Effect of an Experimental Design for Evaluating the Nonlinear Optimal Formulation of Theophylline Tablets Using a Bootstrap Resampling Technique

Hiroaki Arai,*[a] Tatsuya Suzuki,[a] Chosei Kaseda,[b] and Kozo Takayama[c]

[a] Formulation Technology Research Laboratories, Daiichi Sankyo Co., Ltd.; 1–12–1 Shinomiya, Hiratsuka, Kanagawa 254–0014, Japan: [b] Research and Development Headquarters, Yamatake Co.; 1–12–2 Kawana, Fujisawa, Kanagawa 251–8522, Japan: and [c] Department of Pharmaceutics, Hoshi University; 2–4–41 Ebara, Shinagawa-ku, Tokyo 142–8501, Japan. Received January 19, 2009; accepted March 10, 2009; published online March 12, 2009

The optimal solutions of theophylline tablet formulations based on datasets from 4 experimental designs (Box and Behnken design, central composite design, D-optimal design, and full factorial design) were calculated by the response surface method incorporating multivariate spline interpolation (RSM^S). Reliability of these solutions was evaluated by a bootstrap (BS) resampling technique. The optimal solutions derived from the Box and Behnken design, D-optimal design, and full factorial design dataset were similar. The distributions of the BS optimal solutions calculated for these datasets were symmetrical. Thus, the accuracy and the reproducibility of the optimal solutions enabled quantitative evaluation based on the deviations of these distributions. However, the distribution of the BS optimal solutions calculated for the central composite design dataset were almost unsymmetrical, and the basic statistic of these distributions could not be conducted. The reason for this problem was considered to be the mixing of the global and local optima. Therefore, self-organizing map (SOM) clustering was applied to identify the global optimal solutions. The BS optimal solutions were divided into 4 clusters by SOM clustering, the accuracy and reproducibility of the optimal solutions in each cluster were quantitatively evaluated, and the cluster containing the global optima was identified. Therefore, SOM clustering was considered to reinforce the BS resampling method for the evaluation of the reliability of optimal solutions irrespective of the dataset style.

Key words response surface method; formulation; optimization; bootstrap; self-organizing map; theophylline tablet

In recent years, the "Quality by Design (QbD)" concept has been advocated in the International Conference on Harmonisation (ICH) Q8 guidelines,[1] and the establishment of a science-based rationale and a design space in pharmaceutical formulation development is desired. In a pharmaceutical development study based on QbD concepts, it is important to have an enhanced understanding of the relationship between process parameters and quality attributes. In particular, it is necessary to establish a design space to identify multidimensional combinations of the many causal factors that determine target quality. The design of experiments (DOE) is a useful systematic approach for resolving multidimensional problems such as determining the relationship between input factors and process outputs.[2,3] The response data collected DOE are often applied to fit mathematical equations. These equations serve as models to predict the outcome with any given combination of values, and it is possible to calculate optimal solutions. In recent times, scientific approaches such as the response surface method (RSM) and the artificial intelligence (AI) technique have been used for DOE analysis and resolving optimization problems.[4—11] Further, overlapping techniques of some response surface models for multiple quality attributes enable the generation of a common design space with successful operating ranges.[2]

While determining the design space using a predicted model, it is important to evaluate the reliability of the model since it is being used to estimate optimal solutions. The optimal solutions estimated by classical RSM using quadratic polynomial equations can be evaluated by statistical analysis. However, the reliability of optimal solutions estimated by certain nonlinear response surfaces cannot be directly evaluated using a conventional mathematical method. Therefore, we applied a bootstrap (BS) re-sampling technique[12—14] to evaluate the reliability of the optimal solutions predicted by RSM incorporating multivariate spline interpolation (RSM^S),[15] and we previously reported that the novel method was suitable for evaluating the accuracy and precision of the optimal solution.[16] DOE produces certain variations in analytical targets. For example, the central composite design, the Box and Behnken design, *etc.*, are useful for modeling a response surface around continuous factors.[2,3] Therefore, we quantitatively evaluated the effect of the DOE variations on the reliability of the predicted optimal solutions using the BS technique. In this study, we used an experimental dataset of theophylline tablets prepared by the fluidized bed granulation method.

Theoretical

BS for Parameter Estimation The BS technique was introduced by Efron[17] as a computer-based method for estimating the standard error of an empirical distribution of an observed sample. Let $x=(x_1, \ldots, x_n)$ be an n sample with an unknown distribution function F depending on an unknown real parameter θ. The problem is to evaluate the parameter θ by a statistic $\hat{\theta}=s(x)$ from sample x and evaluate the estimation accuracy although distribution F is unknown. In order to evaluate the estimation accuracy, B samples were generated from the initial sample x by resampling. These samples were called BS samples and denoted by x^{*b}.

A BS sample $x^{*b}=(x_1^{*b}, \ldots, x_n^{*b})$ was generated by random resampling by replacing the initial sample x. The distribution function of a BS sample x^{*b} is $\hat{F}$, *i.e.*, the empirical distribution of x. A BS replicate of estimator $\hat{\theta}=s(x)$ is $\hat{\theta}^{*b}=s(x^{*b})$. Therefore, for the mean of sample x, the estima-

* To whom correspondence should be addressed. e-mail: arai.hiroaki.b4@daiichisankyo.co.jp

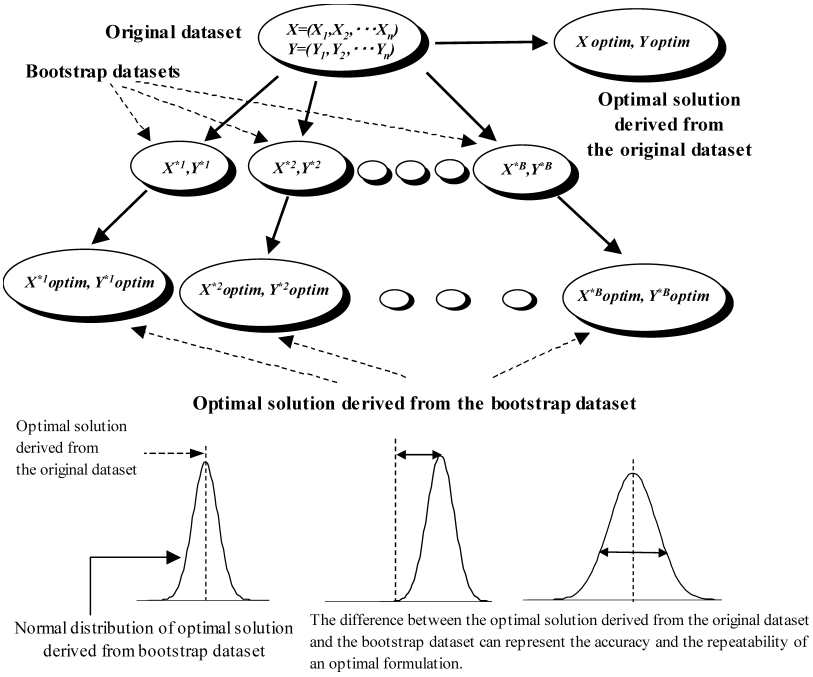


Fig. 1. Evaluation Process for Estimating the Accuracy and Precision (Repeatability) of an Optimal Formulation Based on the Bootstrap Resampling Technique

tor is $s(x)=(1/n)\sum^{n}_{i=1}x_i$, and a BS replicate will be $s(x^{*b})=(1/n)\sum^{n}_{i=1}x_i^{*b}$.

Novel evaluation procedures for an optimal solution have been applied in this technique. The BS evaluation process is shown in Fig. 1, and it has been described as follows.

Step 1. The BS dataset corresponding to the respective original datasets (comprising n data points) is generated by BS resampling that is repeated n times to form an ensemble comprising n results.

Step 2. Step 1 is repeated B times, and B units of the BS dataset are generated.

Step 3. The optimal solution is calculated as X^* *optim*, X^{*1} *optim*, X^{*2} *optim*…X^{*B} *optim* and Y^* *optim*, Y^{*1} *optim*, Y^{*2} *optim*…Y^*B *optim* for each BS dataset, and the distribution of the optimal solution is generated.

Step 4. The optimal solution and standard deviation of the BS analysis are calculated according to Eqs. 1 and 2 as follows.

$$F_{B.m}=\frac{1}{B}\sum_{b=1}^{B}F^{*b} \tag{1}$$

$$SD_B=\left[\frac{1}{B-1}\sum_{b=1}^{B}(F^{*b}-F_{B.m})^2\right]^{1/2} \tag{2}$$

where F^{*b} is the optimal solution of a specified property that is calculated from each BS dataset, $F_{B.m}$ is the BS optimal solution of the same property corresponding to the original solution, and SD_B is the standard deviation of the distribution of F^{*b}.

The accuracy of the original optimal solution, which was calculated from the original dataset, can be evaluated by comparing it to the BS optimal solutions. If the accuracy of the BS optimal solution deviates from that of the optimal original solution, the optimal original solution is considered to have a low reliability with regard to accuracy. In addition, the precision of the optimal original solution can also be evaluated by using BS standard deviation. A large BS standard deviation indicates poor precision of the optimal original solution.

Self-organizing Map A self-organizing map (SOM) is a feedforward neural network that implements a nonlinear projection from high-dimensional input vectors onto a low-dimensional (typically two-dimensional) array of nodes, called a map.[18,19] At time t, each neuron k is characterized by the reference vector $m_k(t)=[m_{i1}(t), m_{i2}(t), \ldots, m_{in}(t)]$ and a position in low-dimensional (typically two-dimensional) nodes represented by the vector $w_i(t)=[w_{i1}(t), w_{i2}(t), \ldots, w_{in}(t)]$. Each input vector $x(t)=[x_1(t), x_2(t), \ldots, x_n(t)]$ is compared with each reference vector, and the winner vector is considered to be that closest to the stimulus. The weights of the weight vector in the neighborhood of the winner vector are adjusted with a strength $\gamma(v,t)$ that is proportional to their distance to the winner vector, v. The neighborhood function $\gamma(v,t)$ is typically Gaussian with standard deviation σ. Winner vectors are adjusted on the basis of the following equation.

$$m_k(t+1)=m_k(t)+\alpha(t)\gamma(v,t)[x(t)-m_k(t)] \tag{3}$$

where $\alpha(t)$ is the algorithm's learning rate. Both $\alpha(t)$ and $\sigma(t)$ will typically decrease exponentially during the ordering phase and then linearly in the convergence phase.

Experimental

Materials Theophylline (Shiratori Pharmaceutical Co., Ltd., Japan), lactose (200-mesh grade, DMV International, The Netherlands), cornstarch (Nihon Shokuhin Kako Co., Ltd., Japan), carmellose calcium (Gotoku Chemical Co., Ltd., Japan), hydroxypropylcellulose (Nippon Soda Co., Ltd., Japan), and magnesium stearate (Nitto Kasei Kogyo K.K., Japan) were all of grades conforming to the current *Japanese Pharmacopoeia* (JP).

Experimental Design The formulations of theophylline tablets are listed in Table 1. The lactose/cornstarch ratio (X_1), percent carmellose calcium (X_2), and percent hydroxypropylcellulose (X_3) were selected as causal

Table 1. Formulation of Theophylline Tablets with the 4 Experimental Designs

Experiment number	Factor level			Formula (mg/tablet)						
	X_1	X_2	X_3	THEO[a)]	LAC[b)]	CS[c)]	CCa[d)]	HPC[e)]	Mg-St[f)]	Total
1	−1	−1	−1	100.0	26.4	61.6	6.0	4.0	2.0	200.0
2	−1	−1	0	100.0	25.8	60.2	6.0	6.0	2.0	200.0
3	−1	−1	1	100.0	25.2	58.8	6.0	8.0	2.0	200.0
4	−1	0	−1	100.0	25.2	58.8	10.0	4.0	2.0	200.0
5	−1	0	0	100.0	24.6	57.4	10.0	6.0	2.0	200.0
6	−1	0	1	100.0	24.0	56.0	10.0	8.0	2.0	200.0
7	−1	1	−1	100.0	24.0	56.0	14.0	4.0	2.0	200.0
8	−1	1	0	100.0	23.4	54.6	14.0	6.0	2.0	200.0
9	−1	1	1	100.0	22.8	53.2	14.0	8.0	2.0	200.0
10	0	−1	−1	100.0	44.0	44.0	6.0	4.0	2.0	200.0
11	0	−1	0	100.0	43.0	43.0	6.0	6.0	2.0	200.0
12	0	−1	1	100.0	42.0	42.0	6.0	8.0	2.0	200.0
13	0	0	−1	100.0	42.0	42.0	10.0	4.0	2.0	200.0
14	0	0	0	100.0	41.0	41.0	10.0	6.0	2.0	200.0
15	0	0	1	100.0	40.0	40.0	10.0	8.0	2.0	200.0
16	0	1	−1	100.0	40.0	40.0	14.0	4.0	2.0	200.0
17	0	1	0	100.0	39.0	39.0	14.0	6.0	2.0	200.0
18	0	1	1	100.0	38.0	38.0	14.0	8.0	2.0	200.0
19	1	−1	−1	100.0	61.6	26.4	6.0	4.0	2.0	200.0
20	1	−1	0	100.0	60.2	25.8	6.0	6.0	2.0	200.0
21	1	−1	1	100.0	58.8	25.2	6.0	8.0	2.0	200.0
22	1	0	−1	100.0	58.8	25.2	10.0	4.0	2.0	200.0
23	1	0	0	100.0	57.4	24.6	10.0	6.0	2.0	200.0
24	1	0	1	100.0	56.0	24.0	10.0	8.0	2.0	200.0
25	1	1	−1	100.0	56.0	24.0	14.0	4.0	2.0	200.0
26	1	1	0	100.0	54.6	23.4	14.0	6.0	2.0	200.0
27	1	1	1	100.0	53.2	22.8	14.0	8.0	2.0	200.0
28	−1.73	0	0	100.0	8.2	73.8	10.0	6.0	2.0	200.0
29	1.73	0	0	100.0	73.8	8.2	10.0	6.0	2.0	20.0
30	0	−1.73	0	100.0	44.5	44.5	3.0	6.0	2.0	200.0
31	0	1.73	0	100.0	37.5	37.5	17.0	6.0	2.0	200.0
32	0	0	−1.73	100.0	42.7	42.7	10.0	2.6	2.0	200.0
33	0	0	1.73	100.0	39.3	39.3	10.0	9.4	2.9	200.0

a) Theophylline, *b*) lactose, *c*) corn starch, *d*) carmellose calcium (ECG-505), *e*) hydroxypropylcellulose (HPC-L), *f*) magnesium stearate.

factors. These variables were assigned according to the (a) Box and Behnken design, (b) central composite design, (c) D-optimal design, and (d) full factorial design, respectively. We prepared 33 types of theophylline tablet formulations.

Preparation Method of Sample Tablets Theophylline was milled using an impact mill (TASM-1CS, Tokyo Atomizer Co., Ltd., Japan), and lactose was sieved through a 60-mesh screen. The milled theophylline (volume mean diameter, approximately 23.6 μm), sieved lactose, cornstarch, and carmellose calcium were blended in a polyethylene bag for 2 min. The mixture was granulated with approximately 6% (w/v) aqueous hydroxypropylcellulose solution in a fluid-bed granulator (MFL.01, Vector Corporation, U.S.A.). The granules were lubricated with magnesium stearate and blended in a bin blender (Turbula unit type T2C, Willy A. Bachofen AG, Switzerland) for 5 min. The final blend was compressed into tablets using a universal testing machine (Autograph AG-5000B, Shimadzu Co., Ltd., Japan) at a compression force of approximately 7.85 kN.

Determination of Response Variables The dissolution ratio of theophylline for the first 15 min (Y_1) and hardness (Y_2) were selected as the response variables that were to be evaluated in the resulting tablets. The values of both variables were the mean of 3 determinations.

a) Dissolution: Dissolution testing was performed by the paddle method according to the recommendations of the current JP at 50 rpm in 900 ml of water at 37 °C. The dissolved theophylline was assayed by an automated flow-through UV spectrophotometric method at 243 nm with a 10-mm-long cell (Automated dissolution apparatus, Toyama Sangyo Co., and Shimadzu Co., Ltd., Japan).

b) Hardness: The hardness of the resulting tablets was measured using a hardness tester (Tablet tester type 6D, Dr. Schleuniger Pharmatron AG, U.S.A.).

Four experimental datasets were prepared according to the 4 DOE; they consisted of 3 causal factors and 2 response variables with 33 theophylline tablet formulations. These datasets are shown in Table 2.

Evaluation of the Optimal Solution by the BS Technique The reliability of the simultaneous optimal solution was evaluated by the BS evaluation technique (described in Fig. 1). In this study, the frequency of BS resamplings was set at 1000.

Evaluation Indices of Accuracy and Reproducibility for Optimal Solutions The d and CV_B values used as evaluation indices of accuracy and repeatability of the optimal solution were calculated using Eqs. 4 and 5.

$$d = \frac{|F - F_{B.m}|}{F} \times 100 \tag{4}$$

$$CV_B = \frac{SD_B}{F_{B.m}} \times 100 \tag{5}$$

where F is the original solution of a specified property, $F_{B.m}$ is the BS optimal solution of the same property corresponding to the original solution, and SD_B is the BS standard deviation.

Software The software used in this study is as follows. JMP®6 (SAS institute Inc., U.S.A.) was used for preparing the DOE and statistical analysis. dataNESIA™ (Yamatake Corporation, Japan) was used for generating the RSM^S and estimating the optimal solution. This software consists of a multidimensional spline interpolation program and a nonlinear optimization program.[20] Viscovery® (Eudaptics Software Gmbh, Austria) was used for SOM clustering. This software can order complex data based on similarity. The ordered data are separated into clusters on the basis of similarity and these clusters are presented in a multi-colored map. The resulting map can be used to extract the features hidden in the data.

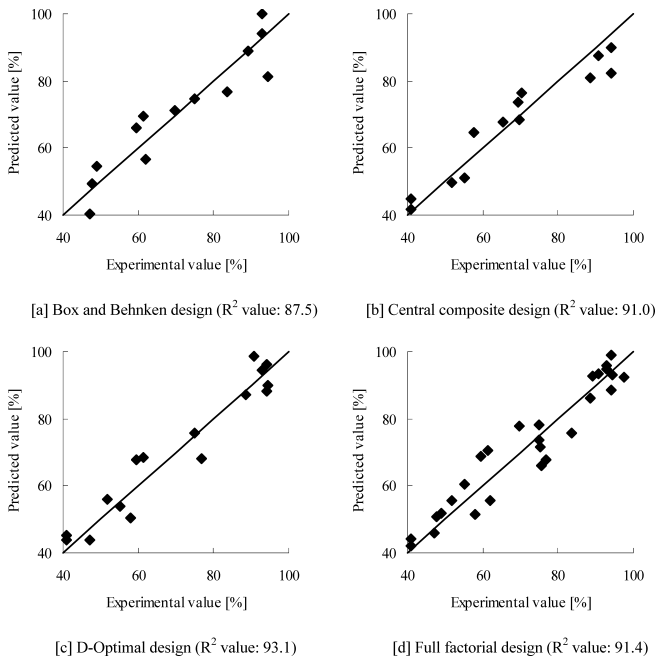
Results and Discussion

Simultaneous Optimization by RSM^S The dissolution

Table 2. Original Experimental Design Datasets of Theophylline Tablet Formulations

	Experiment number	Factor level			% Dissolved theophylline at 15 min, Y_1	Hardness (N), Y_2
		X_1[a] (%)	X_2[b] (%)	X_3[c] (%)		
Full factorial design	11	50	3.0	3.0	75.6	103.0
	13	50	5.0	2.0	97.4	89.2
	23	70	5.0	3.0	75.4	122.6
Box & Behnken design	4	30	5.0	2.0	89.1	108.9
	6	30	5.0	4.0	47.6	110.8
	8	30	7.0	3.0	74.9	93.2
	14	50	5.0	3.0	69.7	114.7
	16	50	7.0	2.0	92.9	93.2
	18	50	7.0	4.0	61.9	117.7
	24	70	5.0	4.0	49.1	144.2
	26	70	7.0	3.0	83.7	129.5
D-optimal design	2	30	3.0	3.0	59.4	80.4
	10	50	3.0	2.0	94.3	93.2
	12	50	3.0	4.0	47.0	118.7
	20	70	3.0	3.0	61.4	112.8
	22	70	5.0	2.0	92.8	112.8
	5	30	5.0	3.0	76.9	93.2
	15	50	5.0	4.0	58.0	74.5
	17	50	7.0	3.0	75.1	100.0
Central composite design	1	30	3.0	2.0	88.5	87.3
	3	30	3.0	4.0	41.0	98.1
	7	30	7.0	2.0	90.6	86.3
	9	30	7.0	4.0	51.9	106.9
	19	70	3.0	2.0	94.1	104.0
	21	70	3.0	4.0	41.0	142.2
	25	70	7.0	2.0	94.2	117.7
	27	70	7.0	4.0	55.0	137.3
	14	50	5.0	3.0	69.7	114.7
	28	10	5.0	3.0	65.4	90.2
	29	90	5.0	3.0	69.3	127.4
	30	50	1.5	3.0	57.7	111.7
	31	50	8.5	3.0	70.2	117.6
	32	50	5.0	1.3	94.7	100.9
	33	50	5.0	4.7	34.2	135.2

a) Lactose/cornstarch ratio (% lactose), *b*) % carmellose calcium, *c*) % hydroxypropylcellulose.



[a] Box and Behnken design (R^2 value: 87.5) [b] Central composite design (R^2 value: 91.0)

[c] D-Optimal design (R^2 value: 93.1) [d] Full factorial design (R^2 value: 91.4)

Fig. 2. Relationships between Experimental and Predicted Values of Y_1 (% dissolved at 15 min)

ratio of theophylline for the first 15 min (Y_1) and the hardness (Y_2) of each formulation are shown in Table 2. The dissolution profiles and hardness varied among the formulations. The response surfaces of Y_1 and Y_2 were generated by RSM[S] as functions of 3 causal factors; the lactose/cornstarch ratio (X_1), percent carmellose calcium (X_2), and percent hydroxypropylcellulose (X_3). The reliability of each response surface was estimated by using a conventional leave-one-out cross-validation (LOOCV) method. The R^2 values defined in Eq. 6[21] for Y_1 and Y_2 were calculated for the 4 DOE datasets.

$$R^2 = 100 \times \left(1 - \frac{SSE}{SST}\right) \tag{6}$$

where *SSE* is the sum of the squared error between the predicted and the measured values. *SST* is the sum of the squared error between each measured value and the average of the measured value.

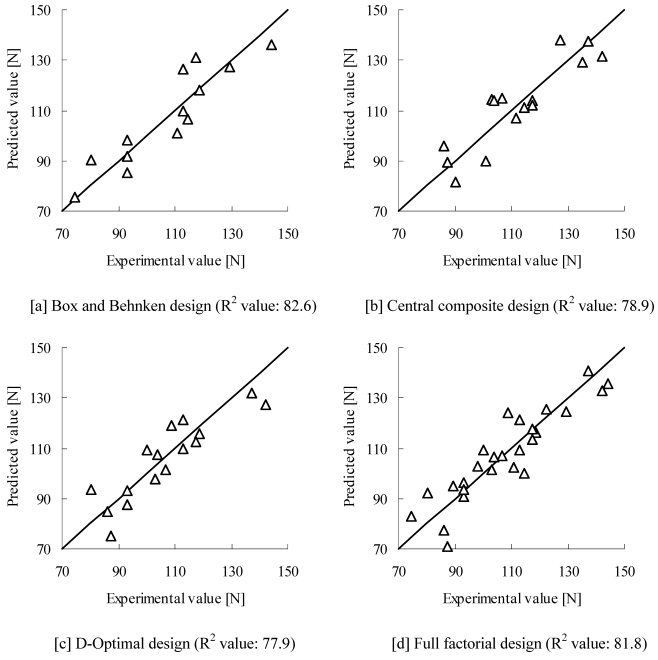
These results are shown in Figs. 2 and 3. All R^2 values were sufficiently high (more than 75); this suggested that all the response surfaces were highly reliable. The simultaneous optimal solutions for X_1, X_2, and X_3 and the predicted values for Y_1 and Y_2 calculated by RSM[S] are shown in Table 3. The optimal solutions of all factors (X_1, X_2, and X_3), except those obtained from the central composite design dataset, were al-

most similar. On the other hand, all the response variables for Y_1 and Y_2 were almost similar regardless of the DOE variation. The fluctuation range of the causal factors in the case of the central composite design is generally wider than that in the case of the other DOE datasets; as a result, the response surface area also becomes larger than that of the other designs.

Evaluation of the Optimal Formulation by the BS Method BS datasets for the 4 DOE datasets were generated by 1000 BS resamplings. The results of the optimal formulations and the predicted responses are shown with a 95% confidence interval in Table 4. Indices corresponding to the accuracy (d) and reproducibility (CV_B) of the optimal solution as defined in Eqs. 4 and 5 are shown in Figs. 4 and 5, respectively. The BS optimal solutions calculated for the Box and Behnken design, the D-optimal design, and the full factorial design datasets were almost the same as the original solution. Moreover, the BS standard deviation calculated for these datasets was sufficiently small, and the 95% confidence intervals of the optimal solutions, calculated by the percentile method, were sufficiently narrow for practical formulation studies.

On the other hand, the BS optimal solution calculated for the central composite design dataset was different from the original solution to a certain extent, and the BS standard deviation was relatively large. The d values that served as an accuracy index and the CV_B values that functioned as a reproducibility index were calculated for the central composite design dataset; these values were apparently larger than those of the other DOE datasets. The distributions of the optimal solutions calculated for the central composite design dataset are shown in Fig. 6. These distributions were almost unsymmetrical; in particular, that of the causal factors X_1 (lactose/cornstarch ratio) exhibited 2 peaks. These results imply that the BS method, which assumes to the normal approximation, could not be directly applied in the case of the central composite design dataset.

The distribution of the optimal solutions did not exhibit



[a] Box and Behnken design (R^2 value: 82.6) [b] Central composite design (R^2 value: 78.9)

[c] D-Optimal design (R^2 value: 77.9) [d] Full factorial design (R^2 value: 81.8)

Fig. 3. Relationships between Experimental and Predicted Values of Y_2 (Hardness).

Table 3. Optimized Formulations and Predicted Responses Estimated by the 4 Experimental Designs

Factors and responses	Experimental design			
	Box & Behnken design	Central composite design	D-optimal design	Full factorial design
X_1 [a)]	66.2	81.1	66.1	66.8
X_2 [b)]	6.06	6.91	5.57	5.84
X_3 [c)]	2.75	2.48	2.48	2.66
Y_1 [d)]	81.7	84.5	83.5	83.5
Y_2 [e)]	122.0	125.6	112.6	118.5

a) Lactose/cornstarch ratio (% lactose), *b*) % carmellose calcium, *c*) % hydroxypropylcellulose, *d*) % dissolved theophylline at 15 min, *e*) hardness [N].

Table 4. Bootstrap Optimal Solutions, Bootstrap Standard Deviations, and 95% Confidence Intervals of Optimal Solutions Generated by the 4 Experimental Designs When Bootstrapping Was Repeated 1000 times

	Optimized formulations			Predicted responses	
	X_1 [c)] (%)	X_2 [d)] (%)	X_3 [e)] (%)	Y_1 [f)] (%)	Y_2 [g)] (N)
Box & Behnken design					
Original solution[a)]	66.2	6.06	2.75	81.7	122.0
Bootstrap solution[b)]	66.2 (2.34)	5.9 (0.337)	2.72 (0.118)	81.9 (2.58)	120.8 (3.09)
95% confidence interval	63.3—68.6	5.18—6.37	2.54—2.93	76.4—87.2	113.5—125.5
Central composite design					
Original solution[a)]	81.1	6.91	2.48	84.5	125.6
Bootstrap solution[b)]	77.2 (7.42)	6.26 (0.720)	2.54 (0.288)	81.0 (5.54)	123.7 (3.40)
95% confidence interval	66.1—86.9	4.60—7.27	2.04—3.06	71.0—91.7	117.1—130.4
D-optimal design					
Original solution[a)]	66.1	5.57	2.48	83.5	112.6
Bootstrap solution[b)]	66.4 (1.12)	5.59 (0.239)	2.53 (0.070)	82.9 (1.87)	113.8 (2.37)
95% confidence interval	64.1—68.4	5.13—6.05	2.40—2.70	79.1—86.7	109.2—118.3
Full factorial design					
Original solution[a)]	66.8	5.84	2.66	83.5	118.5
Bootstrap solution[b)]	66.9 (0.86)	5.76 (0.266)	2.65 (0.093)	83.3 (1.61)	118.0 (2.20)
95% confidence interval	65.1—68.6	5.16—6.14	2.48—2.80	80.8—85.9	111.3—121.8

( ): Bootstrap standard deviation. *a*) Obtained from the original dataset, *b*) bootstrap resampling frequency, approximately 1000 times, *c*) lactose/cornstarch ratio (% lactose), *d*) % carmellose calcium, *e*) % hydroxypropylcellulose, *f*) % dissolved theophylline at 15 min, *g*) hardness.

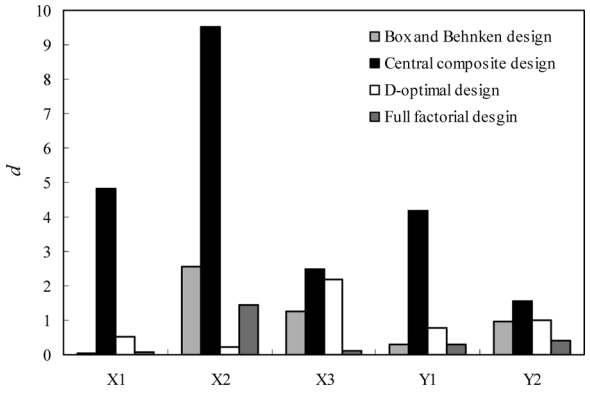


Fig. 4. Comparison of d Index for Accuracy of the Optimal Solution between the 4 Experimental Designs

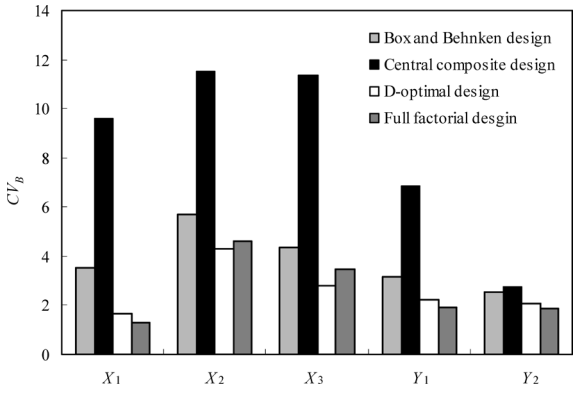


Fig. 5. Comparison of the CV_B Index for Precision (Repeatability) of the Optimal Solution between the 4 Experimental Designs

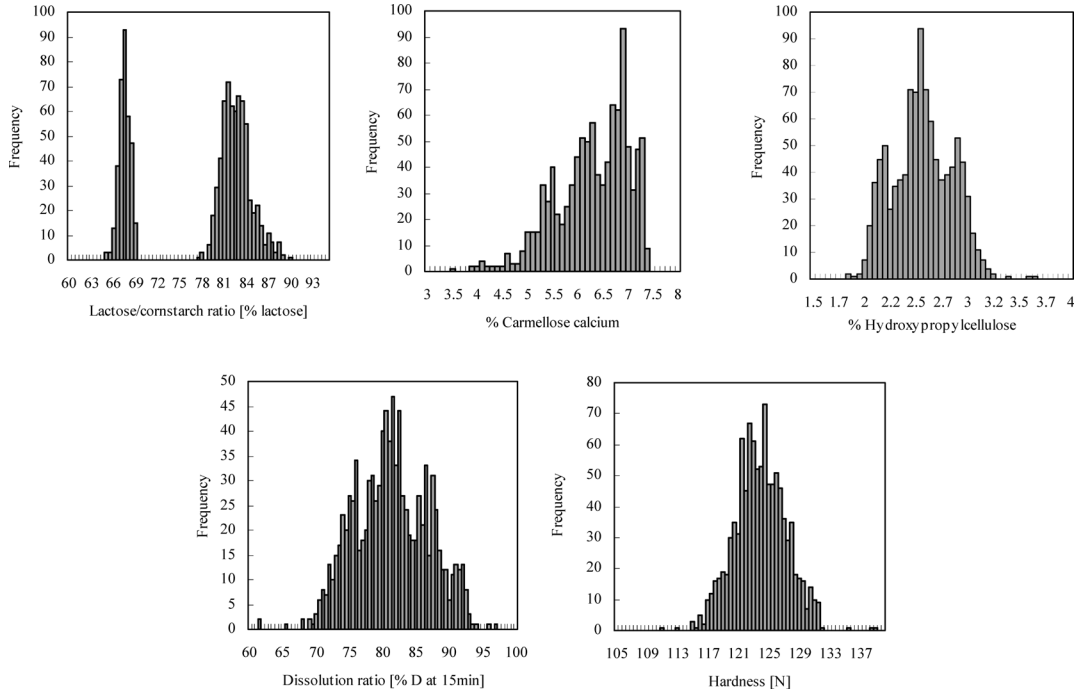


Fig. 6. Histograms of Optimal Solutions Generated by Bootstrap Resampling (N=1000) for the Central Composite Design Dataset

normal distributions due to the change in the calculation ranges of causal factors in each BS resampling dataset, as shown in Fig. 7. As mentioned in the introduction, a response surface is generated using the all data points, and the optimal formulation is estimated inside the inscribed circle on the response surface. This implies that, the maximum–minimum range of each causal factor affects the original optimal solutions. However, BS datasets are generated by random resampling of the original dataset. In the central composite design dataset, axial points exist at both ends of rectangular axes and only 2 axial points are arranged symmetrically at central coordinates. If an axial point is chosen as a BS resampling datum, the calculation range in the response surface generated by the BS dataset is the same as that in the response surface generated by the original dataset; the optimal solution in this case is the global solution. However, if an axial point is not selected as a BS resampling datum, the BS optimal solution estimated for the BS dataset is searched for within the narrowed calculation range; the optimal solution in this case

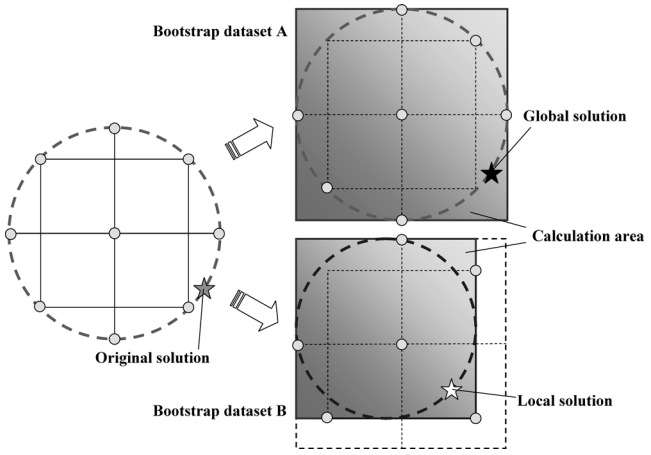


Fig. 7. The 2-Peak Histogram of the Optimal Solution Estimated by the Central Composite Design Dataset

is the local solution. Consequently, 2 peaks are observed in the histogram of the BS optimal solution for the central composite design reflecting a mix of the global and the local solution distributions. To resolve the above problem, SOM clustering was applied for dividing the 2-peak histogram of the BS optimal solutions.

Evaluation of the Optimal Formulation by the BS Method in Combination with SOM Clustering The distributions of the optimal solutions calculated for the central composite design dataset were classified by SOM clustering. The SOMs of the BS optimal formulations of each causal factor are shown in Fig. 8. The BS optimal formulations were classified into clusters as shown in Table 5. The BS optimal solutions were divided into 4 clusters. Then, the *d* index corresponding to the accuracy of each optimal solution was calculated according to Eq. 4 in each cluster (Fig. 9). The BS optimal solution in the case of cluster 1 was similar to the original solution. The *d* values of cluster 1 were apparently smaller than those of the other clusters. Therefore, it was considered that the global optimal solution was contained in the BS optimal solutions of cluster 1.

The SOM of the causal factors X_1 (the lactose/cornstarch ratio) was visually characterized into 2 colored groups; clusters 1, 3, and 4 and cluster 2 (Fig. 9). Further, the distributions of the optimal solutions in the cluster 1 are shown in Fig. 10. These distribution ranges were almost narrower than the original distributions (shown in Fig. 6); in particular, around the distribution of the causal factor X_1, the histogram exhibited 1-peak distributions. This indicated that the SOM could satisfactorily separate the mixture of the BS optimal solutions into several groups.

The frequency of the resampled data points from the individual experiments in each SOM cluster is shown in Table 6. In cluster 1, the frequency of resampling data points is almost even among all the experiments. However, in the other clusters, the resampled frequencies of the axial data points on the rectangular axes, for example, the numbers 29, 31, and 32, were extremely few as compared to other experiments. In the previous study, the number of BS required for the evaluation of the optimal solution was at least around 300. Therefore, we considered cluster 1 to be appropriate for determining the optimal solution from the central composite design dataset.

Conclusion

It was confirmed that a reasonable response surface was

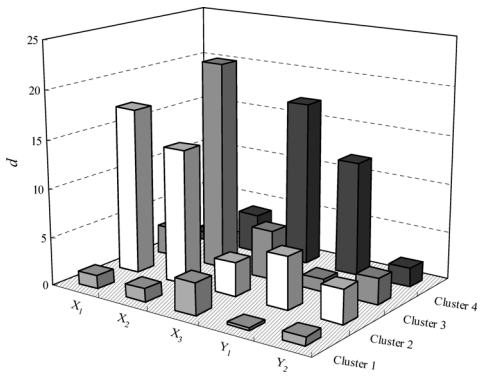


Fig. 9. Comparison of the *d* Index to Determine the Accuracy of the Optimal Solution in the Self-organizing Map Clusters

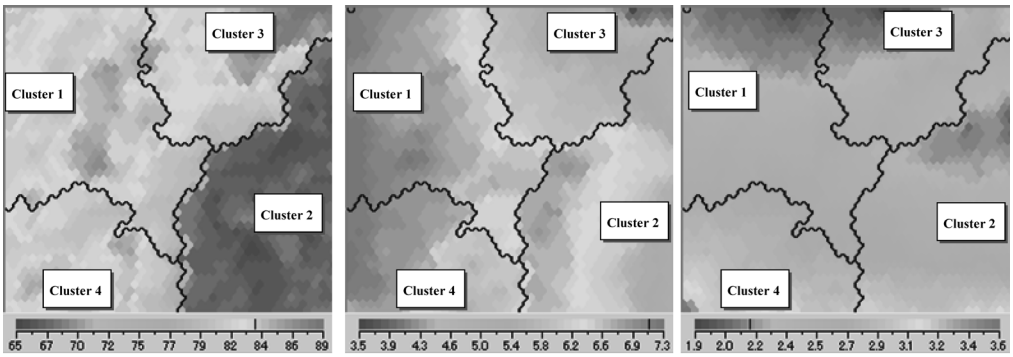


Fig. 8. Self-organizing Maps of Optimal Solutions of Input Factors Estimated from Bootstrap Samples (N=1000) for the Central Composite Design Dataset

Table 5. Bootstrap Optimal Solutions and Bootstrap Standard Deviations in the Self-organizing Map Cluster for Optimal Solutions Estimated from Bootstrap Samples of the Central Composite Design Dataset

	Optimized formulations			Predicted responses	
	X_1[c] (%)	X_2[d] (%)	X_3[e] (%)	Y_1[f] (%)	Y_2[g] (N)
Original solution[a]	81.1	6.91	2.48	84.5	125.6
Bootstrap solution[b]					
Cluster 1	82.2 (1.61)	6.82 (0.30)	2.39 (0.21)	84.8 (4.23)	124.5 (1.75)
Cluster 2	67.2 (0.77)	5.95 (0.57)	2.56 (0.25)	79.9 (4.45)	121.2 (2.81)
Cluster 3	83.5 (2.53)	5.42 (0.61)	2.35 (0.22)	83.4 (4.62)	122.2 (1.91)
Cluster 4	81.8 (1.62)	6.63 (0.50)	2.90 (0.14)	74.5 (2.75)	128.1 (2.21)

( ): Bootstrap standard deviation. *a*) Obtained from the original dataset, *b*) bootstrap resampling frequency, approximately 1000 times, *c*) lactose/cornstarch ratio (% lactose), *d*) % carmellose calcium, *e*) % hydroxypropylcellulose, *f*) % dissolved theophylline at 15 min, *g*) hardness.

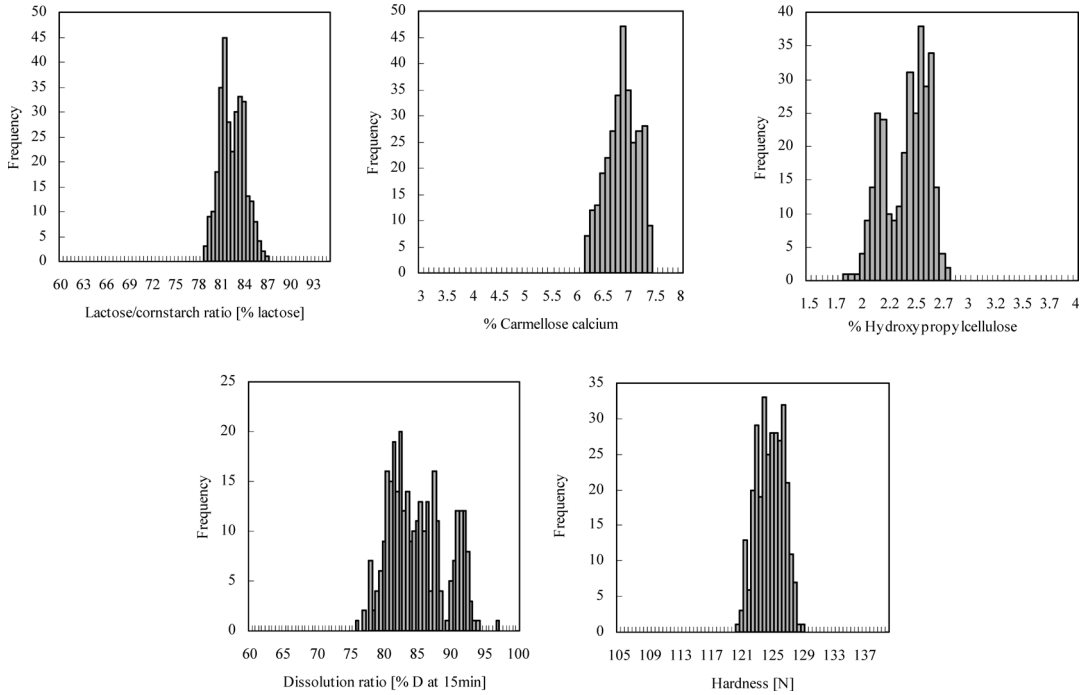


Fig. 10. Histograms of Optimal Solutions in Cluster 1 Separated by Self-organizing Map Clustering of the 1000 Bootstrappings of the Central Composite Design Dataset

Table 6. Frequency of Resampled Data Points in the Self-organizing Map Cluster for Optimal Solutions Estimated from Bootstrap Samples (N=1000) of the Central Composite Design Dataset

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Number of BS samples	305	343	167	185	1000
Experiment number	Frequency of resampled data points				
1	261	341	152	191	943
3	277	385	180	173	1013
7	315	401	159	185	1060
9	261	366	166	169	959
14	281	382	179	190	1031
19	278	404	165	170	1016
21	261	384	187	233	1065
25	297	380	167	157	1001
27	282	370	158	201	1011
28	264	343	240	156	1002
29	453	**2**	253	302	1008
30	307	332	157	165	961
31	449	346	**3**	176	974
32	378	387	220	**35**	1018
33	211	337	119	272	938

generated by RSM^S in all the DOEs studied. However, the method is based only on the BS technique, which was used to estimate the reliability of the optimal solution predicted by RSM^S, and is not applicable to the central composite design dataset. Therefore, SOM was applied to evaluate the reliability of optimal solution, and successful results were obtained using SOM clustering. Thus, SOM clustering was considered to reinforce the BS resampling method in the evaluation of the reliability of the optimal solutions despite the DOE dataset.

Acknowledgements The authors are very grateful to Mr. Hiroaki Tsutsui and Mr. Junji Nishimura, Yamatake Co., for their valuable advice in the data analysis.

References

1) ICH draft guideline, “Pharmaceutical Development Q8,” version 4.3, November, 2005.
2) Montgomery D. C., “Design and Analysis of Experiments,” John Wiley & Sons, U.S.A., 1997.
3) Lewis G. A., Mathieu D., Phan-Tan-Luu R., “Pharmaceutical Experimental Design,” Marcel Dekker, New York, 1999.
4) Rekhi G. S., Nellore R. V., Hussain A. S., Tillman L. G., Malinowski H. J., Augsburger L. L., *J. Controlled Release*, **59**, 327—342 (1999).
5) Paterakis P. G., Korakianiti E. S., Dallas P. P., Rekkas D. M., *Int. J. Pharm.*, **248**, 51—60 (2002).
6) Huang Y. B., Tasai Y. H., Yang W. C., Chang J. S., Wu P. C., *Biol. Pharm. Bull.*, **27**, 1626—1629 (2004).
7) Aikhatib H. S., Sakr A., *Pharm. Dev. Technol.*, **8**, 87—96 (2003).
8) Marengo E., Cavalli R., Rovero G., Gasco M. R., *Pharm. Dev. Technol.*, **8**, 299—309 (2003).
9) Matsumura M., Nakagami H., Yamao T., Takayama K., Nagai T., *Chem. Pharm. Bull.*, **42**, 1902—1908 (1994).
10) Takayama K., Morva A., Fujikawa M., Hattori Y., Obata Y., Nagai T., *J. Controlled. Release.*, **68**, 175—186 (2000).
11) Wu P. C., Obata Y., Fujikawa M., Li C. J., Higashiyama K., Takayama K., *J. Pharm. Sci.*, **90**, 1004—1014 (2001).
12) Ueda N., Nakano R., *IEEE Int. Conf. Neural Netw. Proc.*, **1**, 101—104 (1995).
13) Dupret G., Koda M., *Eur. J. Oper. Res.*, **134**, 141—156 (2001).
14) Zhang J., *Neural Netw.*, **12**, 927—938 (1999).
15) Takayama K., Obata Y., Morishita M., Nagai T., *Pharmazie*, **59**, 392—395 (2004).
16) Arai H., Suzuki T., Kaseda C., Ohyama K., Takayama K., *Chem. Pharm. Bull.*, **55**, 586—593 (2007).
17) Efron B., Tibshirani R. J., “An Introduction to the Bootstrap,” Chapman and Hall, New York, 1993.
18) Kohonen T., *Proc. IEEE*, **78**, 1464—1480 (1990).
19) Noriega G., *Neural Netw.*, **21**, 130—139 (2008).
20) Kaseda C., *Kagakukougaku*, **68**, 315—317 (2004).
21) Bourquin J., Schmidli H., Hoogevest P. V., Leuenberger H., *Eur. J. Pharm. Sci.*, **6**, 287—300 (1998).