

Vehicle Performance Design System Using Data Mining Techniques towards Model-based Development

Tokitomo ARIYOSHI*

Masatoshi HARUMATSU*

Kazushige YANO*

Takahiro FUKUSHIGE*

Kentaro YAMADA*

Kaname NARUKAWA**

ABSTRACT

A vehicle performance design system has been developed that makes it possible to assign target values for different performance demands at the initial stage of the vehicle development process. Based on the concept of model-based development, this system determines appropriate target values through the application of data mining techniques.

The system uses the data available at the time of vehicle development, and comprises the following three techniques:

- (1) A response surface generation technique that helps enable high-accuracy predictions based on small amounts of data;
- (2) A decision-making technique that helps enable high-efficiency data searches that reflect the user's intentions; and
- (3) An impact analysis technique that makes it possible to identify specifications that have the great effect on performance from the relationships between performance and specifications.

It was demonstrated that the application of this system to vehicle dynamic performance simulations enables determination of target performance values and specification values simply and at high speed.

1. Introduction

Today, the automotive design process is expected to satisfy a range of demands proceeding from consideration of factors such as environmental and safety performance, and the process of development of the various elements necessary to the realization of target performance has become large-scale and complex. At the same time, the shortening of the development period and the reduction of cost continue to be important considerations.

A global survey of the automotive market shows that the market expands at a rate of 4% per year, and that factors such as the expansion of markets in emerging nations are changing the sales structure in regional markets. In order to respond to this situation, it is necessary to expand vehicle lineups, leading to modification of the flow of development throughout the automotive industry. The area of digital development is particularly active, and the importance of model-based development (MBD), involving virtual development based on input including simulation and test data and entailing as little creation of actual prototypes as possible, is increasing.

The research discussed in this paper sought to develop a vehicle performance design system that would help to realize MBD in order to increase development efficiency by making it possible to identify multiple target performance factors such as fuel efficiency and acceleration performance, and derive the related specifications, at the initial stage of development.

2. System Configuration

The research discussed here created a vehicle performance design system that will contribute to the development of next-generation vehicles using data mining techniques, realizing a style of development in which optimal target values are set with consideration of the tradeoff between multiple areas of performance.

The developed system (referred to simply as "the system" below) uses data-driven simulations employing a database (DB)^{(1), (2)}; accuracy and responsiveness are therefore important factors. Figure 1 shows an overview of the developed vehicle performance design system.

* Automobile R&D Center

** Fundamental Technology Research Center

As Fig. 1 shows, the system comprises a computer-aided engineering (CAE) simulation section, which simulates vehicle performance, a target performance display section incorporating the DB, and three data mining technologies. The flow of system procedures is as follows.

- (1) The specifications that form the input for the system are sampled using a design of experiments method.
- (2) Training data is formulated by conducting CAE calculations.
- (3) Using the training data, response surfaces and DB are created using machine learning technology.
- (4) Using the DB, optimal target values are determined with consideration of the tradeoff between multiple areas of performance.

The DB is made up of specification values used as input for CAE and performance values obtained from simulation results. Table 1 shows an example of a DB array consisting of five columns of specification values on the left and three columns of performance values on the right. The vehicle performance design system can be operated by using this DB. (1) to (3) below provide an overview of the data mining techniques employed to actualize the vehicle performance design system.

(1) Generation of response surfaces

In order to design target performance, it is necessary for the system to create DB containing large-scale data from the results of CAE calculations.

Figure 2 compares the conventional method and the response surface approximation method in generating data for 100,000 cases and creating a DB, showing the differences in calculation between the conventional method and the proposed method. The use of conventional CAE calculations necessitates an impractical calculation time for the creation of the DB. It was therefore necessary to develop a method of creating DB in the shortest time possible. Using a limited number of CAE calculation results (around 200 cases as training data), the proposed method employs machine learning techniques in order to generate a response surface model (RSM). The use of this RSM to calculate data for 100,000 cases enables a major reduction

Table 1 Database matrix

	Factor					Performance		
	1	2	3	4	5	1	2	3
1	460.91	2026.05	0.42	1418.89	2533.81	6.15	136	4.34
2	448.78	1775.30	0.42	1368.02	2511.51	6.19	137	4.41
3	459.69	2204.39	0.44	1417.41	2500.43	4.81	135	4.26
4	439.90	1924.32	0.44	1346.49	2487.21	5.13	137	4.37
5	458.57	1850.84	0.42	1348.45	2527.38	5.07	136	4.35
:	:	:	:	:	:	:	:	:
200	450.60	1745.68	0.44	1409.86	2590.14	5.38	138	4.39

in calculation time compared to the conventional method. Further details will be provided in Section 2.1.

(2) Decision making (Interactive data selection)

Determining the optimum solution that reflects the user's intentions from the 100,000-case DB is not a simple matter. The system therefore employs the following two data selection methods:

- (a) A direct search technique that searches for specification values directly from performance values; and
- (b) A group search technique that makes it possible to interactively narrow down data while surveying trends in the data using a clustering technique.

Users are able to select and implement the decision-making technique that suits their application from these two search techniques. Further details will be provided in Section 2.2.

(3) Impact analysis (Analysis of impact factor)

Impact analysis is an analytic method that makes it possible to determine the characteristic relations between performance and specification values whether the data is linear or nonlinear or the data environment features multiple specification values.

When nonlinear relationships between variables straddle multiple specifications, visualization of the relationships is challenging. The method visualizes relationships using partial dependence plots⁽³⁾, which are able to display characteristics even for multiple specifications. The degree of effect of each specification value on performance is calculated as impact factor, which is easily understood intuitively, and displayed on a graph. Further details will be provided in Section 2.3.

2.1. Generation of Response Surfaces

A response surface is a mathematical model that correlates a performance value, y , representing system output data, with a specification value, x , representing input

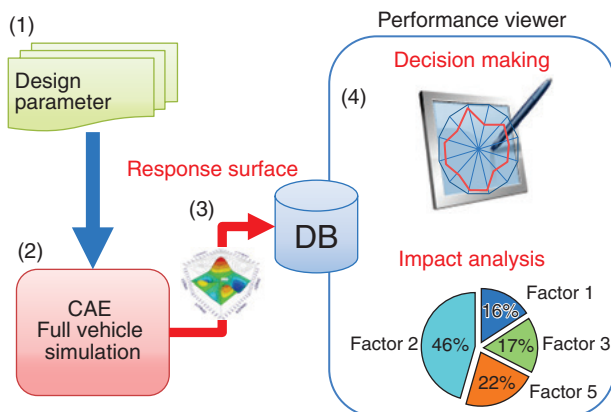


Fig. 1 Vehicle performance design system

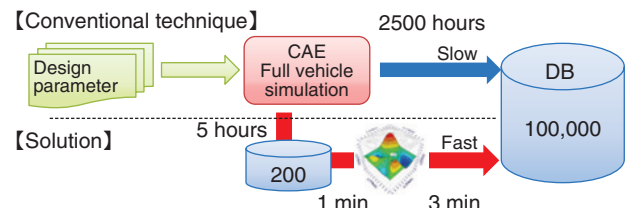


Fig. 2 Effect on response surface

data. This data might be experimental data or data generated by engineering simulations. Because conducting tests and simulations entails costs in terms of time and labor, it is often not possible to obtain data for a large number of cases. If a response surface can be generated for data for a limited number of cases, it is then possible to simply predict an output, $\mathbf{y}$, for a new input, $\mathbf{x}$. The case data necessary for the generation of a response surface is known as training data.

Terming modeling error ε , the relationship between input and output can be expressed by Eq. (1).

$$\mathbf{y} = M(\mathbf{x}) + \varepsilon \quad (1)$$

Here, $M(\mathbf{x})$ is the response surface. Output $\mathbf{y}$ is normally a vector quantity, but will be considered as a scalar quantity below, because modeling can be performed in relation to each component.

Various methods are available for the generation of response surfaces, $M(\mathbf{x})$. Because it was necessary to generate models of high approximation accuracy from limited training data in the research under discussion, the PCE-LAR method⁽⁴⁾ and the Kriging method were combined to produce the PCK method. Terming the function generated by the PCE-LAR method $f(\mathbf{x})$ and the function generated by the Kriging method $Z(\mathbf{x})$, $M(\mathbf{x})$ can be expressed as Eq. (2).

$$M(\mathbf{x}) = f(\mathbf{x}) + Z(\mathbf{x}) \quad (2)$$

As Fig. 3 shows, the PCE-LAR method expresses global changes in output, and the Kriging method expresses local changes in output. The modeling of both global and local changes makes it possible to realize a high level of approximation accuracy⁽⁵⁾. The PCE-LAR method and the Kriging method will be discussed in greater detail below.

2.1.1. PCE-LAR method

The PCE-LAR method combines modeling of global system response using polynomial chaos expansion (PCE) and high-accuracy approximation using least angle regression (LAR). In the PCE method, model $f(\mathbf{x})$ is expressed as a linear combination of orthogonal polynomials. For a scalar input x , $f(x)$ is given by Eq. (3).

$$f(x) = \sum_{i=0}^N a_i \Phi_i(x) \quad (3)$$

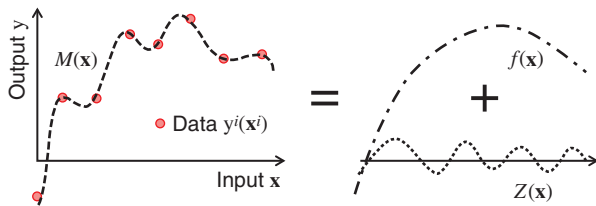


Fig. 3 Response surface model $M(\mathbf{x})$ as sum of global trend $f(\mathbf{x})$ and local fluctuation $Z(\mathbf{x})$

Here, a_i is a coefficient, N is the highest-order polynomial, and $\Phi_i(x)$ is the i -th orthonormal polynomial. In the research under discussion, Legendre polynomials were employed as the orthogonal polynomials. Because the range of the Legendre polynomial input x is $[-1, 1]$, it was necessary to convert the input to the relevant range in advance. In Eq. (3), all Legendre polynomials, from the 1st to the N th, are employed.

The coefficient a_i is determined using the generalized least squares method. If the number of training data points, $(\mathbf{x}, \mathbf{y})$ is termed n , n should be equivalent to or lower than the number of coefficients. An increase in the number of components of input $\mathbf{x}$, M , or the highest order of input $\mathbf{x}$, N , results in an explosive increase in the number of coefficients. For example, when $M=10$ and $N=4$, if all of the Legendre polynomials up to the 4th order are considered, the number of coefficients necessary for modeling exceeds 1,000. This means that the number of necessary training data points increases explosively when M or N increases.

LAR algorithms are employed in order to avoid this issue. LAR algorithms determine polynomials of the necessary order for the generation of the model in order of importance while referring to the training data. This enables considerably reducing the number of $\Phi_i(\mathbf{x})$ functions, i.e. the number of coefficients, employed in the model. This reduction of the number of coefficients employed also increases approximation accuracy. The reason for this is as follows: When an LAR algorithm is not employed, the explosive increase in the number of terms makes it hard to use terms up to the higher orders; when an LAR algorithm is used, the model is generated using polynomials of the necessary order only, making it possible to use high-order terms without the risk of explosive increase in the number of terms.

2.1.2. Kriging method

The Kriging method describes variation from an average. This variation is regarded as conforming to a stochastic process. The variations at two points in the training data, $\mathbf{x}^i$ and $\mathbf{x}^j$, are taken as having correlation with a correlation function, R_{ij} , which is dependent on the distance $|\mathbf{x}^i - \mathbf{x}^j|$. Gaussian-type, exponential-type, and Matérn-type R_{ij} exist. The appropriate type in relation to the training data was distinguished in the research under discussion. Because every type incorporated a parameter, the value of the parameter was determined using optimization computations.

The variation modeled by the Kriging method, $Z(\mathbf{x})$, is given by Eq. (4).

$$Z(\mathbf{x}) = \mathbf{r}(\mathbf{x}) \mathbf{R}^{-1} \Delta \mathbf{y} \quad (4)$$

Here, $\mathbf{r}(\mathbf{x})$ is the row vector of components of n generated by correlating $\mathbf{x}$ and training data $\mathbf{x}^i$, $\mathbf{R}$ is the $n \times n$ matrix making up R_{ij} , and $\Delta \mathbf{y}$ is the column vector of components of n given by subtracting $f(\mathbf{x}^i)$ from training data y^i .

In the actual calculations, the creation of the response surface is divided into first and second stages. In the first stage, the order of precedence of candidates for the polynomial $\Phi_i(\mathbf{x})$ used to construct $f(\mathbf{x})$ is determined using LAR algorithms. In this stage, $f(\mathbf{x})$ remains undetermined. The reason for this is that the function $f(\mathbf{x})$ optimized using the PCE method in isolation will not necessarily correspond to the function $f(\mathbf{x})$ optimized using the combination of the PCE and Kriging methods. In the second stage, the appropriate form of the correlation function is first distinguished. Next, following the order of precedence determined in the first stage, the error of $M(\mathbf{x})$ as determined using the combination of the PCE and Kriging methods is evaluated while adding the $\Phi_i(\mathbf{x})$ functions used to constitute $f(\mathbf{x})$ one by one. The response surface $M(\mathbf{x})$ is determined using the combination of $f(\mathbf{x})$ and $Z(\mathbf{x})$ that produces the least error.

2.1.3. Verification of method

The developed method was applied to data obtained from fuel efficiency and acceleration simulations, and approximation error was studied. Figure 4 shows modeling results for simulations of three fuel economy modes (US-Hot, JC08, and EUDC). Results obtained using a high-accuracy method [anisotropic Kriging (AKR)] provided by a commercially available response surface generation tool are also shown for comparison. The horizontal axes of the graphs show the number of training data points, and the vertical axes show relative error (%). In order to calculate relative error, data differing from the training data was employed, and relative error was determined for each data point and averaged for all data points. For all the fuel economy modes, approximation error becomes smaller in the order AKR→PCE-LAR→PCK. Figure 5

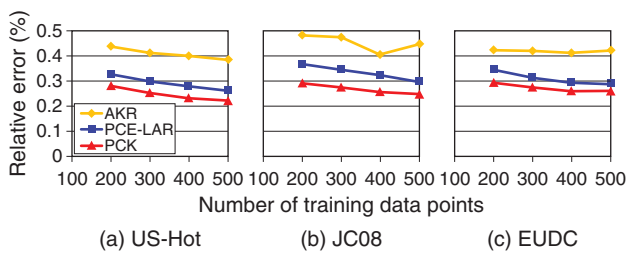


Fig. 4 Relative error plot for three fuel economy modes

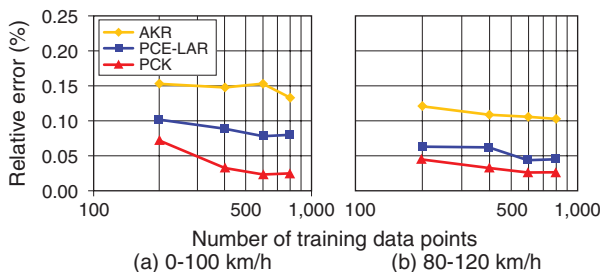


Fig. 5 Relative error plot for two acceleration modes

shows results generated by modeling acceleration time for two acceleration modes (from 0-100 km/h and from 80-120 km/h) when the method was applied to acceleration simulation data. The magnitude of error is half or less than that for the fuel efficiency data. In terms of the difference between the methods, as in the case of fuel economy data, approximation error becomes smaller in the order AKR→PCE-LAR→PCK.

The results discussed above indicated that for both fuel economy and acceleration simulation data, the use of the PCK method reduced error by comparison with the PCE-LAR method. Figure 6 shows models obtained from mode fuel economy simulations, produced in order to assist in understanding the difference between the methods. The horizontal axis shows normalized design values, and the vertical axis shows mode fuel economy values expressed in terms of variation in relation to a reference value. The PCE-LAR method accurately captures the overall trend across the input value range. However, there are considerable local discrepancies between values from data and values predicted by the model. The PCK method, in which the PCE-LAR method is used together with the Kriging method, considerably reduces these discrepancies. Depending on the specific characteristics of the data, the reduction might not be as marked as in this example, but response surfaces generated by the PCK method, combining the PCE-LAR and Kriging methods, tend to display more robust performance than response surfaces generated using the PCE-LAR or Kriging methods in isolation.

2.2. Decision Making

While a user may wish to determine vehicle performance target values using a large-scale DB, the fact that suitable vehicle performance data is not easy to obtain presents an issue. The user therefore employs decision making to conduct a high-speed search for vehicle performance data using a minimal number of operations.

The research under discussion employed two search techniques, direct search and group search.

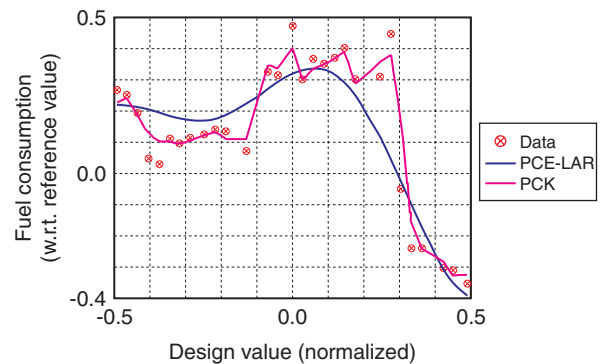


Fig. 6 Comparison of simulation data and response surface model predictions (PCE-LAR and PCK)

2.2.1. Direct search: High-speed search method

The high-speed search method calculates the Euclidean distance between vehicle performance data in the DB and vehicle performance target values, and selects data for which the distance is lowest as solutions. Figure 7 shows the search technique using three-dimensional data as an example. The target for the user is the data marked with a star; because the data space is sparse, the method takes the data that most closely approximates this data as the search solution. Data can be sparse or dense within the formulated DB, and if the appropriate search solution does not exist within the DB, the accuracy of the solution will be low. However, because the search is approximate, calculations can be conducted at high speed.

2.2.2. Direct search: High-accuracy search method

The high-accuracy search method searches for optimum solutions using response surfaces generated in order to create large-scale DB. Figure 8 shows an example of the generation of a response surface from three-dimensional data and its use by the search technique in order to search

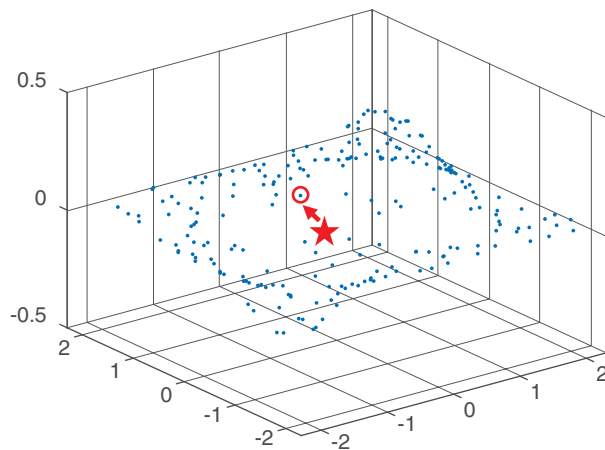


Fig. 7 High-speed search method by data exploration

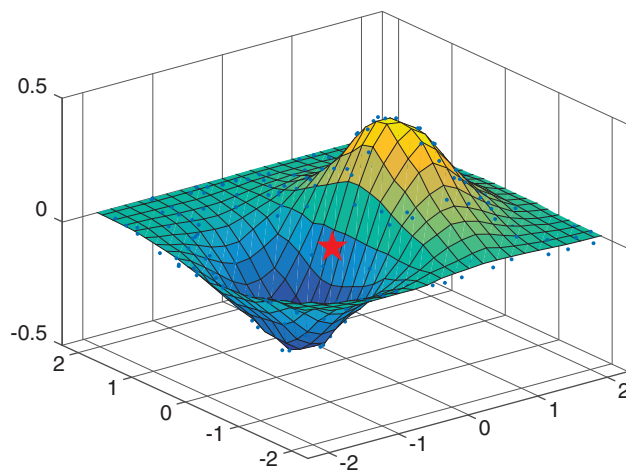


Fig. 8 High-accuracy search method by response surface

for a solution. The target for the user is the data marked with a star. In this method, the data within the DB space is mathematically complemented, and the solution is sought by calculating optimum values based on this complemented data. The method searches for data that approximates an error of 0 between performance values determined by the calculation of response surfaces using the gradient method and vehicle performance target values.

The issue for this method is that local optimization might occur due to the effect of the initial value, and it therefore might not be possible to find the true optimum solution. However, local optimization can be avoided by randomly changing the initial value 100 times in advance, making it possible to realize true optimization. A lengthy calculation time is necessary for the solution of the optimization problem in this method, but the accuracy of the solutions determined is high.

2.2.3. Group search

The group search technique clusters the data within the DB and displays representative data (the centroid for each class) on radar charts. Using these radar charts, the user is able to search for the solution by viewing the data and narrowing it down. Figure 9 shows an example of the process of narrowing down the data. The data within the DB is clustered and expressed as the five classes on the left-hand side of Fig. 9. The user selects one of the classes from among these five. The right-hand side of Fig. 9 shows the result of this selection. Because performance values in the DB (P1-P12) are displayed for each term on each radar chart, the user is able to narrow the data down to the desired performance values while taking the overall performance balance into consideration.

Fuzzy C-Means⁽⁶⁾ was introduced as the clustering method in the research under discussion in order to expand the range of data selection by the user. The Fuzzy C-Means method does not determine class dependency, which is the standard mode of operation for conventional clustering algorithms. Rather than classifying specific data points into a single exclusive class, the method considers each data point to belong to multiple classes, and expresses separation boundaries fuzzily using degrees of assignment, which indicate the degree to which each data point belongs to particular classes. Figure 10 shows the differences in the separation boundaries between three classes for the conventional and the Fuzzy C-Means methods. The left-

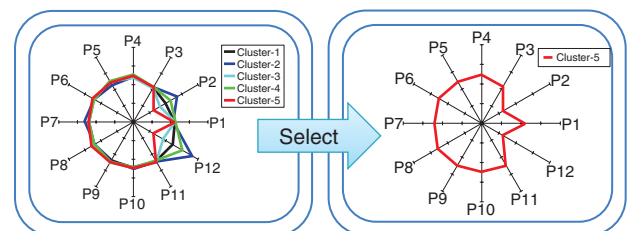


Fig. 9 Decision making

hand image in Fig. 10 shows the result of clustering based on a general determination of dependence, while the right-hand image shows the result of fuzzy clustering using the degree of assignment for data around the separation boundaries between the classes.

Fuzzy C-Means algorithms function as follows:

- (1) The degree of assignment of element i to class j , u_{ij} , is initially randomly set.
- (2) Using the current degree of assignment, u_{ij} , the centroid vector, $\bar{x}_i$, is calculated for each class. m is a fuzzy parameter that determines fuzziness, satisfying the condition $m > 1$. n is the number of feature vectors for machine learning, and element x_j is the aggregation of data. The centroid vector, $\bar{x}_i$, is given by Eq. (5).

$$\bar{x}_i = \frac{\sum_{j=1}^n (u_{ij})^m x_j}{\sum_{j=1}^n (u_{ij})^m} \quad (5)$$

- (3) Using the calculated centroid vector $\bar{x}_i$, the degree of assignment of element x_j to class i , u_{ij} , is updated. $\|\cdot\|$ is the Euclidean norm, and c is the parameter for the number of classes. The degree of assignment, u_{ij} , is given by Eq. (6).

$$u_{ij} = \left(\sum_i^c \left(\frac{\|x_j - \bar{x}_i\|}{\|x_j - \bar{x}_i\|} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (6)$$

- (4) The centroid for each class is recalculated, and clustering is completed when there is no longer any change in the centroids in any class. If there is change in a centroid, the process is recommenced from step (2).

2.3. Impact Analysis

In many cases, performance values are functions of multiple specification values. Because of this, it is not easy to determine the effect of each variable. Partial dependence plots are therefore employed in order to visualize the relationship between performance and each variable. Partial dependence plots are graphs that express the effect of variables other than the variable being focused on as averages. Figure 11 shows examples of partial dependence plots. The figure shows five partial dependence plots in which performance y is expressed in terms of five

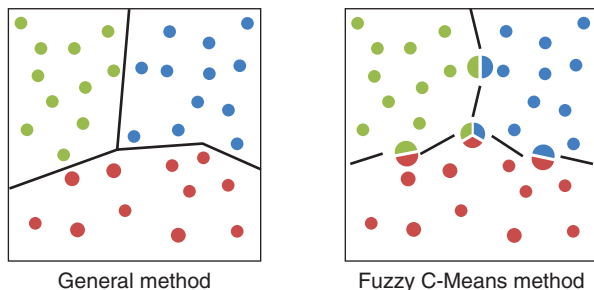


Fig. 10 Difference between clustering methods

specifications. The horizontal axes show the specifications, and the vertical axes show performance y .

Because y is a function of variables termed Factors 1 to 5, even if, for example, Factor 1 was to take a specific fixed value, the value of y would continue to fluctuate if Factors 2 to 5 changed within the scope of their upper and lower limit values. Repeating the operation of determining the average of the fluctuating value of y and plotting this value against the value of Factor 1 makes it possible to obtain the characteristic shown as the solid black line. This expresses the average characteristic of change in y in relation to Factor 1. The two broken red lines show the positions of the 25th percentile and the 75th percentile of the amplitude of fluctuation in y , providing a yardstick for the magnitude of the fluctuation in y due to other variables. The same processing can be used to create partial dependence plots for Factors 2 to 5.

Figure 11 shows that the characteristic of y in relation to Factor 2 is basically linear, indicating a linear relationship between the two. By contrast, the characteristic of y in relation to Factors 1, 3, and 5 describes a curve, indicating a nonlinear relationship with these variables. The characteristic of y in relation to Factor 4 is basically horizontal, indicating that this variable has almost no effect on performance. The intervals between the broken red lines show that all of the variables are affected by the other variables. In the case of variable 2, there are two areas in which the interval between the red broken lines becomes narrower, indicating that in these points the effect of the other variables is diminished.

Impact factor is used to help enable quantitative comparison in order to make the magnitude of the effect of each specification easier to understand. Impact factor is an index that expresses the changes in performance values when specifications are changed, and is determined by partially differentiating performance values with respect to specifications. Terming the response surface function f , the impact factor, g_i ($i=1, 2, \dots$), of Factor x_i ($i=1, 2, \dots$) is expressed as follows:

$$g_i = \frac{\partial y}{\partial x_i} \cong \frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_i} \quad (7)$$

Like performance values, impact factor is a function of multiple variables, making it challenging to compare impact factor between variables. The average value of g_i is therefore employed as the Impact factor of variable x_i .

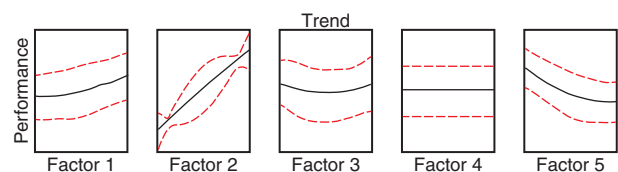


Fig. 11 Trend curves for parameters

Figure 12 shows the calculated Impact factor in descending order as a bar graph. In this example, Factor 2 displays the highest impact factor, two to three times greater than those of Factors 1, 3, and 5, while the impact factor of Factor 4 is basically 0. Unlike partial dependence plots, this method is suited to comparison of the magnitude and ratio of the effect of each variable.

Figure 13 shows the impact factor ratio as a band graph. The impact factor of Factor 2 is similar to the impact factor of Factors 1, 3, and 5 combined.

The selection of variables and the magnitude of change in values when the design is changed are determined on the basis of this impact factor data.

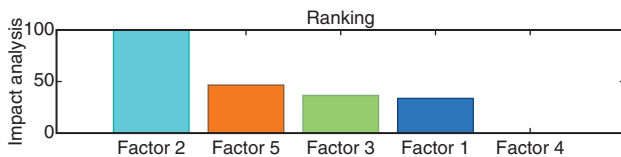


Fig. 12 Ranking of contribution ratio

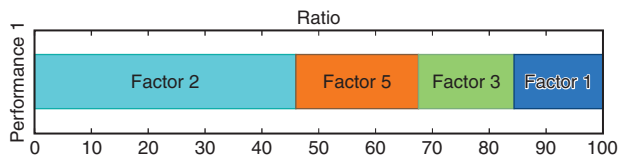


Fig. 13 Contribution ratio expressed as band graph

3. Application

To provide an example of the application of the vehicle performance design system developed in the present research, this section will discuss an analysis of a vehicle dynamic performance simulation using the system.

Data for the five specification values forming the input was sampled using a design of experiments method, and performance values were calculated using CAE in order to generate training data. Data for 400 cases was prepared for this example of system application. Response surfaces were generated using the completed performance value and specification datasets in order to create a large-scale DB.

PCE-LAR, as discussed in Section 2.1., was employed for the generation of response surfaces. In this example, three types of response surfaces were generated for three performance values. The use of these response surfaces not only makes it possible to conduct high-speed calculations for large-scale DB, but also to calculate specifications from performance and analyze impact factor. The results of three analyses will be discussed below.

3.1. Direct Search

Here, the user specifies target values in relation to three aspects of performance. As an example, specification values

will be calculated for the three performance values. Figure 14 shows a radar chart for the three target performance values.

Figures 15 and 16 show the results of calculation of specification values in relation to the three target values shown in Fig. 14. Figure 15 shows the results of the high-speed direct search method, and Fig. 16 shows the results of the high-accuracy search method. Both Figs. 15 and 16 are expressed as radar charts made up of the five specification values, and show five alternative solutions using the target

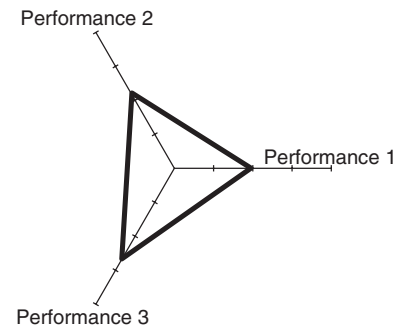


Fig. 14 Target performances

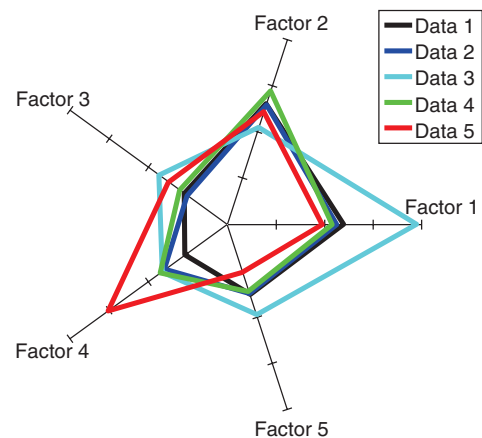


Fig. 15 Direct search results (High-speed search)

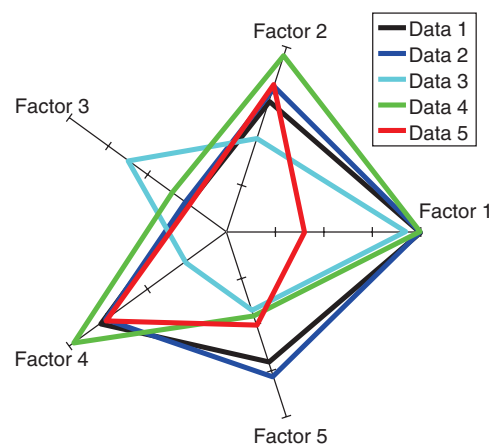


Fig. 16 Direct search results (High-accuracy search)

values specified in Fig. 14.

The user simply has to specify the target performance values, and the system presents a combination of specification values that satisfies the performance targets. The time necessary to produce alternative solutions on the same workstation was 0.043 sec for the high-speed search method and 13.89 sec for the high-accuracy search method.

The user is able to select the desired specification values from the alternative solutions shown in Figs. 15 and 16. In addition, the use of two direct search methods makes it possible to obtain different candidates for solutions even for the same performance values.

3.2. Group Search

Group search is a method of data selection using data in the DB.

The user is able to search for specification values while verifying performance values using radar charts. Because this search method uses a clustering technique, it does not directly identify the desired performance values from among large volumes of data, but uses class data to help enable simpler decision making, in stages. Figure 17 shows the flow of decision making using class data.

- (1) The 100,000-case data in the DB is clustered into five classes and expressed as a radar chart of performance values.
- (2) The user selects the class made up of the desired performance values. As a result, data for 100,000 classes is reduced to data for 21,485 classes.
- (3) The 21,485-class data is reclustered into five classes and expressed as a radar chart of performance values.
- (4) The user selects the desired class from among the five performance value classes. As a result, the data is reduced to data for 7,306 cases. This narrowing-down operation is continued in order to further reduce the amount of case data.

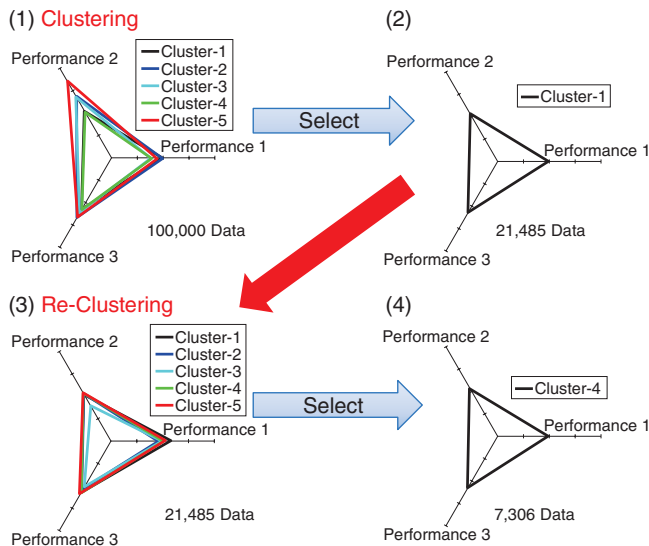


Fig. 17 Decision making flow

The use of data classified into five representative classes makes decision making simpler. Proceeding with decision making reduces the amount of data, advancing the process of narrowing down.

Figure 18 shows the final results of narrowing down the data.

The image on the the left-hand side of Fig. 18 shows the result obtained by narrowing down performance, while the image on the right-hand side shows the result obtained by narrowing down alternative solutions for specification values to five.

In this way, the group search method makes it possible to narrow down the data while verifying the balance between different aspects of performance using radar charts.

3.3. Analysis Using Impact Factor

By making it possible for users to identify specification values in relation to the desired performance through decision making, and to select viable options from data for alternative solutions for specification values, the system is able to increase the efficiency of the development process against conventional methods. The system also makes it possible for the user to set detailed specification values using the results of impact analysis. This process will be explained below.

Figure 19 shows the results of an impact analysis in relation to performance (Performance 1). The upper section of Fig. 19 shows the order of precedence of the Impact factor of each specification, while the lower section shows partial dependence plots for the five specifications

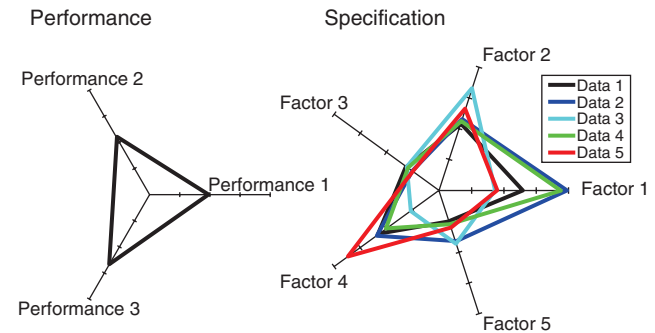


Fig. 18 Decision making results

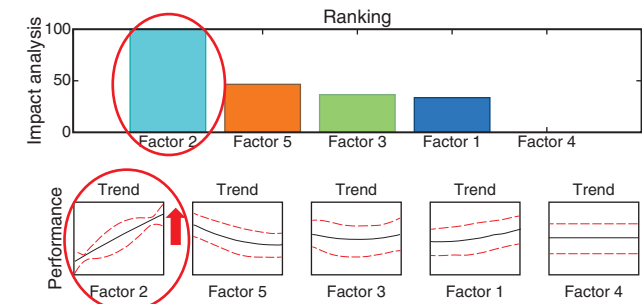


Fig. 19 Results of Impact analysis

in relation to performance. The upper section of the figure indicates that performance could be boosted by changing the parameters of Factor 2. Using the results of the partial dependence plots shown in the lower section of the figure, it is possible to identify the direction for the setting of parameters. The results shown here indicate that increasing the value of Factor 2 will increase performance.

Figures 20 and 21 show changes in performance when the value of Factor 2 is increased. Figure 20 shows a radar chart for the five specifications; Data 2 is produced by changing Factor 2 from its setting in Data 1. Figure 21 shows the change in performance when the value of Factor 2 is changed. The results show that performance is boosted against Data 1 in Data 2.

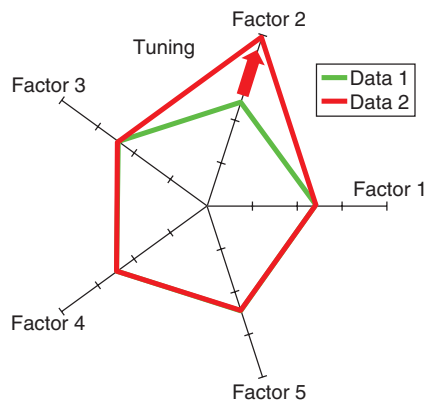


Fig. 20 Factor setting

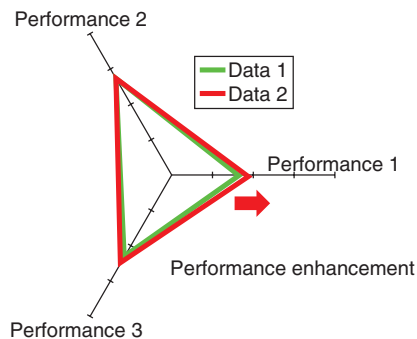


Fig. 21 Performance results

4. Conclusion

The research discussed in this paper developed a vehicle performance design system able to assign target values for different performance demands at the initial stage of the development process using three data mining techniques:

- (1) A response surface generation technique that helps enable high-accuracy predictions based on small amounts of data;
- (2) A decision-making technique that helps enable high-efficiency data searches that reflect the user's

intentions; and

- (3) An impact analysis technique that makes it possible to identify specifications that have the greatest effect on performance from the relationships between performance and specifications.

It was demonstrated that the application of this system to vehicle dynamic performance simulations enables users to conduct analyses simply and to determine performance values and specification values at high speed.

References

- (1) Fukushima, T., Ariyoshi, T., Uda, Y., Okuma, Y., Okabe, T., Yamamoto, K., Graening, L., Menzel, S., Olhofer, M.: Development of Real-time Aerodynamic Design System, Honda R&D Technical Review, Vol. 24, No. 2, p. 107-116
- (2) Tuno, K., Takafumi Kotani, T., Dobashi, M., Koide, F., Fukushima, T., Ariyoshi, T.: Development of Honda Smart Home System, Honda R&D Technical Review, Vol.26, No.1, p.11-17
- (3) Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data mining, Inference, and Prediction, Springer Series in Statistics, p. 369-370, (2009)
- (4) Blatman, G., Sudret, B.: Adaptive sparse polynomial Chaos expansion based on least angle regression, Journal of Computational Physics, Vol. 230, p. 2345-2367, (2011)
- (5) Schobi, R., Sudret, B., Wiart, J.: Polynomial Chaos-based Kriging, ETH Zurich Project Report RSUQ-2015-002, February 13th, (2015)
- (6) Bezdek, J. C., Ehrlich R., Full, W.: FCN: THE FUZZY c-MEANS CLUSTERING ALGORITHM, Computers & Geosciences, vol. 10, p. 2-3, (1984)

■ Author ■



Tokitomo ARIYOSHI



Masatoshi HARUMATSU



Kazushige YANO



Takahiro FUKUSHIGE



Kentaro YAMADA



Kaname NARUKAWA