VIEWING FORMULATION DATA MULTI-DIMENSIONALLY FOR IMPROVED UNDERSTANDING AND OPTIMISATION

Brooks R W, Foo F H L, Rowe R C*

Curvaceous Software
PO Box 43
Gerrards Cross
Buckinghamshire SL9 8UX

*AstraZeneca
Macclesfield
Cheshire SK10 2NA

Introduction

A survey of pharmaceutical preparations manufactured in the UK in the period 1985-6 indicates that tablets form nearly half of all the formulation types (Fig. 1).

The properties of a tablet formulation are determined not only by the ratio in which the ingredients are combined (Fig. 2) but also by the processing conditions (Fig.3).

Although relationships between ingredient levels, processing conditions and product performance may be known anecdotally, rarely can they be precisely quantified. Furthermore, there is little knowledge of how the different variables interact and how this might affect the tablet properties. Most methods of statistical analysis are limited to discovering direct 2-variable correlations, but in reality the relationships between the inputs and outputs are a lot more complex. Models and simulations may be available but in many cases the formulation process has to be carried out in a design space that is multi-dimensional in nature and difficult to conceptualise.

Multi-dimensional geometry such as parallel co-ordinates would allow the visualisation of many variables simultaneously. More importantly it reveals the interactions between them, allowing formulators to explore these, improve their understanding of the process and hence produce better or optimum formulations. This paper describes the concept of parallel co-ordinates and how it can be applied to a tablet formulation data set.

The Theory of Parallel Co-ordinate Mathematics

The most popular way of representing graphical information is using Cartesian co-ordinates. Each axis is at right angles to the other axis or axes, and the position of a point in this space is defined by its position on each of the axes (Fig. 4). The problem with this representation is that the number of axes, or dimensions, is limited to three. By placing the axes parallel, rather than orthogonal, to one another, the number of dimensions that can be represented in a two-dimensional space becomes infinite, limited only by the width of the space (Figs. 4 and 5).

Thus a data point P whose position is defined in multi-dimensional space as P = (4, 3, -6, 2, 7, 8, -2, 1), for example, would be represented as a line connecting the eight axes at the specified points on each axis. For the present, it is sufficient to remember that points on a Cartesian graph become lines on a parallel axes plot. There are further unique properties of such plots but those are beyond the scope of this presentation.

The Curvaceous Visual Explorer (CVE) is a Windows-based software package that generates such plots and is commercially available from Curvaceous Software, PO Box 43, Gerrards Cross, Buckinghamshire SL9 8UX. This package was used to obtain the results presented in the next section.

Application to Tablet Formulation

The data used in the analysis was originally obtained by Bourquin *et al*. The data set consisted of 205 points and was based on experimental design. The variables measured in the experiment are listed in Fig. 6, and the complete data set as plotted in CVE is shown in Fig. 7.

Examination of Previous Results

Fig. 8 shows an example of a correlation that was detected by Bourquin and confirmed using the CVE analysis. The plot shows all formulation and processing variables, and the tablets' dissolution profiles, omitting the other properties for clarity. Using edge markers (the two red triangles), points having high (blue lines) and low (yellow lines) SA concentration are selected separately. Their corresponding values on the other axes are highlighted accordingly. It is evident that the blue lines fall in the higher range of

dissolution while the yellow lines fall in the lower range. This confirms the positive correlation between SA concentration and Dissolution which Bourquin also found.

A table summarising the results of Bourquin's study and the CVE analysis is shown in Fig. 9. The results from both studies compare well, although it should be noted that not very much detail was provided by Bourquin's paper and thus it is not possible to compare the subtle correlations observed using the CVE analysis. One such weak correlation occurs between NaCMC concentration and Disintegration time and is shown in Fig. 10. The blue highlighted lines are points of low NaCMC percentage and correspond to high MC percentage (due to experimental design) and generally result in tablets having low Disintegration time.

Extension of Results

In addition to reproducing the 2-variable correlations that Bourquin had observed, CVE plots also enable the viewer to discern more complex relationships between multiple variables.

Bourquin's experiments and the CVE had both established that SA concentration had a positive effect on the Dissolution profiles of the tablets, but that MgS concentration did not have any direct effect on the same property (Fig. 9). The top diagram in Fig. 11 is a reduced data set showing all the points corresponding to low MgS concentration (0.2%). These points fall under *two distinct bands* of Dissolution, visible at 45 and 60 minutes. Selecting the higher band (in yellow) shows that it corresponds to high SA concentration.

The lower diagram, on the other hand, shows all the points corresponding to high MgS concentration (1.5%). In contrast to the top diagram, these points correspond to a *single continuous band* of Dissolution at 45 and 60 minutes. This observation suggests that the effect of SA on Dissolution is more marked at low MgS concentration as shown in the top diagram. Conversely, this effect is masked at high MgS concentration. The yellow band in the bottom diagram highlights the points corresponding to high SA concentration; while it still corresponds to a higher Dissolution range, there is also much more overlap with the points resulting from lower SA concentrations.

Another multi-variable relationship is shown in Fig. 12. The points highlighted in yellow all show non-zero capping. Although this constitutes a relatively small percentage of the total data set, it is evident that these points correspond generally to lower Dwell time, and to qualities of low Tensile strength, low Disintegration time and high Friability. This result implies that the same variable (or set of variables) that affects tablet capping also affects the other tablet properties.

Using another capability of CVE, a new variable was defined by taking the product of Dwell time and Compression force. Figure 13 shows the highest values of the new variable (the first variable on the plot, labelled DWxCF) selected, which is generally a result of high Dwell times. However, the new variable shows a stronger correlation with high Tensile strength, generally low Friability and almost exclusively zero capping, more so than when Dwell time is individually compared with these properties.

CVE also provides a tool known as clustering, allowing one to group points into discrete clusters if they are within a specified range of each other based on user-specified variables. Figures 14 shows clusters based on the variables Disintegration time, Capping and Dissolution at 15 minutes, with points having a proximity of 25% and clusters a minimum size of 2. The new cluster variable is the first variable in the plot and is designated Cl1.

The top graph highlights examples of 'good' quality product, i.e. low Disintegration time, zero capping and high Dissolution. High SA concentration seems to be a common factor of these points. There is little to suggest any interaction between these properties and Tensile strength, although the corresponding points mostly fall within 1.10 and 1.81 of its range. This may suggest that Tensile strength, although not itself a measure of quality, should be kept within an optimum range rather than extremely low or high. Generally, Friability is also quite low.

The bottom diagram, on the other hand, highlights clusters of high Disintegration time, non-zero capping and low Dissolution rate, i.e. 'bad' product. The different coloured bands show that it is difficult to obtain good results simultaneously in all quality variables. For example, the blue lines show low Disintegration times and high Dissolution rates but non-zero capping. Similarly the yellow lines show low Disintegration times and zero capping but low Dissolution rates, while the green lines are points of zero capping and high Dissolution rates, but also high Disintegration times.

Conclusions

This study has demonstrated how parallel co-ordinate plots may be used to visualise multi-dimensional formulation data. The results show that it is able to reproduce the results previously obtained by traditional methods of analysis. Furthermore, because it possible to view the entire data set simultaneously without first having to 'process' the data statistically, subtle relationships between the variables are preserved. This, together with the tools provided by CVE, makes it easier for the investigator to explore the data visually and to extract new relationships implied by the data. Most importantly, it allows the viewer to discover multi-dimensional relationships which were previously undetectable.

On a wider perspective, the use of such plots would benefit the tablet formulation field by allowing formulators to discover what their most crucial formulation variables are, and thus focus their efforts on improving the control of these factors. They would then be able to make better use the quality measurements and specifications, and adjust their processes accordingly to obtain more consistent product. Discovering new relationships between formulation and processing variables and tablet properties would also allow formulators to troubleshoot their formulations and hence produce optimum formulations, and possibly enable them to improve the efficiency of their manufacturing processes. In short, the benefits of applying parallel co-ordinate analyses to tablet formulation would be more efficient formulation and manufacturing processes and better product quality, hence improving the cost-effectiveness of the entire operation.

References

1. Bourquin, J. *et al*. (1998), "Comparison of artificial neural networks (ANN) with classical modelling techniques using different experimental designs and data from a galenical study on a solid dosage form", *Eur. J. Pharm. Sci.*, **6**, pp. 287-300.
2. Brooks, R. W., "Viewing Process Information Multi-dimensionally for Improved Process Understanding, Operation and Control". Presentation at Aspenworld conference, Boston, 1997.
3. Inselberg, A. and Dimsdale, B. (1990), "Parallel Co-ordinates: A Tool for Visualizing Multi-Dimensional Geometry", *Proc. VISUALIZATION '90*, IEEE Computer Society Press, California, pp. 361-378.
4. Wells, J. (1988), *Pharmaceutical Preformulation - The Physiochemical Properties of Drug Substances*, Ellis Horwood, Chichester.

Fig. 1: Distribution of Formulation Types for Pharmaceuticals in the UK (Wells, 1988)

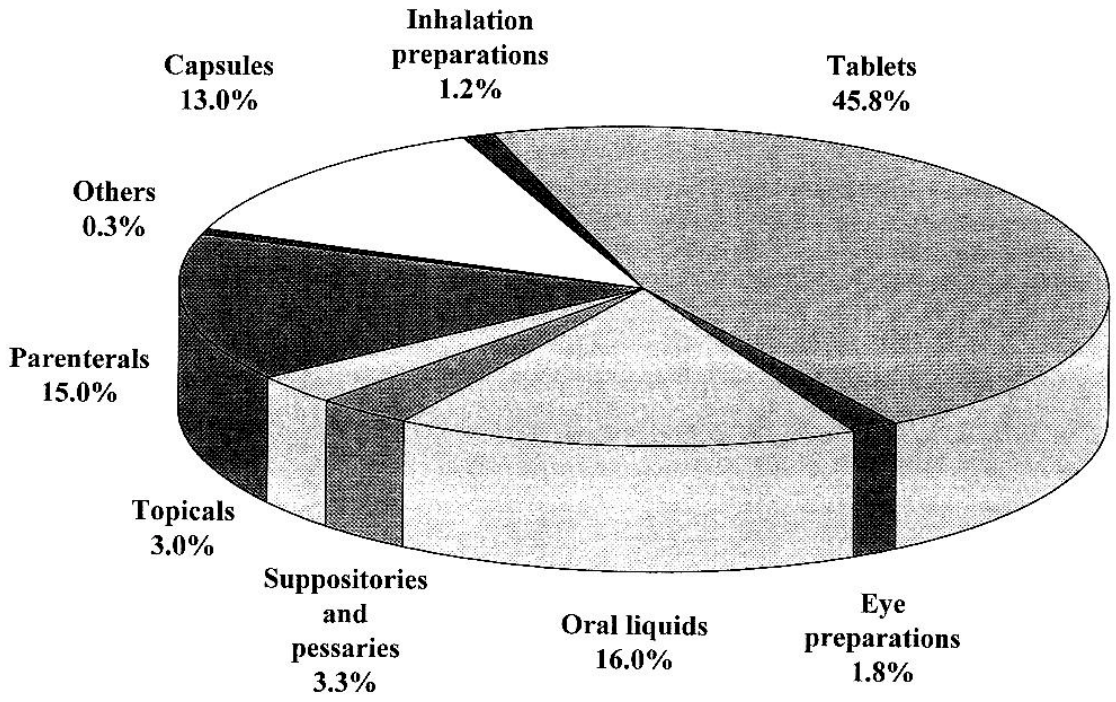


Fig. 2: Representative Tablet Formulation

Drug	}	approx. 90%
Filler/Diluents	}	
Disintegrant	approx. 5%	
Binder	approx. 1 - 4%	
Glidant	approx. 1%	
Lubricant	approx. 0.5 - 2%	

Fig. 3: Processing Conditions and Tablet Properties

Processing Conditions:

- Compression force
- Dwell time
- Rate of compression
- Rate of ejection

Tablet Properties:

- Strength
- Friability
- Disintegration time
- Dissolution profile
- Defects
- Weight control

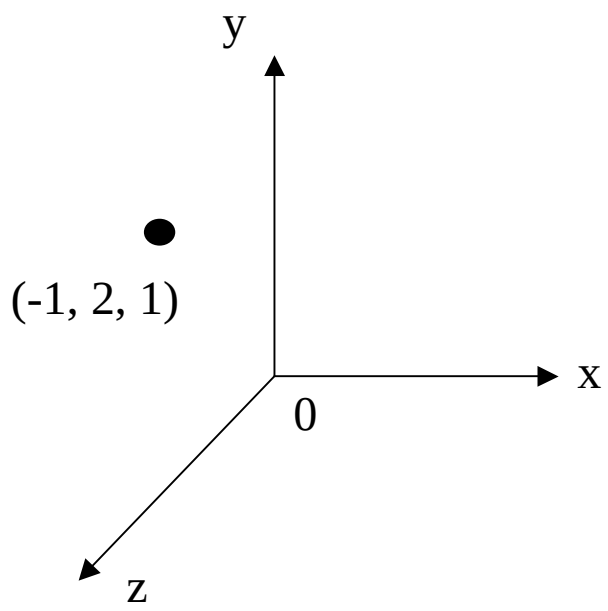
Tablet Formulation -- The Problem

- The design space is multi-dimensional in nature and difficult to conceptualise.
- Standard orthogonal (x-y-z) plots on paper limit visualisation to two or three dimensions at a time.

Using parallel co-ordinates, there is no limit to the number of dimensions that one can see at a time.

Fig. 4: Parallel Co-ordinates Theory

Traditional orthogonal plots:



Parallel co-ordinate representation:

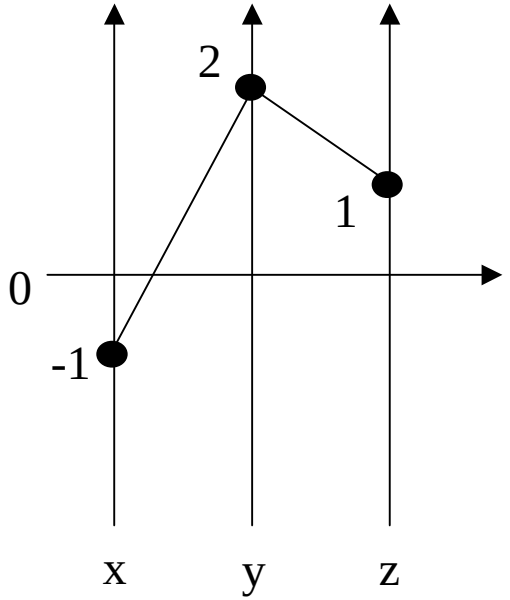


Fig. 5: Parallel Co-ordinates Theory (continued)

- This representation can thus be extended to multiple dimensions, e.g. A point P defined in multi-dimensional space as P = (x_1, x_2, x_3, x_4, x_5, …, x_n) is represented by a line dissecting the parallel axes as follows:

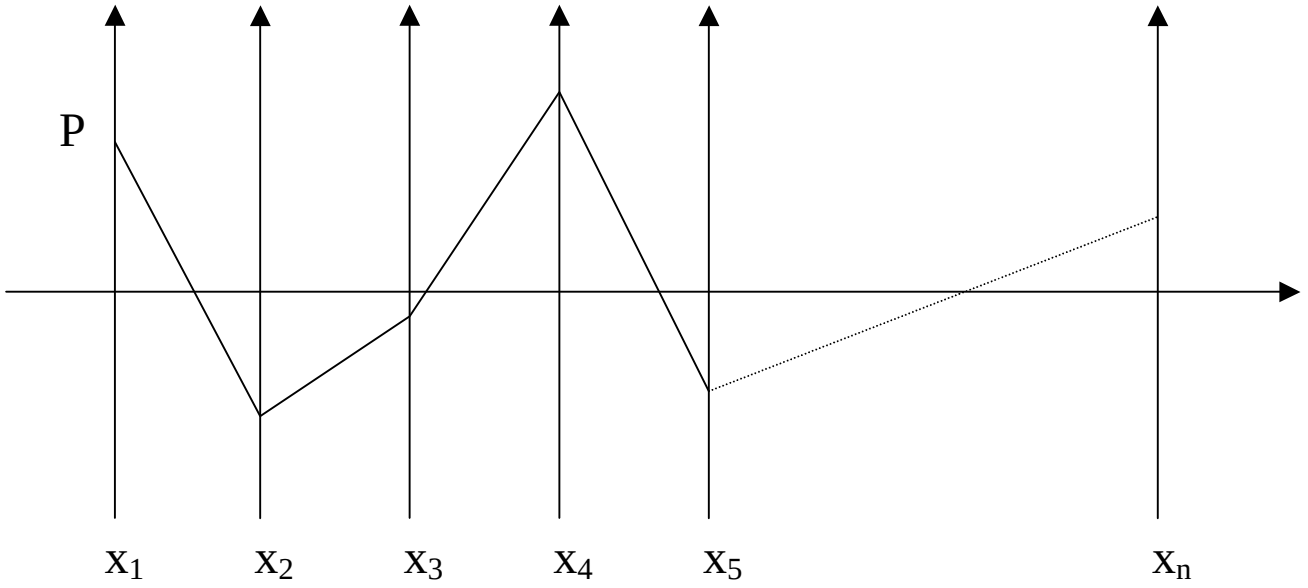


Fig. 6: The Data Set -- Bourquin et al.

- Total number of data points: 205
- Based on experimental design

Formulation Variables:

- Silica aerogel (SA)
- Microcrystalline cellulose (MC)
- Magnesium stearate (MgS)
- Sodium carboxymethylcellulose (NaCMC)

Processing Variables:

- Compression force
- Dwell time

Tablet Properties:

- Tensile strength
- Disintegration time
- Friability
- Capping
- Dissolution at 15, 30, 45 and 60 minutes.

Figure 7: The Data Set represented in Parallel Co-ordinates

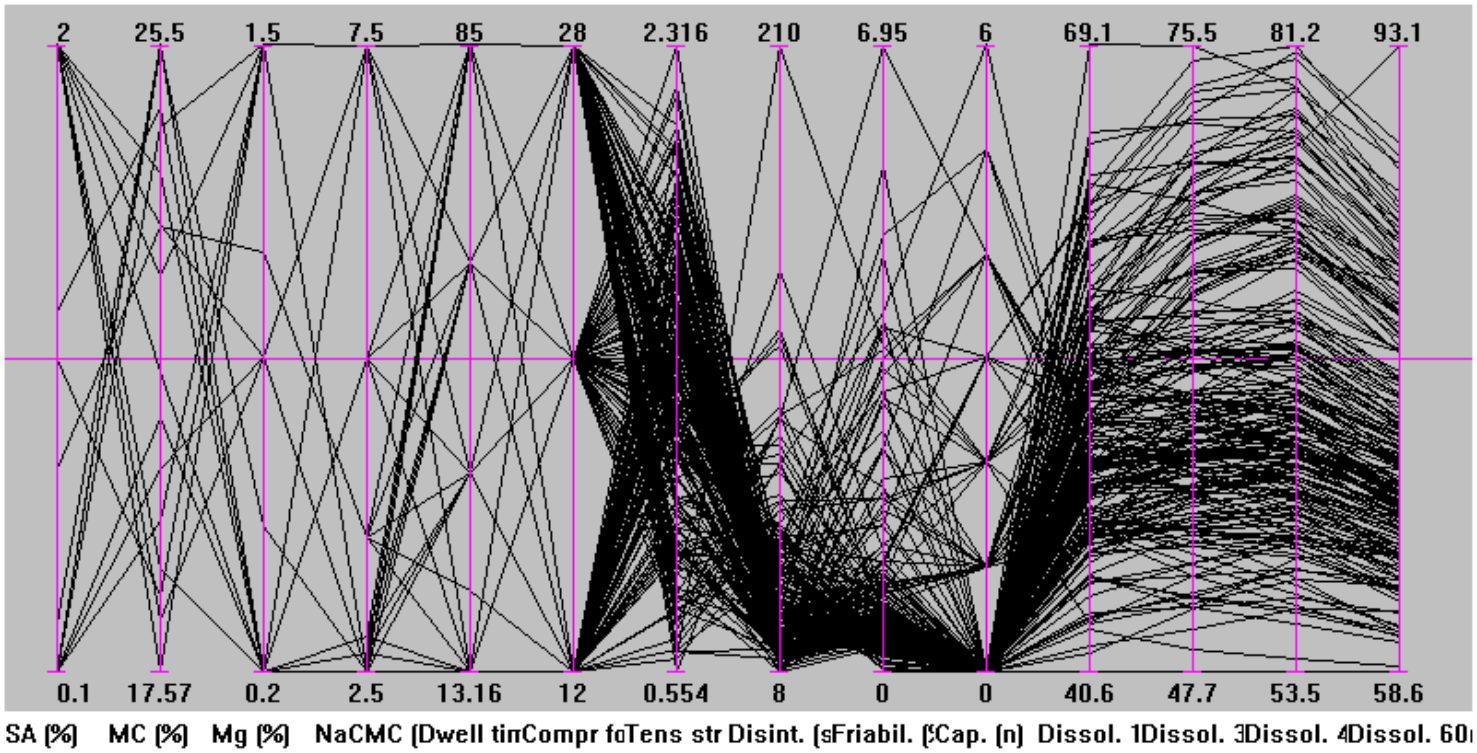
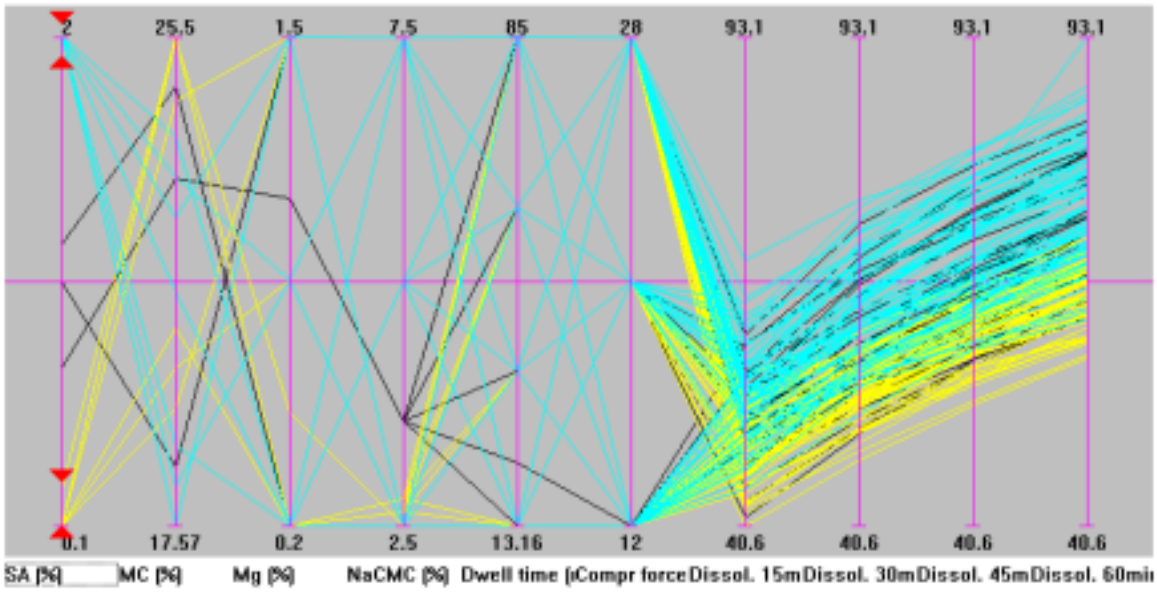


Fig. 8: Plot showing positive correlation between SA concentration and Dissolution profile

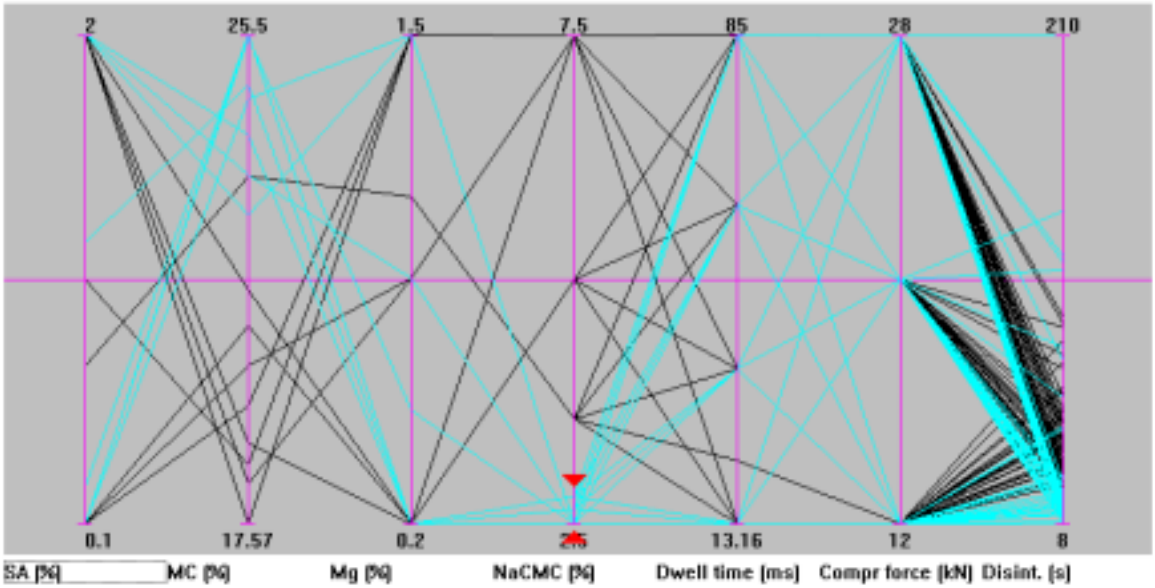


- Highlighting high SA concentration (blue lines) shows that the points correspond to high dissolution.
- Conversely, low SA concentration (yellow lines) corresponds to lower dissolution.

Fig. 9: Table comparing results obtained by Bourquin and Parallel Co-ordinates

	Tensile Strength		**Disintegration Time**		**Dissolution Rates**	
	Bourquin	*CVE*	*Bourquin*	*CVE*	*Bourquin*	*CVE*
SA	No correlation	No correlation	-	No correlation	Positive	Positive
MC	Weak positive	Weak positive	-	Weak negative	-	No correlation
MgS	Negative	Negative	Weak positive	Weak positive	-	No correlation
NaCMC	No correlation	Very weak negative	-	Weak positive	-	No correlation
Dwell Time	Positive	Positive	No correlation	No correlation	-	Very weak negative
Compression Force	No correlation	No correlation	No correlation	Very weak positive	-	No correlation

Fig. 10: Example of a weak correlation not detected by Bourquin's analysis



- Selecting points of low NaCMC concentration shows that they generally have low Disintegration time as a property.

Fig. 11: Plots showing a 3-way correlation

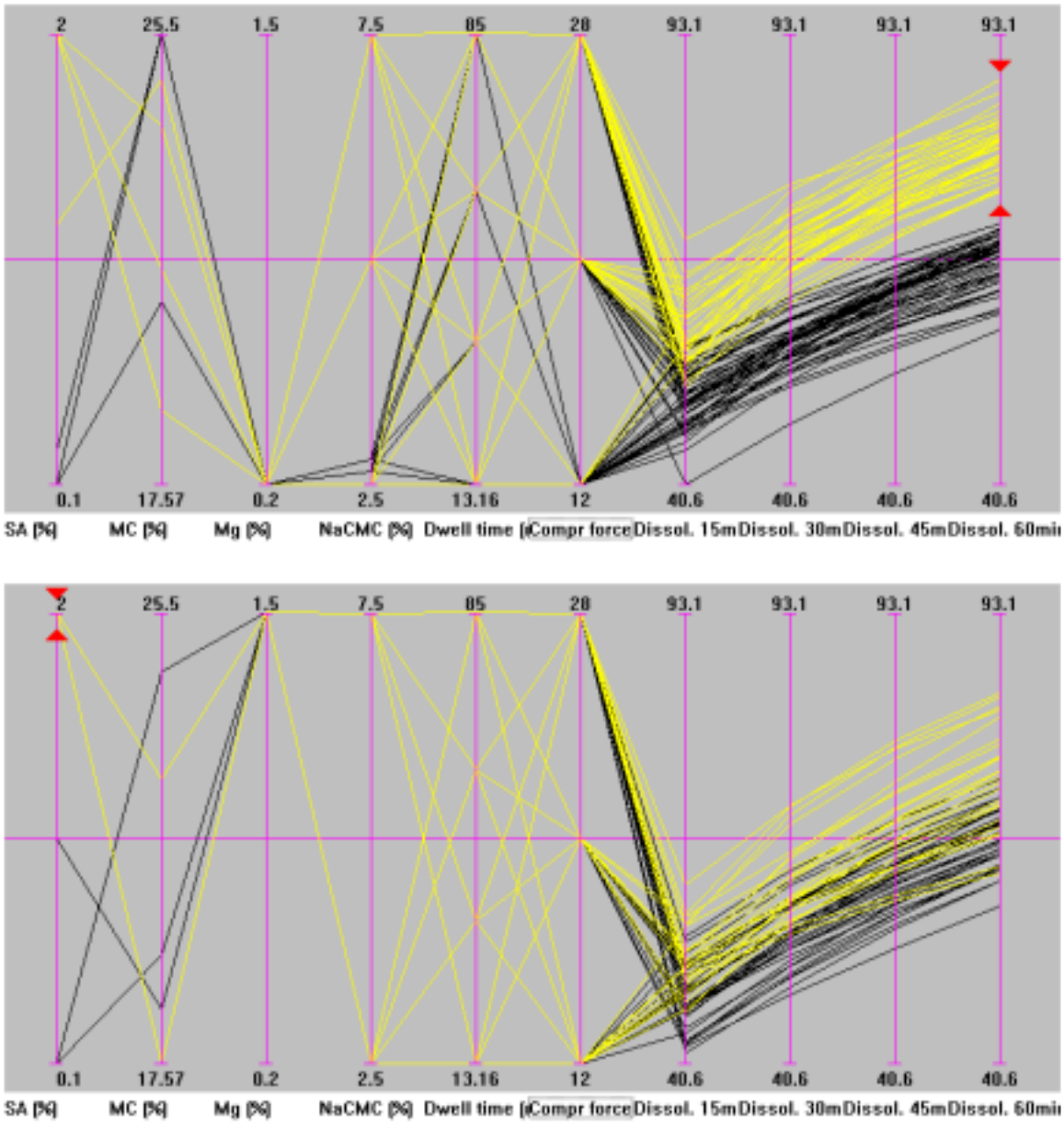


Fig. 12: Points showing non-zero capping

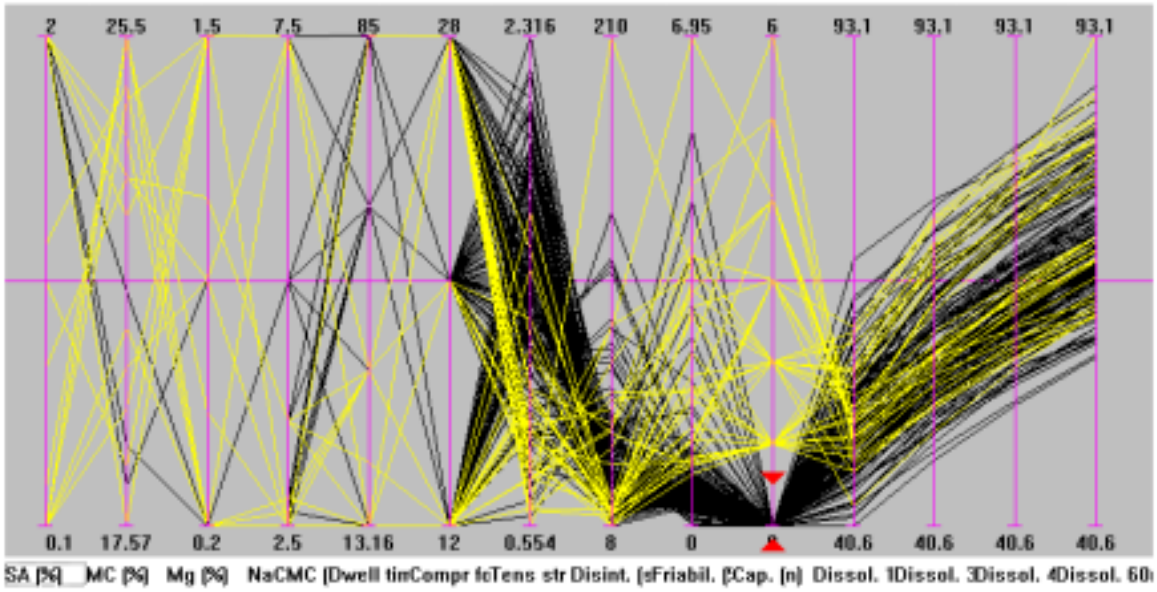


Fig. 13: Defining a new variable for analysis

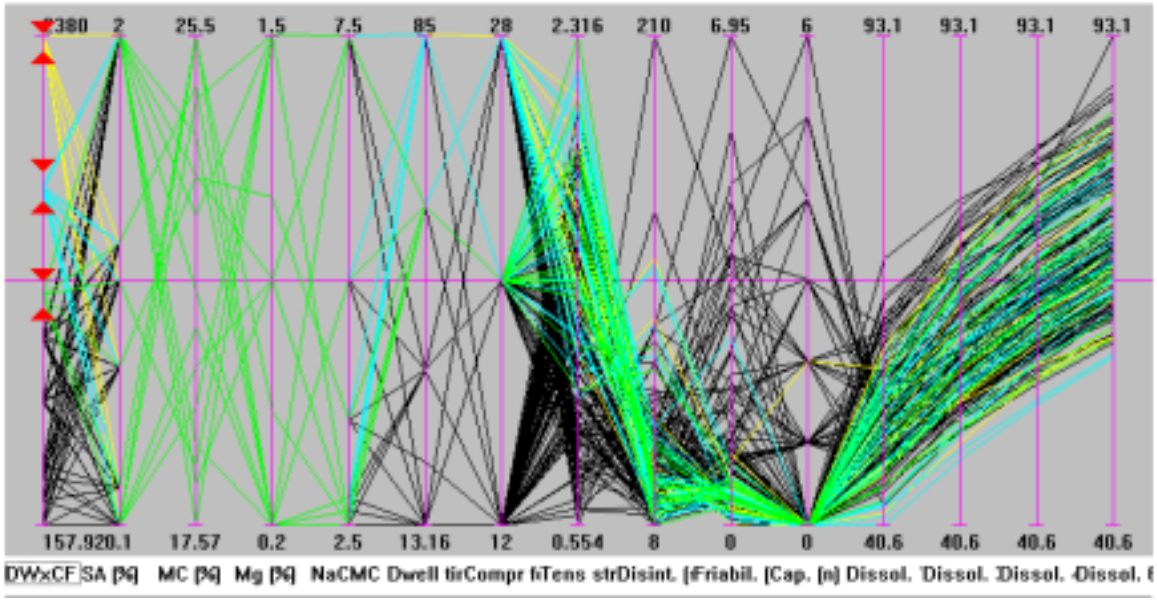
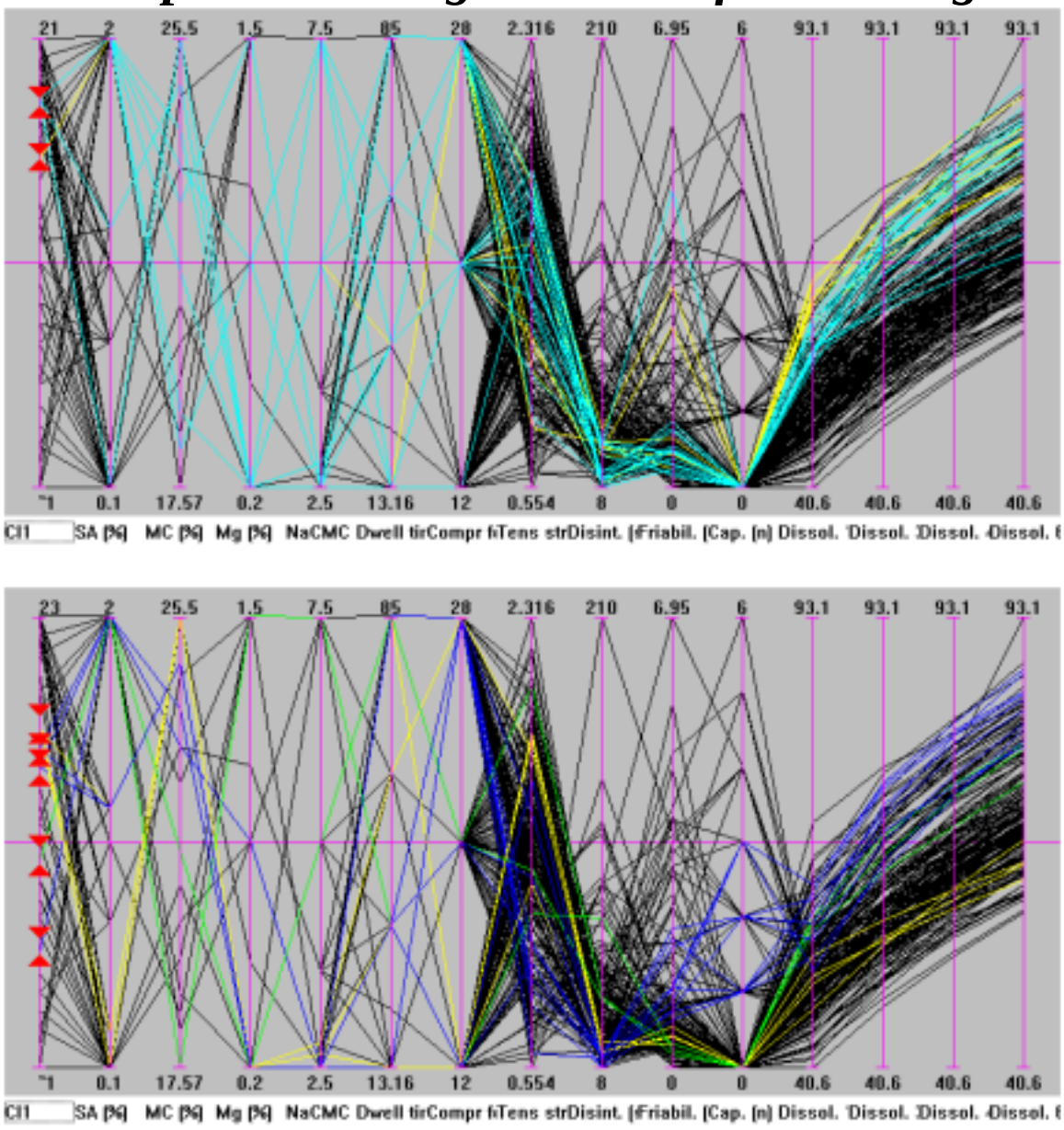


Fig. 14: Points showing 'good' and 'bad' Tablet Properties through the use of Clustering



Conclusions on using Parallel Co-ordinates to Visualise Formulation Data

- Easier to visualise multi-dimensional data.
- Easier to explore data and discover new relationships between variables.
- Able to confirm 2-variable correlations previously obtained using traditional methods.
- Able to detect correlations involving more than 2 variables, of which most statistical methods are not capable.

Potential Benefits to Tablet Formulation Processes

- Discovering the most important variables in formulation.
- Improving consistency and usefulness of quality specifications.
- Discovering contradictions between formulation or processing objectives and product quality.
- Discovering cheaper and more efficient ways of making the same product.