International Journal of Pharmaceutics 396 (2010) 75–82

Contents lists available at ScienceDirect

International Journal of Pharmaceutics

journal homepage: www.elsevier.com/locate/ijpharm

Role of individual test samples in optimal solutions in pharmaceuticals predicted using a nonlinear response surface method

Yoshinori Onuki*, Shingo Kikuchi, Akihito Yasuda, Kozo Takayama

Department of Pharmaceutics, Hoshi University, 2-4-41 Ebara, Shinagawa-ku, Tokyo 142-8501, Japan

ARTICLE INFO

Article history:
Received 9 January 2010
Received in revised form 24 May 2010
Accepted 9 June 2010
Available online 15 June 2010

Keywords:
Formulation optimization
Confidence intervals
Bootstrap resampling technique
Kohonen's self-organizing map
Bayesian estimation
Dermatological patch

ABSTRACT

Establishing a method to evaluate the reliability of an optimal solution is an exciting challenge for the nonlinear response surface method. We reported previously that the bootstrap (BS) resampling technique and Kohonen's self-organizing map are promising tools for meeting this challenge. To understand the usefulness of these techniques further, we employed a formulation optimization study of photocrosslinked polyacrylic acid (PAA) hydrogel as a case study. In a series of experiments, a large number of optimal solutions were generated with BS resampling and they were classified into three distinct clusters with SOM clustering. Using analysis of Bayesian estimation, we clarified the mode of generating clusters; e.g., cluster 2 was distinguished by the difference in features between the BS optimal solutions and the original optimal solution, whereas cluster 3 was distinguished by the substantial change in the shape of the response surfaces. We concluded that cluster 1 represents the global optimal solution, and then estimated 95% confidence intervals of the optimal solutions using the BS optimal solutions. These findings prove that our method is a valid approach for evaluating nonlinear optimal solutions. This method has applications for establishing a science-based rationale for, and a design space in, pharmaceutical formulation development.

1. Introduction

The international conference on harmonisation of technical requirements for registration of pharmaceuticals for human use (ICH) Q8 guideline recently propounded the establishment of a science-based rationale. The concept of quality by design described in the ICH Q8 guideline states that "quality cannot be tested into products; i.e., quality should be built in by design". Response surface methodology (RSM) is one recommended method for establishing "the design space". RSM is used widely to optimize formulations of pharmaceuticals (Khuri and Cornel, 1987; Myers and Montgomery, 1995). The general procedures for determining optimal solutions via RSM include the collection of experimental data, the generation of a response surface between the causal factors and response variables, and the search for individual and simultaneous optimal solutions (Onuki et al., 2008b). Selection of the methods used to generate the response surface is the most important for optimizing formulation because, for the most part, the estimation of an accurate optimal solution depends significantly on the extent to which the generated model approximates the actual relationship.

We developed an RSM that incorporates multivariate spline interpolation, RSM-S. Multivariate spline (MVS) interpolation is used to generate the response surface. The basic concept of MVS interpolation involves a boundary element method (Sandwell, 1987). MVS interpolation can be used to estimate nonlinear relationships between factors and response variables with high accuracy. Green functions are used for the minimum curvature interpolation of multidimensional data points. MVS interpolation estimates multidimensional data using a thin-plate spline that represents the sum of interpolations made with a Green function and a linear polynomial equation ("thin-plate estimation") (Wahba, 1990). Thus, it enables the incorporation of observational data, including experimental errors, naturally. We have applied RSM-S to the formulation optimization of various pharmaceuticals. Our findings to date suggest that RSM-S is an effective tool for optimizing the formulation of pharmaceuticals (Arai et al., 2007; Kikuchi and Takayama, 2009; Nishikawa et al., 2008; Onuki et al., 2005, 2004, 2008b; Takayama et al., 2004).

For nonlinear RSMs, such as RSM-S, establishing a method to evaluate the reliability of the optimal solution estimate has remained a challenge to overcome. In the case of a classical linear RSM, the reliability of optimal solution can be evaluated by a statistical technique. In contrast, no satisfactory method has yet been established for nonlinear RSM. We recently devised a novel method to address this issue (Kikuchi and Takayama, 2009; Onuki et al., 2008b). The method is achieved by making use of bootstrap (BS) resampling and Kohonen's self-organizing map (SOM) with an RSM-S. We previously applied this method to the preparation of

* Corresponding author. Tel.: +81 3 5498 5783; fax: +81 3 5498 5783.
E-mail address: onuki@hoshi.ac.jp (Y. Onuki).

0378-5173/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpharm.2010.06.013

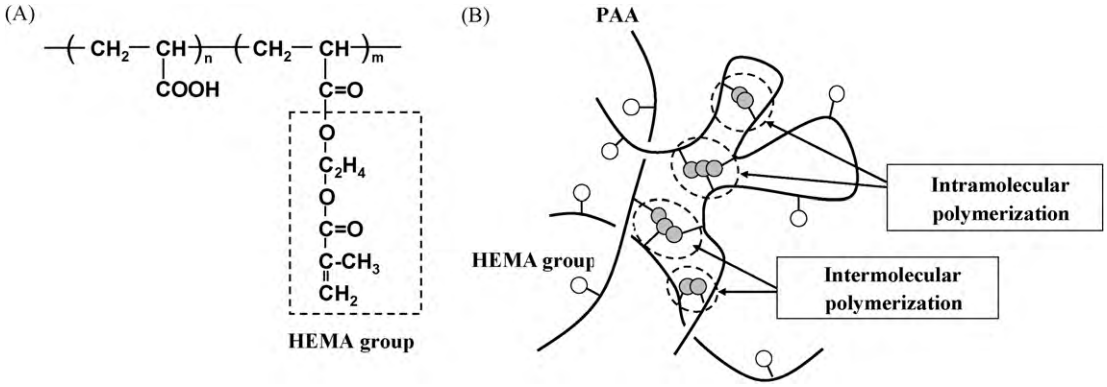

Fig. 1. Formulation optimization study of photocrosslinked PAA hydrogel for use in dermatological patch adhesive. (A) Chemical structure of HEMA-derivatized PAA. (B) Photopolymer mechanism by formation of crosslinked PAA networks by intermolecular and intramolecular polymerization of HEMA groups in PAA molecules.

a theophylline tablet as a case study (Onuki et al., 2008b). A large number of BS samples were generated from the original data set using BS resampling, and simultaneous optimal solutions for each BS sample, BS optimal solutions, were then estimated. Because the BS optimal solutions seemed to have both global and local optimal solutions, we performed SOM clustering to extract the global optimal solutions from the whole BS optimal solutions and we estimated the 95% confidence intervals of the variables.

From the findings, we thought that the method was promising for evaluating the reliability of nonlinear optimal solutions. In this study, we applied it to the other optimization study of a hydrogel used in a dermatological patch adhesive. Recently, our laboratory has developed a photocrosslinked polyacrylic acid (PAA) hydrogel, which is made from a PAA modified with 2-hydroxyethyl methacrylate (HEMA) (Fig. 1A) (Nishikawa et al., 2008; Onuki et al., 2005, 2008a). For the preparation of the hydrogel, a photopolymerization technique was employed. When the polymer aqueous solution is exposed to ultraviolet (UV) light, HEMA groups in the PAA molecules react, leading to the construction of inter- and intramolecular crosslinking structures with covalent bonding and gel formation (Fig. 1B). Compared with conventional hydrogels made by crosslinking with ionic bonds, the hydrogel is superior in several aspects. For example, because our hydrogel can be set at a lower crosslinking density than other gels, we expected to obtain a hydrogel that retains more water without losing mechanical strength. Further, because photopolymerization enables the rapid conversion of a polymer solution into a hydrogel under physiological conditions, this hydrogel is also attractive as a biomaterial. In a previous study, we investigated the relationships between formulation factors and physical properties (Onuki et al., 2005). We also performed a formulation optimization study using RSM-S and then decided the optimal conditions to prepare the preferred hydrogel for use as a dermatological patch adhesive (Onuki et al., 2005).

In this study, we evaluated the reliability of the optimal solution by using the novel method incorporating BS resampling technique and SOM clustering. We also investigated the mechanism for generating global and local optimal solutions in BS optimal solutions. For the investigation, Bayesian estimation was employed. The findings offer a scientific rationale showing that our strategy is a valid approach for evaluating the reliability of nonlinear optimal solutions.

2. Theoretical

2.1. BS resampling

BS resampling is a computer-based method of statistical inference. The basic concept of BS resampling is random sampling from the original data (Efron and Tibshirani, 1993). A BS sample, $x^* = (x_1^*, x_2^*, \ldots, x_n^*)$, is sampled randomly, with a replacement, from the original data points $x_1, x_2, \ldots, x_n$. As for the BS resampling, duplication of the same data in one BS sample is allowed; therefore, a part of the original data may be selected several times in each BS sample. By repeating the BS resampling procedure, a large number of BS samples are generated from the original data. BS resampling is commonly used for estimating confidence intervals and the bias and variance of an estimator. The BS resampling technique has been applied recently to various research fields (Efron and Tibshirani, 1993). The BS resampling process for evaluating the reliability of simultaneous optimal solutions estimated by RSM-S is shown in Fig. 2. Further details are described in Section 3.

2.2. SOM clustering

The SOM is a feedforward-type neural network model (Kohonen, 1995). The typical structure of the SOM comprises one input layer and one output layer, and the array of nodes is located in the output layer. Every node, $m_i = (m_{i1}, m_{i2}, \ldots, m_{in})$, has the same number of parametric reference vectors as the input vector (x). The SOM algorithm is based on unsupervised, competitive learning. When an input vector is given to the network, the Euclidean distances from the input vector to all the nodes are calculated (Takayama et al., 2000, 2004). By comparing the Euclidean distances, every node competes for similarity to the input vector, and the winner node (m_c) is defined as the node that is closest to the input vector. After competitive learning, the weight vectors of the winner node and the neighborhood area are updated. The update formula for a node with a reference vector (m_k) is calculated as:

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

where $\alpha(t)$ is a monotonically decreasing learning coefficient and $x(t)$ is the input vector. The neighborhood function $\gamma(v,t)$ depends on the lattice distance between the winner node and the node (v). This process is repeated for each input vector for a (usually large) number of cycles. The network ultimately associates the output nodes with groups or patterns of input vectors. SOM clustering requires no human intervention during learning. The SOM has received much attention recently as a promising tool for data mining. In this study, SOM clustering was used to sort the cluster of global optimal solutions from the widely distributed BS solutions.

2.3. Bayesian estimation

Bayesian estimation is a method to estimate the posterior probability of the hidden variable from observable data under an assumption of a model of the dependency of hidden and observable variables (Cooper and Herskovitz, 1992). It is based on Bayes's The-

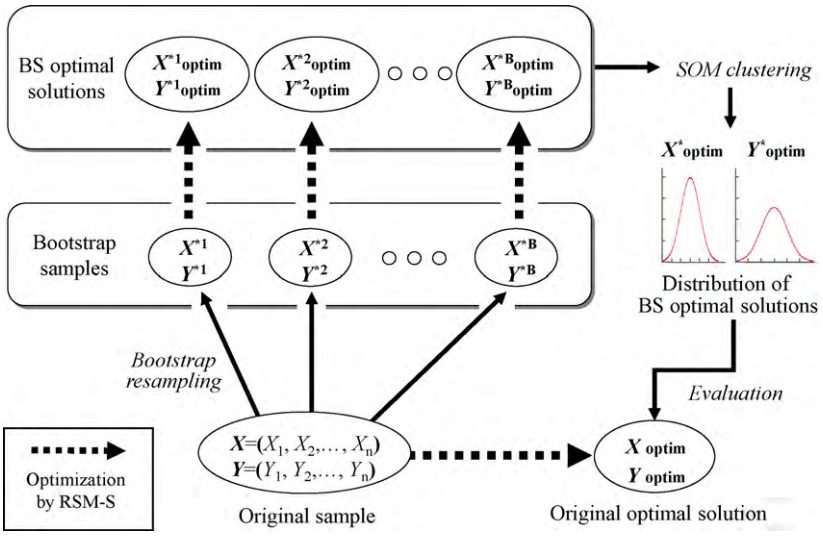

Fig. 2. Procedure for evaluating the reliability of simultaneous optimal solutions estimated by RSM-S.

orem, which is a simple mathematical formula used for calculating conditional probabilities as follows:

$$P\left(\frac{H}{D}\right) = \frac{P(D/H)P(H)}{P(D)},$$

where H is the hypothesis, D is the data, $P(H)$ is the prior probability of H, $P(D/H)$ is the likelihood of H, $P(D)$ is the marginal probability of D, and $P(H/D)$ is the posterior probability. $P(H)$ is the probability that H is correct before the data D are seen. $P(D/H)$ is the conditional probability of seeing the data D given that the hypothesis H is true.

$P(D)$ is the prior probability of witnessing the data D under all possible hypotheses, and is expressed as follows:

$$P(D) = \sum P\left(\frac{D}{H}\right) P(H).$$

The Bayesian interpretation of probability can be seen as an extension of logic that enables reasoning with uncertain statements. The rules of Bayesian statistics can be justified by requirements of rationality and consistency. At present, Bayesian probability is one of the most popular interpretations of the concept of probability. This probabilistic knowledge is used widely in various fields such as artificial intelligence, applied statistics, and, more recently, bioinformatics (Jansen et al., 2003; Vadim and Artem, 2005; Wade et al., 2009; Xue et al., 2006).

3. Materials and methods

3.1. Model data

The data dealt with in this study correspond with Table 7 in our previous article (Onuki et al., 2005). This experiment was performed using an orthogonal experimental design (Table 1). The initiator concentration and modification of the base polymer with HEMA were selected as formulation factors, and 10 kinds of model formulations were prepared. Their physical properties such as gel fraction, degree of swelling, turbidity, and probe tack were measured as crucial response variables for a dermatological patch adhesive. In this study, the experimental data set, including model formulations and their response variables, is called the "original sample". As a result of the experiment, RSM-S was used to construct a reliable model of the correlations between the formulation factors and response variables. The leave-one-out cross-validation showed very good correlation coefficients (r) of more than 0.931 between the predicted and experimental values (data not shown).

Table 1
Model formulations for photocrosslinked PAA hydrogel based on an orthogonal experiment design.

Formulation number	Initiator concentration (wt% of polymer amount)	Modification with HEMA (mol%)
Rp.1	0.29	8.7
Rp.2	0.29	26.3
Rp.3	1.71	8.7
Rp.4	1.71	26.3
Rp.5	1.00E−03	17.5
Rp.6	2.00	17.5
Rp.7	1.00	5.0
Rp.8	1.00	30.0
Rp.9	1.00	17.5
Rp.10	1.00	17.5

Further explanations of the experiments were described fully in the previous article (Onuki et al., 2005).

3.2. Data analysis

A procedure for evaluating the reliability of optimal solutions is shown in Fig. 2. To begin, we estimated the optimal solution from the original sample with RSM-S, called "the original optimal solution". We defined a hydrogel having sufficient mechanical strength and high water content as being preferable. The conditions for seeking a simultaneous optimal solution was the following: gel fraction and probe tack value must be more than 80% and 200 mN/5 mm Φ, and a lower degree of swelling and turbidity were the preferred characteristics. Further information about the original optimal formulation estimated from the previous study is shown in Table 2.

Table 2
Predicted and observed values of a simultaneous optimal solution estimated from the original data.

	Predicted values	Observed values
Formulation factor		
Initiator concentration (%)	0.90	–
Modification with HEMA (mol%)	16.1	–
Characteristics		
Gel fraction (%)	80.3	85.9 ± 0.3
Degree of swelling	254.6	245.3 ± 8.2
Probe tack (mN/5 mm Φ)	391.1	483.5 ± 39.6
Turbidity (ABS at 505 nm)	0.401	0.462 ± 0.018

These data are quoted from our previous article (Onuki et al., 2005).

The response variables of the hydrogel prepared according to the formulation were coincident with the predicted values. Thus, a reliable original optimal solution was thought to have been obtained from the study. Details of the optimization procedure in RSM-S have been described in full (Onuki et al., 2005).

One thousand BS samples were generated from the original sample using BS resampling, and the optimal solutions of each BS sample, "the BS optimal solutions", were estimated with RSM-S. dataNESIA® Version 3.0 (Yamatake Corp., Tokyo, Japan) was used for RSM-S and BS resampling.

To classify the BS optimal solutions into distinct clusters, SOM clustering was performed. BS optimal solutions and their estimated response variables were used as the input vectors. The number of nodes in the output was set at 2000. Viscovery® (Eudaptics Software GmbH, Vienna, Austria) was used for SOM clustering. This software offers several clustering techniques such as SOM-Ward, Ward, and SOM-Single-Linkage. Of these techniques, SOM-Ward was used for clustering the BS optimal solution because it is considered the most efficient in general.

The data sets of 1000 BS samples were arranged for analysis of Bayesian estimation. First, the compositions of the BS samples of each cluster (e.g., formulations resampled to a BS sample and their duplication number) were identified. Next, the formulations of the BS samples were sorted and coded. Formulations that were not resampled to a BS sample were coded as 0, and the resampled formulations were coded as 1 regardless of the number of duplications in the BS resampling process. We used the data sets of each cluster as the input data set and then analyzed them according to the Bayesian estimation. BeyoNet 4.0 software (AIST, Tsukuba, Japan) was used for the Bayesian estimation.

4. Results

4.1. Classification of BS optimal solutions into distinct clusters

Fig. 3 shows histograms of the formulation factors and the response variables of the BS optimal solutions. In advance of preparing the histograms, the data were transformed to a logit form, because they have lower and upper limits. The white line in the graphs represents the values of the original optimal solution. Their distributions differed from the normal distribution. By contrast, some conditions such as those involving modification with HEMA and the degree of swelling, and turbidity showed several peaks. This result is similar to that observed in our previous study (Onuki et al., 2008b). In our previous study, we attributed this to the coexistence of global and local optima in the whole BS optimal solutions.

To classify whole BS optimal solutions into distinct clusters, SOM clustering was conducted. SOM feature maps show that BS optimal solutions were classified into three clusters with distinct response variables (Fig. 4A). We also acquired the centroid BS optimal solutions of clusters from the reference vectors of the SOM, and we regarded them as the typical data sets of each cluster (Table 3 and Fig. 4B). Cluster 2 showed more modification with HEMA; e.g., the centroid value was 18.9 mol% compared with 16.1 mol% of the original optimal solution. The greater HEMA modification was accompanied by a lower degree of swelling, higher probe tack, and higher turbidity (Fig. 4A). By contrast, the centroid BS optimal solutions of clusters 1 and 3 were very close to the original optimal solution. Apart from the gel fraction, the response variables of cluster 3 were similar to those of cluster 1 in every way (Fig. 4A).

Table 3
Typical BS optimal solutions of each cluster.

	Cluster 1	Cluster 2	Cluster 3
Formulation factor			
Initiator concentration (%)	0.90	0.88	0.83
Modification with HEMA (mol%)	15.7	18.9	16.2
Characteristics			
Gel fraction (%)	80.2	80.0	81.9
Degree of swelling	263.6	214.8	253.3
Probe tack (mN/5 mm Φ)	383.0	430.1	404.1
Turbidity (ABS at 505 nm)	0.391	0.456	0.403

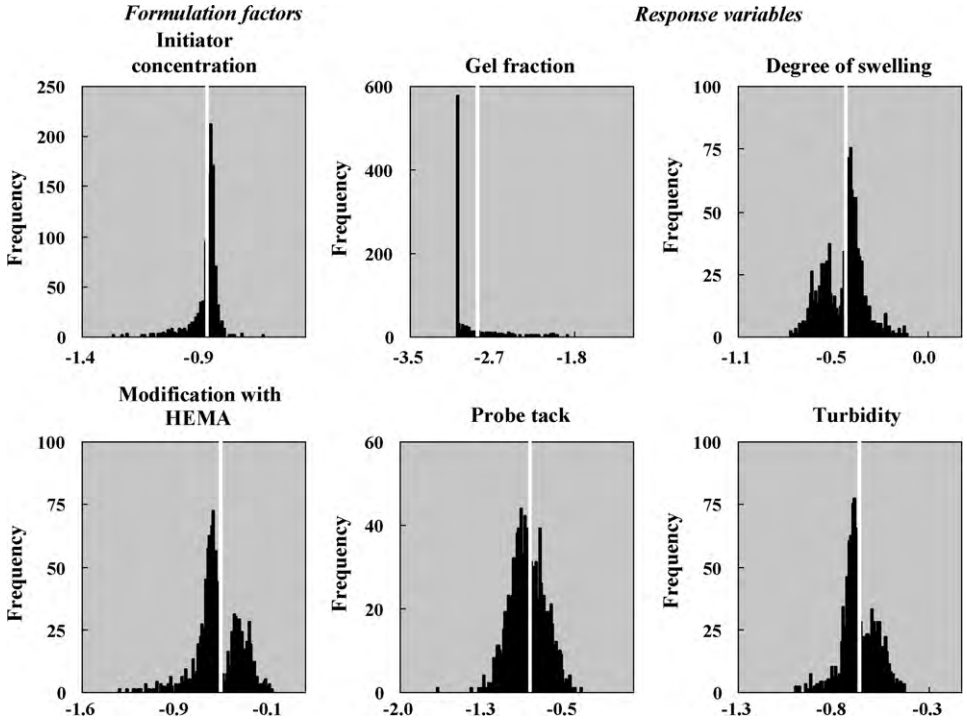

Fig. 3. Histograms of formulation factors and response variables generated with BS resampling. BS resampling was repeated 1000 times. These histograms represent 1000 optimal solutions estimated from the BS samples. The white lines represent the values of the original optimal solutions. Original data were transformed to a logit form prior to preparation of histogram.

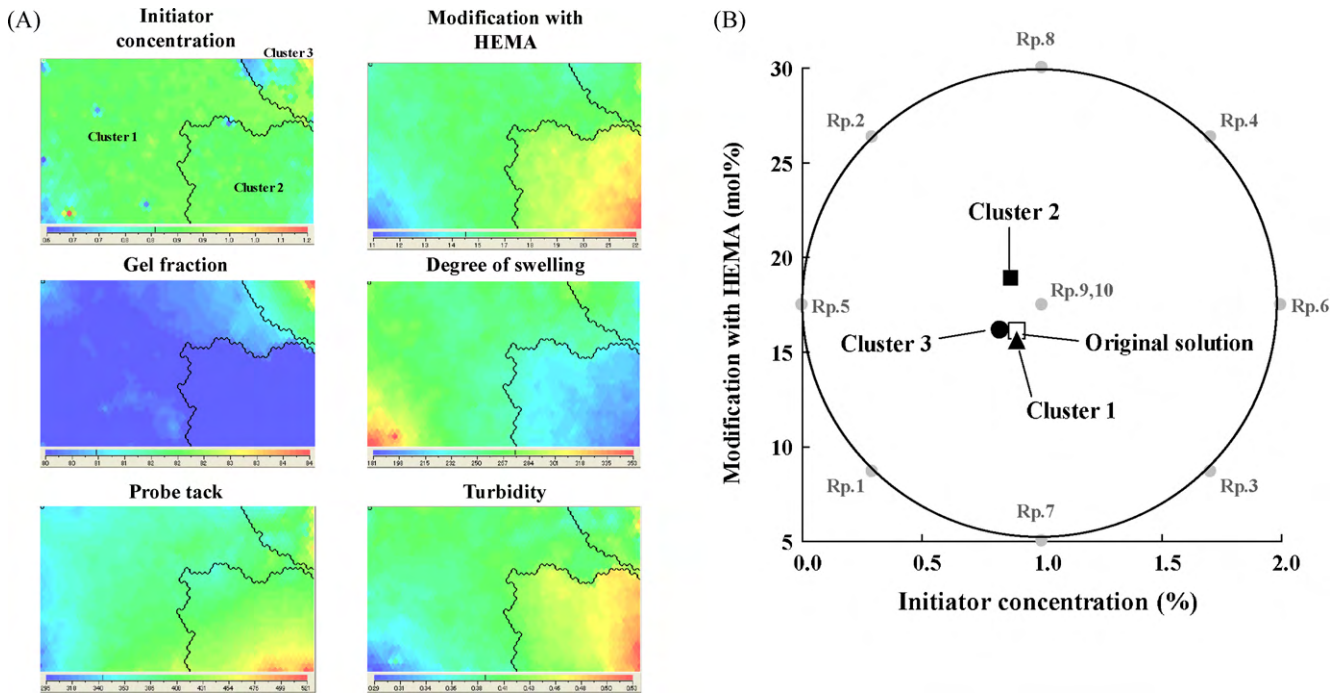

Fig. 4. Clusters of BS optimal solutions classified by SOM clustering. (A) SOM feature maps of formulation factors and response variables. (B) Centroid BS optimal solutions of clusters. (□) Original optimal solution, (▲) cluster 1, (■) cluster 2, and (●) cluster 3.

4.2. Elucidating the mode of generation of distinct clusters

To investigate how the SOM clustering detected the three clusters, we analyzed 1000 data sets of BS samples according to Bayesian estimation. Fig. 5 shows the Bayesian probabilities according to whether each formulation was resampled to the BS samples of clusters. In cluster 3, the resampled probability of Rp.5 was very low, 46.4%, indicating that the cluster was constructed by BS samples having less Rp.5. As for cluster 2, although the impact was not as strong as that for Rp.5 in cluster 3, the resampled probability of Rp.7 was the lowest, 88.1%. In contrast to clusters 2 and 3, the values of resampled probability of cluster 1 were very high (more than 93.3%) and changed little, suggesting that cluster 1 was composed of BS samples with a homogeneous proportion of formulations.

We next investigated the contribution of model formulations to the decision of optimal formulation using the leave-one-formulation-out (LOFO) approach. The data points corresponding with each formulation were removed from the original sample, and they were used as the LOFO samples. Their optimal solutions, "LOFO optimal solutions", were estimated by the same method using RSM-S (Table 4). The weights of the contribution of the formulation factors were normalized, and the distances from the original optimal solution to each LOFO optimal solution were evaluated. Removing the data points of Rp.7 caused the LOFO optimal solution to move furthest away from the original optimal solution. Rp.5, a crucial formulation for cluster 3, also made the optimal solution change markedly.

We further examined the similarity of the response variables estimated from two different response surfaces. One surface was generated from the original sample, "original response surfaces", and the other one was generated from the BS samples, "BS response surfaces". Because every BS sample differs from the original sample, this results in the generation of BS response surfaces with different shapes to the original surfaces. If the change in the shape of

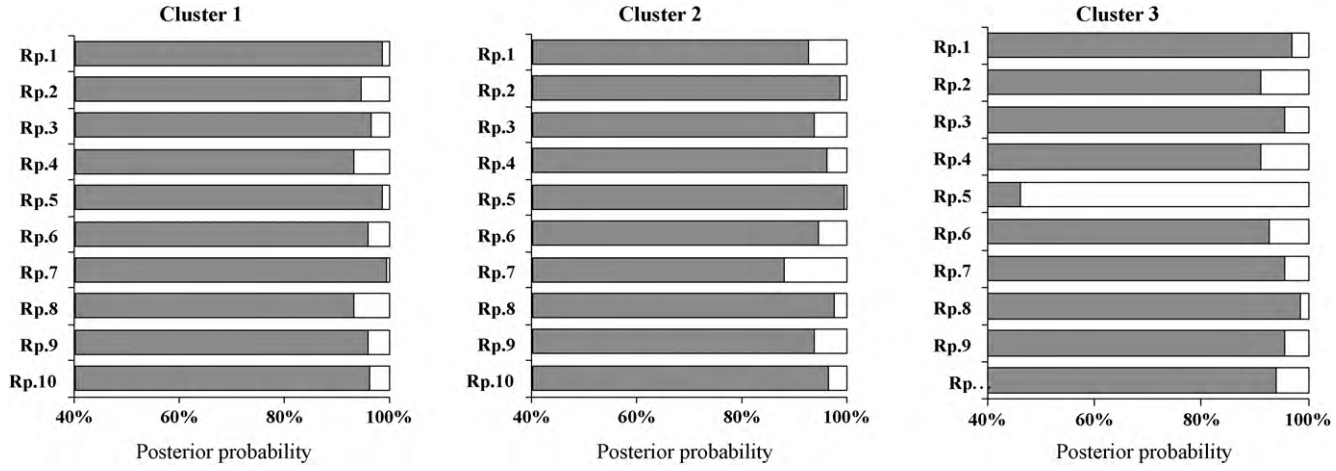

Fig. 5. Resampled probability of formulations to BS samples in clusters according to Bayesian estimation.

Table 4
Optimal solutions estimated according to the leave-one-formulation-out (LOFO) approach using RSM-S.

Formulation removed from original sample	Initiator concentration (wt% of polymer amount)	Modification with HEMA (mol%)	Normalized distance from the original optimal solution[a]
Rp.1	0.95	15.9	2.76
Rp.2	0.90	16.5	1.71
Rp.3	0.93	15.9	1.71
Rp.4	0.89	16.0	0.54
Rp.5	0.75	15.3	8.17
Rp.6	0.85	16.0	2.34
Rp.7	0.91	18.9	11.19
Rp.8	0.92	15.8	1.56
Rp.9	0.90	17.5	5.62
Rp.10	0.89	15.7	1.59

[a] These values indicate distances from the original optimal solution to LOFO optimal solutions on the coordinate in which the formulation factors of model formulations were coded as 0–100%.

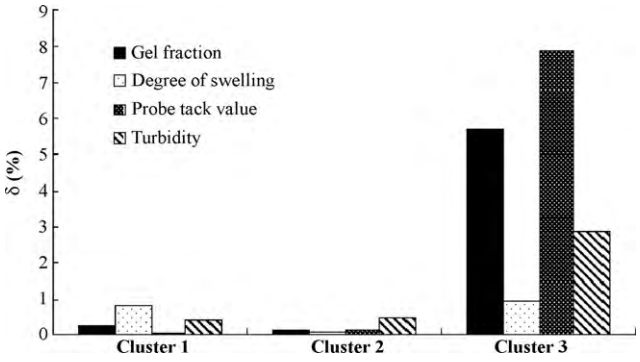

Fig. 6. Difference between the response variables estimated from the BS response surfaces and original response surfaces. A similarity index, δ, was defined so that higher δ values represent lower predictive accuracy of the response variables.

the BS response surfaces is substantial, the predicted values from the BS response surfaces also change significantly from those of the original response surfaces. In this experiment, we investigated the response variables of the centroid BS optimal solutions shown in Table 3. The centroid BS optimal solutions were regarded as typical optimal solutions of each cluster. Their response variables were estimated from the BS response surfaces. In addition, because they are not the model formulations shown in Table 1, we can regard these BS optimal formulations as untested formulations for the original surfaces. To compare the values of response variables, a degree of similarity, δ, was defined as follows:

$$\delta(\%) = \frac{\left|F_o - F^i_{B.m}\right|}{F_o} \times 100$$

where F_o signifies the response variables of the centroid BS optimal solutions estimated from the original response surfaces, and $F^i_{B.m}$ signifies the response variables estimated from the BS response surfaces shown in Table 3. As a result, the δ values of the response variables of cluster 3 were very high, whereas those of clusters 1 and 2 were very low. This suggests that the BS response surfaces of cluster 3 differed markedly from the original surfaces (Fig. 6).

5. Discussion

In order to evaluate the reliability of nonlinear optimal solutions, we recently developed a novel method based on integrating the BS resampling technique and SOM clustering into an RSM-S (Kikuchi and Takayama, 2009; Onuki et al., 2008b). We note that the methodology is similar to a parametric approach. RSM-S is a nonlinear method by nature; however, the processes of BS resampling and SOM clustering enable us to deal with the nonlinear BS optimal solutions as if they were parametric parameters. Although percentile method has been reported as a nonparametric method for inferring confidence intervals (Nourissat et al., 2010; Tang et al., 2010), we thought that the parametric approach is more realistic and universal. A previous case study of the preparation of theophylline tablets suggested that the BS resampling technique and SOM clustering are promising tools for evaluating the reliability of nonlinear optimal solutions (Onuki et al., 2008b). In this study, we applied the method to a formulation optimization study of a hydrogel for use as a dermatological patch adhesive.

One thousand BS optimal solutions were generated using the BS resampling technique. The histograms of the formulation factors and response variables were far from a normal distribution, and some showed several peaks (Fig. 3). Because the BS optimal solutions were also thought to contain global and local optimal solutions, as in the previous study, we performed SOM clustering, which showed that BS optimal solutions could be classified into three distinct clusters. The BS optimal solutions of clusters 1 and 3 were similar to the original optimal solution, whereas those of cluster 2 were very different. In particular, the degree of modification with HEMA of cluster 2 was higher than that of the original optimal solution. Polymers with a large degree of modification with HEMA possess many crosslinkable points, producing a rigid hydrogel with high crosslinking density. We have already clarified that the hydrogel with high crosslinking density is accompanied by a lower degree of swelling, higher turbidity, and higher probe tack values (Nishikawa et al., 2008; Onuki et al., 2005, 2008a). Such relationships are well represented by the feature maps (Fig. 4A).

SOM clustering is a powerful tool in terms of classification into distinct clusters. However, we are yet to find evidence to explain how the clusters are generated. One possible reason for the generation of clusters relates to the risk involved in the BS resampling, which is a random resampling technique. Because the data points resampled in a BS sample differ between samples, different BS optimal solutions will be estimated. We think some model formulations act as a crucial factor in deciding the optimal solution. That is, the BS optimal solutions change substantially according to whether the BS sample contains the crucial formulations, and the optimal solutions induce the generation of distinct clusters corresponding to one global solution and several local solutions.

To test our hypothesis, we firstly examined the relationships between the compositions of BS samples and their optimal solutions according to Bayesian estimation (Fig. 5). In cluster 3, the resampled probability of Rp.5 was obviously low, whereas, in cluster 2, the lowest resampled probability was observed from Rp.7. On the basis of this result, we thought that generation of clusters 2 and 3 was provoked by deleting Rp.7 and Rp.5 from the original sample. By contrast, because the resampled probability of formulations changed little, cluster 1 represented BS samples having a uniform population of formulations.

The LOFO approach was performed next to investigate the involvement of the data points of the model formulations in decid-

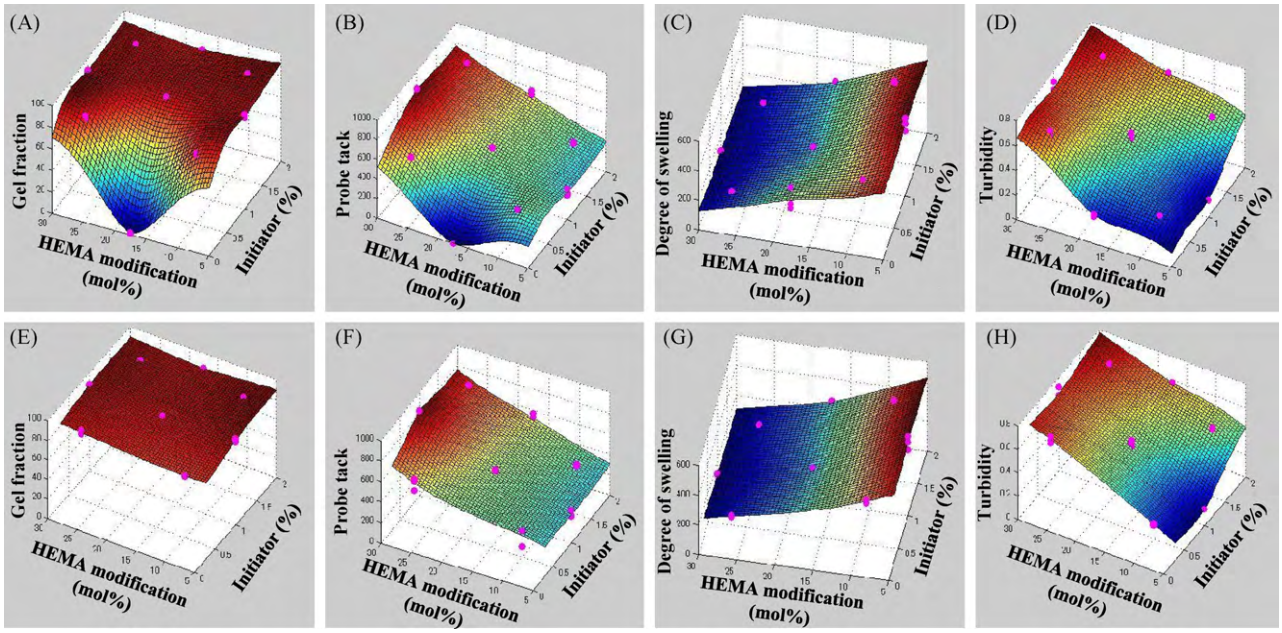

Fig. 7. The risk of missing the predictive accuracy of the response surfaces caused by removing Rp.5 from the original data. The response surfaces (A–D) were generated from the original data set, and the response surfaces (E–H) were generated from the LOFO sample after removal of Rp.5. (A) and (E) show the gel fraction; (B) and (F), the degree of swelling; (C) and (G), the probe tack; and (D) and (H), turbidity. The points represent the model formulations of the hydrogel. The original response surfaces (A–D) are taken from the previous study (Onuki et al., 2005).

ing on the optimal solution. By removing Rp.7, the LOFO optimal solution moved furthest from the original solution (Table 4), implying that Rp.7 has the most influence on the features of the optimal solution. Interestingly, the LOFO optimal solution (initiator concentration, 0.91%; modification with HEMA, 18.9 mol%) was similar to the centroid of cluster 2 shown in Table 2 (initiator concentration, 0.88%; modification with HEMA, 18.9 mol%). Taken together, cluster 2 was generated by BS samples missing formulations that exert a dominant influence on the features of the optimal formulation, such as Rp.7.

By removing Rp.5, a crucial formulation for cluster 3, the LOFO optimal solution also moved away from the original optimal solution (Table 4). However, the LOFO optimal solution (initiator concentration, 0.75%; modification with HEMA, 15.3 mol%) was not close to the centroid of cluster 3 (initiator concentration, 0.83%; modification with HEMA, 16.2 mol%). We think that Rp.5 affected the generation of cluster 3 by a mode different from that of Rp.7 in generating cluster 2. Rp.5 is a particular model formulation; i.e., no hydrogel was formed from the formulation because of its low initiator concentration (Onuki et al., 2005). Its response variables differed markedly from those of the other model formulations. For instance, except for Rp.5, every gel fraction value was very high, more than 75%. This probably relates to the generation of cluster 3; that is, if a BS sample misses being resampled in Rp.5, the shape of the response surfaces will differ considerably from the original surfaces like those shown in Fig. 7. Such substantially changed response surfaces will no doubt lead to significant reduction in the prediction accuracy of RSM-S.

To confirm this issue, we compared the similarity of the response variables estimated from the two different response surfaces: the original response surfaces and the BS response surfaces. If these response surfaces differ considerably in shape, the δ values should increase. As we anticipated, large δ values were observed from cluster 3, whereas the δ values from clusters 1 and 2 were very small (Fig. 6). The BS response surfaces of cluster 3 are thought to differ significantly in shape from those of the original surfaces; thus, cluster 3 was distinguished by its low predictive accuracy because of a substantial change in the shape of the BS response surfaces.

Table 5
95% confidence intervals of the simultaneous optimal solution estimated by RSM-S in the optimization of photocrosslinked PAA hydrogel.

	Lower	Upper
Formulation factor		
Initiator concentration (%)	0.78	0.98
Modification with HEMA (mol%)	13.4	17.7

Our study allowed us to clarify the mode of generation of clusters. Our findings lead us to conclude that clusters 2 and 3 represent local optimal solutions coexisting in BS optimal solutions and that cluster 1 is the cluster representing the global optimal solutions. Finally, we estimated 95% confidence intervals of the optimal solution using the BS optimal solutions of cluster 1 (Table 5). The original optimal solution was within the 95% confidence intervals.

6. Conclusions

Our method of integrating the BS resampling technique and SOM clustering into RSM-S is a promising tool for evaluating the reliability of nonlinear optimal solutions. Using the BS resampling technique and SOM clustering, we successfully extracted the clusters representing the global optimal solution, and we calculated the 95% confidence intervals of the optimal solutions. In this study, we also clarified the mode of generation of distinct clusters in BS optimal solutions. The findings indicate that our strategy is a valid approach for evaluating the reliability of nonlinear optimal solutions. This evaluation method should offer insights into developing a science-based rationale for, and a design space in, pharmaceutical formulation development.

Acknowledgments

This study was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan. The authors are grateful to Yamatake Corporation for providing us with dataNESIA version 3.0.

References

Arai, H., Suzuki, T., Kaseda, C., Ohyama, K., Takayama, K., 2007. Bootstrap re-sampling technique to evaluate the optimal formulation of theophylline tablets predicted by non-linear response surface method incorporating multivariate spline interpolation. Chem. Pharm. Bull. (Tokyo) 55, 586–593.

Cooper, E.G., Herskovitz, E., 1992. A Bayesian method for the induction of probabilistic networks from data. Mach. Learn..

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman and Hall/CRC, London.

Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N.J., Chung, S., Emili, A., Snyder, M., Greenblatt, J.F., Gerstein, M., 2003. A Bayesian networks approach for predicting protein–protein interactions from genomic data. Science 302, 449–453.

Khuri, A.I., Cornel, J.A., 1987. Response Surface: Design and Analysis. Marcel Dekker Inc., New York, NY.

Kikuchi, S., Takayama, K., 2009. Reliability assessment for the optimal formulations of pharmaceutical products predicted by a nonlinear response surface method. Int. J. Pharm. 374, 5–11.

Kohonen, T., 1995. Self-organizing Maps. Springer Series in Information Sciences, Berlin.

Myers, R.H., Montgomery, D.C., 1995. Response Surface Methodology: Process and Product Optimization Using Designed Experiments. Wiley Series in Probability and Statistics, New York, NY.

Nishikawa, M., Onuki, Y., Isowa, K., Takayama, K., 2008. Formulation optimization of an indomethacin-containing photocrosslinked polyacrylic acid hydrogel as an anti-inflammatory patch. AAPS PharmSciTech 9, 1038–1045.

Nourissat, A., Bairati, I., Samson, E., Fortin, A., Gelinas, M., Nabid, A., Brochet, F., Tetu, B., Meyer, F., 2010. Predictors of weight loss during radiotherapy in patients with stage I or II head and neck cancer. Cancer 116, 2275–2283.

Onuki, Y., Hoshi, M., Okabe, H., Fujikawa, M., Morishita, M., Takayama, K., 2005. Formulation optimization of photocrosslinked polyacrylic acid modified with 2-hydroxyethyl methacrylate hydrogel as an adhesive for a dermatological patch. J. Control. Release 108, 331–340.

Onuki, Y., Morishita, M., Takayama, K., 2004. Formulation optimization of water-in-oil-water multiple emulsion for intestinal insulin delivery. J. Control. Release 97, 91–99.

Onuki, Y., Nishikawa, M., Morishita, M., Takayama, K., 2008a. Development of photocrosslinked polyacrylic acid hydrogel as an adhesive for dermatological patches: involvement of formulation factors in physical properties and pharmacological effects. Int. J. Pharm. 349, 47–52.

Onuki, Y., Ohyama, K., Kaseda, C., Arai, H., Suzuki, T., Takayama, K., 2008b. Evaluation of the reliability of nonlinear optimal solutions in pharmaceuticals using a bootstrap resampling technique in combination with Kohonen's self-organizing maps. J. Pharm. Sci. 97, 331–339.

Sandwell, D.T., 1987. Biharmonic spline interpolation of GEOS-3 and Seasat altimeter data. Geophys. Res. Lett. 14, 139–142.

Takayama, K., Morva, A., Fujikawa, M., Hattori, Y., Obata, Y., Nagai, T., 2000. Formula optimization of theophylline controlled-release tablet based on artificial neural networks. J. Control. Release 68, 175–186.

Takayama, K., Obata, Y., Morishita, M., Nagai, T., 2004. Multivariate spline interpolation as a novel method to optimize pharmaceutical formulations. Pharmazie 59, 392–395.

Tang, N.S., Li, H.Q., Tang, M.L., 2010. A comparison of methods for the construction of confidence interval for relative risk in stratified matched-pair designs. Stat. Med. 29, 46–62.

Vadim, A., Artem, C., 2005. Prediction of HLA-A2 binding peptides using Bayesian network. Bioinformation 1, 58–63.

Wade, P.S., Jason, D., Jürgen, M., Ira, J.K., Mark, H.P., 2009. A decision aid for intensity-modulated radiation-therapy plan selection in prostate cancer based on a prognostic Bayesian network and a Markov model. Artif. Intell. Med. 46, 119–130.

Wahba, G., 1990. Spline Models for Observational Data. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Xue, Y., Chen, H., Jin, C., Sun, Z., Yao, X., 2006. NBA-Palm: prediction of palmitoylation site implemented in naive Bayes algorithm. BMC Bioinform. 458, 1–10.