Visualization and Data Mining of Pareto Solutions Using Self-Organizing Map

2 authors:

Shigeru Obayashi
Tohoku University
502 PUBLICATIONS   **5,772** CITATIONS

SEE PROFILE

Daisuke Sasaki
Kanazawa Institute of Technology
82 PUBLICATIONS   **861** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Design optimization of vortex generators for transonic aircraft View project

Visualization and Data Mining of Pareto Solutions Using Self-Organizing Map

Shigeru Obayashi and Daisuke Sasaki

Institute of Fluid Science, Tohoku University,
Sendai, 980-8577 JAPAN
obayashi@ieee.org, sasaki@reynolds.ifs.tohoku.ac.jp

Abstract. Self-Organizing Maps (SOMs) have been used to visualize tradeoffs of Pareto solutions in the objective function space for engineering design obtained by Evolutionary Computation. Furthermore, based on the codebook vectors of cluster-averaged values of respective design variables obtained from the SOM, the design variable space is mapped onto another SOM. The resulting SOM generates clusters of design variables, which indicate roles of the design variables for design improvements and tradeoffs. These processes can be considered as data mining of the engineering design. Data mining examples are given for supersonic wing design and supersonic wing-fuselage design.

1 Introduction

Multiobjective Evolutionary Algorithms (MOEAs) are getting popular in many fields because they will provide a unique opportunity to address global tradeoffs between multiple objectives by sampling a number of non-dominated solutions. To understand tradeoffs, visualization is essential. Although it is trivial to understand tradeoffs between two objectives, tradeoff analysis in more than three dimensions is not trivial as shown in Fig. 1. To visualize higher dimensions, Self-Organizing Map (SOM) by Kohonen [1,2] is employed in this paper.

SOM is one of neural network models. SOM algorithm is based on unsupervised, competitive learning. It provides a topology preserving mapping from the high dimensional space to map units. Map units, or neurons, usually form a two-dimensional lattice and thus SOM is a mapping from the high dimensions onto the two dimensions. The topology preserving mapping means that nearby points in the input space are mapped to nearby units in SOM. SOM can thus serve as a cluster analyzing tool for high-dimensional data. The cluster analysis of the objective function values will help to identify design tradeoffs.

Design is a process to find a point in the design variable space that matches with the given point in the objective function space. This is, however, very difficult. For example, the design variable spaces considered here have 72 and 131 dimensions, respectively. One way of overcoming high dimensionality is to group some of design variables together. To do so, the cluster analysis based on SOM can be applied again.

Based on the codebook vectors of cluster-averaged values of respective design variables obtained from the SOM, the design variable space can be mapped onto

another SOM. The resulting SOM generates clusters of design variables. Design variables in such a cluster behave similar to each other and thus a typical design variable in the cluster indicates the behaviour/role of the cluster. A designer may extract design information from this cluster analysis. These processes can be considered as data mining for the engineering design.

At first, SOM is applied to map objective function values of non-dominated solutions in four dimensions. This will reveal global tradeoffs between four design objectives. The multipoint aerodynamic optimization of a wing shape for SST at both supersonic and transonic cruise conditions has been performed by using MOEAs previously [3]. Both aerodynamic drags were to be minimized under lift constraints, and the bending and pitching moments of the wing were also minimized instead of imposing constraints on structure and stability. A high fidelity Computational Fluid Dynamics (CFD) code, a Navier-Stokes code, was used to evaluate the wing performance at both conditions. In this design optimization, planform shapes, camber, thickness distributions and twist distributions were parameterized in total of 72 design variables. To alleviate the required computational time, parallel computing was performed for function evaluations. The resulting 766 non-dominated solutions are analyzed to reveal tradeoffs in this paper. The resulting SOM is also used to create a new SOM of the cluster-averaged design variables.

Second, SOM is applied to map entire solutions evaluated during the evolution of two-objective optimization. Based on the wing design system mentioned above, an aerodynamic optimization system for SST wing-body configuration was developed in [4]. To satisfy severe tradeoff between high aerodynamic performance and low sonic boom, the present objectives were to reduce C_D at a fixed C_L as well as to satisfy the equivalent area distribution for low boom design proposed by Darden [5]. Wing shape and fuselage configuration were defined in total of 131 design variables. The SOM of the objective function values indicates the non-dominated front as edges of the map and the SOM of the cluster-averaged design variables reveals the role of the design variables for design tradeoffs.
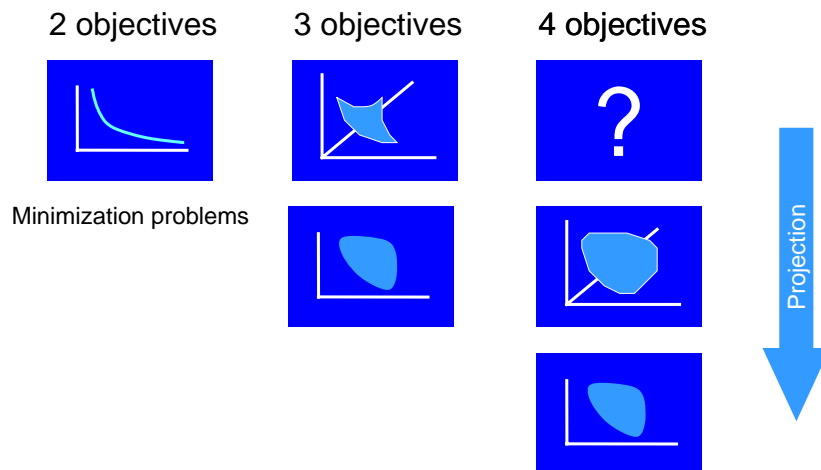
Fig. 1. Visualization of Pareto front

2 Evolutionary Multiobjective Optimization

2.1 MOGAs

The genetic operators used here are based on MOGAs [6,7]. Selection is based on the Pareto ranking method and fitness sharing. Each individual is assigned to its rank according to the number of individuals that dominate it. A fitness sharing function is used to maintain the diversity of the population. To find non-dominated solutions more effectively, the so-called best-N selection is employed.

For real function optimizations like the present research, however, it is more straightforward to use real numbers for encoding. Thus, the floating-point representation is used here. Accordingly, blended crossover (BLX-α) [8] is adopted at the crossover rate of 100%. This operator generates children on a segment defined by two parents and a user specified parameter α. The disturbance is added to new design variables within 10% of the given range of each design variable at a mutation rate of 20%. Crossover and mutation rates are kept high because the best-N selection gives a very strong elitism. Details for the present MOGA were given in Refs. 3, 4 and 7.

2.2 CFD Evaluation

To evaluate the design, a high fidelity Euler/Navier-Stokes code was used. Taking advantage of the characteristic of GAs, the present optimization is parallelized on SGI ORIGIN2000 at the Institute of Fluid Science, Tohoku University. The system has 640 Processing Elements (PE's) with peak performance of 384 GFLOPS and 640 GB of memory.

A simple master-slave strategy was employed: The master PE manages the optimization process, while the slave PE's compute the Navier-Stokes code. The parallelization became almost 100% because almost all the CPU time was dominated by CFD computations. The population size used in this study was set to 64 so that the process was parallelized with 32-128 PE's depending on the availability of job classes. The present optimization requires about six hours per generation for the supersonic wing case when parallelized on 128 PE's.

2.3 Neural Network and SOM

SOM [1,2] is a two-dimensional array of neurons:

$$\mathbf{M} = \left\{ \mathbf{m}_1 \quad \cdots \quad \mathbf{m}_{p \times q} \right\} \tag{1}$$

One neuron is a vector called the codebook vector:

$$\mathbf{m}_i = \left[m_{i_1} \quad \cdots \quad m_{i_n} \right] \tag{2}$$

This has the same dimension as the input vectors (*n*-dimensional). The neurons are connected to adjacent neurons by a neighbourhood relation. This dictates the topology, or the structure, of the map. Usually, the neurons are connected to each other via rectangular or hexagonal topology. One can also define a distance between the map units according to their topology relations.

The training consists of drawing sample vectors from the input data set and "teaching" them to SOM. The teaching consists of choosing a winner unit by means of a similarity measure and updating the values of codebook vectors in the neighbourhood of the winner unit. This process is repeated a number of times.

In one training step, one sample vector is drawn randomly from the input data set. This vector is fed to all units in the network and a similarity measure is calculated between the input data sample and all the codebook vectors. The best-matching unit is chosen to be the codebook vector with greatest similarity with the input sample. The similarity is usually defined by means of a distance measure. For example in the case of Euclidean distance the best-matching unit is the closest neuron to the sample in the input space.

The best-matching unit, usually noted as $\mathbf{m}_c$, is the codebook vector that matches a given input vector $\mathbf{x}$ best. It is defined formally as the neuron for which

$$\|\mathbf{x} - \mathbf{m}_c\| = \min_i \left\{\|\mathbf{x} - \mathbf{m}_i\|\right\} \tag{3}$$

After finding the best-matching unit, units in SOM are updated. During the update procedure, the best-matching unit is updated to be a little closer to the sample vector in the input space. The topological neighbours of the best-matching unit are also similarly updated. This update procedure stretches the best-matching unit and its topological neighbours towards the sample vector. The neighbourhood function should be a decreasing function of time. In the following, SOMs were generated in the hexagonal topology by using Viscovery® SOMine 4.0 Plus [9].

2.4 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 [10].

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 (§3.2), while a large number will be used for generation of codebook vectors for respective design variables (§3.3).

3 Four-Objective Optimization for Supersonic Wing Design

3.1 Formulation of Optimization

Four objective functions used here are
1. Drag coefficient at transonic cruise, $C_{D,t}$
2. Drag coefficient at supersonic cruise, $C_{D,s}$
3. Bending moment at the wing root at supersonic cruise condition, M_B
4. Pitching moment at supersonic cruise condition, M_P

In the present optimization, these objective functions are to be minimized. The transonic drag minimization corresponds to the cruise over land; the supersonic drag minimization corresponds to the cruise over sea. Lower bending moments allow less structural weight to support the wing. Lower pitching moments mean less trim drag.

The present optimization is performed at two design points for the transonic and supersonic cruises. Corresponding flow conditions and the target lift coefficients are described as
1. Transonic cruising Mach number, $M_{\infty,t} = 0.9$
2. Supersonic cruising Mach number, $M_{\infty,s} = 2.0$
3. Target lift coefficient at transonic cruising condition, $C_{L,t} = 0.15$
4. Target lift coefficient at supersonic cruising condition, $C_{L,s} = 0.10$
5. Reynolds number based on the root chord length at both conditions, Re=1.0 x 10^7

Flight altitude is assumed at 10 km for the transonic cruise and at 15 km for the supersonic cruise. To maintain lift constraints, the angle of attack is computed for each configuration by using $C_{L\alpha}$ obtained from the finite difference. Thus, three Navier-Stokes computations per evaluation are required. During the aerodynamic optimization, wing area is frozen at a constant value.

Design variables are categorized to planform, airfoil shapes and the wing twist. Planform shape is defined by six design variables, allowing one kink in the spanwise direction. Airfoil shapes are composed of its thickness distribution and camber line. The thickness distribution is represented by a Bézier curve defined by nine polygons. The wing thickness is constrained for structural strength. The thickness distributions are defined at the wing root, kink and tip, and then linearly interpolated in the spanwise direction. The camber surfaces composed of the airfoil camber lines are defined at the inboard and outboard of the wing separately. Each surface is represented by the Bézier surface defined by four polygons in the chordwise direction and three in the spanwise direction. Finally, the wing twist is represented by a B-spline curve with six polygons. In total, 72 design variables are used to define a whole wing shape. A three-dimensional wing with computational structured grid and the corresponding CFD result are shown in Figs. 2 and 3. See Ref. 3 for more details for geometry definition and CFD information.
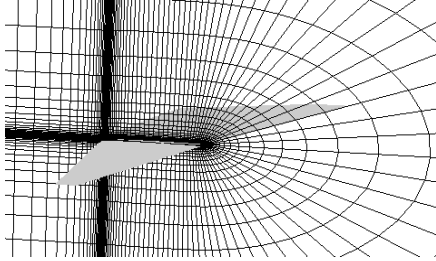
Fig. 2. Wing grid in C-H topology
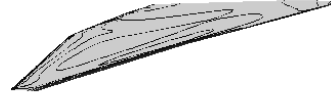
Fig. 3. Pressure contours on the upper surface of a wing computed by the CFD code

### 3.2 Visualization of Design Tradeoffs: SOM of Tradeoffs

The evolution was computed for 75 generations until all individuals become non-dominated. An archive of non-dominated solutions was also created along the evolution. After the computation, the 766 non-dominated solutions were obtained in the archive as a three-dimensional surface in the four-dimensional objective function space. By examining the extreme non-dominated solutions, the archive was found to represent the Pareto front qualitatively.

The present non-dominated solutions of supersonic wing designs have four design objectives. First, let's project the resulting non-dominated front onto the two-dimensional map. Figure 4 shows the resulting SOM with seven clusters. For better understanding, the typical planform shapes of wings are also plotted in the figure. Lower right corner of the map corresponds to highly swept, high aspect ratio wings good for supersonic aerodynamics. Lower left corner corresponds to moderate sweep angles good for reducing the pitching moment. Upper right corner corresponds to small aspect ratios good for reducing the bending moment. Upper left corner thus reduces both pitching and bending moments.

Figure 5 shows the same SOM contoured by four design objective values. All the objective function values are scaled between 0 and 1. Low supersonic drag region corresponds to high pitching moment region. This is primarily because of high sweep angles. Low supersonic drag region also corresponds to high bending moment region because of high aspect ratios. Combination of high sweep angle and high aspect ratio confirm that supersonic wing design is highly constrained.
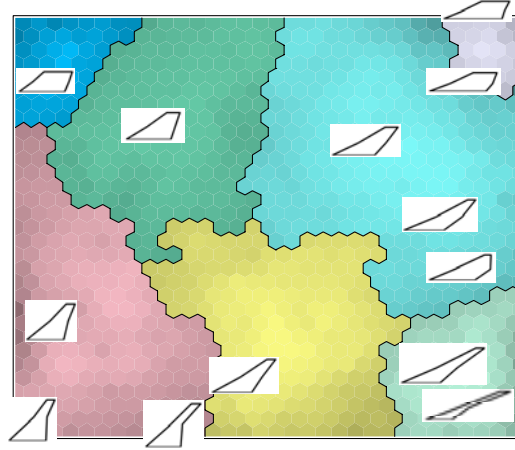
Fig. 4. SOM of the objective function values and typical wing planform shapes
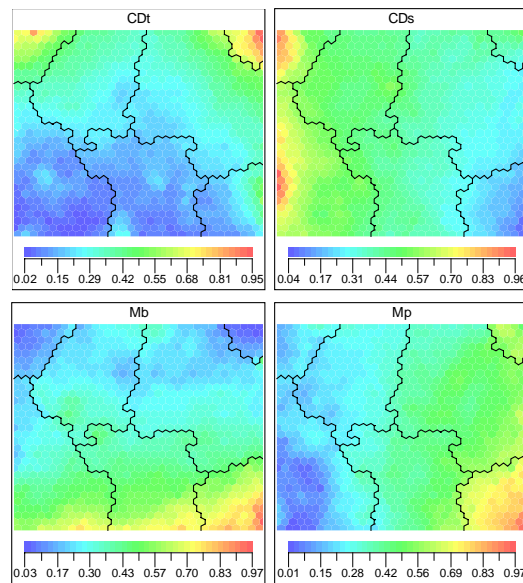
Fig. 5. SOM contoured by each design objective

3.3 Data Mining of Design Space: SOM of Design Variables

The previous SOM provides clusters based on the similarity in the objective function values. The next step is to find similarity in the design variables that corresponds to the previous clusters. To visualize this, the previous SOM is first revised by using larger number of clusters of 49 as shown in Fig. 6. Then, all the design variables are

averaged in each cluster, respectively. Now each design variable has a codebook vector of 49 cluster-averaged values. This codebook vector may be regarded to represent focal areas in the design variable space. Finally, a new SOM is generated from these codebook vectors as shown in Fig. 7.

This process can be done for encoded design variables (genotype) and decoded design variables (phenotype). In the earlier study, the genotype was used for SOM. However, the genotype and phenotype generated completely different SOMs. A possible reason is because the various scaling appears in phenotype, for example, one design variable is between 0 and 1 and another is between 35 to 70. The difference of order of magnitude in design variables may lead to different clusters. To avoid such confusion, the genotype is used for SOM here.

In Fig. 7, the labels indicate 72 design variables. DVs 00 to 05 correspond to the planform design variables. These variables have dominant influence on the wing performance. DVs 00 and 01 determine the span lengths of the inboard and outboard wing panels, respectively. DVs 02 and 03 correspond to leading-edge sweep angles. DVs 04 and 05 are root-side chord lengths. DVs 06 to 25 define wing camber. DVs 26 to 32 determine wing twist. Figure 7 contains seven clusters and thus seven design variables are chosen from each cluster as indicated. Figure 8 shows SOM's of Fig. 4 contoured by these design variables.

The sweep angles, DVs 02 and 03, make a cluster in the lower left corner of the map in Fig. 7 and the corresponding plots in Fig. 8 confirm that the wing sweep has a large impact on the aerodynamic performance. DVs 11 and 51 in Fig. 8 do not appear influential to any particular objective. By comparing Figs. 8 and 5, DV 01 has similar distribution with the bending moment Mb, indicating that the wing outboard span has an impact on the wing bending moment. On the other hand, DV 00, the wing inboard span, has an impact on the pitching moment. DV 28 is related to transonic drag. DV 04 and 05 are in the same cluster. Both of them have an impact on the transonic drag because their reduction means the increase of aspect ratio. Several features of the wing planform design variables and the corresponding clusters are found out in the SOMs and they are consistent with the existing aerodynamic knowledge.
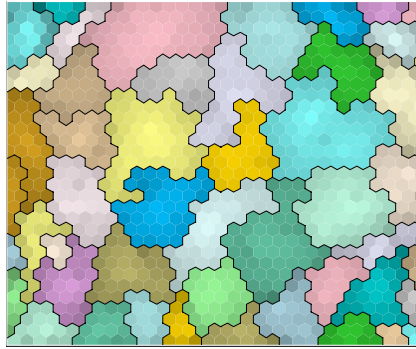
Fig. 6. SOM of objective function values with 49 clusters
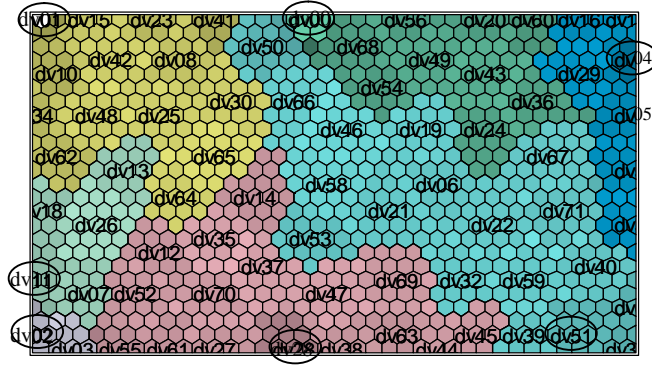
Fig. 7. SOM of cluster-averaged design variables
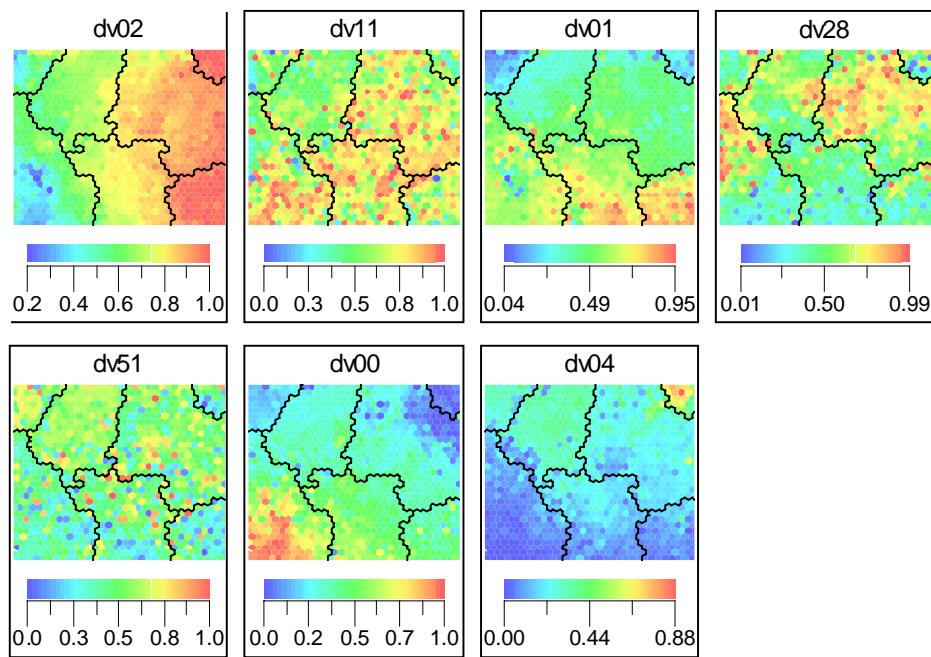
Fig. 8. SOM contoured by design variables selected from clusters in Fig. 7

4 Two-Objective Optimization for Supersonic Wing-Fuselage Design

4.1 Formulation of Optimization

In this study, SST wing-body configurations are designed to improve the aerodynamic performance and to lower the sonic boom strength. Therefore, design objectives are to reduce C_D at Mach number 2.0 at a fixed C_L (=0.10) and to match Darden's equivalent area distribution that can achieve low sonic boom. Multiblock Euler calculation was used to evaluate aerodynamic performance [11]. For the evaluation of sonic boom strength, an equivalent area distribution is matched to Darden's equivalent area distribution for 300 ft fuselage SST at Mach number 1.6 at $C_L = 0.125$.

To evaluate aerodynamic performances, aerodynamic evaluation has to be automatically performed for a given SST wing-body configuration. The wing definition was almost same as the previous wing optimization. Then, 55 additional design variables were used to define nonsymmetric fuselage configuration. Four more design variables represented the wing lofting. The total number of design variables is 131.

As body length and wing area is fixed to 300 ft and 9,000 ft^2, respectively, body volume, minimum diameter of body and wing volume must be greater than values given in Table 1. The other constraints are implemented to design variables as boundaries. As a result, the present SST wing-body design problem has two objective functions of minimization, three constraints and 131 design variables, and is optimized by real-coded MOGAs. Master-slave type parallelization was again performed to reduce the large computational time of each CFD evaluation in the optimization process. Figures 9 and 10 show typical computational grid and corresponding CFD result, respectively. See Ref. 4 for more details for geometry definition and CFD information.

Table 1. Constraints of SST wing-body configuration

Body volume $\geq$ 30,000 ft^3
Minimum diameter $\geq$ 11.8 ft (0.23$\leq$x/L$\leq$0.70)
Wing volume $\geq$ 16,800 ft^3
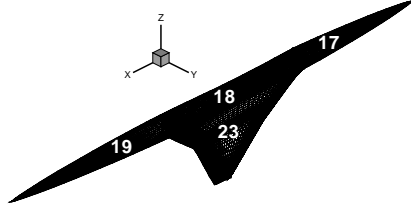
Fig. 9. Surface grid for SST wing-fuselage configuration (numbers indicate corresponding multiblock grids)
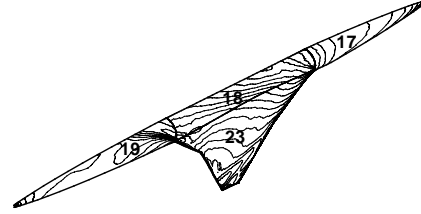
Fig. 10. Computed pressure distribution on the upper surface of SST wing-fuselage configuration

4.2 Visualization of Design Tradeoffs: SOM of Function Landscape

First, all the solutions obtained during the present evolutionary computation were mapped onto SOM according to the scaled objective function values. The resulting SOM is shown in Fig. 11. Several non-dominated solutions are indicated by * in the figure. The map consists of eight clusters. The lower left cluster contains the extreme non-dominated solution of the minimum drag. The upper right cluster contains the extreme non-dominated solution of the minimum boom. The corresponding objective functions values are then plotted in Fig. 12. Because only two objectives are used here, the map coordinates approximately matches to the objectives. The vertical direction corresponds to the drag and the horizontal axis corresponds to the sonic boom. The lower edge and the right edge of the map indicate the non-dominated front. Although the mapping is not essential to visualize tradeoffs here, the cluster analysis may be used to generate clusters of design variables.
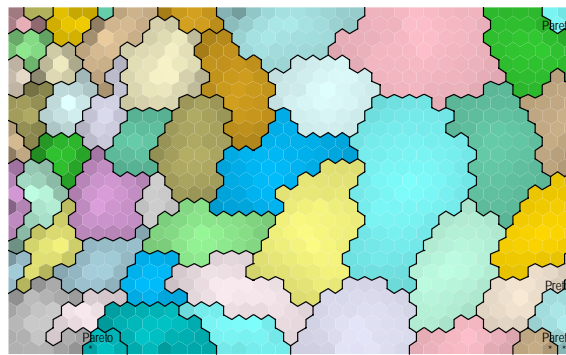

**Fig. 11.** SOM of the objective function values
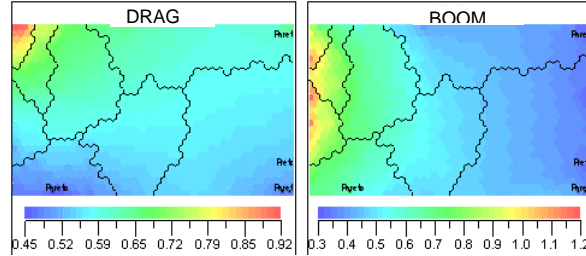
Fig. 12. SOM coloured by each design objective

### 4.3 Data mining of Design Space: SOM of Design Variables

To generate SOM of the design variables, Fig. 11 was divided into 50 clusters as Fig. 13. Then, Fig. 14 was generated from codebook vectors of cluster-averaged design variables in Fig. 13. Figure 14 shows SOM of the design variables in five clusters. In Fig. 14, the labels indicate 131 design variables. Figure 14 can be interpreted from the behaviors of the design variables representing the corresponding clusters. Figure 15 shows the map of Fig. 11 contoured by the five design variables indicated in Fig. 14. A trend of the design variables in the left cluster of Fig. 14 is represented by DV 123 in Fig. 15. Its distribution appears the inverse of the sonic boom in Fig. 12. DV 123 determines the twist angle at the wing tip. It has an impact on the list distribution, leading to influences on the equivalent cross sectional distribution and thus on the sonic boom strength. The center cluster in Fig. 14 is represented by DV 2 and its distribution in Fig. 15 appears the inverse of the drag in Fig. 12. DV 2 is one of the design variables that define the sharpness of the nose of the fuselage. Blunt nose is known to increase drag for supersonic aircraft. The right cluster in Fig. 14 is represented by DV 28 and the corresponding distribution in Fig. 15 has a local minimum in the middle of the left, upper edge of the map. This is one of the design variables that determine the body radius distribution at the side of the fuselage, but it does not seem primarily related to either objective here. DV's 89 and 91 have opposite trends, but they are not influential to the non-dominated front, either.

Fig. 13. SOM of objective function values with 50 clusters
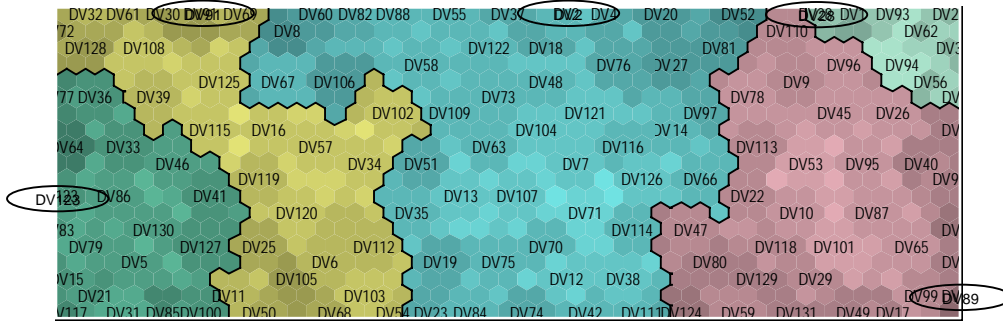
Fig. 14. SOM of cluster-averaged design variables
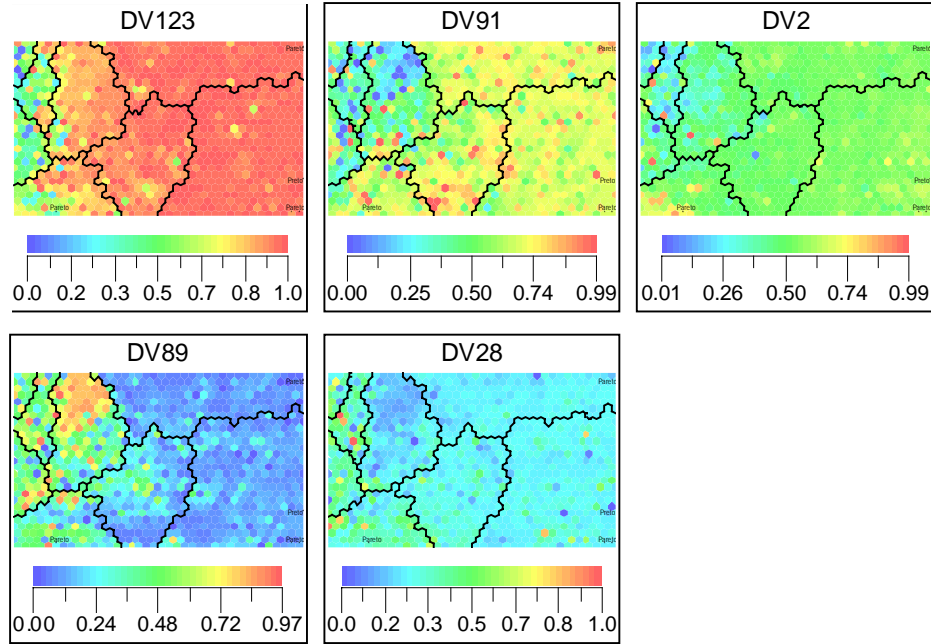
Fig. 15. SOM contoured by design variables selected from clusters in Fig. 14

5 Concluding Remarks

Design tradeoffs have been investigated for two multiobjective aerodynamic design problems of supersonic transport by using visualization and cluster analysis of the non-dominated solutions based on SOMs. The first optimization is to design

supersonic wings defined by 72 design variables with four objectives to be minimized. The second optimization is to design supersonic wing-body configurations represented by in total 131 design variables with drag and boom minimization. Design data were gathered by MOGAs.

SOM is first applied to visualize tradeoffs between design objectives. In the first design case, four objective functions were employed and 766 non-dominated solutions were obtained. Three-dimensional non-dominated front in the objective function space has been mapped onto the two-dimensional SOM where global tradeoffs are successfully visualized. In the second design case, entire solutions during the evolution have been mapped onto SOM to visualize function landscape, and the non-dominated front was found at the edges of the map. The resulting SOMs are further contoured by each objective, which provides better insights into design tradeoffs.

Furthermore, based on the codebook vectors of cluster-averaged values for respective design variables obtained from the SOMs, the design variable space is mapped onto another SOM. Design variables in the same cluster are considered to have similar influences in design tradeoffs. Therefore, by selecting a member (design variable) from a cluster, the original SOM in the objective function space is contoured by the particular design variable. It reveals correlation of the cluster of design variables with objective functions and their relative importance. Because each cluster of design variables can be identified influential or not to a particular design objective, the optimization problem may be divided into subproblems where the optimization will be easier to lead to better solutions.

These processes may be considered as data mining of the engineering design. The present work demonstrates that MOGAs and SOMs are versatile design tools for engineering design.

References

1   Kohonen T.: *Self-Organizing Maps*. Springer, Berlin, Heidelberg (1995)
2   Hollmen J.: Self-Organizing Map, http://www.cis.hut.fi/~jhollmen/dippa/node7.html, last access on October 3, 2002
3   Sasaki D., Obayashi S. and Nakahashi K.: Navier-Stokes Optimization of Supersonic Wings with Four Objectives Using Evolutionary Algorithm. Journal of Aircraft Vol.39, No.4 (2002) 621-629
4   Sasaki D., Yang G. and Obayashi S.: Automated Aerodynamic Optimization System for SST Wing-Body Configuration. AIAA Paper 2002-5549 (2002)
5   Darden, C. M.: Sonic Boom Theory: Its Status in Prediction and Minimization. Journal of Aircraft, Vol.14, No.6 (1977) 569-576
6   Fonseca C. M. and Fleming P. J.: Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization. Proc. of the 5th ICGA (1993) 416-423
7   Obayashi S., Takahashi S. and Takeguchi Y.: Niching and Elitist Models for MOGAs. Parallel Problem Solving from Nature – PPSN V, Lecture Notes in

Computer Science, Springer, Vol.1498, Berlin Heidelberg New York (1998) 260-269

8   Eshelman L. J. and Schaffer J. D.: Real-Coded Genetic Algorithms and Interval Schemata. Foundations of Genetic Algorithms 2, Morgan Kaufmann Publishers, Inc., San Mateo (1993) 187-202

9   Eudaptics software gmbh. http://www.eudaptics.com/technology/somine4.html, last access on October 3, 2002

10   Vesanto, J. and Alhoniemi, E.: Clustering of the Self-Organizing Map, IEEE Transactions on Neural Networks, Vol.11, No.3 (2000) 586 -600

11   Yang, G., Kondo, M. and Obayashi, S.: Multiblock Navier-Stokes Solver for Wing/Fuselage Transport Aircraft. JSME International Journal, Series B, Vol.45, No.1 (2002) 85-90