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

In this paper, analysis of variance (ANOVA) and self-organizing map (SOM) are 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 effect of interaction between design variables on objective functions and SOM can visualize the trade-off among objective functions. This information will be helpful for designer to determine the final design from the non-dominated solutions of multi-objective problem. These methods are applied to a fly-back booster of reusable launch vehicle design which has 4 objective functions and 71 design variables, and a transonic airfoil design performed with adaptive search region method.

I.   Introduction

Recent advances in computing power and the development of high fidelity Computation Fluid Dynamics (CFD) codes, design optimization using CFD become one of inevitable tools in the field of aircraft designs. There are several approaches for design optimization using CFD: Adjoint method, which is gradient-based method, is very efficient for design problems having a lot of design variables[1-2]; Genetic Algorithm (GA), which is population-based method, is very powerful for multi-objective design problems[3-5].

However, researches on the optimization so far are concentrated only on finding the optimal solution or the non-dominated 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 problem. Thus, it is preferable for a designer to provide the non-dominated solutions with some useful information for the decision of the final design. The information about the design space, such as trade-off relations between objective functions and the relations between design variables and objective functions will be one of the useful information for decision of the final design. Furthermore, this information will be useful to understand why the final design has a good performance and makes it possible to simplify the design problem by eliminating the design variables which do not have a large influence on the objective functions.

The process to find the information from the design results may be called 'data mining'. In this study, Analysis of Variance (ANOVA)[6] and Self-Organizing Map (SOM)[7] are used for the data mining. The former uses a 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 interaction between design variables on objective function.

---

[*] 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

ANOVA expresses the information in a quantitative way. On the other hand, SOM employs a nonlinear projection algorithm from high to low-dimension and a clustering technique. This method can visualize not only the relation between the design variables and the objective functions but also the trade-off between objective functions. The method expressed the information in the qualitative way.

These techniques are applied to the results of two design problems: one is fly-back booster of the reusable launch vehicle design which treats 4 objective functions and 71 design variables. Results show that the present data-mining techniques are useful to analyze the problem with a large number of design variables. The other is a transonic airfoil design using the adaptive search region method. It showed that the result of data mining largely depends on the definition of the design space.

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 in order to calculate the variance. The response surface model used in this study is the Kriging model[8].

A. Kriging Model

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

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

where $\mathbf{x} = \{x_1, x_2, \cdots, x_n\}$ denotes the vector of design variables, $\mathbf{y}$ is the column vector of sampled response data and $\mathbf{1}$ is unit column vector. $\mathbf{R}$ is the correlation matrix whose (i, j) element is

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

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

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

The $\hat{\mu}$ can be calculated as follow equation:

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

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

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

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

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

Maximization of the likelihood function is an m-dimensional unconstrained non-linear optimization problem. The alternative method[9] is used to solve this problem.

B. Decomposition of Variance

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 the design variable. The 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 model are as follows:

$$\hat{\mu}_{total} \equiv \int \cdots \int \hat{y}(x_1, \ldots, x_n) dx_1 \cdots dx_n \tag{7}$$

$$\hat{\sigma}_{total}^2 = \int \cdots \int [\hat{y}(x_1,......,x_n) - \hat{\mu}_{total}]^2 dx_1 \cdots dx_n \tag{8}$$

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

$$\hat{\mu}_i(x_i) \equiv \int \cdots \int \hat{y}(x_1,\cdots,x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n - \hat{\mu}_{total} \tag{9}$$

$$\hat{\mu}_{i,j}(x_i,x_j) \equiv \int \cdots \int \hat{y}(x_1,......,x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_{j-1} dx_{j+1} \cdots dx_n - \hat{\mu}_i(x_i) - \hat{\mu}_j(x_j) - \hat{\mu}_{total} \tag{10}$$

$\hat{\mu}_i(x_i)$ 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

$$\int [\hat{\mu}_i(x_i)]^2 dx_i \tag{11}$$

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

$$\frac{\int [\hat{\mu}_i(x_i)]^2 dx_i}{\int \cdots \int [\hat{y}(x_1,......,x_n) - \hat{\mu}_{total}]^2 dx_1 \cdots dx_n} \tag{12}$$

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 an unsupervised learning, nonlinear projection algorithm[10] 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 is the response of two neighboring neurons in the low-dimensional space. Thus, SOM reduces the dimension of input data while preserving their features.

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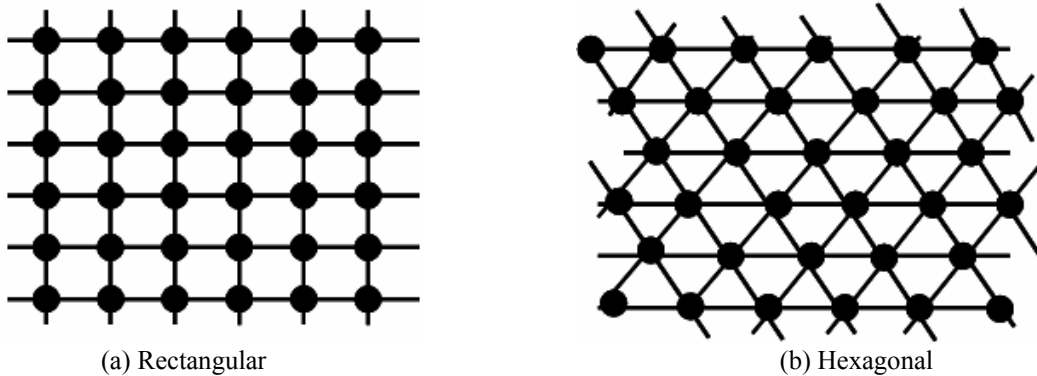

(a) Rectangular   (b) Hexagonal

Figure 1 Topology used in SOMs

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

$$\left\| \mathbf{x} - \mathbf{m}_c \right\| = \min \left\| \mathbf{x} - \mathbf{m}_k \right\| \qquad (k = 1, \cdots\cdots, M) \qquad (13)$$

Once the best-matching unit is determined, the weight adjustments are performed not only for the best-matching unit but also for its neighbors. 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 become closer to the input vector as shown in Fig. 2. The weight vectors are situated in the cross of the solid lines. The best-matching unit is the weight vector who is 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 with dashed lines. Repeating 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 the corresponding neighboring neurons in the two-dimensional map.
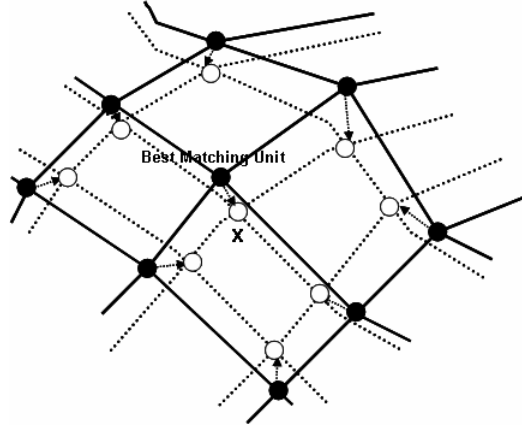
Figure 2 Adjustment of the best-matching unit and its neighbors

B. Kohonen's Batch-SOM

In this investigation, SOMs are generated by using commercial software Viscovery® SOMine plus 4.0[11] 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. 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 $\{\mathbf{m}_j\}$ near the best-matching unit. The Batch-SOM algorithm can be formulated as follows:

$$c_i = \arg\min_j \left\| \mathbf{x}_i - \mathbf{m}_j \right\| \qquad (14)$$

$$\mathbf{m}_j^* = \frac{\sum_i h_{jc_i} \mathbf{x}_i}{\sum_{i'} h_{jc_{i'}}} \qquad (15)$$

where $\mathbf{m}_j^*$ is the adjusted weight vector. The neighborhood relationship between two neurons j and k is defined by the following Gaussian-like function:

$$h_{jk} = \exp\left(-\frac{d_{jk}^2}{r_t^2} \right) \qquad (16)$$

where d_{jk} denotes the Euclidean distance between the neuron k and the neuron j 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 all each record is read and matched. On the contrary, the 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. The Batch-SOM is a more robust approach, since it mediated over a large number of learning steps. In the 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, 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. The trained SOM is systematically converted into visual information.

C. Cluster Analysis

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

Hierarchical agglomerative algorithm is used for the clustering here. 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[11]. The number of clusters is determined by the hierarchical sequence of clustering. A relatively small number of clusters are used for visualization, while a large number are used for the generation of weight vectors for respective design variables.

IV.   Results

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

Geometry of the fly-back booster[12] used in this design is shown in Fig. 3(a). In this design, fuselage shape is fixed and only wing shape design is considered, because the fuselage is filled with the liquid propellant rocket engines, so little change to its size is possible. The design variables used to define wing shape are related to planform, airfoil, wing twist and relative wing position to 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, respectively, by using thickness and camber distributions. Both distributions are parameterized by using Bezier curves and linearly interpolated in the spanwise direction. Wing twist in refined by using a B-spline curve with six control points. Relative position of the wing root to the fuselage is parameterized by x and z coordinates of the leading edge, angle of attack and dihedral angle. Total 71 design variables are used to wing geometry definition.
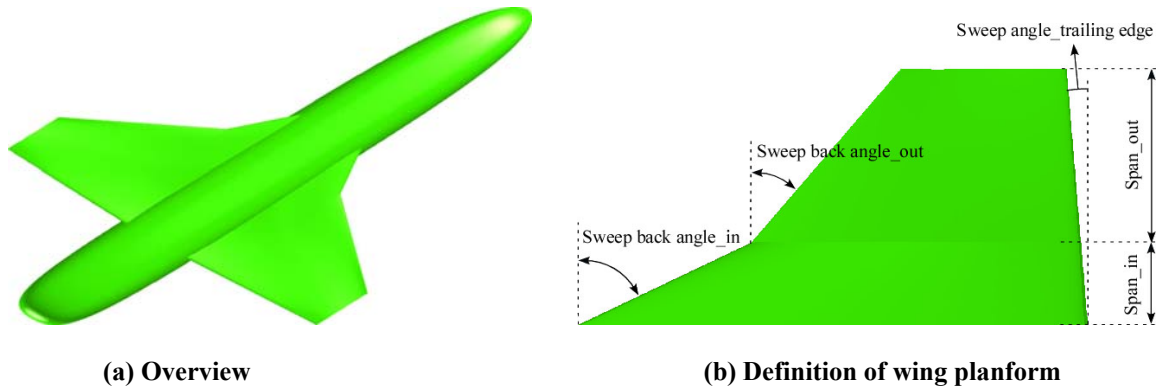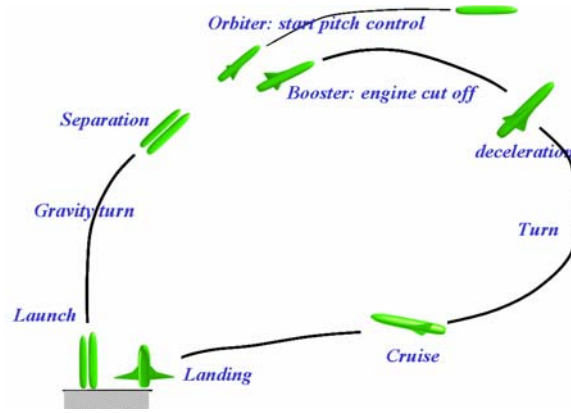

(a) Overview   (b) Definition of wing planform

Figure 3 Geometry of the fly-back booster

Figure 4 Typical flight sequence for TSTO fly-back booster

According to the trajectory analysis[13], the separation of the booster and orbiter takes places around Mach 3 and the booster turns over, slows down, cruise at transonic speed and lands at subsonic speed as shown in Fig. 4. In order to maintain good aerodynamic performances in 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.

$$F_1 = \left| C_{M_p}^{SUPERSONIC} - C_{M_p}^{TRANSONIC} \right| \qquad (17)$$

A significant control problem related to the RLV flight may originate in a large variation of the aerodynamic center between supersonic and transonic flight conditions. It is, then, desirable to design wing shapes with a less variation in the aerodynamic center. It will yield easier stability control.

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

$$F_2 = \left| C_{M_p}^{TRANSONIC} \right| \qquad (18)$$

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

3. Minimization of the drag at the transonic flight conditions

$$F_3 = C_D^{TRANSONIC} \qquad (19)$$

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

4. Maximization of the lift at the subsonic flight conditions

$$F_4 = C_L^{SUBSONIC} \qquad (20)$$

To reduce the required runway distance, the lift obtained at the subsonic flight conditions should be minimized.

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

As the optimizer, Adaptive Range Multi-Objective Genetic Algorithm (ARMOGA)[14] is used. The population size of the present ARMOGA is 8 and 40 generations are performed. Figure 5 shows 102 non-dominated solutions obtained by ARMOGA. However, it is difficult to understand the feature of design space from the Fig. 5. For the better understand of the design space, ANOVA and SOM are performed with 102 non-dominated solutions.
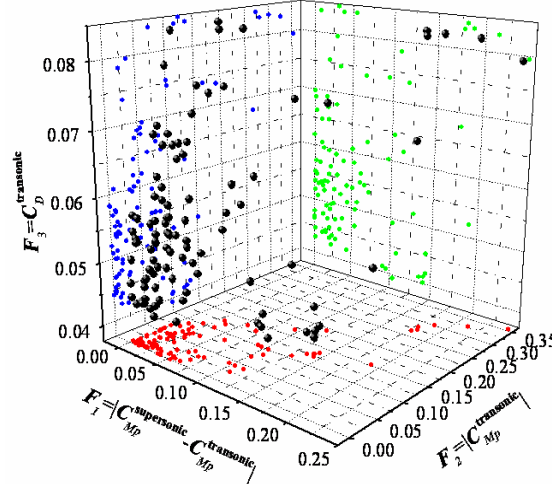

Figure 5 Non-dominated solutions projected onto three-dimensional objective function space

ANOVA

ANOVA is performed for 4 objective functions. Variance of design variables and their interactions whose proportion to the total variance is over than 1.0% are shown in Fig. 6.
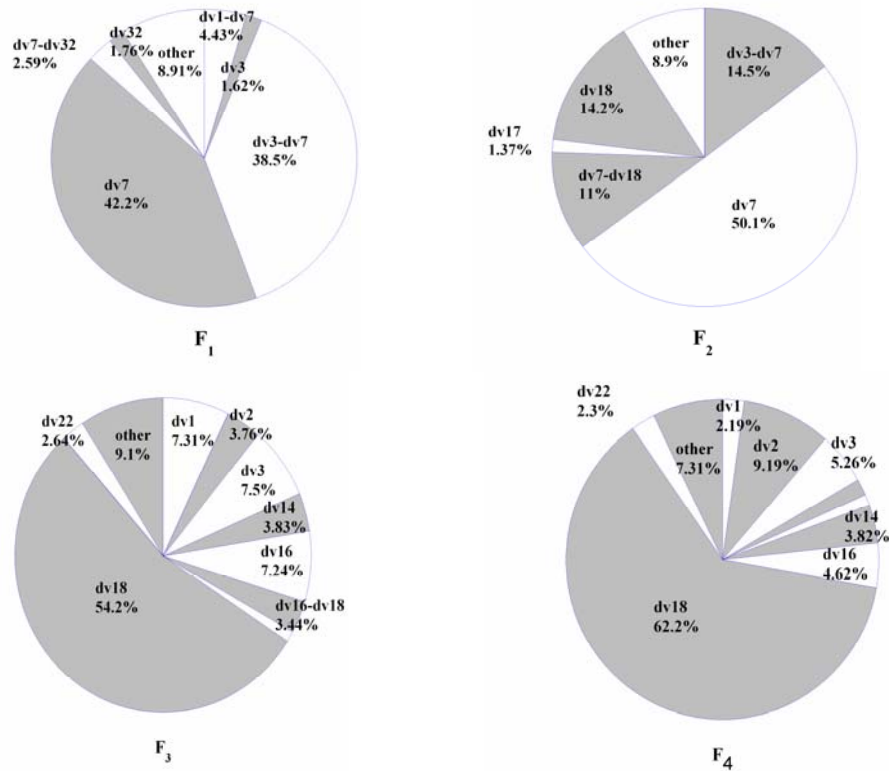

Figure 6 ANOVA results

7

American Institute of Aeronautics and Astronautics

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. About F_3 and F_4, dv18 (rearward camber height at kink) gives the largest effect. dv7 and dv18 are illustrated in the Fig. 7. These findings correspond to the aerodynamic knowledge.
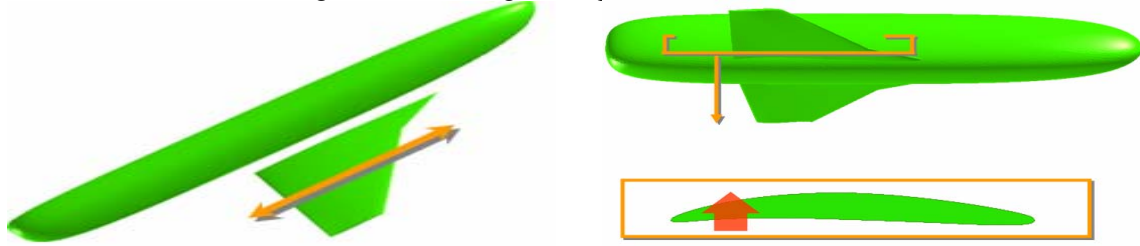
(a) x coordinate of relative wing position to fuselage (dv7)   (b) rearward camber height at kink (dv18)

Figure 7 Illustrations of dv7 and dv18

SOM

SOM is applied to the non-dominated solutions. Figure 8 is SOMs colored by 4 objective functions. According to results, SOMs colored by F_1 and F_2 show similar color patterns. This means that F_1 and F_2 are not in the trade-off relation because both objective functions should be minimized. On the other hand, though SOMs colored by F_3 and F_4 show the similar distribution of color, F_3 and F_4 are in a severe trade-off relation because F_3 should be minimized but F_4 should be maximized.
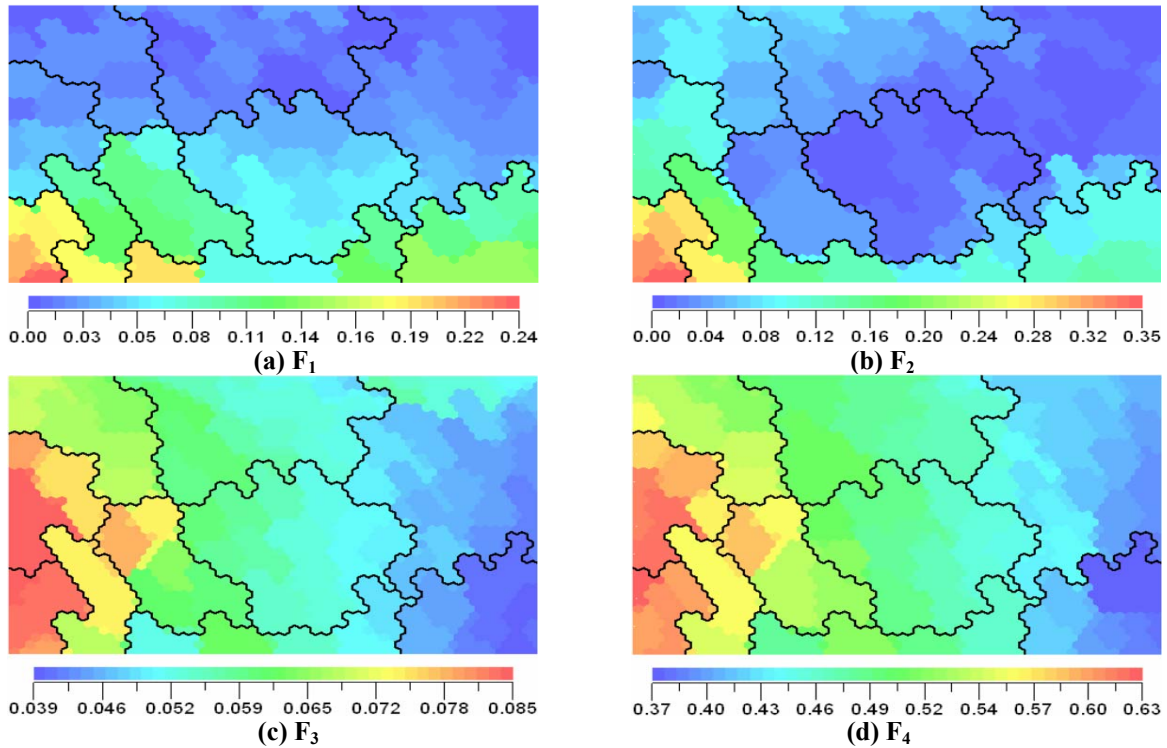

0.00 0.03 0.05 0.08 0.11 0.14 0.16 0.19 0.22 0.24
(a) F_1

0.00 0.04 0.08 0.12 0.16 0.20 0.24 0.28 0.32 0.35
(b) F_2

0.039 0.046 0.052 0.059 0.065 0.072 0.078 0.085
(c) F_3

0.37 0.40 0.43 0.46 0.49 0.52 0.54 0.57 0.60 0.63
(d) F_4

Figure 8 SOMs colored by objective functions

Figure 9 shows SOMs colored by three design variables (dv7, dv18 and dv15). In Fig. 9(a), colored by dv7, the clusters, whose dv7 value is large, are situated at the lower left corner. In this area, the clusters, whose F_1 and F_2 values are large, exist as shown in Figs. 8(a) and 8(b). This means that large dv7 value brings about bad performances of F_1 and F_2. In Fig 9(b), colored by dv18, the clusters, whose dv18 is large, exist in the left-hand side. This color pattern is very similar to those for F_3 and F_4 as shown in Figs. 8(c) and 8(d). This means that large dv18

value related with large F_3 and F_4 values. From the results, it can be said 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. 9(c), colored by dv15 (x coordinate of forward at kink), there is no noticeable trend of color distribution. This means that dv15 has little influence on the objective functions. These results are coincident with the results of ANOVA.
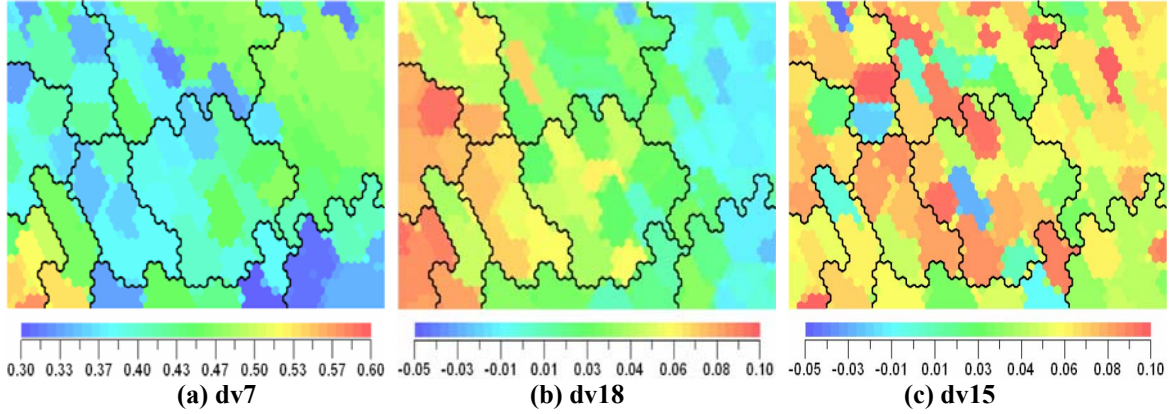

0.30 0.33 0.37 0.40 0.43 0.47 0.50 0.53 0.57 0.60 -0.05 -0.03 -0.01 0.01 0.03 0.04 0.06 0.08 0.10 -0.05 -0.03 -0.01 0.01 0.03 0.04 0.06 0.08 0.10

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

Figure 9 SOMs colored by design variables

V.   Transonic Airfoil Design

In this design, geometry of airfoil is defined by using PARSEC[15]. Figure 10 shows 11 basic parameters for PARSEC airfoil. In this design, only a sharp trailing-edge airfoil was considered, therefore, ΔZ_{TE} was set to zero. A total of 10 design variables were used to define the geometry of airfoil. The design problem is defined as follows:

Minimize C_d

subject to   a) $C_l \geq C_{l\_RAE2822}$

               b) $Area \cong Area_{RAE2822}$

at the flow condition of Mach = 0.73 and an angle of attack (AOA) = 2.7°. C_l and C_d are drag and lift coefficient of designed airfoil and $C_{l\_RAE2822}$ and $C_{d\_RAE2822}$ are those of RAE2822 airfoil. *Area* is the cross-sectional area of the airfoil.
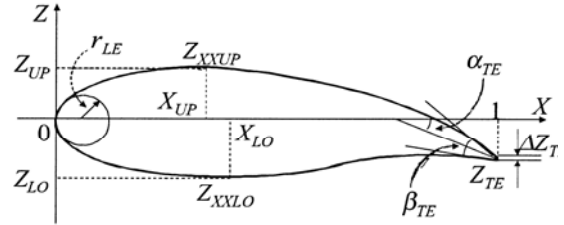

Figure 10 PARSEC airfoil and its parameters

In this design, the search region of optimization problem is successively changed by investigating the probabilistic distribution of design variables. Namely, adaptive search region method[16] is used. The design method consists of two stages: i) the optimal search and ii) the change of search region. The procedure of optimal search is shown in Fig. 11.

1. Kriging models are constructed for C_l and C_d with N sample points. In this study, a number of initial sample points are 50. These points are selected by using Latin hypercube design method[17] to spread the points uniformly in the search region.
2. GA operations
   - Generation of initial population and evaluation.
   - Selection of parents
   - Crossover and mutation
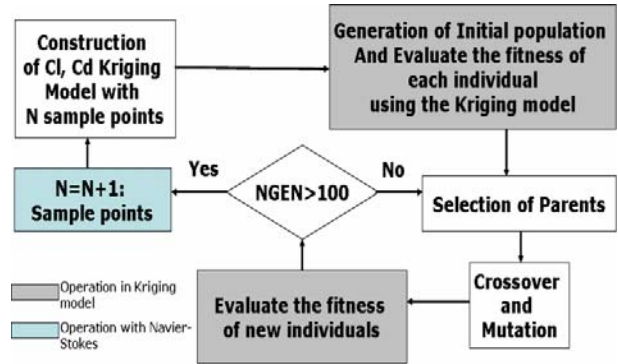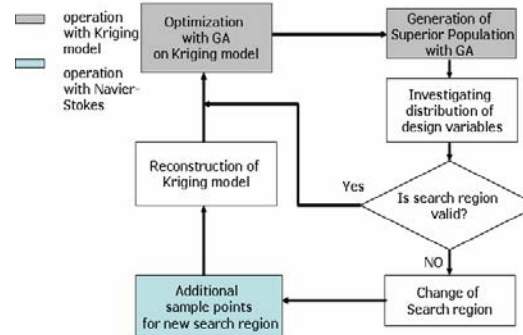   - Evaluation of new individuals in the Kriging models


Figure 11 Optimal search procedures

When the generation exceeds 100, the point which gives maximum EI is selected as an additional sample point. This routine is iterated until the termination criterion is reached. In this study, termination criterion is the maximum number of additional sample points.

Once the optimization is over, the validity of the search region is examined. This procedure is shown in Fig. 12. First, the superior population (SP) is generated by GA. The meaning of 'SP' is that its individual satisfies all design constraints and the objective function values of its individual are larger than certain values. The distribution of design variable 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 by using 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 represented in Fig. 13. 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 were performed. For this design, data mining

Figure 12 Change of search region

techniques are performed in both the initial and final search regions to examine the dependency on the search region.
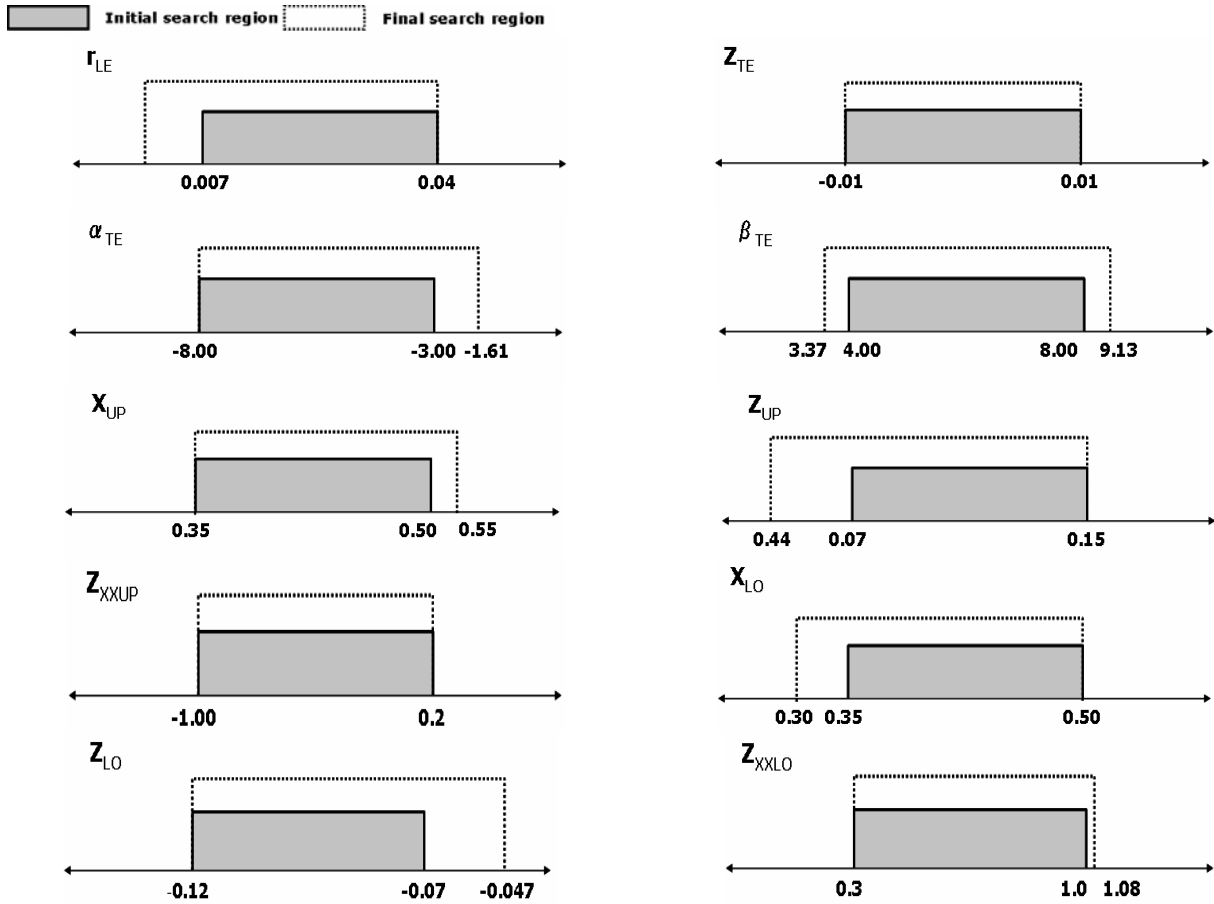
Figure 13 Comparison of the initial search region and the final search region

ANOVA

First, ANOVA is done in the initial search region with 56 sample data. First 50 points are selected by the Latin hypercube design and the last 6 data are selected by the optimal search algorithm shown in Fig.11. The variance of design variables and there interactions whose proportion to the total variance is over than 2.0% are shown in Fig. 14. According to Fig. 14, Z_{UP} and Z_{LO} have a comparatively large effect on C_l, and Z_{UP} have the largest effect on C_d. These parameters are related with definition of airfoil thickness. These findings coincide with the aerodynamic knowledge.
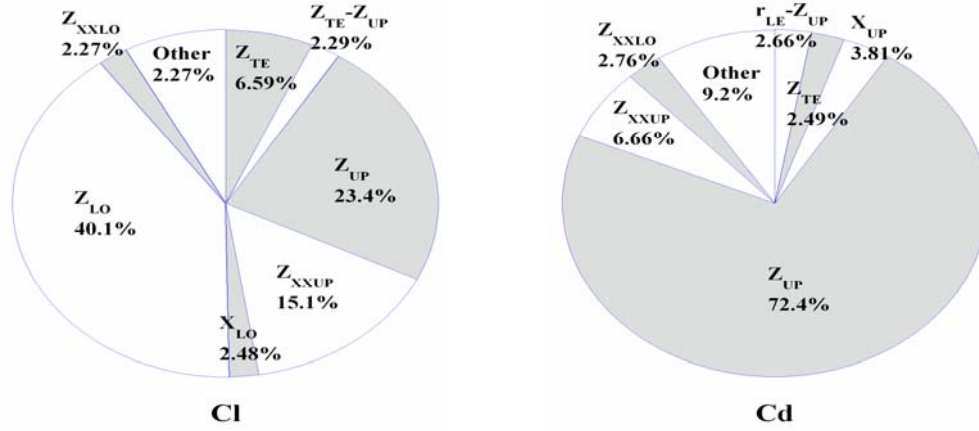

Figure 14 ANOVA results in initial search region

The ANOVA is also performed in the final search region with 110 sample data. First 56 points are the same points used in the initial search region case and the rest of the points are selected during the optimization process using the adaptive search region method. According to Fig. 15, Z_{UP} and Z_{LO} give a large effect on C_l, and Z_{UP} and Z_{UP}-Z_{LO} have a comparatively large effect on C_d. The proportion of r_{LE} which is small in the initial search region becomes 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 effect of design variables on objective functions may differs according to definition of the search region. Thus, the elimination of a design variable, which has little effect on objective functions, without validity check of the search region may lead to an undesirable design result.
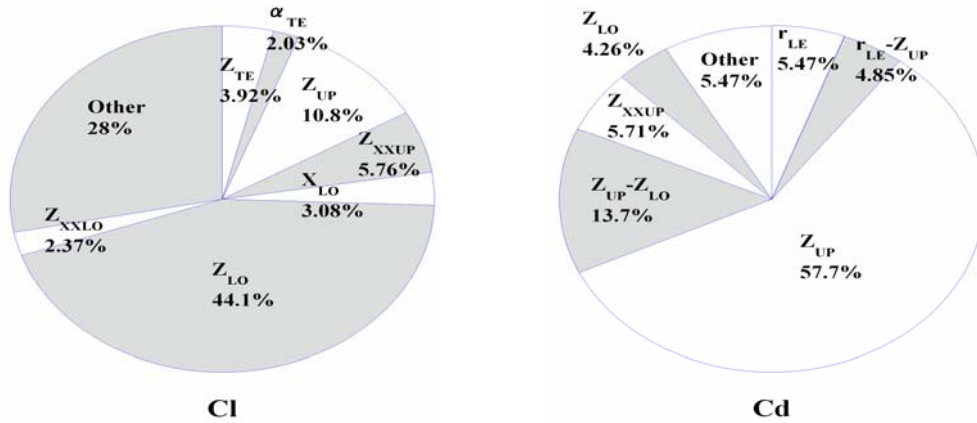

Figure 15 ANOVA results in final search region

SOM

SOM is also applied to the result of the transonic airfoil design. Figure 16 shows the SOMs colored by C_l and C_d in initial search region. In the upper left corner and the lower right corner in Fig. 16 (a) and 16(b), color patterns are very similar. On the other hand, in the upper right corner and the lower left corner, color patterns of the clusters are

opposite. From these observations, we can classify 56 data into 4 groups: i) Both C_l and C_d are large, ii) Both C_l and C_d are small, iii) C_l is large and C_d is small and iv) C_l is small and C_d is large. This means that the Latin hypercube design used for the initial sample point selections generated a wide variety of airfoils with various C_l and C_d performances successfully
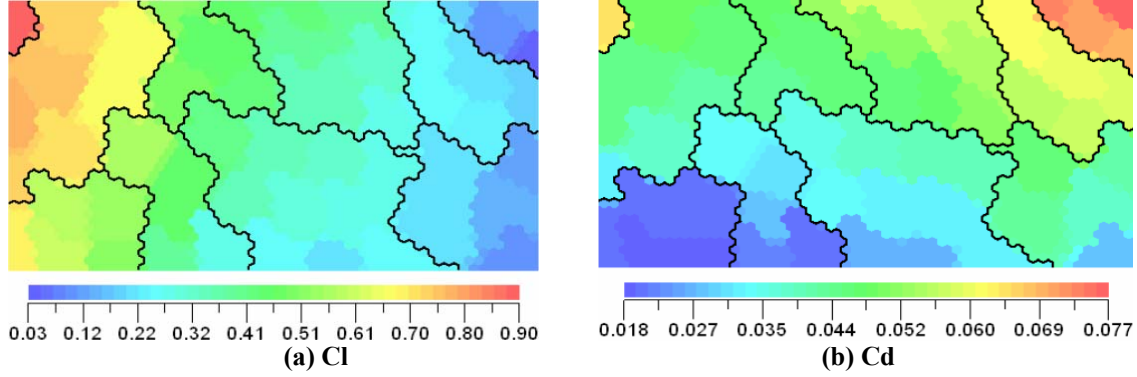

(a) Cl   **(b) Cd**

Figure 16 SOMs colored by objective functions in initial search region

(a) Z_{UP}   **(b) Z_{LO}**   **(c) r_{LE}**
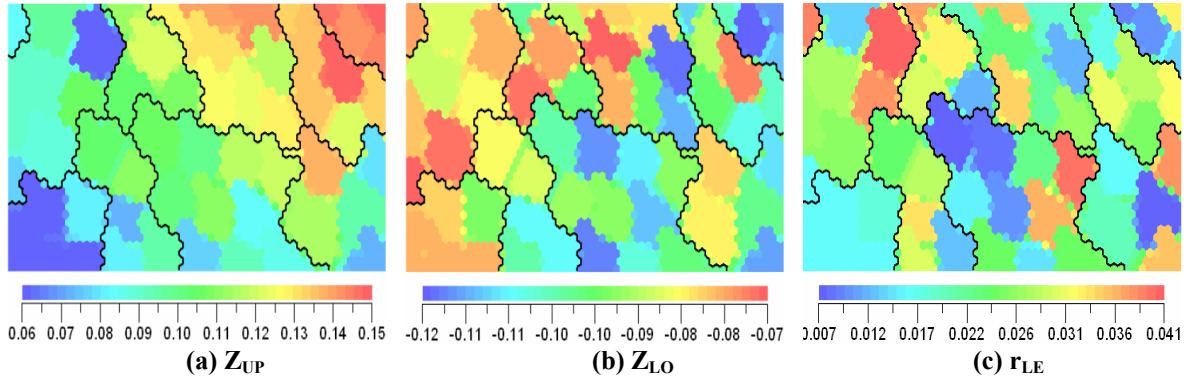
Figure 17 SOMs colored by design variables in the initial search region

Figure 17 shows the SOMs colored by three design variables. In Fig. 17(a), the clusters having 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 of color is similar to that of SOM colored by C_d. In Fig. 17(b), though the distribution of color is not so clear, it is slightly similar to that of SOM colored by C_l. In Fig 17(c), colored by r_{LE}, there is no noticeable color pattern. From the results, Z_{UP} and Z_{LO} have some effect on C_d and C_l, respectively and r_{LE} give little effect on C_l and C_d. The results coincide with results of ANOVA applied for the same search region.
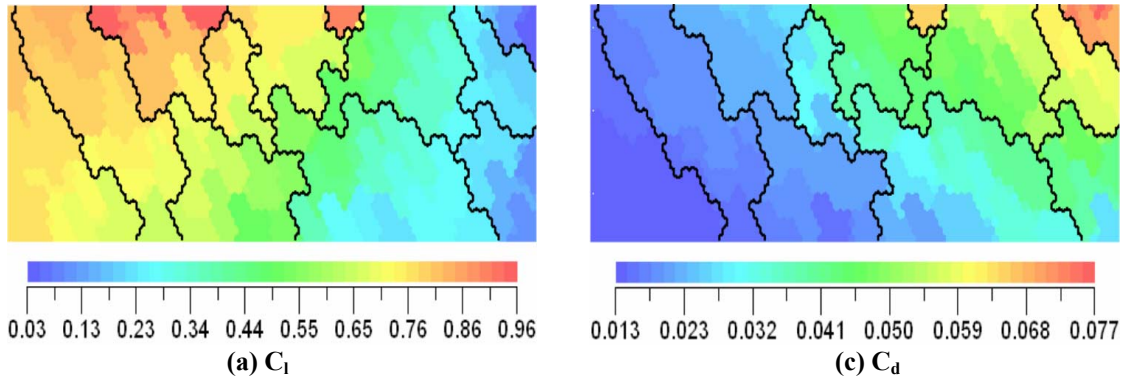

(a) C_l   **(c) C_d**

Figure 18 SOMs colored by objective function in the final search region

SOM is also applied to the final design space with 110 data. Figure 18 shows the SOMs colored by C_l and C_d in the final search region. In Fig. 18(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. 18(b), the clusters having large C_d values are located in the right side and the cluster having small C_d values are located in the left side. From these SOMs, 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_l performances are not so good. Figure 19 shows the SOM colored by the generated order. The data, whose generated order is first to 50th, are generated by the Latin hypercube design and the data, whose generated order is later than 50th, are generated through the optimization

Figure 19 SOM colored by generated order

process. The color distribution of this map is similar with that of Fig 18(a). The clusters with high generated order have relatively good C_l and C_d performances. This means that the optimization method used in this investigation works well.
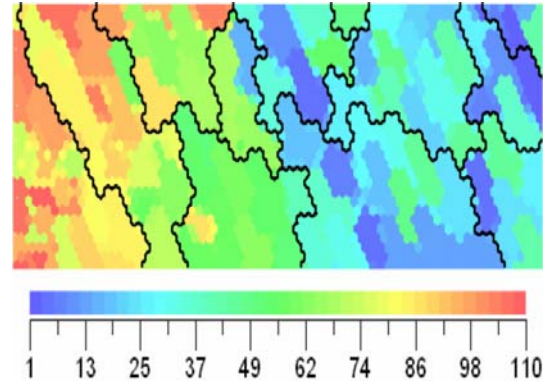
Figure 20 shows the SOMs colored by four design variables. The color pattern of Fig. 20(a) is similar to that of Fig. 18(b). The color distributions of Fig. 20(b) and Fig. 20(c) are similar to that of Fig. 18(a). This means that small Z_{UP} value is related with a good C_d performance of airfoil. In Fig. 20(d), clusters having small r_{LE} values are located in the left-hand side. This is the same with Fig. 18(b). Small r_{LE} value is related with a good C_d performance. The results coincide with the results of ANOVA in the final search region.
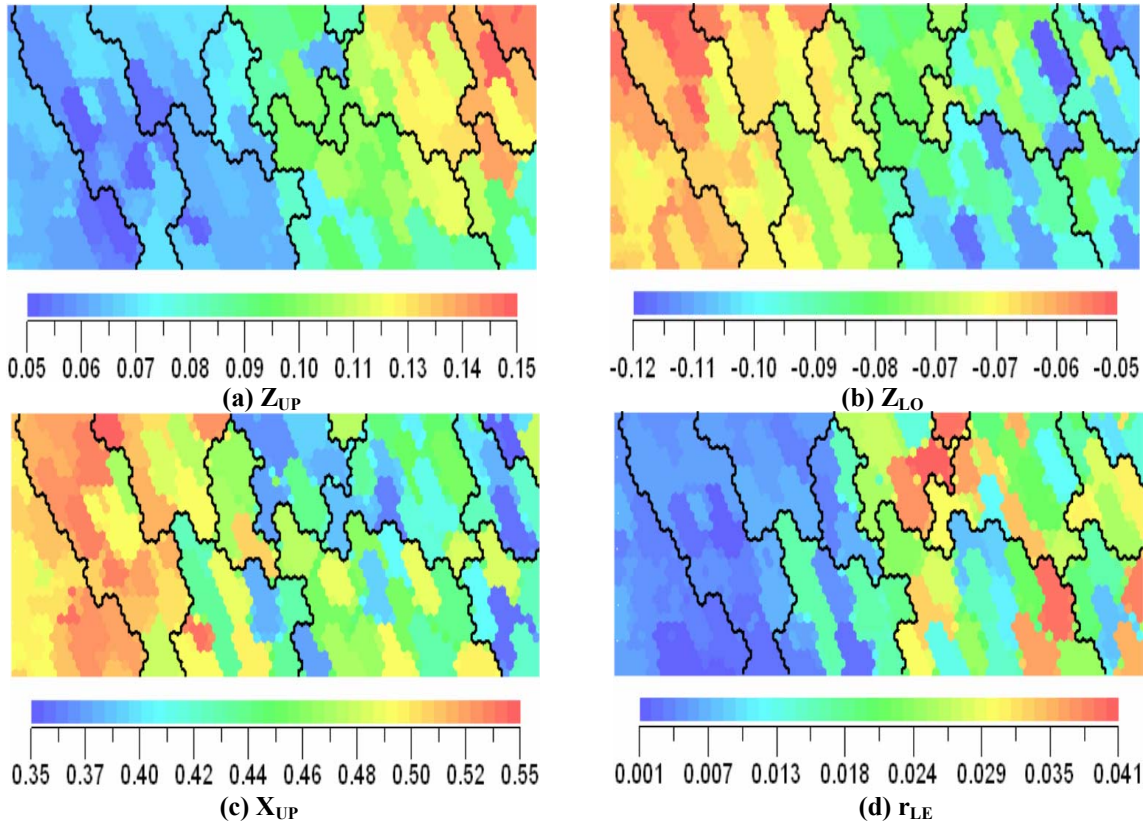

(a) Z_{UP}

(b) Z_{LO}

(c) X_{UP}

(d) r_{LE}

Figure 20 SOMs colored by design variables in final search region

VI. Conclusion

In this paper, 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 variables on objective functions quantitatively and SOM shows the information qualitatively. Furthermore, ANOVA can show the effect of interaction of design variables on objective functions. On the other hand, SOM can show the trade-off between objective functions. The acquired information helps the designer to determine the final design from the non-domination solutions of multi-objective problems and also can be used to identify why the obtained optimum solution has a good performance. Furthermore, the information will makes it possible to simplify the design space by eliminating design variables which have little effect on the design problem.

However, as shown in transonic airfoil design case using the adaptive search region method, the acquired information depends on the definition of search region. In order to obtain the correct information, the search region should be selected carefully.

As a future work, the method of selecting the data for data mining should be investigated. Results of most data mining techniques largely depend on the data they used. For the consistency of information obtained from data mining, the robust data selection method is necessary.

References

[1] J. Reuther and A. Jameson, "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 2991, pp. 1011-1020.

[3] Coello Coello, C. A. 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] 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.

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

[8] 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.

[9] Mardia, K. V. and R. J. Marshall, "Maximum likelihood estimation of models for residual covariance in Spatial regression," *Biometrika* Vol. 71, pp. 135-146, 1984.

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

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

[12] 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 appeared)

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

[14] Sasaki, D. and Obayashi, S., "Efficient Search for Trade-Offs by Adaptive Range Multi-Objective Genetic Algorithms," Journal of Aerospace Computing, Information, And Communication, Vol. 2, 2005, pp. 44-64.

[15] 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.

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

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