JOURNAL OF SPACECRAFT AND ROCKETS
Vol. 45, No. 5, September–October 2008

Knowledge Discovery for Flyback-Booster Aerodynamic Wing Design Using Data Mining

Kazuhisa Chiba*
Japan Aerospace Exploration Agency, Tokyo 182-8522, Japan
and
Shigeru Obayashi†
Tohoku University, Sendai 980-8577, Japan

DOI: 10.2514/1.28511

Data mining has been performed on the results of an aerodynamic design optimization of a two-stage-to-orbit reusable launch vehicle flyback-booster wing. Three data mining techniques were compared, including self-organizing map, functional analysis of variance, and the rough set theory. The optimization problem had four aerodynamic objective functions and 71 wing shape design variables. The hypothetical design database resulting from the optimization contained a total of 302 solutions which included 102 nondominated solutions. Consequently, the acquired knowledge of the design space consisted of general design characteristics, correlation between objective functions, and the effects of these design variables on the objective functions, for both nondominated as well as all solutions. The comparison also revealed the similarities and differences among the three data mining techniques used in this study. Even though all three techniques discovered detailed design knowledge and the results produced by the combination of all three methods compensated disadvantages of each method when applied individually, it was discovered that the self-organizing map produced the overall best results. Moreover, this study has also shown that the knowledge acquired from both nondominated solutions and from all solutions found was consistent despite the differences between the design spaces. Furthermore, it was shown that data mining is essential for visualizing results of an evolutionary multi-objective optimization problem and extracting useful design knowledge from these results.

Nomenclature

A	=	attribute
$\underline{A}X$	=	lower approximation of X using A
$\overline{A}X$	=	upper approximation of X using A
C	=	condition attribute
C_D	=	drag coefficient
C_L	=	lift coefficient
C_{M_p}	=	pitching-moment coefficient
C1, C2, ...	=	cluster 1, cluster 2, ...
D	=	decision attribute
dv1, dv2, ...	=	design variable 1, design variable 2, ...
F_1	=	objective function 1; absolute value of shift of aerodynamic center between supersonic and transonic flight conditions
F_2	=	objective function 2; absolute value of transonic C_{M_p}
F_3	=	objective function 3; transonic C_D
F_4	=	objective function 4; subsonic C_L
$R_A(x)$	=	indiscernibility set
S	=	design space
U	=	universe
X	=	subset
x	=	object
$\hat{y}$	=	estimated value of unknown function y

Presented as Paper 7992 at the 14th AIAA/AHI Space Planes and Hypersonic Systems and Technologies Conference, Canberra, Australia, 6–9 November 2006; received 24 October 2006; revision received 5 October 2007; accepted for publication 5 October 2007.

*Project Researcher, Aviation Program Group. Member AIAA.
†Professor, Institute of Fluid Science. Associate Fellow AIAA.

I. Introduction

ALTHOUGH solving design optimization problems is important for many disciplines of engineering, the most significant part of the process is the extraction of useful knowledge of the design space from results of optimization runs. The results produced by evolutionary multi-objective optimization algorithms are not individual optimal solutions but rather an entire set of solutions. That is, the result of a multi-objective optimization is not sufficient from the practical point of view as designers need a conclusive shape and not the entire selection of possible optimal shapes. On the other hand, this set of optimal solutions produced by an evolutionary multi-objective optimization algorithm can be considered a hypothetical design database. Now, data mining techniques can be applied to this hypothetical design database to acquire useful design knowledge [1,2].

In this study, three data mining techniques, including the self-organizing map (SOM), functional analysis of variance (ANOVA), and the rough set theory (RST) were applied to the aerodynamic design optimization problem of a two-stage-to-orbit (TSTO) reusable launch vehicle (RLV) flyback booster. This problem addresses the current need of a space transport system with a substantially reduced cost to extend the space utilization. One of the focused areas of research is the RLV system [3,4], suggested as a replacement for present expendable launch vehicle systems. Because of the difficult requirements, such as a higher performance propulsion system and greater reduction of its structural weight, current proposals for the introduction of reusable components in space transportation involve the TSTO configuration with winged flyback booster powered by liquid rocket engines for vertical takeoff and horizontal landing [5]. As the wing geometry of the flyback booster generates the aerodynamic performance, it is one of the most important elements during the return phase. Therefore, the correlations among aerodynamic characteristics, such as lift, drag, and moment, are important design parameters. Moreover, it is important to find design variables sensitive to the aerodynamic performance; that is, acquisition of knowledge in the design space is essential to improve the aerodynamic performance of the winged flyback booster. The objective of this study is to gain useful

knowledge of the design space by applying data mining techniques. As an added benefit of this study, a comparison of advantages and disadvantages of the applied data mining techniques has been conducted and reported.

II. Design Optimization Problem

The reference mission for the considered design optimization problem is the TSTO RLV to transport a 10-t payload into low earth orbit, which is similar to the present H-IIA mission [6]. The booster sizing is obtained by preliminary computation using the set of empirically derived equations. In the formulation of the design optimization problem, the fuselage geometry was assumed fixed with a given size and only the wing shape was allowed to be optimized.

Trajectory analysis around a typical TSTO configuration based on the present mission [7] showed that the separation of the booster and orbiter took place roughly at a Mach number of 3.0. Then, the flyback booster turned over, slowed down, cruised at transonic speeds, and landed at subsonic speed, shown in Fig. 1. Note that the major part of its cross range was in the transonic region. The four considered objective functions were to minimize the shift of the aerodynamic center between supersonic and transonic conditions F_1, transonic C_{M_p} F_2, transonic C_D F_3, as well as to maximize subsonic C_L F_4.

In this study, TAS-Code [8,9], based on the three-dimensional Reynolds-averaged Navier–Stokes computation using the modified Spalart–Allmaras one-equation model [10] on unstructured hybrid mesh, was employed in the aerodynamic evaluation. Adaptive range multi-objective genetic algorithm (ARMOGA) [11] was used as an optimizer.

The design variables were related to the planform, the airfoil shape, the wing twist, and the position relative to the fuselage. The wing planform was determined by five design variables shown in Fig. 2. A kink was placed on the leading edge. Airfoil shapes were defined at the wing root, kink, and the tip using thickness distributions and camber lines. The thickness distributions were described by Bézier curves using 11 control points and linearly interpolated in the spanwise direction. The camber line distributions were parameterized using Bézier curves with four control points and incorporated linearly in the spanwise direction. Wing twist was refined using B splines with six control points. The position of the wing root relative to the fuselage was parameterized by x and z coordinates of the leading edge, angle of attack, and dihedral angle. The entire wing shape was thus defined using 71 design variables, summarized in Table 1. Once the wing was defined, the junction line between the wing and fuselage was extracted and the final wing-fuselage geometry was derived by omitting the part of the wing inside the fuselage.

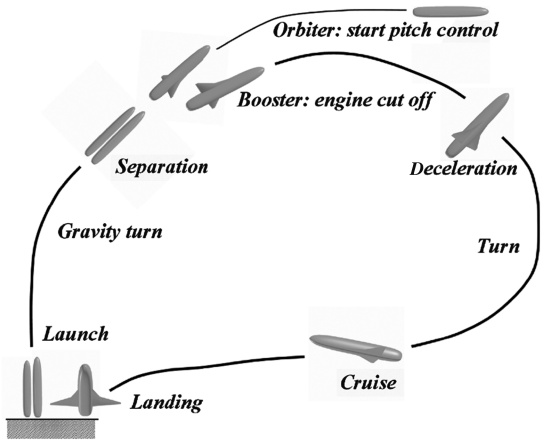


Fig. 1 Typical flight sequence for TSTO flyback booster.

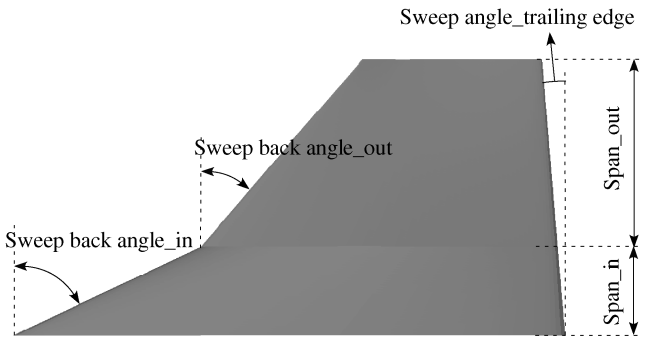


Fig. 2 Wing planform shape definition along with some of the major design parameters.

Table 1 Detail of design variables

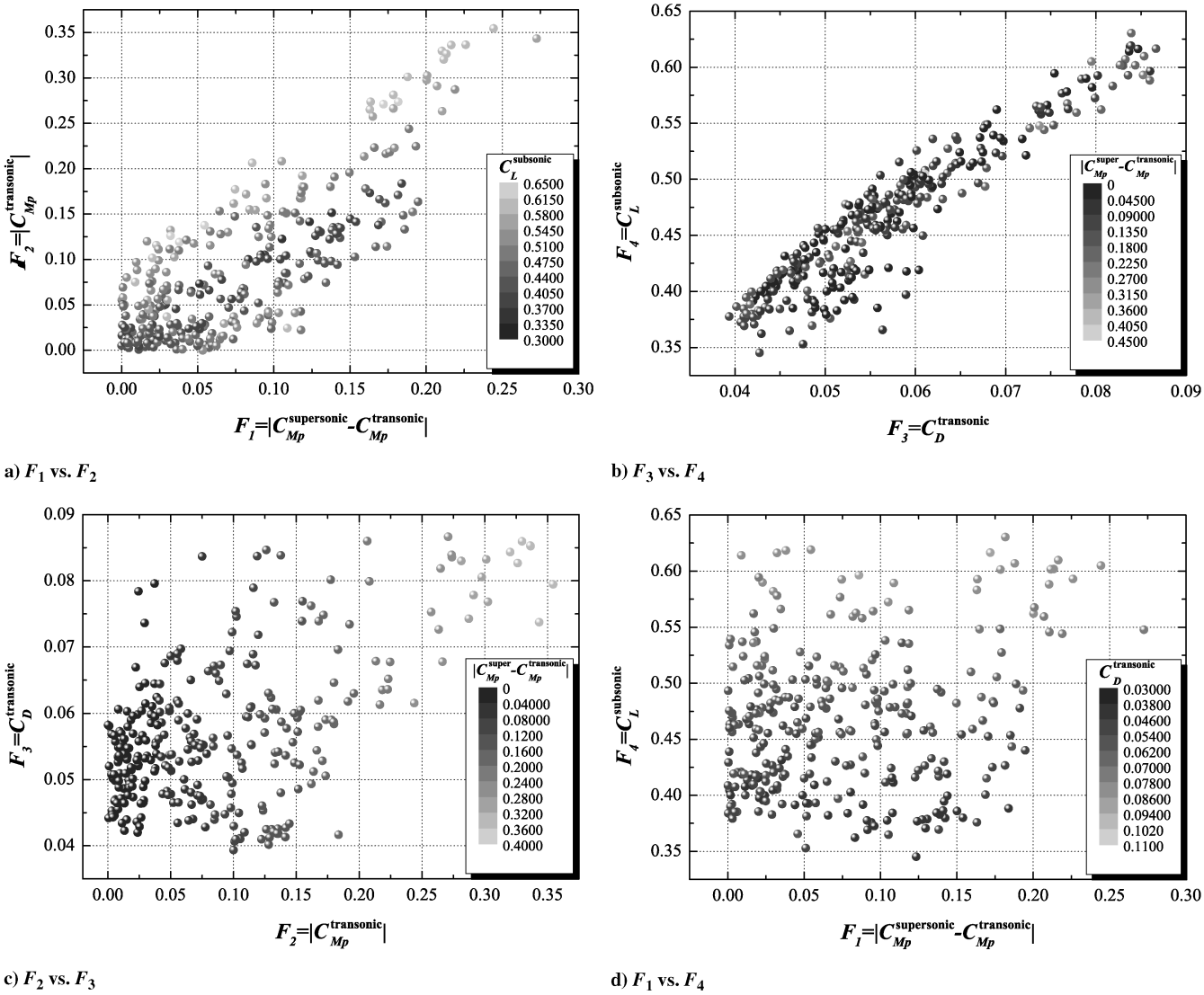
Serial number	Correspondent design variable	
1, 2	Span length	Inner, outer
3, 4	Sweepback angle of leading edge	Inner, outer
5	Sweep angle of trailing edge	——
6	Wing size relative to fuselage	——
7, 8	Wing position relative to fuselage	x, z
9	Angle of attack of wing relative to fuselage	——
10	Dihedral	——
11–22	Airfoil camber line	Root, kink, tip
23–64	Airfoil thickness distribution	Root, kink, tip
65–71	Wing twist	——

In the ARMOGA algorithm, the population size was set to eight and the evolutionary runs were terminated after 40 generations. Consequently, a total of 102 nondominated solutions extracted from all 302 solutions of 40 generations were obtained. Although single-objective problems may have a unique optimal solution, multi-objective problems have a set of optimal solutions, largely known as the tradeoff, Pareto-optimal solutions, or nondominated solutions (That is, all solutions are composed of a set of nondominated solutions and a set of dominated solutions.) These solutions are optimal in the sense that no other solutions in the search space are superior to them when all objectives are considered. Figures 3 and 4, respectively, show the four two-dimensional projections of all solutions and the nondominated solutions to better understand the tradeoffs among the four objective functions. The optimal values of F_1 and F_2 are zero, and the nondominated solutions reached the origin, that is, the optimal values of F_1 and F_2 in Fig. 4a. As the plots in Fig. 4a represent the nondominated solutions for all four objective functions, it is clear that there is a tradeoff surface spread among the nondominated solutions near the origin. Figure 4a shows that there is no tradeoff between the shift of the aerodynamic center and the transonic C_{Mp}. The Pareto front can be seen clearly between F_3 and F_4 in Fig. 4b. Thus, Fig. 4b indicates that there is a tradeoff between the transonic C_D and the subsonic C_L. This result indicates that a high-lift device may be needed for an RLV booster for landing, similar to an aircraft. Although the Pareto front for F_2 attained the optimum front, the Pareto front for F_3 did not reach C_D of zero, shown in Fig. 4c. Thus, there is a slight tradeoff between F_2 and F_3, and the transonic C_D can be improved while the transonic C_{Mp} increases. Moreover, although the Pareto front for F_1 also attained the optimum front, the Pareto front for F_4 did not have an apparent limit, shown in Fig. 4d. Therefore, there was also a slight tradeoff between F_1 and F_4. This indicated that shift of the aerodynamic center and the transonic C_{Mp} was optimized simultaneously, while the subsonic C_L was reduced slightly [12].

III. Data Mining Techniques

A. Self-Organizing Map

The SOM [13] is an unsupervised learning, nonlinear projection algorithm from high- to low-dimensional space, employing neural



a) F_1 vs. F_2

b) F_3 vs. F_4

c) F_2 vs. F_3

d) F_1 vs. F_4

Fig. 3 All solutions on two-dimensional plane.

networks. 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. In this study, SOMs were generated by using commercial software Viscovery SOMine 4.0 Plus[‡] produced by Eudaptics, GmbH. Although SOMine is based on the general SOM concept and algorithm, it employs an advanced variant of unsupervised neural networks, that is, Kohonen's Batch SOM [14,15]. The algorithm consists of two steps that are iterated until no more significant changes occur: search of the best-matching unit for all input data and adjustment of weight vector near the best-matching unit. The trained SOM is systematically converted into visual information [16,17].

B. Functional Analysis of Variance

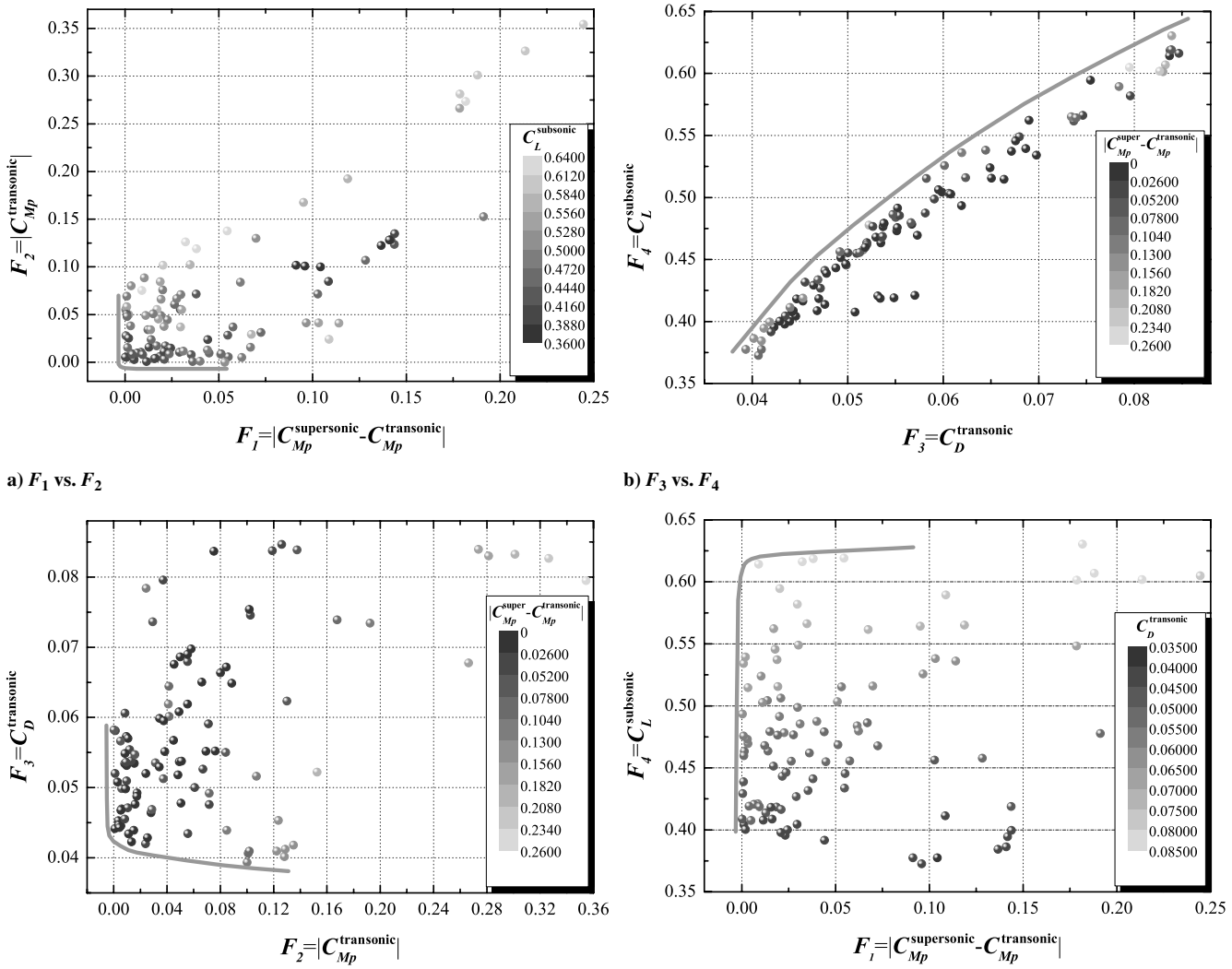
ANOVA [18,19] is a data mining technique showing the effect of each design variable to the objective and the constraint functions in a quantitative manner. ANOVA uses the variance of the model due to design variables on the approximation function. By decomposing the total variance of the model into the variance due to each design variable, the influence of each design variable on the objective function can be calculated. The decomposition is accomplished by integrating out the variables of model $\hat{y}$.

C. Rough Set Theory

RST is a new approach to address the issue of vagueness [20,21] which has become a topic of interest for many researchers in computer science. It is an important method for decision support systems and data mining tools. In fact, it is a new mathematical approach to analyze data.

The basic idea behind RST is to construct approximations of sets using the binary relation R_A. The indiscernibility sets $R_A(x)$ form basic building blocks from which subsets $X \subseteq U$ can be assembled. If X cannot be defined in a crisp manner using attributes A, we can circumscribe them through lower and upper approximations $\underline{A}X$ and $\overline{A}X$. The lower approximation consists of those objects that certainly belong to X, whereas the upper approximation consists of the objects that possibly belong to X. The boundary region is defined as the difference between the upper and the lower approximation, and consists of the objects that we cannot decisively assign as being either members or nonmembers of X. The outside region is defined as the complement of the upper approximation, and consists of the objects that are definite nonmembers. An RST is any subset $X \subseteq U$ defined through its lower and upper approximations. Figure 5 shows these ideas graphically. In the data mining by RST, as a minimum

[‡]"Eudaptics," available online at http://www.eudaptics.com [cited 16 June 2004].



a) F_1 **vs.** F_2 **b)** F_3 **vs.** F_4

c) F_2 **vs.** F_3 **d)** F_1 **vs.** F_4

Fig. 4 Derived nondominated solutions on two-dimensional plane.

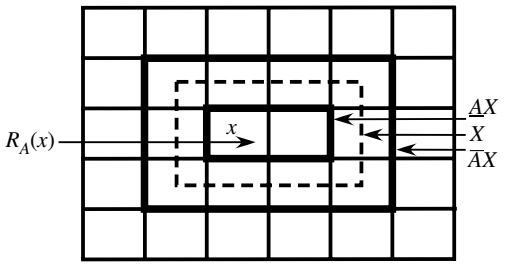


Fig. 5 Illustrated ideas of RST.

combination of attributes is found by using reduction to determine the decision attribute D, decision rules are extracted. In this study, ROSETTA [22][§] was used.

IV. Data Mining Results

A. Design Space Formed by Nondominated Solutions

The knowledge of the design space generated by nondominated solutions gives the overview of the tradeoffs between the objective functions and the influence of the design variables on these tradeoffs. As tradeoff information is in the hypothetical design database formed by the optimum solutions, the data mining methods can be used by designers to directly design knowledge.

1. *Knowledge Acquired by Using Self-Organizing Map*

The resulting 102 nondominated solutions were projected onto the two-dimensional map. Figure 6 shows the resulting SOM with 10 clusters, taking the four objective functions into consideration.

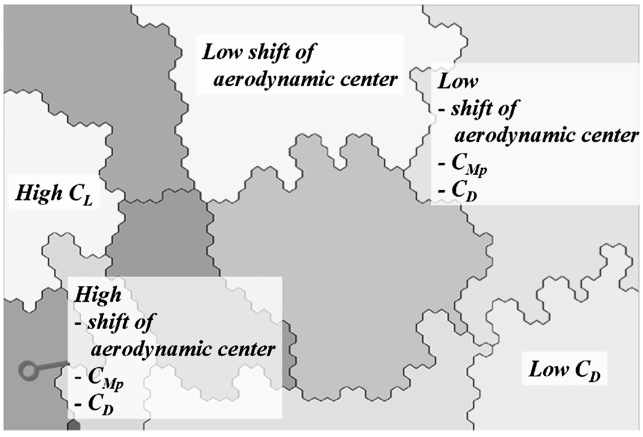


Fig. 6 SOM of the nondominated solutions in the four-dimensional objective function space.

[§]"ROSETTA Technical Reference Manual," available online at http://www.idi.ntnu.no/~aleks/thesis/ [cited 2 March 2006].

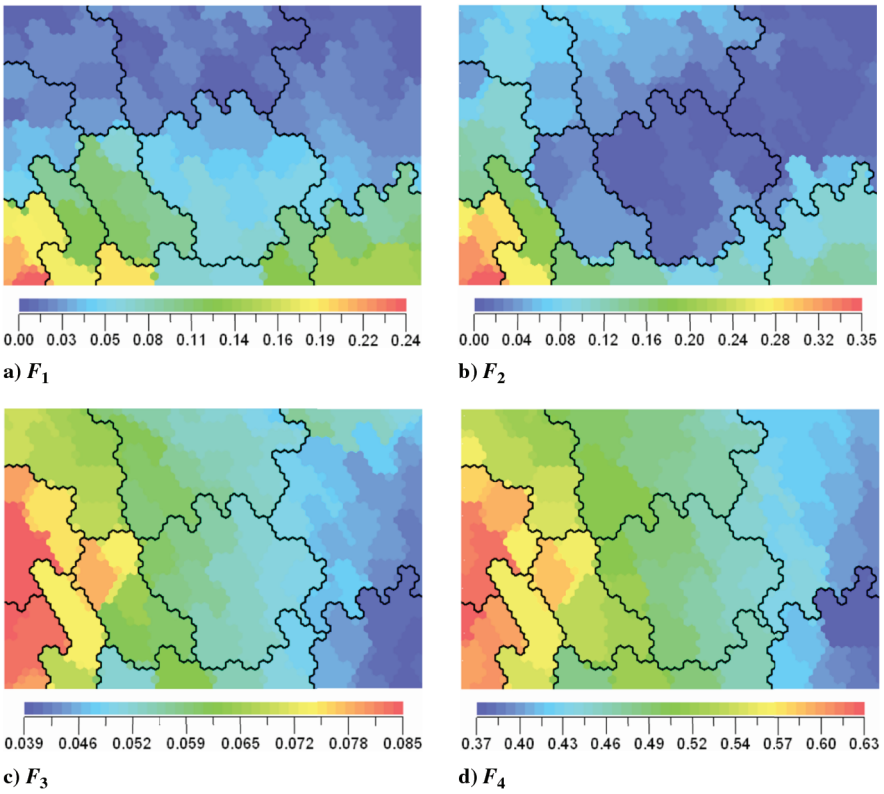


Fig. 7 SOM from the nondominated solutions colored by the four objective functions.

Figure 7 shows the SOMs colored by the four objective values, respectively. These colored figures show that the SOM can be grouped as follows: upper center area on the SOM corresponds to the designs with the low shift of aerodynamic center; upper right corner corresponds to the designs with the low shift of aerodynamic center, transonic C_{Mp}, and transonic C_D; lower right corner corresponds to the designs with the low transonic C_D; lower left corner corresponds to the high shift of aerodynamic center, transonic C_{Mp}, and transonic C_D; left center region corresponds to the high subsonic C_L.

The comparison between these colored SOMs reveals the tradeoffs and correlation between objective functions. Figures 7a and 7b show high-value regions for the shift of aerodynamic center and the transonic C_{M_p} coincide with each other. That is, when one objective function increases, another objective function is also strictly increased. Furthermore, because Fig. 7c is very similar to Fig. 7d, there is a strong tradeoff between transonic C_D and subsonic C_L. This knowledge from the SOM corresponds to the results of Fig. 4b.

The SOM can also be contoured by the 71 design-variable values. The SOM colored by the three important design variables and their sketches are shown in Fig. 8. Figure 8a shows the SOM colored by the design variable of the x coordinate of wing position to fuselage (dv7), illustrated in Fig. 7b. It is notable that "dv$x$" denotes the serial number of the design variable which corresponds to the number in Table 1. Here, the x coordinate is held on the fuselage. Higher values are located in the lower left corner in Figs. 7a and 7b. High values of the shift of the aerodynamic center, transonic C_{Mp}, and transonic C_D are clustered in this area. Thus, this means that the values of the shift of the aerodynamic center, transonic C_{Mp}, and transonic C_D become worse when the wing position is behind the fuselage.

Figure 8c shows the SOM colored by dv18, the rearward camber height at kink, illustrated in Fig. 8d. This map reveals that the solution with a higher value of the rearward camber height at kink has high subsonic C_L and also has high transonic C_D. Figure 8e shows the SOM colored by dv22, the rearward camber height at tip, illustrated in Fig. 8f. This figure reveals that the solution with a lower value of the rearward camber height at tip has low transonic C_D and also has low subsonic C_L. Because there is no design variable which is simply effective to subsonic C_L (or transonic C_D), there is a strong tradeoff between subsonic C_L and transonic C_D.

As the SOMs colored using several other design variables have no clear pattern, those design variables have no effect on determining tradeoffs among the four objective functions. That is, it means that the sorting of the design variables can be also performed from the SOM.

2. *Knowledge Acquired by Using Functional Analysis of Variance*

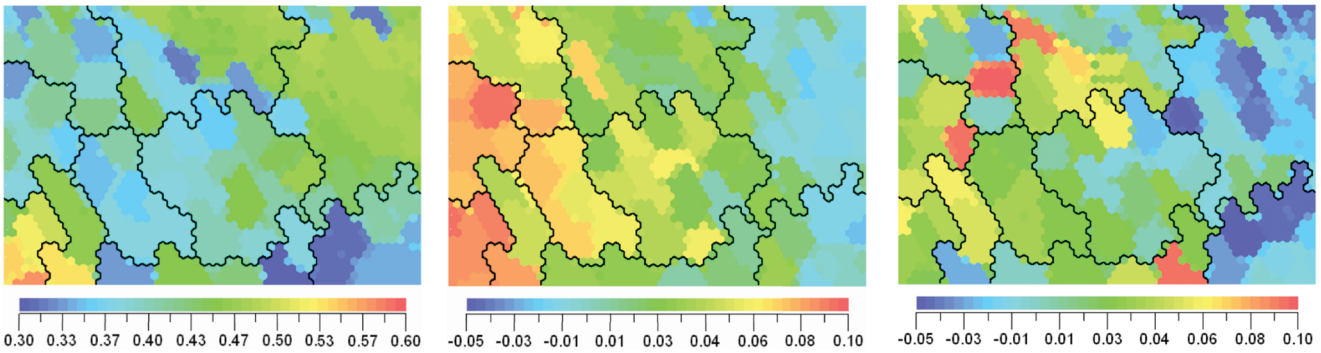
The variance of the design variables and their interactions are shown in Fig. 9. The proportion which is larger than 1% to the total variance is shown. Note that "dv" indicates design variable and "-" indicates interactions between two design variables.

The results reveal that dv7, which is the x coordinate of relative wing position to fuselage, gives the largest effect on F_1 and F_2, and dv18, which is the rearward camber height at wing tip, gives the largest influence for F_3 and F_4. When the wing position relative to the fuselage is shifted, varying of the aerodynamic center also breaks out, and this design variable gives the effect on the transonic C_{Mp}. In addition, the variation of the camber-line curvature generally gives effects on the aerodynamic performance of the wing. The knowledge obtained by ANOVA corresponds to general knowledge regarding aerodynamics.

When the results from ANOVA are compared with the results from the SOM, the influence regarding dv7 and dv18 closely match. However, the results from ANOVA do not show much influence regarding dv22. In this case, dv22, with a specific smaller value, influences the reduction of the transonic C_D. As the decrease of the dv22 value reduces the induced drag at tip, the result from the SOM is appropriate. Therefore, it is revealed that ANOVA works best for design variables having a global effect on an the objective function.

3. *Knowledge Acquired by Using Rough Set Theory*

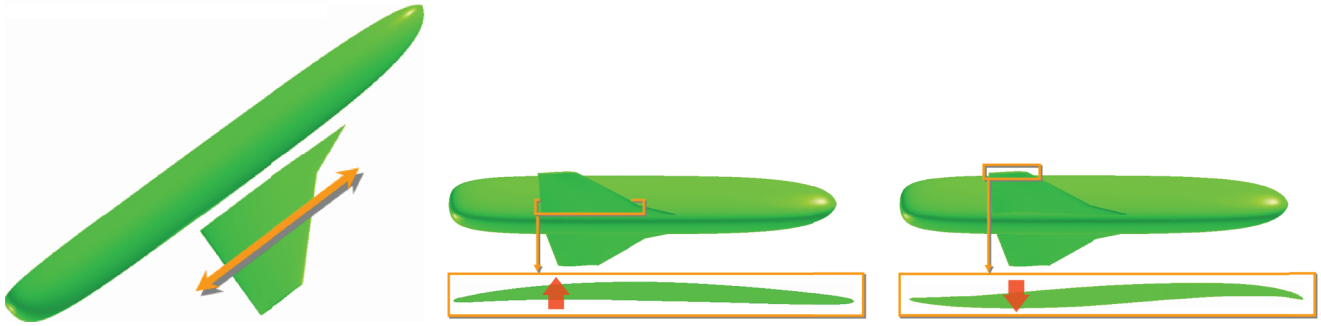
The procedure for data mining used by RST is summarized as follows: 1) preparation of data, 2) dispersion of data, 3) generation of *reduct* data (a reduct denotes a minimally sufficient subset of features [21]), 4) generation of rules, 5) filtering, and 6) construction of rules.



a) dv7 as the *x* coordinate of wing position to fuselage

c) dv18 as a rearward camber height at kink

e) dv22 as a rearward camber height at tip

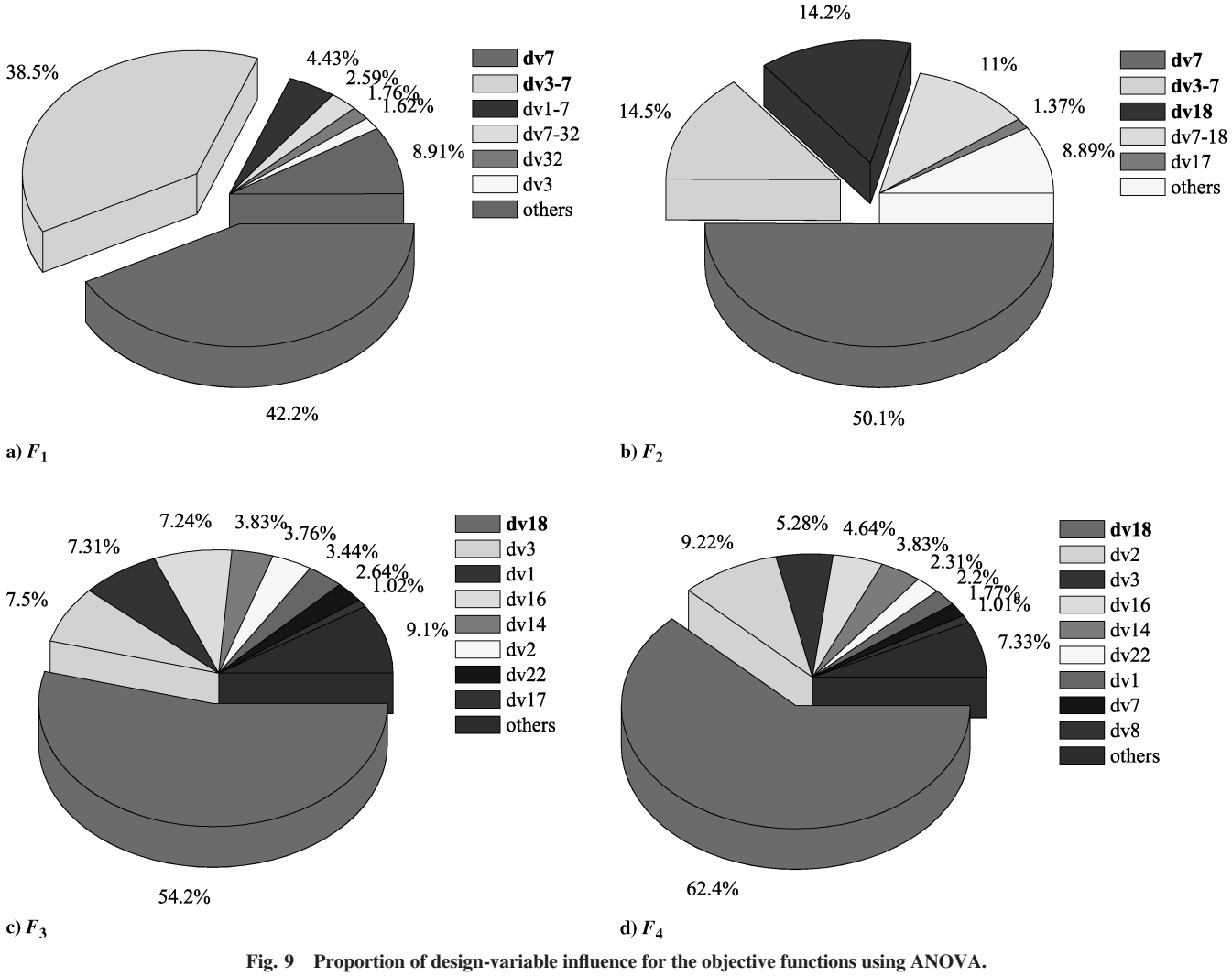


b) Sketch of dv7

d) Sketch of dv18

f) Sketch of dv22

Fig. 8 SOM colored by important design variables and their sketches.



a) F_1

b) F_2

c) F_3

d) F_4

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

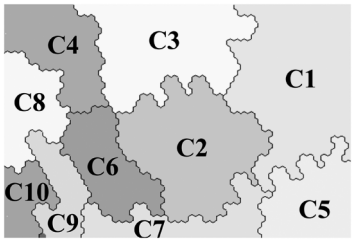


Fig. 10 Serial number of cluster on SOM from the nondominated solutions.

Table 2 Serial number for four clusters including the extreme solutions in the nondominated solutions and the correspondent objective functions

Cluster	Particularity of performance improvement
C1	Shift of aerodynamic center
C2	Transonic C_{Mp}
C5	Transonic C_D
C10	Subsonic C_L

The data consisted of 102 nondominated solutions with four objective function values and 71 design variables. Thus, the object x denotes the nondominated solutions, the condition attribute C corresponds to the design variable, and the decision attribute D is the objective function.

In the present study, each cluster classified by the SOM was employed as decision attribute D. The name of each cluster is shown in Fig. 10. Four clusters included in the extreme solutions for each objective function are each defined as decision attribute D. Table 2 summarizes the correlations between the four clusters as decision attributes and the corresponding four objective functions. The rules obtained for each decision attribute are summarized in Tables 3–6. The range of a design variable is decided when the optimization problem is defined. The values of a and b contained in the range of a design variable in dv$x$($[a, b]$) show the extent of effects on a decision attribute D. The lower or upper limit determined during the problem definition phase is denoted by $*$. That is, for example, the top rule in Table 3 means the value of $0.462112 \leq \text{dv7} \leq *$ and that of $* \leq \text{dv61} \leq 0.978189$ have effects on the shift of the aerodynamic center. In this study, the rules for the top 10 applicable data points are shown. The rule for which the number of applicable data points is equal to one is ignored. These results show that a high value of dv7 has an influence on the shift of the aerodynamic center, the characteristic value of dv7 has influence on transonic C_{Mp}, and dv18 has influence on transonic C_D and subsonic C_L. As this is consistent

Table 3 Rules for C1 generated by RST using the nondominated solution data

Rule	Number of data
dv7([0.462112,*)) and dv61([$*$, 0.978189)) → C1	17
dv7([0.462112,*)) and dv69([$*$, −4.687520)) → C1	16
dv18([0.014534,*)) and dv46([0.913306, $*$)) → C1	16
dv18([0.014534,*)) and dv23([0.002630, $*$)) → C1	16
dv18([0.014534,*)) and dv43([0.025468, $*$)) → C1	16
dv18([0.014534,*)) and dv71([−6.687150, $*$)) → C1	16
dv22([$*$, −0.010476)) and dv68([0.749406, 0.799224)) → C1	15
dv15([0.172055,0.198975)) and dv18([$*$, 0.014534)) → C1	15
dv22([$*$, −0.010476)) and dv46([0.913306, $*$)) → C1	14
dv22([$*$, −0.010476)) and dv40([0.039847, 0.047539)) → C1	14

Table 4 Rules for C2 generated by RST using the nondominated solution data

Rule	Number of data
dv11([0.259898,*)) and dv20([0.070858,*)) and dv23([0.001868, 0.002630)) → C2	5
dv7([0.385122,0.462112)) and dv10([*,2.145600)) and dv46([0.889645, 0.913306)) → C2	4
dv18([0.014534,0.044527)) and dv43([0.023762,0.025468)) and dv46([0.913306, $*$)) → C2	4
dv17([*,0.642730)) and dv29([*,0.026988)) and dv61([0.984426, $*$)) → C2	4
dv23([0.001868,0.002630)) and dv24([*,0.001607)) and dv65([1.291140, 1.363690)) → C2	4
dv13([0.703171,0.732237)) and dv35([*,0.994927)) and dv65([1.291140, 1.363690)) → C2	4
dv20([0.070858,*)) and dv21([0.707026,*)) and dv34([$*$, 0.007870)) → C2	4
dv11([0.259898,*)) and dv30([*,0.500564)) and dv64([0.006496, 0.006996)) → C2	4
dv12([*,0.015801)) and dv25([0.012418,*)) and dv66([0.327994, 0.351801)) → C2	4
dv11([0.259898,*)) and dv23([0.001868,0.002630)) and dv30([$*$, 0.500564)) → C2	4

Table 5 Rules for C5 generated by RST using the nondominated solution data

Rule	Number of data
dv11([*,0.192853)) and dv18([*,0.014534)) and dv53([$*$, 0.010481)) → C5	4
dv5([3.205220,4.976520)) and dv18([*,0.014534)) and dv45([$*$, 0.622205)) → C5	4
dv37([*,0.001320)) and dv40([*,0.039847)) and dv49([0.994318, 0.995602)) → C5	4
dv40([*,0.039847)) and dv49([0.994318,0.995602)) and dv55([$*$, 0.015348)) → C5	4
dv22([$*$, −0.010476)) and dv31([*,0.588815)) and dv35([$*$, 0.994927)) → C5	3
dv21([*,0.601322)) and dv29([0.026988,0.028812)) and dv51([*,0.001391)) → C5	3
dv2([*,0.410756)) and dv33([0.972706,0.984433)) and dv38([$*$, 0.001229)) → C5	3
dv5([3.205220,4.976520)) and dv8([−0.050745, −0.044714)) and dv54([$*$, 0.038919)) → C5	3
dv40([*,0.039847)) and dv43([0.023762,0.025468)) and dv59([$*$, 0.686297)) → C5	3
dv29([0.026988,0.028812)) and dv40([*,0.039847)) and dv43([0.023762, 0.025468)) → C5	3

Table 6 Rules for C10 generated by RST using the nondominated solution data

Rule	Number of data
dv13([0.732237,*)) and dv22([0.030320,*)) and dv39([0.011883, ∗)) → C10	3
dv33([0.984433,*)) and dv48([0.007403,*)) and dv56([0.423662, ∗)) → C10	2
dv20([0.049659,0.070858)) and dv44([0.496361,0.512868)) and dv51([0.001885, ∗)) → C10	2
dv23([0.002630,*)) and dv33([0.984433,*)) and dv51([0.001885, ∗)) → C10	2
dv10([2.14560,2.967960)) and dv32([0.875713,*)) and dv59([0.721385, ∗)) → C10	2
dv13([0.732237,*)) and dv18([0.044527,*)) and dv51([0.001885, ∗)) → C10	2
dv13([0.732237,*)) and dv34([0.008203,*)) and dv51([0.001885, ∗)) → C10	2

with the previous results obtained by the SOM and ANOVA, there is some confidence that the rules have been generated appropriately. As ANOVA shows the total intensity in the whole design space directly, it cannot show the intensity in a particular design variable. For example, as Table 3 shows, when dv18 has a strong intensity to F_1, the local region of dv18 has an effect on the shift of the aerodynamic center. The SOM and ANOVA do not reveal this knowledge, thus these rules obtained by RST are useful in extracting knowledge from the design space at a much more detailed level than the previous two methods. The knowledge obtained by RST provides the relationship between individual design variables and their effects on the tradeoffs.

On the other hand, a physical construction for a rule obtained by RST is difficult when compared with the SOM and ANOVA. The order of rules does not always correspond to the intensity of effect on the objective functions. A large number of solutions needs to be analyzed to generate more accurate rules. When only a small number of solutions with certain characteristics satisfies a decision attribute, the rule does not always produce satisfactory results. In summary, RST is an effective method when there are many solutions which satisfy a decision attribute as well as a considerable number of the design variables.

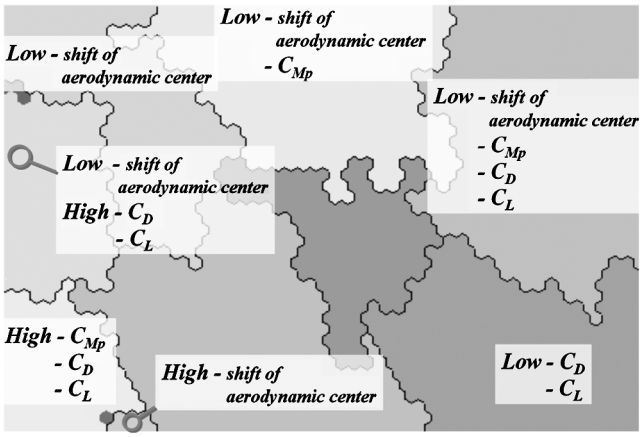


Fig. 11 SOM of all solutions in the four-dimensional objective function space.

B. Design Space Formed by All Solutions

The data mining analysis of the design space spanning all solutions produced in multi-objective optimization provides some knowledge regarding the sensitivity of design variables to objective functions, and hence gives a direction for searching for better designs. In addition, it can reveal a favorable tradeoff region in the design space.

1. *Knowledge Acquired by Using Self-Organizing Map*

The set of all 302 solutions has been projected onto the two-dimensional map of the SOM. Figure 11 shows the resulting SOM with nine clusters, taking the four objective functions into

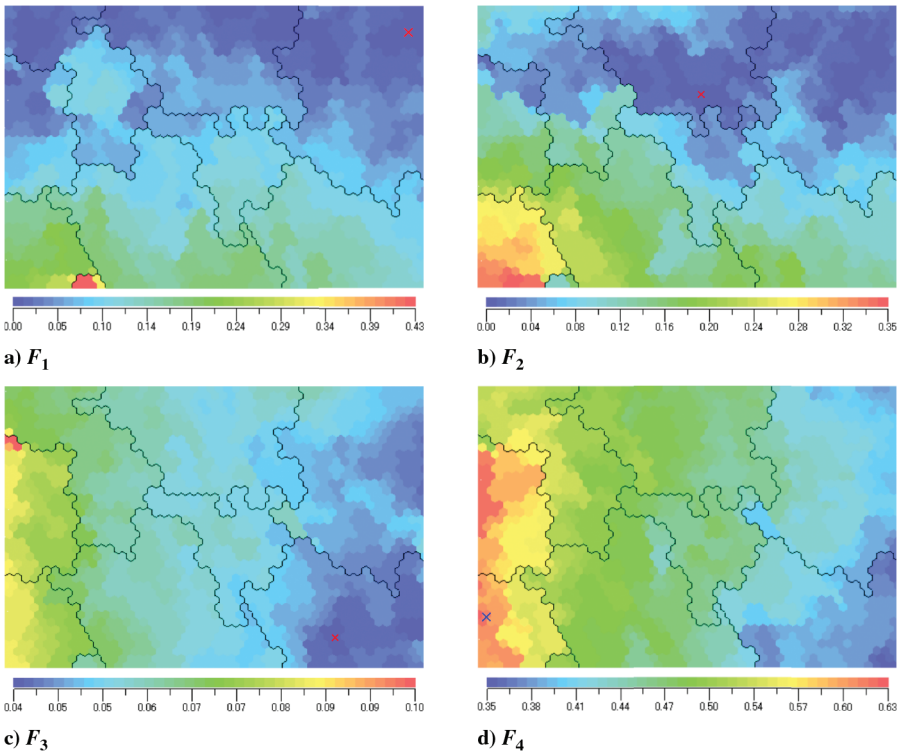


Fig. 12 SOMs of all solutions colored by the objective functions. The symbol × denotes the respective extreme solutions.

consideration. Figure 12 shows the SOMs colored by the four objective function values, respectively. Figures 12a and 12b show that F_1 and F_2 can achieve the lower values simultaneously. These figures also reveal the tendency in which both objectives can acquire the high values simultaneously. The region giving high values of F_1 and F_2 is small. This fact shows that ARMOGA generates a large number of better solutions for each objective function due to range adaptation. It is notable that 302 solutions are biased in the design space, giving better objective function values. Figures 12c and 12d show that there is a severe tradeoff between the transonic C_D and the subsonic C_L in the design space, and no favorable tradeoff region exists for all objectives. However, this design space can have a favorable tradeoff region when the subsonic C_L is tolerable. High-lift devices are one of the potential possibilities for solving this problem.

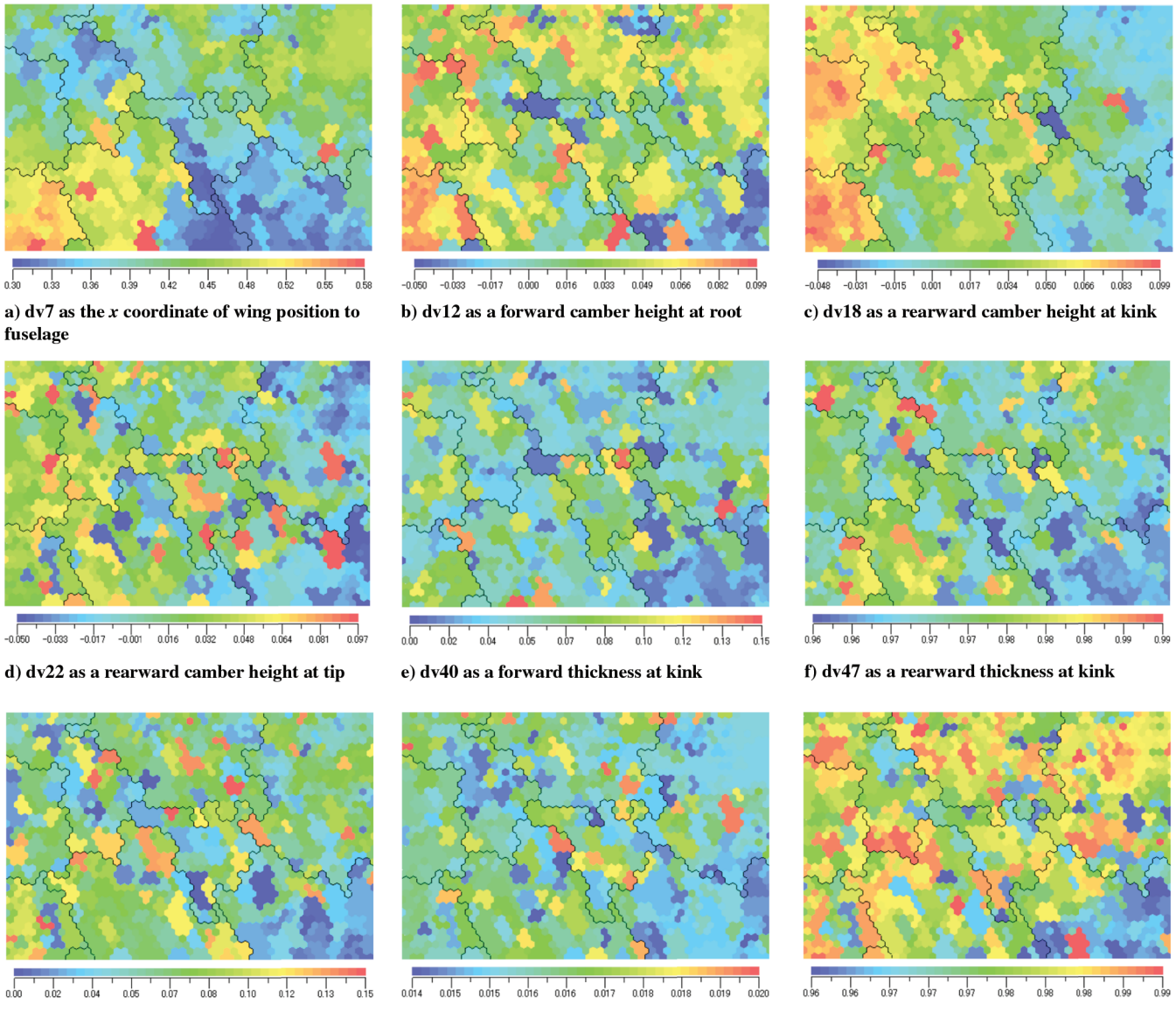
Figure 13 shows the SOMs contoured by design variables with a colored pattern. Figures 13a, 13c, and 13d, colored by dv7, dv18, and dv22, show similar design knowledge compared with Fig. 8. Therefore, these design variables are effective in both design spaces generated by nondominated solutions and by all solutions. That is, the knowledge in the design space spanned by all solutions is conserved. However, the knowledge revealed by Fig. 13 is obscure because of the solution diversity. Although the other design variables, such as dv12, dv40, dv47, dv54, dv55, and dv61, seem to have influence on the transonic C_D (and also subsonic C_L) in Figs. 13b and 13e–13i, there are no specific characteristics on the colored SOM as a whole. Consequently, the information from all solutions is similar to the knowledge from the nondominated solutions.

2. *Knowledge Acquired by Using Functional Analysis of Variance*

Figure 14 shows the ANOVA results for all 302 solutions. This reveals that the influences on design variables obtained from all solutions are similar to the one obtained for the nondominated solutions shown in Fig. 9. This result also corresponds to knowledge acquired using the SOM. F_1 and F_2 have the subordinate relation to each other. Design variable 18 is effective to F_3 and F_4. Therefore, there is always a strong tradeoff among them in the design space spanning all solutions. As was the case with the SOM, the results obtained for all solutions closely correspond to results produced for nondominated solutions.

3. *Knowledge Acquired by Using Rough Set Theory*

The rules generated by RST are obtained using a similar procedure as the one described for nondominated solutions. The name of each cluster is shown in Fig. 15. Table 7 summarizes the correlations



a) dv7 as the *x* coordinate of wing position to fuselage

b) dv12 as a forward camber height at root

c) dv18 as a rearward camber height at kink

d) dv22 as a rearward camber height at tip

e) dv40 as a forward thickness at kink

f) dv47 as a rearward thickness at kink

g) dv54 as a forward thickness at tip

h) dv55 as a forward thickness at tip

i) dv61 as a rearward thickness at tip

Fig. 13 SOMs colored by characteristic design variables.

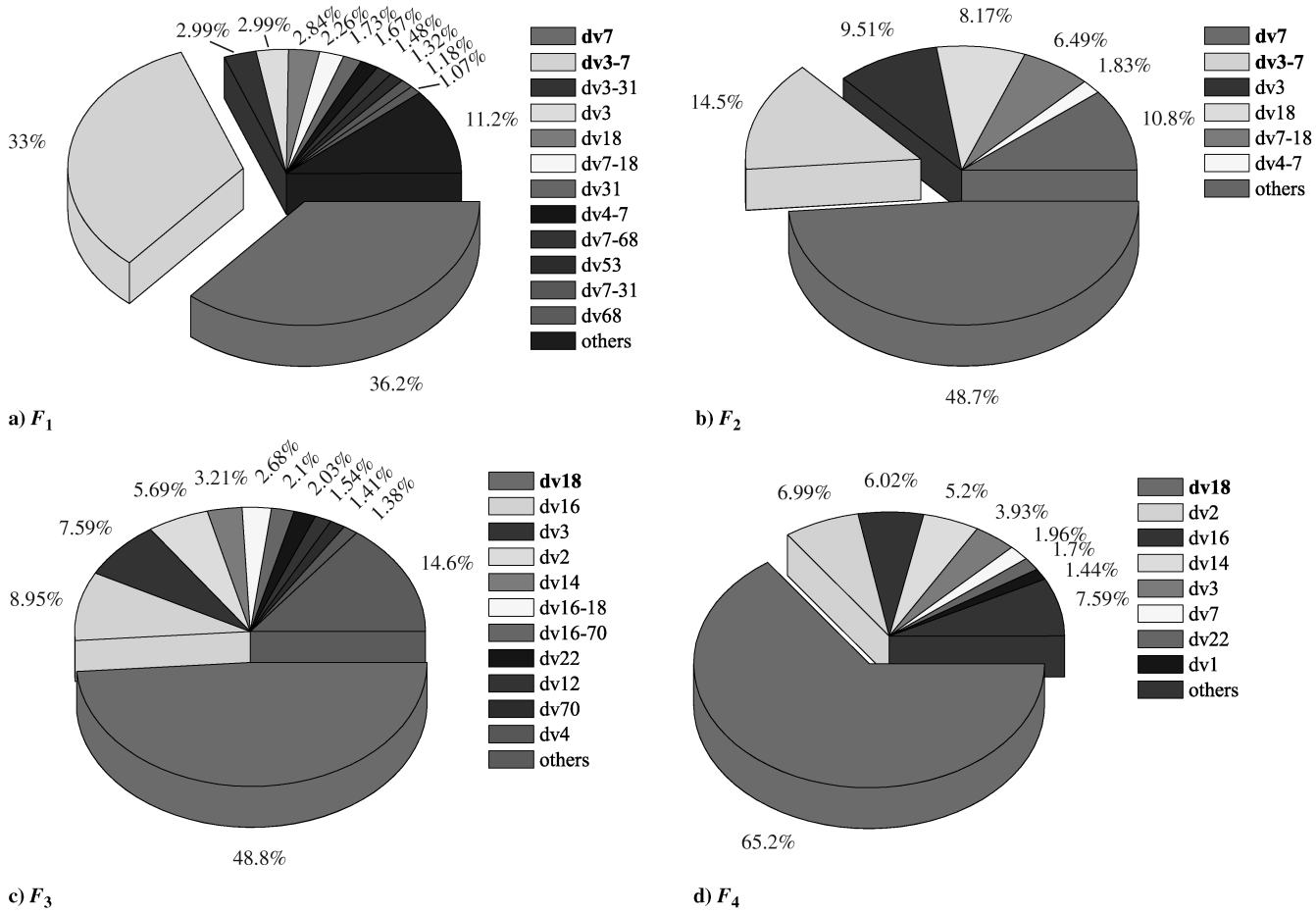


Fig. 14 Proportion of design-variable influence for the objective functions in the all-solution space using ANOVA.

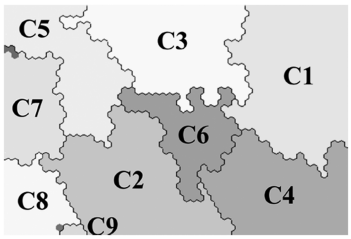


Fig. 15 Serial number of cluster on SOM from all solutions.

between four clusters, the decision attributes, and the corresponding four objective functions. The obtained rules are summarized in Tables 8–11 for each objective function. In particular, the knowledge extracted from cluster C1 shown in Fig. 15 is important because C1 reveals a favorable tradeoff region when C_L is tolerable, as shown in Fig. 11. Table 12 shows the design variables occurring earlier in ANOVA analysis as well as their number appearing in the rules. The design variables showing only in RST analysis and their number occurring in the rules are also shown in Table 12. Because the result by ANOVA only shows proportion, the effect which is given by a design variable with low proportion cannot be interpreted. When the occurrence number in rules becomes zero, then the applicable number of the design variables is zero or one. That is, these design variables have little effect on each objective function.

On the other hand, the applicable number of the design variables having strong effects is small. Therefore, the large applicable number does not always correspond to the strength of the effect. It is important to address all design variables, and then physical constructions are also needed. Table 12 reveals that dv7, dv3, and dv18 are essential for C1 and C3, and dv18 is important for C4 and C8. The affinity of effective design variables also demonstrates the correlation between the objective functions.

The design variables occurring only in RST analysis are also important. For example, dv42 appears many times for the result regarding C1 by RST in Table 12, despite the fact that there is no occurrence of dv42 in the result regarding C1 by ANOVA. RST can address locally effective design variables, whereas the SOM and ANOVA reveal global effects. RST can show the effective design variables only in a limited design space. In fact, the rules shown in Table 8 reveal that dv42 of $[0.402889, *)$ is effective. As the design space is defined in $0.2 \leq \text{dv42} \leq 0.46$, only 22% of the region has some effect on C1. There is a possibility of oversight for the SOM and ANOVA. The parameter dv42 controls the position of the maximum thickness at kink. When dv42 is large, the aerodynamic center shifts rearward. Therefore, it is physically correct that a small range of dv42 with high value gives effects on F_1 and F_2.

Table 7 Serial number for four clusters including the extreme solutions in all solutions and the correspondent objective functions

Cluster	Particularity of performance improvement
C1	Shift of aerodynamic center; favorable tradeoff region
C3	Transonic C_{Mp}
C4	Transonic C_D
C8	Subsonic C_L

C. Comparison of the Three Data Mining Techniques

First of all, the knowledge acquired in the analysis is similar when 102 nondominated solutions are used and when the set of all 302 solutions is considered. This fact shows the possibility that the

Table 8 Rules for C1 generated by RST using all-solution data

Rule	Number of data
dv7([0.379569,0.463235)) and dv18([*,0.06781)) and dv35([0.995614, ∗)) → C1	20
dv18([*,0.006781)) and dv22([*,-0.011248)) and dv42([0.402889, ∗)) → C1	20
dv18([*,0.006781)) and dv42([0.402889,*)) and dv57([0.025121, 0.027792)) → C1	19
dv18([*,0.006781)) and dv25([0.011960,0.012410)) and dv42([0.402889, ∗)) → C1	18
dv3([*,55.205799)) and dv18([*,0.006781)) and dv24([0.001757, ∗)) → C1	17
dv18([*,0.006781)) and dv37([0.001476,0.002020)) and dv42([0.402889, ∗)) → C1	17
dv22([∗, −0.011248)) and dv42([0.402889,*)) and dv60([0.848821, 0.870887)) → C1	16
dv3([*,55.205799)) and dv40([0.040085,0.054506)) and dv68([0.734865, 0.798422)) → C1	16
dv3([*,55.205799)) and dv40([0.040085,0.054506)) and dv71([−6.333930, ∗)) → C1	15
dv18([*,0.006781)) and dv19([0.156790,0.235667)) and dv42([0.402889, ∗)) → C1	15

Table 9 Rules for C3 generated by RST using all-solution data

Rule	Number of data
dv28([*,0.336057)) and dv62([0.009258,*)) and dv69([−4.622460, ∗)) → C3	7
dv7([0.379569,0.463235)) and dv14([*,0.021668)) and dv32([0.856722, 0.881603)) → C3	7
dv7([*,0.379569)) and dv32([*,0.856722)) and dv37([0.002020, ∗)) → C3	7
dv3([62.753700,*)) and dv34([*,0.007820)) and dv52([0.001690, 0.001762)) → C3	6
dv23([*,0.001821)) and dv51([0.002018,*)) and dv52([0.001690, 0.001762)) → C3	6
dv50([0.008145,0.008867)) and dv51([0.002018,*)) and dv52([0.001690, 0.001762)) → C3	5
dv6([302.364990,*)) and dv7([*,0.379569)) and dv11([0.264078, ∗)) → C3	5
dv1([0.183163,0.196521)) and dv32([*,0.856722)) and dv48([0.007955, ∗)) → C3	5
dv18([0.006781,0.041033)) and dv22([0.032237,*)) and dv30([∗, 0.501708)) → C3	5
dv16([0.001324,0.038344)) and dv33([0.970830,0.982542)) and dv61([0.983121, ∗)) → C3	5

Table 10 Rules for C4 generated by RST using all-solution data

Rule	Number of data
dv18([*,0.006781)) and dv54([*,0.43352)) and dv58([∗, 0.515947)) → C4	14
dv3([*,55.205799)) and dv51([*,0.001529)) and dv67([∗, −1.647800)) → C4	13
dv40([*,0.040085)) and dv51([*,0.001529)) and dv67([∗, −1.647800)) → C4	12
dv18([*,0.006781)) and dv20([0.041139,0.067107)) and dv58([∗, 0.515947)) → C4	11
dv20([0.041139,0.067107)) and dv51([*,0.001529)) and dv67([∗, −1.647800)) → C4	9
dv6([294.427002,302.364990)) and dv9([∗, −1.487150)) and dv10([∗, 2.088340)) → C4	9
dv6([294.427002,302.364990)) and dv10([*,2.088340)) and dv21([∗, 0.614866)) → C4	9
dv22([∗, −0.011248)) and dv61([*,0.974320)) and dv70([−5.336230, −5.236000)) → C4	9
dv31([*,0.598579)) and dv40([*,0.040085)) and dv58([∗, 0.515947)) → C4	8
dv31([*,0.598579)) and dv57([*,0.025121)) and dv58([∗, 0.515947)) → C4	8

Table 11 Rules for C8 generated by RST using all-solution data

Rule	Number of data
dv29([0.028516,*)) and dv48([0.007955,*)) and dv51([0.001529, 0.002018)) → C8	7
dv1([0.196521,*)) and dv29([0.028516,*)) and dv32([0.881603, ∗)) → C8	7
dv29([0.028516,*)) and dv32([0.881603,*)) and dv65([1.246800, 1.424610)) → C8	7
dv29([0.028516,*)) and dv30([0.501708,0.523840)) and dv48([0.007955, ∗)) → C8	6
dv18([0.041033,*)) and dv30([0.501708,0.523840)) and dv35([0.995614, ∗)) → C8	6
dv37([0.002020,*)) and dv51([0.001529,0.002018)) and dv61([0.983121, ∗)) → C8	6
dv7([0.463235,*)) and dv24([0.001757,*)) and dv25([∗, 0.011960)) → C8	4
dv18([0.041033,*)) and dv20([0.067107,*)) and dv58([0.515947, 0.525239)) → C8	4
dv10([2.088340,3.344390)) and dv20([0.041139,0.067107)) and dv48([0.007955, ∗)) → C8	4
dv20([0.067107,*)) and dv29([0.028516,*)) and dv48([0.007955, ∗)) → C8	4

knowledge obtained from space S is conserved into the knowledge from other space included in space S. The knowledge extracted by three data mining techniques is summarized next.

The knowledge acquired by SOM includes the following:

1) Although all objective functions were not optimized simultaneously, there was a favorable tradeoff region if subsonic C_L was tolerable. However, subsonic C_L is an important performance. Therefore, this knowledge suggests the necessity of high-lift devices.
2) There was a similar relation between F_1 and F_2. When one of these two objective functions is considered, the other tendency can be predicted.
3) There was a strong tradeoff between F_3 and F_4.
4) The design variables affecting each objective function were found and their qualitative effects were revealed.

The knowledge obtained by ANOVA is as follows:

The important design variables with global effects were found (this information could be used by the SOM), and their

Table 12 Effective design variables addressed by ANOVA and RST, and their applicable number in rules

	By ANOVA						By RST		
C1									
Serial number of design variable	**dv7**	**dv3**	**dv18**	dv31	dv57	dv68	**dv42**		
Applicable number in rules	1	3	7	0	1	1	6		
C3									
Serial number of design variable	**dv7**	**dv3**	**dv18**				dv32	dv52	
Applicable number in rules	3	1	1				3	3	
C4									
Serial number of design variable	**dv18**	dv16	dv3	dv2	dv14	——	dv58	dv51	dv67
Applicable number in rules	2	0	1	0	0	——	4	3	3
C8									
Serial number of design variable	**dv18**	dv2	dv16	dv14	dv3	——	dv29	dv48	dv20
Applicable number in rules	2	0	0	0	0	——	5	4	3

quantitative effects on them for each objective function were revealed. Design variables 7, 18, and 3 have effects on F_1 and F_2. Also, dv18 has effect on F_3 and F_4.

The knowledge obtained by RST includes the following:

1) The effective region of the design space for the design variables has global effects.
2) The design variables for which global effects could be ignored were found (this information was used by ANOVA).
3) The design variables with local effects only, which the SOM and ANOVA could not find, and their effective regions were found.
4) Result by RST is sensitive to the set of solutions used in the analysis.

As the rules by RST are not generated considering the actual physical phenomena, it is difficult to directly use these rules to guide the choice of the desired values of design variables which improve the overall design. Thus, the results presented in this study should rather be used as guidelines for identifying important design variables (i.e., variables having the largest impact on the design quality) rather than specific recommendations for design-variable values.

The combination of the three techniques compensated the deficiencies discovered when applying any of these methods individually.

V. Conclusions

The three data mining techniques have been carried out for the evolutionary-based aerodynamic design optimization of a flyback-booster wing. The conducted analyses revealed the knowledge in the design space regarding tradeoffs, correlations among the objective functions, globally and locally effective design variables, and a favorable tradeoff region. In addition, the benefits and drawbacks of the three investigated data mining techniques were shown. The distinguishing feature of a self-organizing map is the generation of a qualitative description. The advantage of this method includes the intuitive visualization of two-dimensional colored maps of design space using bird's-eye-like views. As a result, the self-organizing map reveals the tradeoffs among objective functions. Moreover, self-organizing maps roughly address the effective design variables and also reveal how a specific design variable affects objective functions and other design characteristics. However, the self-organizing map is subjective due to color cognizance. There is also a possibility of oversight because of a large number of objective functions and design variables. On the other hand, the distinguishing property of functional analysis of variance is the quantitative description. The advantage of this method is the fact that it directly finds globally effective design variables. That is, functional analysis of variance explicitly addresses the effective design variables. However, functional analysis of variance cannot directly identify the effects of design variables on objective functions. Finally, rough set theory generates rule description. The advantage of this method is that it finds locally effective design variables. This knowledge can be used to narrow down the search and optimization process to a local region of the design space. The disadvantage is the difficulty in capturing of effects for relevant physical phenomena in the set of generated rules. Also, rough set theory is sensitive to the set of analyzed solutions. When all three methods are combined together, the results obtained can compensate the disadvantages of the individual methods.

Acknowledgments

We would like to thank Rafal Kicinger, a postdoctoral fellow in the School of Information Technology and Engineering, George Mason University, and Taro Imamura, a researcher in the aviation program group, Japan Aerospace Exploration Agency, for their useful suggestions and help.

References

[1] Holden, C. M. E., and Keane, A. J., "Visualization Methodologies in Aircraft Design," AIAA Paper 2004-4449, 2004.

[2] Obayashi, S., Jeong, S., and Chiba, K., "Multi-Objective Design Exploration for Aerodynamic Configurations," AIAA Paper 2005-4666, 2005.

[3] Powell, R. W., Lockwood, M. K., and Cook, S. A., "Road from the NASA Access-to-Space Study to a Reusable Launch Vehicle," International Astronautical Federation Paper 98-V.4.02, 1998.

[4] Staniszewski, E. A., "Semireusable Launch Vehicle: A Next-Generation Launch Vehicle?," *Journal of Spacecraft and Rockets*, Vol. 38, No. 3, 2001, pp. 368–373.

[5] Sippel, M., Herbertz, A., Kauffmann, J., and Schmid, V., "Analysis of Liquid Fly-Back Booster Performance," AIAA Paper 99-4827, 1999.

[6] Watanabe, A., and Shimizu, T., "System Design and Program Status of the H-IIA Rocket," International Astronautical Federation Paper 97-V.1.04, 1997.

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

[8] Ito, Y., and Nakahashi, K., "Direct Surface Triangulation Using Stereolithography Data," *AIAA Journal*, Vol. 40, No. 3, 2002, pp. 490–496.

[9] Ito, Y., and Nakahashi, K., "Improvements in the Reliability and Quality of Unstructured Hybrid Mesh Generation," *International Journal for Numerical Methods in Fluids*, Vol. 45, No. 1, 2004, pp. 79–108.
doi:10.1002/fld.669

[10] Dacles-Mariani, J., Zilliac, G. G., Chow, J. S., and Bradshaw, P., "Numerical/Experimental Study of a Wingtip Vortex in the Near Field," *AIAA Journal*, Vol. 33, No. 9, 1995, pp. 1561–1568.
doi:10.2514/3.12826

[11] 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, No. 1, 2005, pp. 44–64.
doi:10.2514/1.12909

[12] Chiba, K., Obayashi, S., and Nakahashi, K., "Design Exploration of Aerodynamic Wing Shape for Reusable Launch Vehicle Flyback Booster," *Journal of Aircraft*, Vol. 43, No. 3, 2006, pp. 832–836.
doi:10.2514/1.12782

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

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

[15] Pampalk, E., Rauber, A., and Merkl, D., "Content-Based Organization

and Visualization of Music Archives," *Proceedings of the 10th International Conference on Multimedia*, Assoc. for Computing Machinery Press, New York, 2002, pp. 570–579.

[16] Obayashi, S., and Sasaki, D., "Visualization and Data Mining of Pareto Solutions Using Self-Organizing Map," *2nd International Conference on Evolutionary Multi-Criterion Optimization*, LNCS 2632, Springer–Verlag Heidelberg, Germany, 2003, pp. 796–809.

[17] Chiba, K., Oyama, A., Obayashi, S., Nakahashi, K., and Morino, H., "Multidisciplinary Design Optimization and Data Mining for Transonic Regional-Jet Wing," *Journal of Aircraft*, Vol. 44, No. 4, 2007, pp. 1100–1112. doi:10.2514/1.17549

[18] Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol. 13, No. 4, 1998, pp. 455–492. doi:10.1023/A:1008306431147

[19] Jeong, S., Murayama, M., and Yamamoto, K., "Efficient Optimization Design Method Using Kriging Model," *Journal of Aircraft*, Vol. 42, No. 2, 2005, pp. 413–420. doi:10.2514/1.6386

[20] Pawlak, Z., "Rough Sets," *International Journal of Computer and Information Sciences*, Vol. 11, No. 5, 1982, pp. 341–356. doi:10.1007/BF01001956

[21] Kusiak, A., "Rough Set Theory: A Data Mining Tool for Semiconductor Manufacturing," *IEEE Transactions on Electronics Packaging Manufacturing*, Vol. 24, No. 1, 2001, pp. 44–50. doi:10.1109/6104.924792

[22] Øhrn, A., "Discernibility and Rough Sets in Medicine: Tools and Applications," Ph.D. Dissertation, Dept. of Computer and Information Science, Norwegian Univ. of Science and Technology, Trondheim, Norway, 1999.

J. Korte
Associate Editor