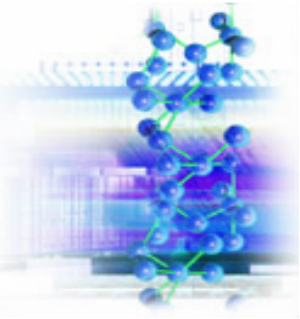
Advancing process solutions
Pharmaceutical Technology EUROPE

Graphical representation of formulation data for analysis and optimization

Feb 1, 2007
By: Phil Plumb, Ray Rowe, Peter York



Pharmaceutical formulations are examples of high-dimensional problems in which a number of physical ingredients are processed to yield a product with numerous properties. In the development of a commercial formulation, it is necessary to investigate the interactions between the ingredients, processing conditions and properties. Typically, an orthogonal experimental design, such as central composite, is used to generate a set of test formulations. The results are then subjected to statistical analysis to generate a model that may be used to identify an optimum formulation.

Traditional orthogonal designs map only the centroid and surfaces of the design space. Where this space is large and property response surfaces are highly curved, there is a significant probability that the performance of the proposed optimum does not match that predicted. It then becomes necessary to perform further experiments to map the region around the putative optimum more completely. However, it is difficult to visualize the boundaries of the optimal region in high-dimensional space and, consequently, a rational design for subsequent formulation experiments may be difficult to achieve.

In this article, an indomethacin (IMC) topical gel formulation[1] is used to demonstrate the Windows-based graphical computer programs Curvaceous Visual Explorer (CVE [Curvaceous Software Ltd, Gerrards Cross, UK]) and Curvaceous Response Surface Visualiser (CRSV [Curvaceous Software Ltd]), and their ability to investigate the interactions between formulation variables and identify the boundaries of the optimal zone.

Indomethacin topical gel formulation

The development of an IMC topical gel formulation has been reported by Takayama *et al*.[1] The ingredients were IMC (X_1), carboxyvinyl polymer (X_2), triethanolamine (X_3), ethanol (X_4) and *d*-limonene (X_5). Seven properties were

determined: plasma concentrations of IMC at 3, 6 and 24 h (Y_1–Y_3); spreadability (Y_4); IMC stability (Y_5); and two discrete variables, irritancy (L_1) and appearance (L_2). Irritancy (L_1) was assigned a score of 1 if there was no change in skin appearance or 2 if mild erythema was observed. Similarly, L_2 was assigned scores of 1 or 2 dependent upon whether or not the product formed a yellowish, clear gel.

A central composite experimental design was used to prepare a set of 27 distinct formulations. Of these, one formulation containing 60% ethanol (Y_4) failed to form a gel and is excluded from the study. The authors developed a second order polynomial model for each property; the coefficients of which are fully described in the original paper.[1]

CVE analysis

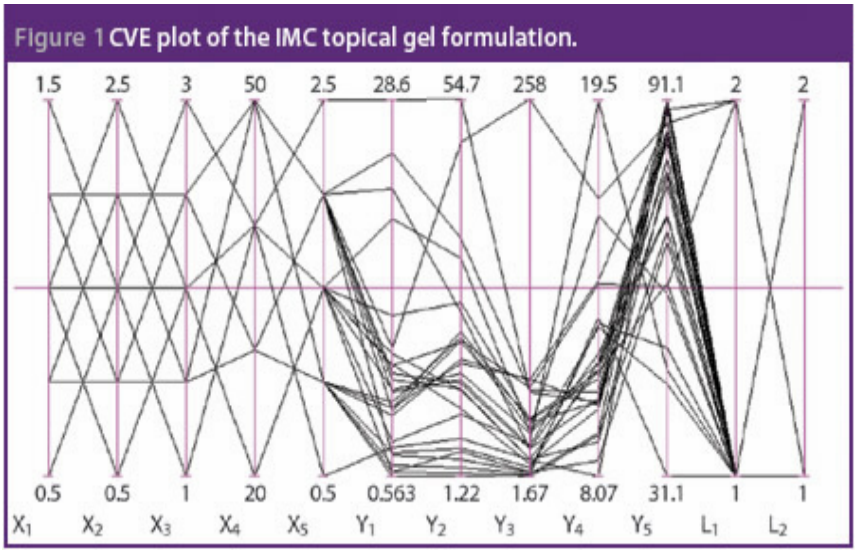


Figure 1 **CVE plot of the IMC topical gel formulation.**

The most popular means of displaying graphical data is the use of Cartesian coordinates. All axes are considered to lie mutually perpendicular and the position of a point in space is represented by its displacement along each axis. Cartesian graphical display is limited to three dimensions. In contrast, CVE uses parallel coordinate geometry (PCG) to display the dimensions as a series of parallel abscissae. In this format, a point in high-dimensional space is represented as a line cutting each axis at the appropriate value. The number of dimensions that can be displayed in this manner is essentially unlimited. The use of CVE to explore the relationships between the ingredients/processing conditions and properties of an immediate release tablet formulation has been described by Brooks *et al.* [2,3]

Figure 1 shows the topical gel formulation data set presented in PCG format. Skin irritancy is an ethically unacceptable property of a topical formulation. Those formulations exhibiting irritancy (L_1=2) are highlighted in Figure 2. It is evident from this plot that irritancy is associated with ethanol (X_4) and *d*-limonene (X_5) concentrations of 50% and 2%, respectively. Interestingly, nonirritant formulations of IMC could be prepared with *d*-limonene concentrations up to 2.5% provided that ethanol was maintained at concentrations no greater than 40%. In the original paper,[1] the polynomial model for irritancy, only the terms X_4, X_5 and X_4X_5 possessed significant coefficients. Thus, CVE correctly identified ethanol and *d*-limonene as the key contributory factors for irritancy.

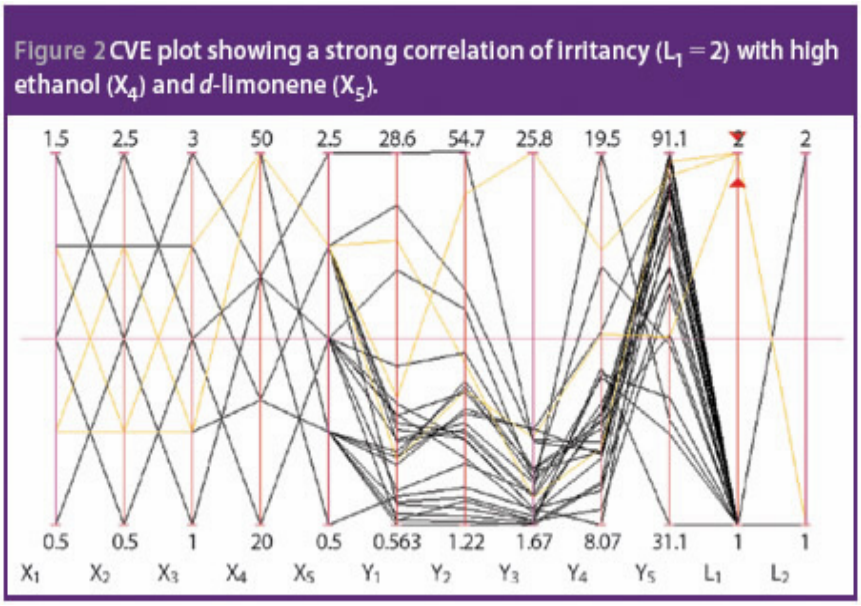


Figure 2 CVE plot showing a strong correlation of irritancy (L_1 = 2) with high ethanol (X_4) and *d*-limonene (X_5).

Figure 2

Poor appearance (L_2=2) is undesirable in a commercial formulation. Figure 3 highlights such unacceptable formulations. The plot indicates a correlation between poor appearance and low to medium IMC solubility (Y_1 to Y_3) and spreadability (Y_4). However, there were no obvious correlations with the formulation ingredients. The polynomial model contained three linear and ten quadratic terms with significant coefficients. Consequently, both CVE and the polynomial model

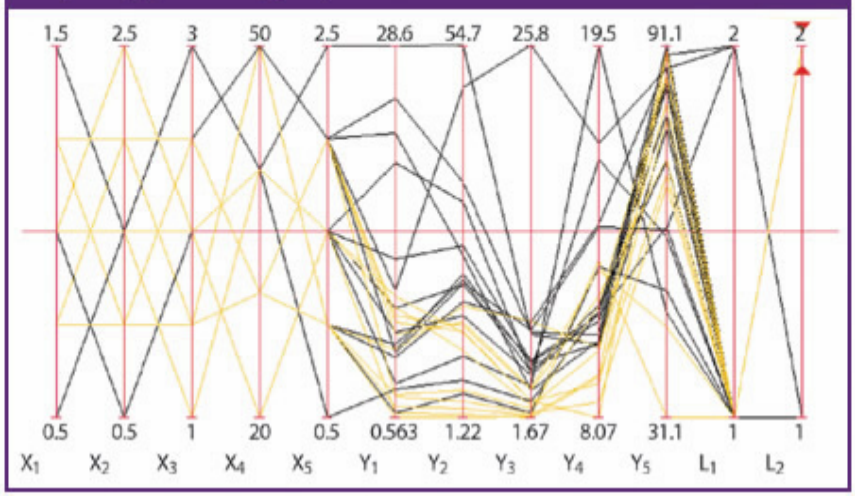


Figure 3 CVE plot showing weak correlation between low spreadability (Y_4) and poor appearance (L_2 = 2).

Figure 3

demonstrate that appearance is controlled by numerous complex interactions that are not easily interpretable from either graphical or statistical analysis.

Similarly, correlations between IMC concentration (Y_1 to Y_3), spreadability (Y_4) or stability (Y_5) with the formulation ingredients were difficult to visualize. Takayama's polynomial models for these properties contain 9–16 linear and quadratic terms with significant coefficients.

From these observations, it was concluded that CVE is capable of identifying simple correlations between ingredients and properties such as irritancy. This conclusion is in accord with the observations made by Brooks *et al.* [3] Where simple correlations exist, CVE has the potential to identify the boundaries of an optimal region and guide subsequent experimental design. However, complex relationships are more difficult to interpret, and, in such cases, visualization of the optimal region(s) remains difficult.

CRSV analysis

CVE operates on experimental data only. In contrast, CRSV uses the multidimensional envelope of all experimental points in the data set to predict the effect of setting some variables at untried values on the feasible ranges of all other variables. The data are presented in PCG format and the boundaries (envelopes) of model space determined. Figure 4 displays the points of a central composite experimental design for two ingredient variables, X_1 and X2, in both Cartesian and PCG formats. It is noticeable that the envelopes (red) do not form straight lines between $X_1=X_2=\pm\sqrt{2}$ but form notches between the X_1 and X_2 axes. This is a result of the central composite design in which experimental points lie at the centroid and surface of an *n*-dimensional sphere. Thus, a point at $\sqrt{2},\sqrt{2}$ lies outside the design space and is not displayed.

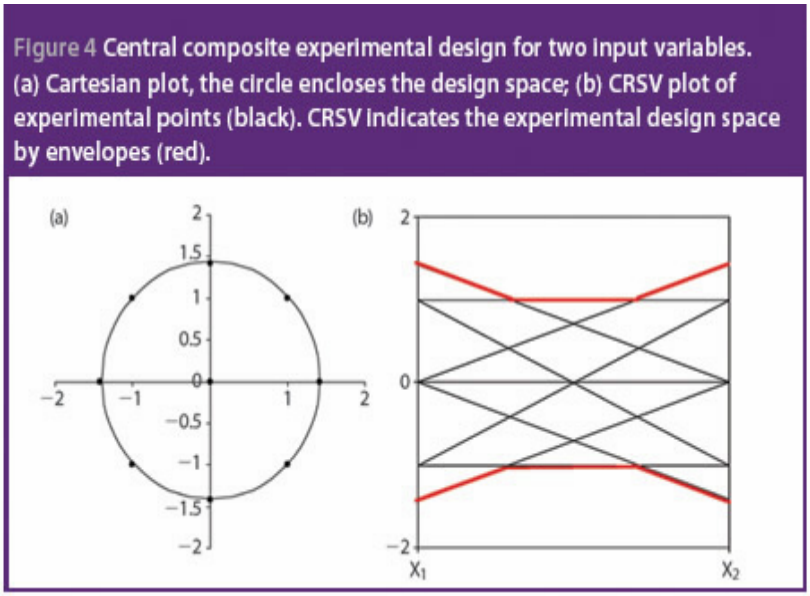


Figure 4 Central composite experimental design for two input variables. (a) Cartesian plot, the circle encloses the design space; (b) CRSV plot of experimental points (black). CRSV indicates the experimental design space by envelopes (red).

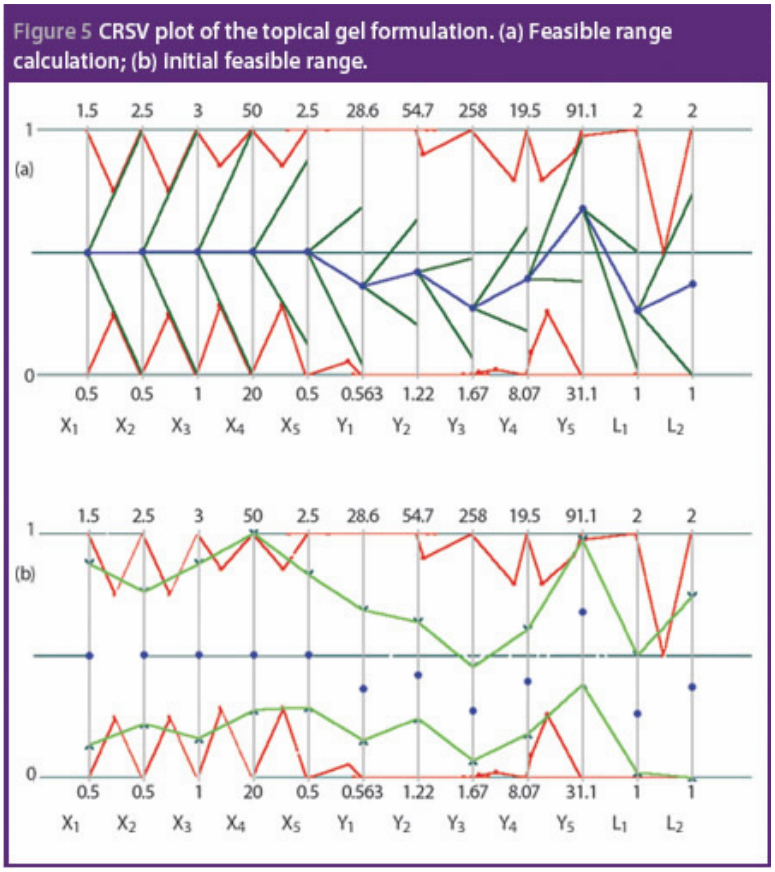


Figure 5 CRSV plot of the topical gel formulation. (a) Feasible range calculation; (b) initial feasible range.

Upon loading the experimental data, CRSV places a point at the centre of the left-most dimension and uses the data set to calculate the range of feasible values of the second dimension. The feasible range of the third dimension is calculated from the midpoint of the second dimension's feasible values. This process is continued for all dimensions as shown in Figure 5(a). The feasible value midpoints, termed "actual' values, are displayed as blue lines (Figure 5(b)) and the upper and lower feasible values are then connected to mark the feasible range (green). Current upper limit, actual and lower limit values may be tabulated below the graphical display. The position of the actual value points may be adjusted by the user to any point within the limits defined by the feasible range. Following which, new feasible ranges may be calculated.

CRSV may be used in modelling and optimization modes. In modelling mode, the actual values of the input variables may be adjusted to any combination lying within the feasible range. Following recalculation, the intersections of the upper and lower limits with the output variable dimensions indicate the range of potential property values that the formulation may take. For example, Takayama *et al.* proposed the optimum formulation for the IMC topical gel recorded in Table 1. Actual values for X_1–X_4 were set to the values proposed in the original paper. It was noted, however, that the proposed value for X_5 (*d*-limonene) of 2.5% could not be achieved by CRSV. In fact, the maximum X_5 value that could be attained was 2.3%. It is concluded that Takayama's optimum lies outside the design space and, therefore, could not be mapped by CRSV. Table 1 and Figure 6 record the feasible ranges of the formulation properties with these X_i settings; the property values were adjusted to the approximate midpoint of their respective ranges. Although CRSV did not permit the proposed optimum to be set up exactly, there was a close

correspondence between the property range predicted by CRSV and the performance observed by the authors.

CRSV may also be operated in optimization mode by adjusting the actual values of the property variables to desired target values and determining feasible ranges for the ingredient variables. The adjustment of any one formulation parameter leads to the generation of a new set of upper and lower feasible limits for all variables. As a result, adjustment of any one property may limit the feasible range of another, conflicting, property. Therefore, it is important to optimize properties in order of decreasing importance to overall product performance.

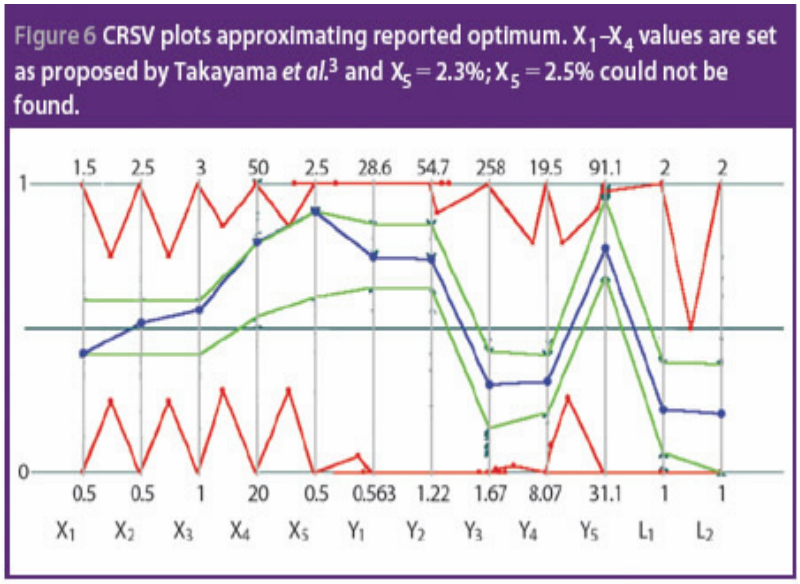


Figure 6 CRSV plots approximating reported optimum. X_1–X_4 values are set as proposed by Takayama *et al.*[3] and $X_5 = 2.3\%$; $X_5 = 2.5\%$ could not be found.

Figure 6

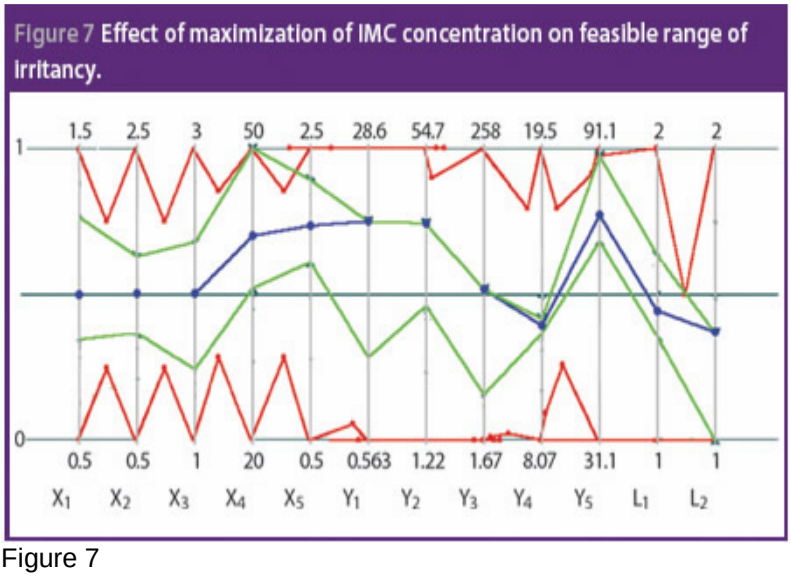


Figure 7 Effect of maximization of IMC concentration on feasible range of irritancy.

Figure 7

The optimization objectives were to maximize IMC concentration (Y_1–Y_3) and stability (Y_5) and minimize spreadability (Y_4), irritancy (L_1) and appearance (L_2). As noted from CVE analysis, maximal values of Y_3 were only obtained from irritant formulations. Therefore, maximizing Y_1–Y_3 will result in L_1 greater than 1 (Figure 7) and hence, increase the risk of irritancy.

It is ethically unacceptable to produce an irritant topical formulation. Poor appearance may be considered to be commercially unacceptable, provided that sufficiently high IMC concentrations are achievable. Therefore, it was decided to assign L_1 and L_2 as the highest priority properties. Plasma concentrations, Y_1–Y_3 were assigned second priority.

A stepwise process (Figure 8) was used in which the actual points for L_1 and L_2 were first adjusted to values of 1 while the actual values of the remaining variables were kept within their feasible ranges (Figure 8(a)). Plasma concentrations of IMC (Y_1–Y_3) were then adjusted upwards as far as possible and new feasible ranges calculated. After this adjustment, it was noted that the actual value and lower feasible limit of *d*-limonene (X_5) coincided at this point, thus limiting the feasible range of IMC concentration.

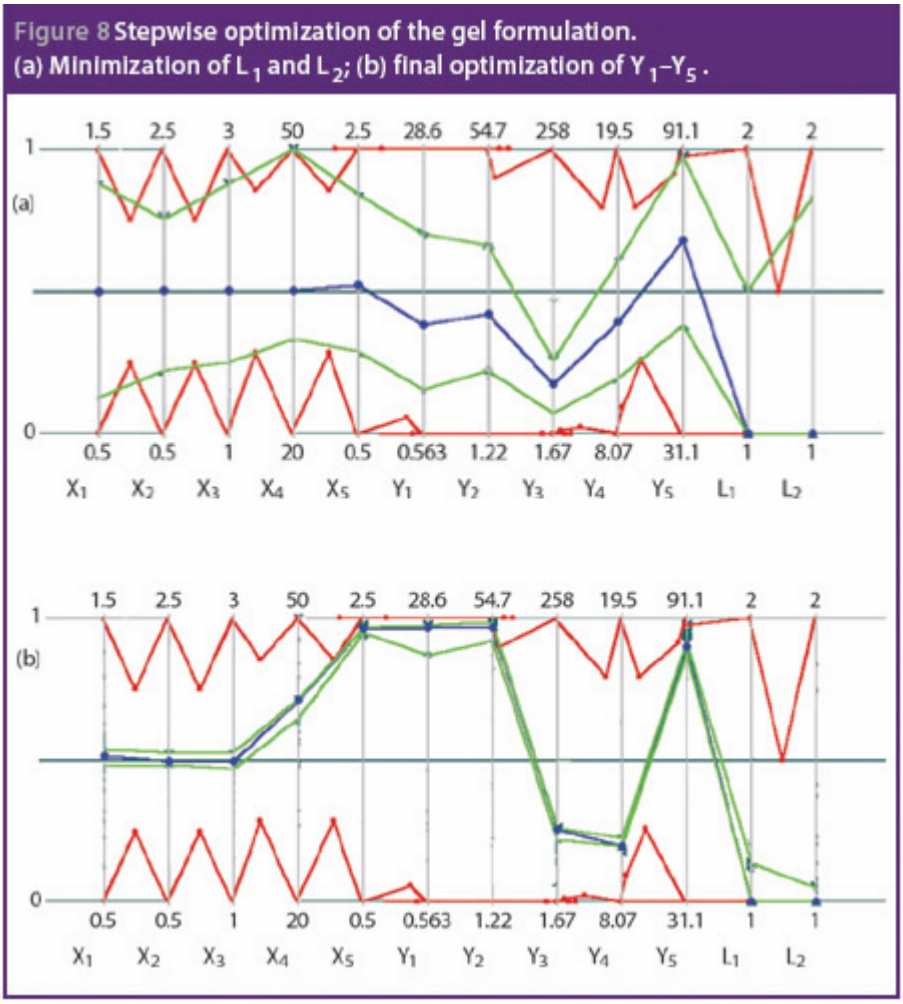


Figure 8 **Stepwise optimization of the gel formulation. (a) Minimization of L_1 and L_2; (b) final optimization of Y_1–Y_5.**

Adjusting the actual values of X_1–X_5 and Y_4 and Y_5 to the midpoint of their feasible ranges and recalculating the feasible limits resulted in an increase in the upper limits for IMC concentration, allowing further improvement of the IMC concentrations. This process was continued until no further increase in IMC concentration could be obtained (Figure 8(b)).

The potential optimum suggested by CRSV analysis is compared with the composition and observed performance of the optimum proposed by the original authors in Table 2. CRSV proposed a lower ethanol content (X_4) than was proposed by Takayama's group. Significantly, the CRSV optimum predicted an IMC concentration at 24 h (Y_3) that was approximately 25 μg mL^{-1} lower than observed for Takayama's optimum. CVE demonstrated that ethanol and *d*-limonene are critical for achieving high Y_3 values while the envelopes calculated by CRSV indicated that formulations containing high values of both ethanol and *d*-limonene lie outside model space.

Table 2 Optimum formulation proposed by CRSV.

Variable	Takayama (observed)	CRSV		
		Lower limit	Actual	Upper limit
X_1	0.914	0.972	1.015	1.036
X_2	1.54	1.47	1.49	1.54
X_3	2.13	1.93	1.99	2.05
X_4	44.0	39.4	41.4	41.4
X_5	2.50	2.39	2.43	2.43
Y_1	25.7	24.8	27.7	27.8
Y_2	49.0	50.3	53.0	53.8
Y_3	85.1	56.5	67.4	67.7
Y_4	11.8	10.3	10.3	10.6
Y_5	86.2	83.7	85.2	85.9
L_1	1	1.0	1.0	1.1
L_2	1	1.0	1.0	1.0

Thus, the comparatively lower X_4 and Y_3 values predicted by CRSV are a consequence of its inability to extrapolate beyond model boundaries. The feasible range for ingredients and properties was narrow for all variables. This observation implies that the optimal zone is very small with respect to the model space and that small variations in the composition of the formulation may have significant effects on its properties. In contrast, a broad feasible range for the ingredients, coupled with narrow ranges for the properties would imply a more robust formulation.

Conclusions

This study has demonstrated the value of PCG, as implemented in CVE and CRSV, as a means of interpreting formulation data without recourse to statistical analysis. CVE revealed that high concentrations of ethanol and *d*-limonene were critical causal factors for irritancy. In contrast, the causal factors for the remaining formulation properties were less easy to identify. The polynomial models derived by Takayama *et al.* indicated that ethanol and *d*-limonene were the sole ingredients governing irritancy, while the remaining properties were governed by nine or more significant factors. Thus, the observations made in CVE fully accorded with those made by Takayama's group.

Despite the presence of complex ingredient/property relationships, CRSV was able to identify a formulation similar in both composition and performance to that reported by the original authors. Thus, CRSV has shown its capability to identify rapidly the region of design space in which the true optimum is likely to lie. Indeed, the CRSV solution may even be suitable for use in preclinical and early clinical studies. The feasible ranges provide a rational set of boundaries for further experiments designed to map the optimal region in greater detail, and thus increase understanding of the formulation and provide greater confidence in the predicted optimum solution.

CVE and CRSV are easy-to-use graphical packages that require no statistical skills on the part of the formulator. CVE enables formulators to identify the critical relationships and interactions within a formulation. CRSV permits them to explore the effect of varying one or more components on the properties of the formulation and identify an optimal region for further experimentation. Taken together, the two programs represent valuable tools in the formulators' armoury and offer the potential to increase the understanding of a formulation, the efficiency of its development and confidence in its performance.

References

1. K. Takayama *et al., Int. J. Pharm.*, 61, 225–234 (1990).

2. J. Bourquin *et al., Eur. J. Pharm. Sci.,* **6**, 287–300 (1998).

3. R. Brooks *et al., Pharm. Technol. Eur.,* **12**(10), 40–50 (2000).

Phil Plumb is a senior scientist in the department of pharmaceutical and analytical R&D at AstraZeneca R&D Charnwood (UK).
Ray Rowe is professor of industrial pharmaceutics
Peter York is professor of physical pharmaceutics, both at the Institute of Pharmaceutical Innovation, University of Bradford (UK).