. When the number of SOM units is large, to facilitate quantitative analysis of the map and the data, similar units need to be grouped, *i.e.*, clustered. The two-stage procedure — first using SOM to produce the prototypes which are then clustered in the second stage — was reported to perform well when compared to direct clustering of the data[11].

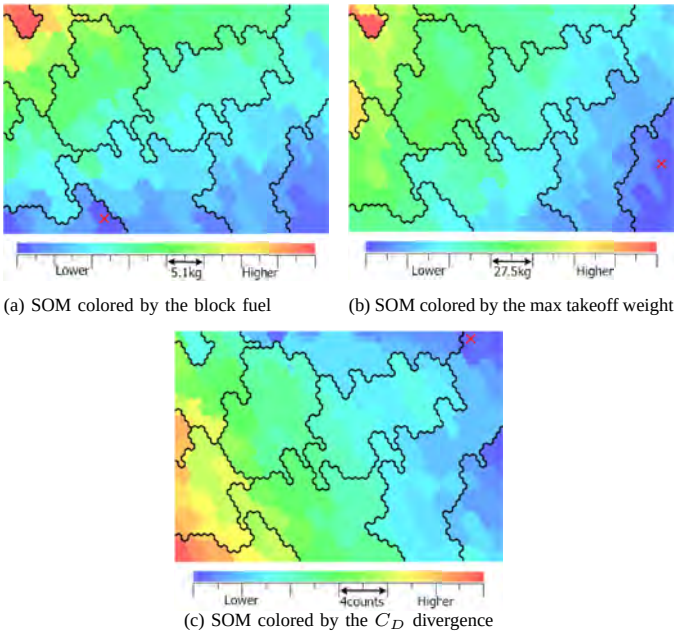
Hierarchical agglomerative algorithm is used for clustering here. The algorithm starts with a clustering where each node by itself forms a cluster. In each step of the algorithm two clusters are merged: those with minimal distance according to a special distance measure, the SOM-Ward distance[9]. This measure takes into account whether two clusters are adjacent in the map. This means that the process of merging clusters is restricted to topologically neighbored clusters. The number of clusters will be different according to the hierarchical sequence of clustering. A relatively small number will be chosen for visualization, while a large number will be used for generation of codebook vectors for respective design variables.

B. Knowledge in the Design Space by SOM

1) Tradeoff Analysis of the Design Space: All of the solutions have been projected onto the two-dimensional map of SOM. Figure 2 shows the resulting SOM with 11 clusters considering the three objectives. Furthermore, Fig. 3 shows the SOMs colored by the three objectives. These color figures show that the SOM indicated in Fig. 2 can be grouped as follows: The upper left corner corresponds to the designs with high block fuel and maximum takeoff weight. The left center area corresponds to designs with high maximum takeoff weight and C_D divergence. The lower left corner corresponds to designs with low block fuel and high C_D divergence. Figure 3(a) and Fig. 3(c) show that there is a tradeoff between these two objective functions. The lower center area corresponds to designs with low block fuel. The right hand side corresponds to designs with low C_D divergence. As the coloring in Fig. 3(a) is similar to that in Fig. 3(b), there was not a severe tradeoff between the block fuel and the maximum takeoff weight. The lower right corner corresponds to designs with low value of all objectives. Extreme non-dominated solutions are indicated in Fig. 3(a) to (c). As they are in different clusters, the simultaneous optimization of the three objectives is impossible. However, the lower right cluster has relatively low values for all three objectives. Thus, this region of the design space may provide a sweet spot for the present design problem.

2) Effects of Aerodynamic Performance on Objective Functions: Figure 4 shows the SOMs colored by the aerodynamic performance under transonic cruising flight condition. Figures 4(a) and (b) show the SOMs colored by C_L and C_D, respectively. As these figures show similar coloring, the L/D increase is not so easy. Lower C_D values are located in the lower right corner in Fig. 4(b). As this area clusters designs with low value of all objectives, this observation suggests that when all objectives are optimized simultaneously, the C_D under the cruising flight condition is also reduced. Furthermore, as the clusters of lower values of the maximum takeoff weight shown in Fig. 3(b) appears on the right hand side of the map, C_D can be decreased

Fig. 2. SOM of all solutions in the three-dimensional objective function space.



(a) SOM colored by the block fuel (b) SOM colored by the max takeoff weight

(c) SOM colored by the C_D divergence

Fig. 3. SOM colored by the objective functions. The symbol $\times$ denotes the respective extreme non-dominated solutions.

simultaneously with the maximum takeoff weight. As the area with higher C_D shown in Fig. 4(b) generally coincide with the area with higher objective function values, C_D is a very important performance index.

Figure 4(c) shows the SOM colored by L/D; lower values are located in the upper left corner. As the higher values of the block fuel shown in Fig. 3(a) are present at the same location, lower L/D makes the block fuel worse. Furthermore, higher L/D values are located in the lower area shown in Fig. 4(c). As the lower values of the block fuel shown in Fig. 3(a) are present at the same area, higher L/D was effective to decrease the block fuel. However, higher transonic L/D values were not necessarily effective to reduce the block fuel in Fig. 4(c) because not only the cruise condition but also the complete flight profile from takeoff to landing were considered in the present study.

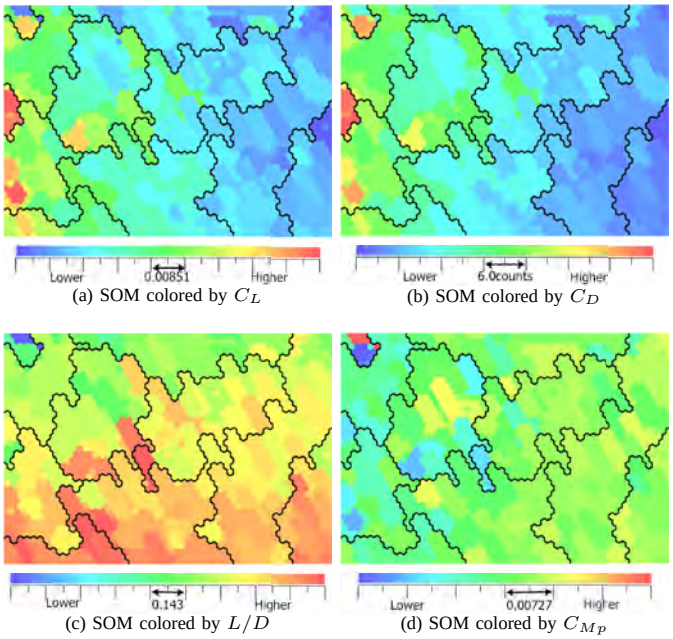
Figure 4(d) shows the SOM colored by C_{Mp}. When C_{Mp} increases and C_L decreases and L/D is reduced. C_L and C_D increase with decreasing C_{Mp}. That is, a decrease in C_{Mp} makes the objective function values worse.

As the resulting SOMs, colored by C_L and C_D under subsonic flight condition, appear similar to transonic C_L and C_D shown in Fig. 4(a) and (b), their influences to the objective functions were also the same. That is, the effects of subsonic aerodynamic performance on objective functions might be predicted from the effects of transonic aerodynamic performance in the present study.

3) Additional Characteristics: Figure 5 shows the SOM colored by three other characteristic values. Figure 5(a) shows the SOM colored by the constraints of the evaluated fuel mass. The colored values are defined as follows:

$$Value = \text{Volume}_{\text{required fuel}} - \text{Volume}_{\text{fuel capacity}} \quad (3)$$

where, $\text{Volume}_{\text{required fuel}}$ denotes the fuel volume required to fly the given range, and $\text{Volume}_{\text{fuel capacity}}$ denotes the fuel capacity volume that can actually be carried in the wing. When this value is greater than zero, the aircraft cannot fly the given range. As the area with



(a) SOM colored by C_L (b) SOM colored by C_D

(c) SOM colored by L/D (d) SOM colored by C_{Mp}

Fig. 4. SOM colored by aerodynamic performance under transonic cruising flight condition.

values of over zero corresponds to the area with high maximum takeoff weight, the aerodynamic characteristics and design values that have effects on maximum takeoff weight dominate this constraint.

Figure 5(b) shows the SOM colored by the ranking in the optimizer based on Pareto ranking. As the upper left region has a poorer ranking, larger block fuel and maximum takeoff weight as objective functions 1 and 2 dominate the poor ranking. In contrast, the lower left area with higher C_D divergence does not have poor ranking. These observations indicate that improvement in C_D divergence is not dominated by the specific aerodynamic performance and design variables, and further improvement cannot be achieved by the present problem easily.

Figure 5(c) shows the SOM colored by the angle between inboard and outboard on the upper wing surface for the gull-wing at the kink location. Angles greater and less than 180 deg correspond to gull and inverted gull-wing, respectively. The locations of higher values of this angle as shown in Fig. 5(c) correspond to positions of higher C_D under the transonic cruising flight condition shown in Fig. 4(b). However, at angles less than 180 deg, there was little correlation between Fig. 4(b) and Fig. 5(c). The inverted gull-wing did not affect aerodynamic performance. The inverted gull-wing is known to have a structural weight increase, which is also observed in the present results. Indeed, the locations of higher angles in Fig. 5(c) had higher maximum takeoff weights as shown in Fig. 3(b). Therefore, non-gull wings should be designed in future.

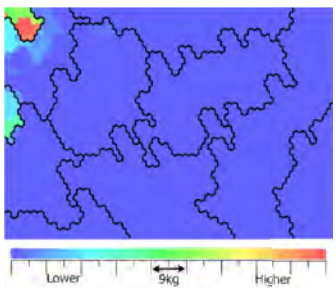
4) Effects of Design Variables: Finally, Fig. 6 and Fig. 7 show the SOMs colored by the selected design variables with regard to the PARSEC airfoil parameters at 35.0% and 55.5% spanwise locations, respectively. Moreover, Fig. 8 shows the SOM colored by the design variable, twist angle. The design variables can be summarized as follows, taking into consideration the effects on each objective function and aerodynamic performance.

There are no design variables that show large effects on objective function 1 as block fuel. The large twist angles at the 35.0% spanwise location makes objective function 2 as maximum takeoff weight worse. In addition, large twist angles at the 55.5% spanwise location increase objective function 3 as C_D divergence. However, no design variable of the PARSEC airfoil had apparent effects on any objective functions by itself. As shown later, PARSEC design variables have direct effects on aerodynamic performances. However, the present objective functions are not pure aerodynamic characteristics. Therefore, effects of the design variables on the objective functions were not trivial. There were no design variables and no aerodynamic characteristics that were effective on the sweet spot with relatively low values for all three objective functions. Therefore, the individual that resides in the sweet spot cannot be generated by hand. A correlation between objective function and design variable is desirable when the sensitivity of the design variable is to be investigated; this is one of the important aspects in optimization problems in general.

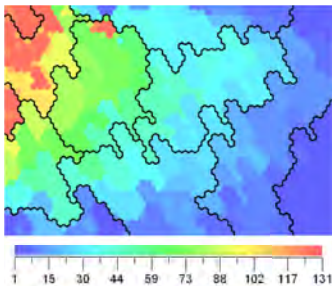
Next, the effects of design variables on aerodynamic performance were investigated. From the correspondence between Figs. 4, 6, 7, and 8, the effects of respective design variables are summarized in Tables II to IV. These tables indicate that the design variables of the PARSEC airfoil have effects on aerodynamic performance directly. It is noted that the effects of design variables to C_D can be predicted from the above results because Figs. 4(a) and (b) are similar. Furthermore, the effects of design variable on aerodynamic performance under the subsonic flight condition can be predicted because the SOMs appeared similar at the transonic and subsonic flight conditions. The leading-edge curvature of PARSEC airfoil at 35.0% spanwise location was effective to L/D and C_{Mp}.

The geometry near the 55.5% spanwise location was not changed markedly with regard to twist angle, as shown in Fig. 8(b). The geometry near the 96.0% spanwise location was changed to upward twisting. Conversely, the geometry near the 35.0% spanwise location was changed to downward twisting. The improvement in the vicinity of the 35.0% spanwise location restrained the shock wave, reducing the wave drag. When the drag decreases, the lift may decrease simultaneously. The lift was increased to compensate for the reduction in the vicinity of the kink so that the angle of attack of the outboard wing was increased although the wing is still twisted down. It should be noted that the angle of attack near the kink had an effect on the transonic drag, especially as shown in Fig. 8(a). This corresponds to the phenomena shown in the CFD visualization. Specifically, the shock wave in the vicinity of the kink is weakened. The angle of attack near the kink with downward twisting is replaced from the initial geometry and the lost lift is made up to replace the angle of attack at the outboard wing with upward twisting so that the wave drag is reduced near the kink. Upward twisting at the outboard wing has no influence on transonic drag, as shown in Fig. 8(c). This corresponds to the CFD prediction. The other design variables were not effective to reduce the objective functions or to increase aerodynamic performance as C_D and L/D under transonic cruise flight condition. Data Mining techniques using SOM were found to be able to classify the design variables considering their influence on the objectives and aerodynamic performance.

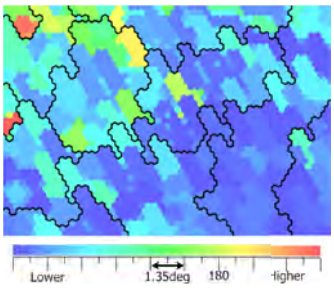
Design knowledge regarding block fuel, which is the most important element of the present optimization problem, will be considered. The following two points are the keys to improve block fuel: 1) L/D increase, 2) $dC_D/d\alpha$ increase, at any Mach number. However, there were no single design variable in the present design space capable of satisfying them simultaneously. In fact, this was confirmed by the SOMs. Although PARSEC design variables correspond to aerodynamic performances, there are no direct effects on other objective functions. It would be easier to understand the design space if the design variables have direct influences on the objective



(a) SOM colored by the constraint as wing box volume

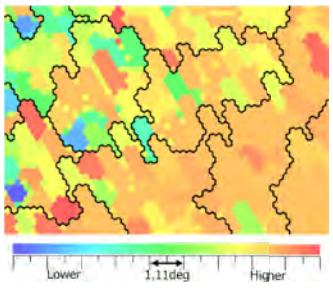


(b) SOM colored by the ranking in the optimizer

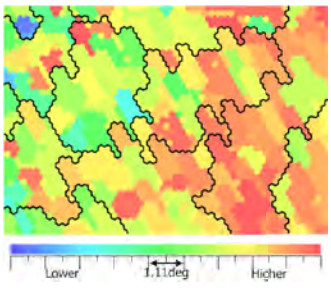


(c) SOM colored by the angle on upper surface expressing the gull-wing at kink location

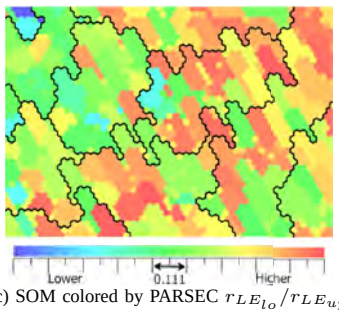
Fig. 5. SOM colored by the characteristic values.



(a) SOM colored by PARSEC α_{TE}

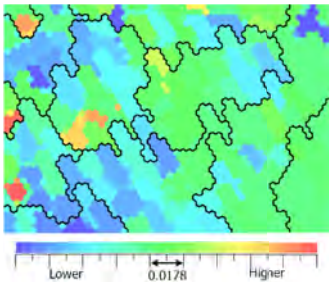


(b) SOM colored by PARSEC β_{TE}

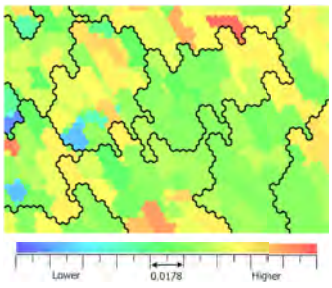


(c) SOM colored by PARSEC $r_{LE_{lo}}/r_{LE_{up}}$

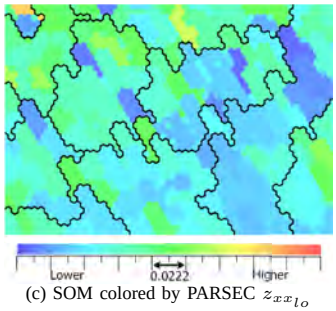
Fig. 6. SOM colored by characteristic design variables regarding the PARSEC airfoil at 35.0% spanwise location. The minimum and maximum values of color bar are set using the minimum and maximum values of each design variable in optimizer, respectively.



(a) SOM colored by PARSEC x_{up}

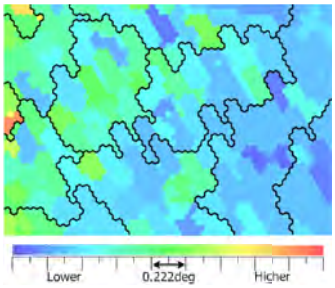


(b) SOM colored by PARSEC x_{lo}

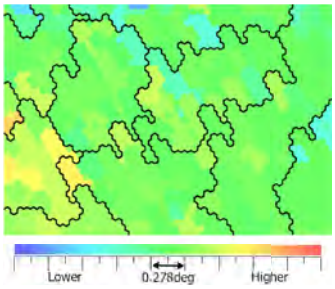


(c) SOM colored by PARSEC $z_{xx_{lo}}$

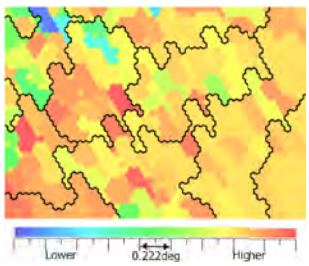
Fig. 7. SOM colored by the characteristic design variables regarding the PARSEC airfoil at 55.5% spanwise location. The minimum and maximum values of color bar are set using the minimum and maximum values of each design variable in optimizer, respectively.



(a) SOM colored by the twist angle at 35.0% spanwise location



(b) SOM colored by the twist angle at 55.5% spanwise location



(c) SOM colored by the twist angle at 96.0% spanwise location

Fig. 8. SOM colored by the characteristic design variables involving wing twist. The minimum and maximum values of color bar are set using the minimum and maximum values of each design variable in optimizer, respectively.

TABLE II

EFFECTS OF DESIGN VARIABLES TO C_L UNDER TRANSONIC CRUISING FLIGHT CONDITION.

design variable		C_L
PARSEC α_{TE} @ 35.0%	decrease	increase
PARSEC x_{up} @ 55.5%	increase	increase
PARSEC x_{lo} @ 55.5%	decrease	increase
Twist @ 35.0%	increase	increase
Twist @ 55.5%	increase	increase

TABLE III

EFFECTS OF DESIGN VARIABLES TO L/D UNDER TRANSONIC CRUISING FLIGHT CONDITION.

design variable		L/D
PARSEC $r_{LE_{lo}}/r_{LE_{up}}$ @ 35.0%	decrease	decrease
PARSEC $z_{xx_{lo}}$ @ 55.5%	increase	decrease

TABLE IV

EFFECTS OF DESIGN VARIABLES TO C_{Mp} UNDER TRANSONIC CRUISING FLIGHT CONDITION.

design variable		C_{Mp}
PARSEC α_{TE} @ 35.0%	decrease	decrease
PARSEC β_{TE} @ 35.0%	decrease	decrease
PARSEC $r_{LE_{lo}}/r_{LE_{up}}$ @ 35.0%	decrease	increase
PARSEC x_{up} @ 55.5%	increase	decrease
PARSEC x_{lo} @ 55.5%	decrease	decrease
PARSEC $z_{xx_{lo}}$ @ 55.5%	increase	increase

functions.

C. Functional Analysis of Variance

Analysis of Variance (ANOVA)[3] uses the variance of the objective functions due to the design variables on the response surface models. Thus, the response surface model should first be constructed for each objective function to calculate the variance. The response surface model employed in the present study is the Kriging model[12]. The Kriging model, developed in the field of spatial statistics and geostatistics, predicts the distribution value of the known point by using stochastic processes. The Kriging model is expressed as follows:

$$\hat{y}(\boldsymbol{x}) = \hat{\mu} + \boldsymbol{r}'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{I}\hat{\mu}) \tag{4}$$

where $\boldsymbol{x} = \{x_1, x_2, \cdots, x_n\}$ denotes the vector of design variables, $\boldsymbol{y}$ is the column vector of sampled response data, and $\boldsymbol{I}$ is unit column vector. $\boldsymbol{R}$ is the correlation matrix whose (i, j) element is

$$R\left(\boldsymbol{x}^i, \boldsymbol{x}^j\right) = \exp\left[-\sum_{k=1}^{n} \theta_k \left|x_k^i - x_k^j\right|^2\right] \tag{5}$$

The correlation vector between $\boldsymbol{x}$ and the m sampled data is expressed as

$$\boldsymbol{r}'(\boldsymbol{x}) = \left[R\left(\boldsymbol{x}, \boldsymbol{x}^1\right), R\left(\boldsymbol{x}, \boldsymbol{x}^2\right), \cdots, R\left(\boldsymbol{x}, \boldsymbol{x}^m\right)\right] \tag{6}$$

The value $\hat{\mu}$ is estimated using the generalized least squares method as

$$\hat{\mu} = \frac{\boldsymbol{I}'\boldsymbol{R}^{-1}\boldsymbol{y}}{\boldsymbol{I}'\boldsymbol{R}^{-1}\boldsymbol{I}} \tag{7}$$

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

$$\hat{\mu}_{\text{total}} \equiv \int \cdots \int \hat{y}\left(x_1, x_2, \cdots, x_n\right) dx_1 dx_2 \cdots dx_n \tag{8a}$$

$$\hat{\sigma}^2_{\text{total}} = \int \cdots \int \left[\hat{y}\left(x_1, x_2, \cdots, x_n\right) - \hat{\mu}\right]^2 dx_1 dx_2 \cdots dx_n \tag{8b}$$

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

$$\begin{aligned}&\hat{\mu}\left(x_i\right)\\ &\equiv \int \cdots \int \hat{y}\left(x_1, x_2, \cdots, x_n\right) dx_1 dx_2 \cdots \\ &\qquad\qquad dx_{i-1} dx_{i+1} \cdots dx_n\\ &\quad - \hat{\mu}\end{aligned} \tag{9}$$

$$\begin{aligned}&\hat{\mu}_{i,j}\left(x_i, x_j\right)\\ &\equiv \int \cdots \int \hat{y}\left(x_1, x_2, \cdots, x_n\right) dx_1 dx_2 \cdots dx_{i-1} dx_{i+1} \cdots \\ &\qquad\qquad dx_{j-1} dx_{j+1} \cdots dx_n\\ &\quad - \hat{\mu}_i\left(x_i\right) - \hat{\mu}_j\left(x_j\right) - \hat{\mu}\end{aligned} \tag{10}$$

$\hat{\mu}(x_i)$ 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 obtained as follows:

$$\hat{\sigma}^2_{x_i} = \int \left[\hat{\mu}_i\left(x_i\right)\right]^2 dx_i \tag{11}$$

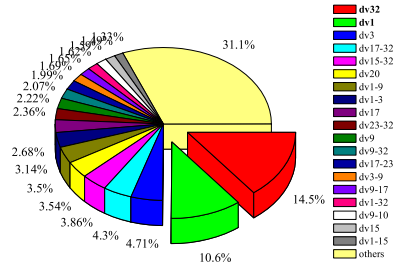
The proportion of the variance P due to design variable x_i to total variance of model can be expressed by dividing Eq. (11) with Eq. (8b).

$$\begin{aligned}P &= \frac{\hat{\sigma}^2_{x_i}}{\hat{\sigma}^2_{\text{total}}}\\ &= \frac{\displaystyle\int \left[\hat{\mu}_i\left(x_i\right)\right]^2 dx_i}{\displaystyle\int \cdots \int \left[\hat{y}\left(x_1, x_2, \cdots, x_n\right) - \hat{\mu}\right]^2 dx_1 dx_2 \cdots dx_n}\end{aligned} \tag{12}$$

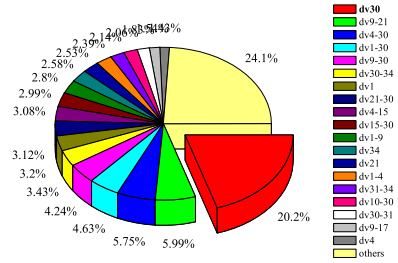
This value indicates the effect of design variable x_i on the objective function[13].

1) Effects of Design Variables: Figures 9 and 10 show the proportion of the influence of design variables on the objective functions and aerodynamic performance obtained by ANOVA. The influence of the design variables for each objective function obtained by ANOVA and SOM is summarized in Table V. ANOVA and SOM predicted similar influence for the two objective functions, the maximum takeoff weight and C_D divergence. As the design variables correspond to aerodynamic performance, these two objective functions have correlation with aerodynamic performance. However, the block fuel did not have a correspondent result between ANOVA and SOM. As the block fuel is computed from the wing structural weight and L/D at subsonic, transonic, and off-design conditions, it is sensitive to various elements. That is, the present design variables do not have direct influence on the block fuel. When the influence of design variable is investigated, the correlation is needed between objective function and design variable.

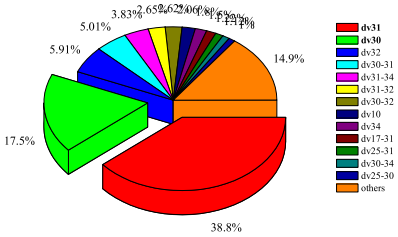
Here, the disadvantages of ANOVA and SOM will be investigated. The disadvantage of ANOVA is the following. Although it reveals that "which" design variable influences, it is unclear that "how" that design variable influences. Whereas, the disadvantages of SOM is the following; 1) qualitatively and subjective. 2) it is possible to fail finding of the design knowledge due to a large number of



(a) Influence for block fuel

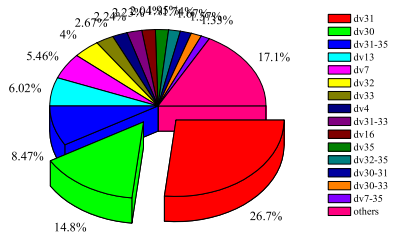


(b) Influence for maximum takeoff weight

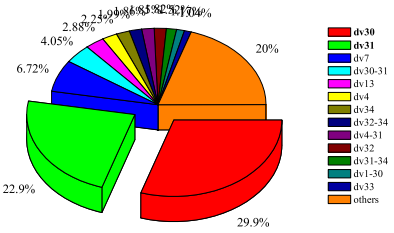


(c) Influence for C_D divergence

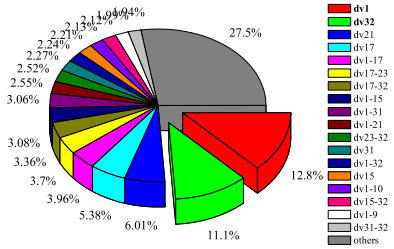
Fig. 9. Proportion of design-variable influence for the objective functions using ANOVA.



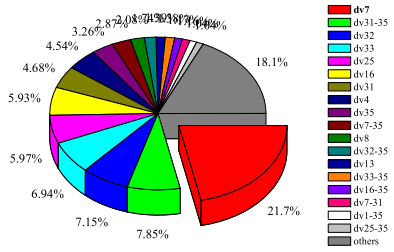
(a) Influence for C_L



(b) Influence for C_D



(c) Influence for L/D



(d) Influence for C_{Mp}

Fig. 10. Proportion of design-variable influence for aerodynamic performance at the transonic cruising condition using ANOVA.

objective functions and design variables. 3) the interaction between the design variables cannot be investigated directly. ANOVA and SOM compensate with the respective disadvantages, then knowledge regarding the design space can be obtained more clearly by the combination between them such as Data Mining is performed by SOM after sensitive design variables are addressed by ANOVA.

D. Evaluation of an Improved Geometry

The design knowledge obtained by Data Mining shows that a non-gull wing should be designed. Therefore, we modified the optimized wing shape (called as '*optimized*' shown in Fig.11) which achieved the higher improvement in the block fuel to the non-gull wing shape (called as '*optimized_mod*') to verify the design knowledge obtained by the previous Data Mining.

The evaluated results are shown in Figs. 11 to 13. These figures show that *optimized_mod* improves both block fuel and maximum takeoff weight. Moreover, by comparison of the polar curves at constant C_L for cruising condition, C_D of *optimized_mod* was found to be reduced by 10.6 counts over the initial geometry. Due to the improvement of drag, the block fuel of $optimized\_mod$ was reduced by 3.6 percent.

In the present MDO system, surface spline function of the geometry deviation was used for the modification of the wing shape (surface mesh), and then the volume mesh was modified by the unstructured dynamic mesh method. However, this process made the surface mesh distorted around the leading edge and highly limited the design space. This mesh generation process might be the primary reason for the difficulty in finding the non-gull geometry with better block fuel performance. The secondary reason is that only the small number of the generations has been performed. However, this result reveals that

TABLE V

COMPARISON OF THE MOST INFLUENTIAL DESIGN VARIABLE FOR THE OBJECTIVE FUNCTIONS BETWEEN ANOVA AND SOM.

	ANOVA	*SOM*
block fuel	Twist @ 77.5%	—
max takeoff weight	Twist @ 35.0%	Twist @ 35.0%
C_D divergence	Twist @ 55.5%	Twist @ 55.5%

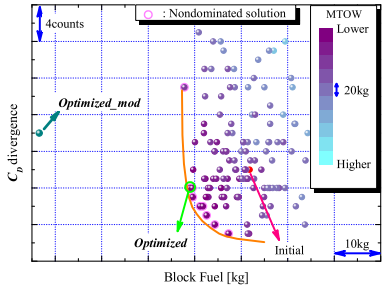


Fig. 11. Comparison of *optimized_mod* and all solutions on two-dimensional plane between block fuel and C_D divergence.

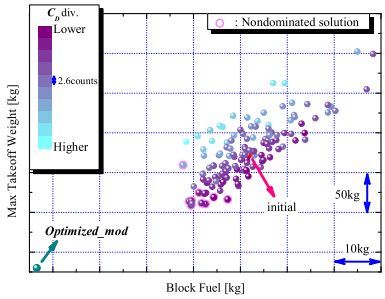


Fig. 12. Comparison of *optimized_mod* and all solutions on two-dimensional plane between block fuel and maximum takeoff weight.

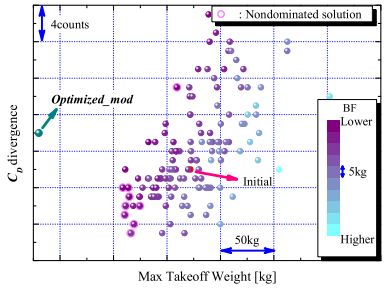


Fig. 13. Comparison of *optimized_mod* and all solutions on two-dimensional plane between maximum takeoff weight and C_D divergence.

Data Mining technique salvages the information. It is demonstrated that the knowledge discovery by Data Mining regarding design space is an important aspect in the practical optimization.

IV. Conclusion

Data Mining for the design space was performed using a SOM and an ANOVA for a large-scale, real-world MDO problem to provide the design knowledge. As a result, SOM reveals that "which" and "how" design variable influences the objective functions and aerodynamic performances. The higher value of 35% twist angle increases the maximum takeoff weight. The higher value of 55.5% twist angle increases the drag divergence. No design variable has direct influence regarding the block fuel. Detailed observations of SOM revealed that there is a sweet spot in the design space where the three objectives become relatively low. Whereas, ANOVA shows that "which" design variable influences. Here, the result of the influence for the block fuel by ANOVA does not correspond to on by SOM. As the block fuel is computed from various variables, the reliability of results by ANOVA decreases. SOM and ANOVA compensate with the respective disadvantages, then design knowledge is acquired more clearly by the combination between them.

Although the present MDO results showed the inverted gull-wings as non-dominated solutions, one of the key features found by Data Mining was the non-gull wing geometry. When this knowledge was applied to one optimum solution, the resulting design was found to have better performance compared with the original geometry designed in the conventional manner. The Data Mining technique provides knowledge regarding the design space and may salvage lost information during the optimization operation, which will be an important facet of solving practical optimization problems.

References

[1] S. Obayashi and D. Sasaki, "Visualization and data mining of Pareto solutions using self-organizing map," in *Proceedings of the 2nd International Conference on Evolutionary Multi-Criterion Optimization*, *Lecture Notes in Computer Science 2632*, Springer-Verlag Heidelberg, Faro, Portugal, 2003, pp. 796–809.

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

[3] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, vol. 13, no. 4, pp. 455–492, 1998.

[4] K. Chiba, S. Obayashi, K. Nakahashi, and H. Morino, "High-fidelity multidisciplinary design optimization of wing shape for regional jet aircraft," in *Proceedings of the 3rd International Conference on Evolutionary Multi-Criterion Optimization*, *Lecture Notes in Computer Science 3410*, Springer-Verlag Heidelberg, Guanajuato, Mexico, 2005, pp. 621–635.

[5] A. Oyama, S. Obayashi, K. Nakahashi, and N. Hirose, "Aerodynamic wing optimization via evolutionary algorithms based on structured coding," *Computational Fluid Dynamics Journal*, vol. 8, no. 4, pp. 570–577, 2000.

[6] M. Murayama, K. Nakahashi, and K. Matsushima, "Unstructured dynamic mesh for large movement and deformation," 2002, aIAA Paper 2002-0122, 2002.

[7] W. Yamazaki, K. Matsushima, and K. Nakahashi, "Aerodynamic optimization of NEXST-1 SST model at near-sonic regime," 2004, aIAA Paper 2004-0034, 2004.

[8] D. Sasaki and S. Obayashi, "Efficient search for trade-offs by adaptive range multi-objective genetic algorithms," in *Journal of Aerospace Computing, Information, and Communication*, Vol. 2, 2005, pp. 44–64.

[9] "Eudaptics website," uRL: http://www.eudaptics.com [cited 16 June 2004].

[10] G. Deboeck and T. Kohonen, *Visual Explorations in Finance with Self-Organizing Maps*. London, Springer Finance, 1998.

[11] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Transactions on Neural Networks*, vol. 11, no. 3, pp. 586–600, 2000.

[12] A. J. Keane, "Wing optimization using design of experiment, response surface, and data fusion methods," *Journal of Aircraft*, vol. 40, no. 4, pp. 741–750, 2003.

[13] S. Jeong, K. Yamamoto, and S. Obayashi, "Kriging-based probabilistic method for constrained multi-objective optimization problem," 2004, aIAA Paper 2004-6437, 2004.