Visualising Design Space, Control Space and Operating Envelopes make QBD and PAT Easy

Robin Brooks, John Wilson, Na Zhao, Curvaceous Software Limited, PO Box 43,
Gerrards Cross, Buckinghamshire SL9 8BZ, United Kingdom.
www.curvaceous.com

Summary

The fundamental concepts of Quality by Design are Design Space and Control Space. They are directly related to Operating Space which is the equivalent fundamental concept of PAT.

These concepts are relatively easy to understand in two or three-variable problems because they can be visualised. However in real-world situations when there are four or more variables these have not been able to be visualised until now. Previous methods such as chemometrics, neural networks and multi-variate statistics have used advanced mathematical methods for dimensionality reduction in attempts to satisfy the human brain's craving for visual images of the problem. But these methods introduce considerable complexity and require an advanced mathematical education so that their use is restricted to a minority.

This article explains the use of the n-dimensional geometric methods of Geometric Process Control (GPC) to allow visualisation of the n-dimensional Design and Control Spaces that are the foundation of QBD and their extension into the real-time Operating Spaces that are the implicit foundation of PAT. Geometry is inherently visual and so provides solutions that are much easier to understand which in turn allows even more understanding to be gained. There are no equations or calculations for the user hence an advanced mathematical education is not required. This means that the problem owners themselves can develop and assess their own Design, Control and Operating Spaces and communicate them and their conclusions easily to others.

In QBD applications, where a new process or product is being designed, the GPC method starts with the very small set of points obtained from the subset of first-round screening experiments which delivered acceptable results. The multi-dimensional envelope (often referred to as a Response Surface) is constructed from those few points. This is the first iteration of the Design Space and it encloses many possible Control Spaces. It is non-linear so half-factorial experiment designs should not be used. Particular Control Spaces within the Design Space are found by fixing the value of one or more variables and projecting the envelope into the subspaces (hyperplanes) in which these variables are constant at their fixed values. Moving the fixed values then shows how the Control Space is affected by changes in formulation or recipe or, indeed, how to Scale-Up manufacture of product to achieve the same specifications in different sizes of equipment.

In PAT applications the Design Space is referred to as the Operating Space and is equipment specific. The multi-variable graph is utilised to choose a set of 'Best Practice' operating points. These are multi-dimensional points whose coordinates are

the values of all the variables which gave good results in the past. The data required to make these decisions is obtained from visually analysing past process operations during which process knowledge and understanding is dramatically increased in consequence of the much greater simplicity and hence speed of the visual analysis process. The multi-dimensional envelope of the chosen points can now be constructed to form a solid object. The operating objective to re-achieve in the future the same 'Best Practice' objective by which the original points were chosen is transformed into the geometric objective of remaining inside the particular Control Space defined by the current operating point. A new-format Operator Display is an inherent part of the method so that no separate project is required to design and implement a Human-Machine Interface.

BASICS of the Geometric Process Control METHOD

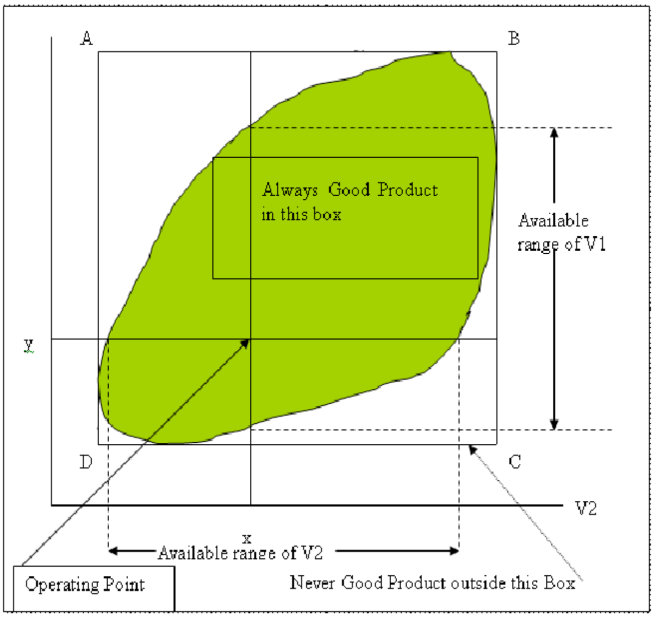


Figure 1 Design Space and Control Space in two dimensions

Geometric Process Control (GPC) is based upon the use of envelopes of future and past process operation. These are modelled as n-dimensional solids where the dimensions are the ranges of values of the independent and dependent variables that constitute the model. The envelopes are found by visually identifying a cloud of points in the multi-dimensional space where the desired objective, such as making product within specification, has been achieved in the past (Operating Space / Control Space) or is predicted to be achieved in the future (Design Space / Control Space).

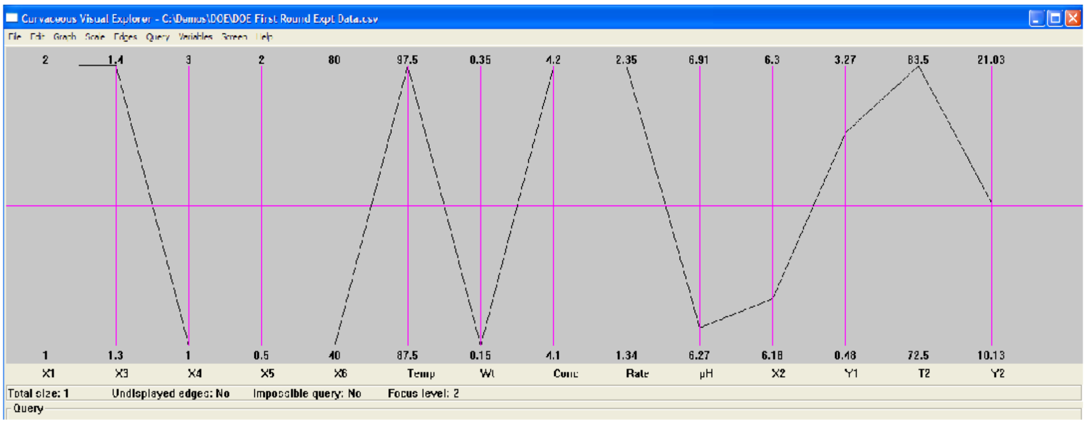
The green "sausage" region in Figure 1 is the Design Space or Operating Space. The Control Space is the available region about a current or chosen set of values of the variables. Figure 1 shows the available ranges on two variables V1 and V2 when the current operating point is (x,y). These ranges define the Control Space in 2-D.
This is not really a hypercube because if the value of one variable is changed the available ranges of the other variables change.

Primary objectives, such as making product in specification, are achieved in the Control Spaces. In just two dimensions the limits of language have been met in our inability to quantitatively describe the ‘sausage’. Traditionally this forced the simplification of the definition of Control Space by imagining it to be a much smaller rectangular area, enclosed within the sausage, such as the interior box in Figure 1, within which the assumption of independence between variables would still give always-good results. But rectangular areas (hypercubes) as shown are often too small to use in practice and the sausage could not be visualised in more than 3 dimensions, which inhibited engineers from considering it as a Design Space or Operating Space.

Geometric Process Control uses the Inselberg Parallel Coordinate Transformation which makes it possible to see a multi-variable graph containing several hundred variables (such as temperatures, pressures, flows and product qualities) and thousands of different observations in a single picture, and hence to visualise the “feasible sausage” in its full dimensionality.

Parallel Dimensions

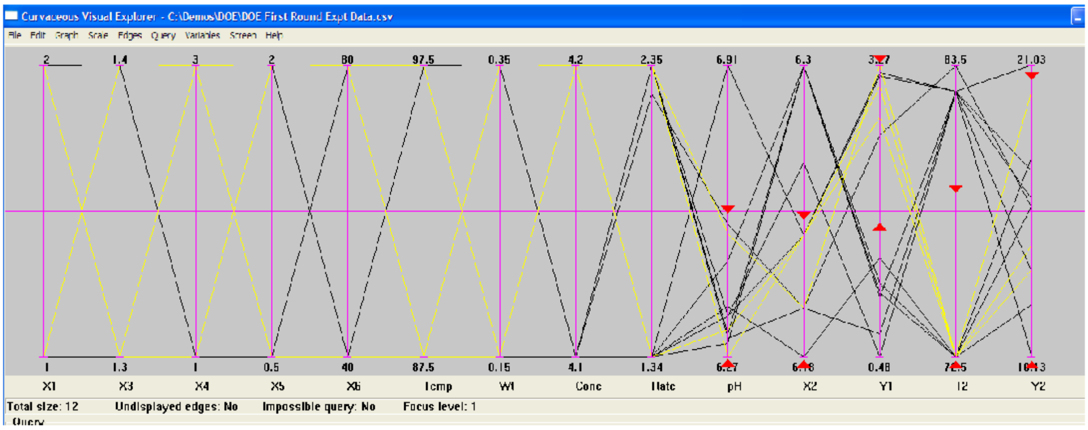
Recognising that the key problem is visualisation, GPC uses a new type of graph in which the axes are drawn vertically and parallel to each other thus allowing many axes. A related set of variable values are represented by a polygonal line connecting the values of each variable plotted on its own axis. For an example see Figure 2 where the values of all the experimental variables “X1” through “Rate” for one experiment and the Result variables “pH” through “Y2” are shown as one polygonal line or multi-dimensional point.



Total size: 1 Undisplayed edges: No Impossible query: No Focus level: 2

Figure 2 The nine experimental variables and five result variables for a single experiment

The graph is more useful when many points are shown and in Figure 3 the whole of a set of 12 first-round screening experiments for a 2-stage process are present. The quality specifications of the ideal product are shown as red triangles delineating the desired ranges of the specifications and all polygonal lines passing through all five ranges have been coloured yellow. In the terms of Figure 1 the subset of yellow points defines the “sausage” or the Design Space within which a desired objective is achievable. The whole set of points can be referred to as the Experiment Space or the Knowledge Space.



Total size: 12 Undisplayed edges: No Impossible query: No Focus level: 1

Figure 3 The whole set of 12 first-round screening experiments

Much can be deduced from Figure 3 by simple visual inspection. For instance, the desired specifications are not achieved when “Wt” and ”Temp” are both at maximum; the results for “pH” and “Y1” cluster at one end of their ranges suggesting much greater sensitivity to the experimental values in those ranges and also that the Control Space is off-centre in the Design Space so that a differently-shaped Design Space might be advantageous.

Different Design Spaces to achieve the same Control Space also occur in Scale-Up problems where the same product is to be made in two or more differently-sized plants or vessels. Some additional experiments are necessary in the new plant but by superimposing the different Design Spaces for the two plants on the same graph considerable understanding of how measurable variables are involved in Scale-Up can be obtained.

The Design Space (the sausage) is defined by the envelope of the set of yellow selected points in Figure 3. Separating these points gives the cloud of points that populate the Design Space. These are used to document and model a multi-dimensional solid object. The operating objective is transformed into the geometric objective of remaining at all times as an interior point of this object thus ensuring that the objective is achieved by which the points were selected. The beauty of this type of modelling is that it is wholly data-driven.

The underlying geometric methods are independent of the type of problem so are the same for modelling Design Space and Control Space in pre-Registration activities such as Product Design and Formulation as they are for modelling the Operating Space and Control Space in a post-registration PAT context of an operational production plant. The difference between the two is a matter of scale as typically there are very few data points available in the Product Design and pre-production environment while there may be hundreds or even thousands of data-points in a production environment such as in the example in Figure 4. The methods are the same for Batch and Continuous processes and thus apply across the whole of the process industries from food and pharmaceuticals to oil refining.

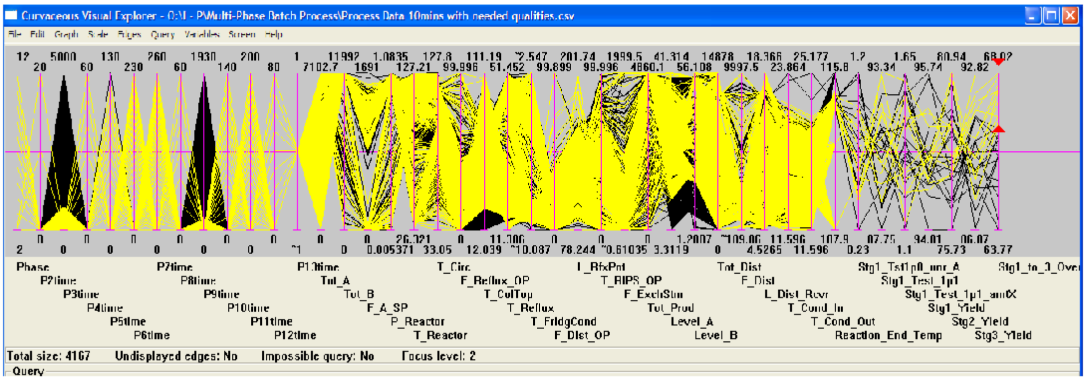


Figure 4 The first Stage of a multi-Stage Batch Production process with the yellow points showing operation in Stage 1 that led to the best results at the end of Stage 3

VISUALISING DESIGN and CONTROL SPACES for QBD

Figure 5 shows in red the envelope obtained from the cloud of points of Figure 3. The method has deduced that the experimental variables are truly independent of each other because it has constructed a box around their limiting values. The result values are not independent and there the envelope is a much more complex shape. The points themselves have been discarded. A new value has been chosen for “X1” and with the requirement to remain as an interior point of the object the envelope of all possible interior points passing through the fixed value of X1 has been calculated and shown as the series of green ranges on each variable. The blue line is the mid-point of each green range. Thus at the chosen value of X1 values of “pH” are projected to be in the bottom third of the pH range in Figure 5 but a small reduction in X1 value increases the pH range to the bottom half of scale as can be seen in Figure 6.

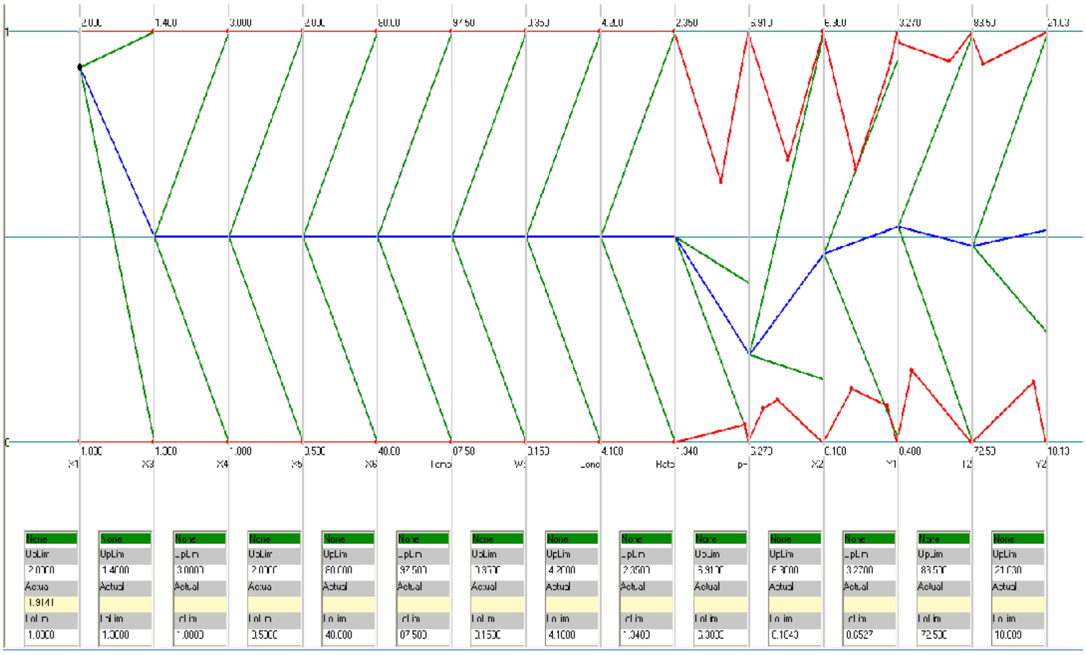


Figure 5 Visual Exploration of a Design Space

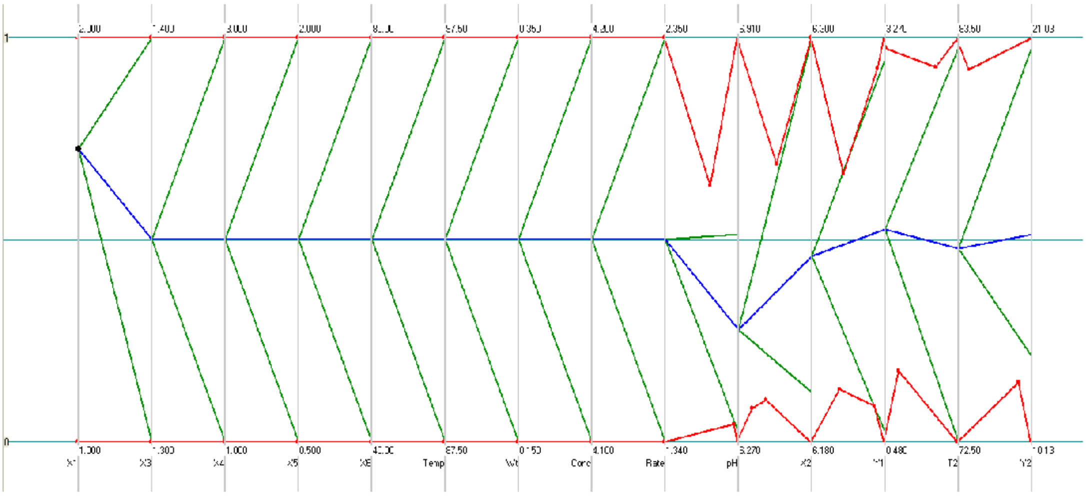
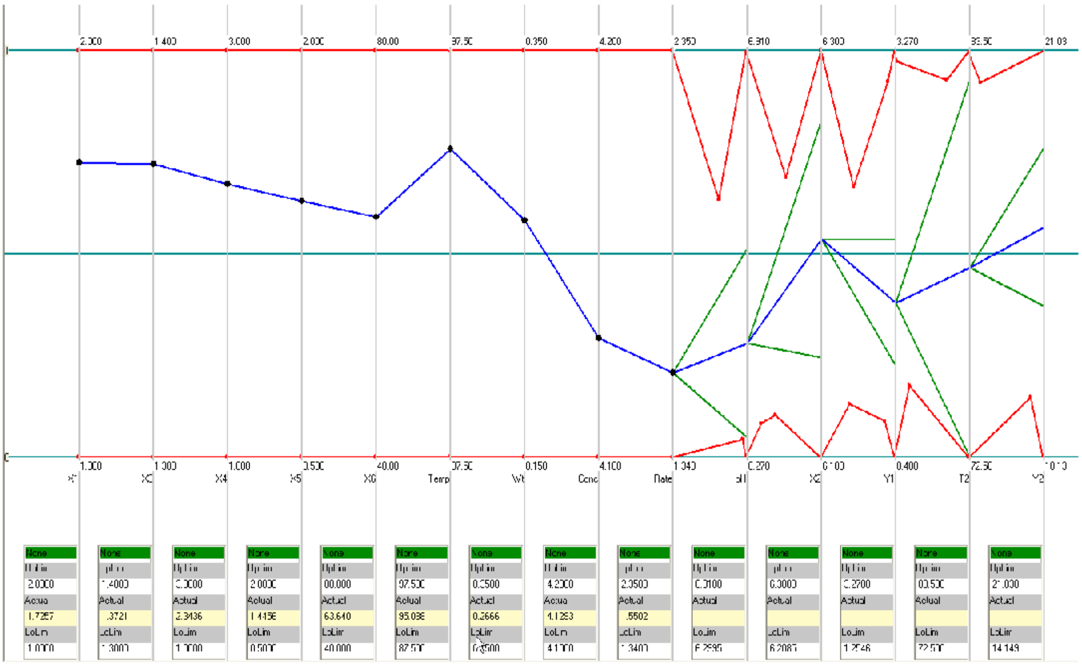


Figure 6 A slightly lower value for X1 is seen to move the range of pH upwards

In Figure 7 fixed values have been assigned to all the experimental variables and the green ranges on the result variables predict the ranges in which the results will lie when the experiment is performed. The experimenter would reach this position by moving fixed values until the predicted results are where they are required to be. The experiment would then be performed and the new point added to the dataset, so that he is adding information specifically in his area of interest, and would then reconstruct the Design Space envelope and repeat the exploration until satisfied. This application is also known as "Response Surface Visualisation".



	X1	X3	X4	X5	X6	Temp	Wt	Conc	Rate	pH	X2	Y1	T2	Y2
	None	None	None	None	None	None	None	None	None	None	None	None	None	None
UpLim	2.0000	1.4000	3.0000	2.0000	80.000	97.500	0.3500	4.2000	2.3500	6.9100	6.3000	3.2700	83.500	21.030
Actual	1.7257	.3721	2.3436	1.4156	63.640	95.038	0.2666	4.1293	.5502					
LoLim	1.0000	1.3000	1.0000	0.5000	40.000	87.500	0.1500	4.1000	1.3400	6.2595	6.2087	1.2546	72.500	14.149

Figure 7 All the experimental variables have now been assigned fixed values and the green ranges on the result variables show the ranges within which the results are predicted to lie when the experiment is performed.

VISUALISING DESIGN and CONTROL SPACES for PAT

In a PAT application it is necessary to predict or infer product qualities from the available process measurements. This can be done by finding the appropriate Operating Space as in QBD. But it is more useful for the process operator to invert the Operating Space and show the working ranges available around the current operating point which will keep predicted qualities within their specification ranges – the currently available Control Space. The alternative form of the display for plant operations use is shown in Figure 8 where the numbers along the top and bottom grey lines show the limits of the Design Space (the red Design Space outline is normally hidden in this view to minimise visual clutter). The current operating point is shown by black dots and the space available around the current operating point within which specifications can be achieved is shown by the green envelope. Variables highlighted in turquoise are predicted quality result ranges.

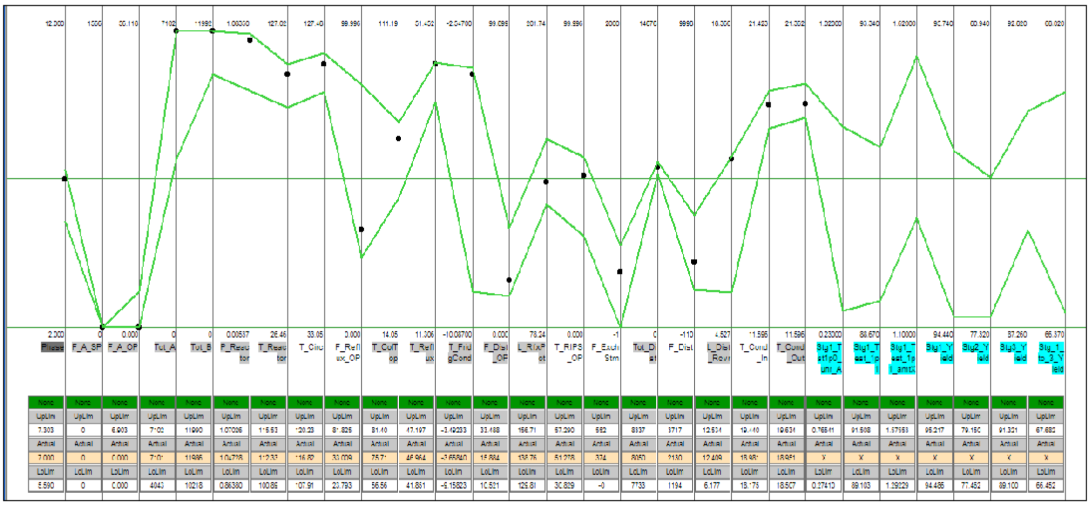


Figure 8 The Operators view of the currently usable part of the Control Space in green within the fixed limits of the Design Space shown by the values along the upper and lower grey lines. Variables highlighted in turquoise are quality results being predicted.

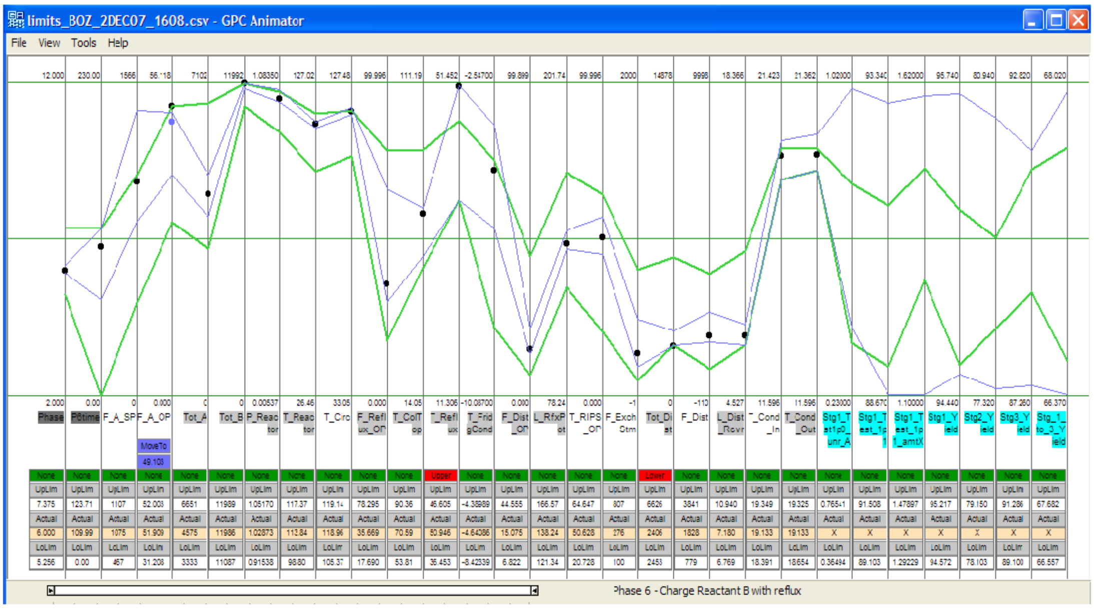
Obviously as the process values alter from one time-step to the next so the green ranges change and the shape of the green envelope alters. Production must stay inside this green envelope at all times and this is the defined role of process control. Operators rapidly appreciate that the green envelope shows them the usable range of each process variable relative to their operating objective and to the current operating point of the process. This is information that they have never had before but which they immediately understand as it complies with and quantifies their experience-based qualitative mental ‘model’ of the process.

Operating Advice for the Operator without a Rules Base

Alarms are shown to the operator on the display. Further use is then made of geometry to generate the smallest process movement advice to the operator that will clear the alarm. Each process variable is identified as Manipulable or non-Manipulable (A flow is Manipulable whereas a temperature is not in a simple manual control system. In an automated control system Setpoints, Targets and even Limits or Constraints of the control scheme may be considered Manipulable). The Advisory

Algorithm attempts to find a set of moves on the Manipulable variables only that will change the shape of the green envelope to clear the alarm condition.

Figure 9 shows an example of operating advice. The process alarms on non-Manipulable variables can be cleared by making moves on one Manipulable variable from the present black dot to the new blue dot. Implementing this advice would cause a change from the present green Control Space to the new blue Control Space which changes process operation sufficiently to clear the two alarms on the non-Manipulable variables while still predicting a quality value within specification. It is important to note that this advice is generated by a same-for-everyone algorithm. An Artificial Intelligence (AI) Rule Base is not required thus reducing the cost of implementation by one or two orders of magnitude.



	Phase	P6time	F_A_SP	F_A_OP	Tot_A	Tot_B	P_Reactor	T_Reactor	T_Circ	F_Reflux_OP	T_ColTop	T_Reflux	T_FridgCond	F_Dist_OP	L_RfxPot
	12.000	230.00	1566	56.118	7102	11992	1.08350	127.02	127.48	99.996	111.19	51.452	-2.54700	99.899	201.74
	2.000	0.00	0	0.000	0	0	0.00537	26.46	33.05	0.000	14.05	11.306	-10.08700	0.000	78.24
				MoveTo 49.103											
	None	None	None	None	None	None	None	None	None	None	None	Upper	None	None	None
UpLim	7.375	123.71	1107	52.003	6651	11989	1.05170	117.37	119.14	78.295	90.36	46.605	-4.38989	44.555	166.57
Actual	6.000	109.99	1075	51.909	4575	11986	1.02873	113.84	118.96	35.669	70.59	50.946	-4.64086	15.075	138.24
LoLim	5.256	0.00	467	31.208	3333	11087	0.91538	98.80	105.37	17.690	53.81	36.453	-8.42339	6.822	121.34

	T_RIPS_OP	F_ExchStm	Tot_Dist	F_Dist	L_Dist_Rcvr	T_Cond_In	T_Cond_Out	Stg1_Test1p0_unr_A	Stg1_Test_1p1	Stg1_Test_1p1_amtX	Stg1_Yield	Stg2_Yield	Stg3_Yield	Stg_1_to_3_Yield
	99.996	2000	14878	9998	18.366	21.423	21.362	1.02000	93.340	1.62000	95.740	80.940	92.820	68.020
	0.000	-1	0	-110	4.527	11.596	11.596	0.23000	88.670	1.10000	94.440	77.320	87.260	66.370
	None	None	Lower	None	None	None	None	None	None	None	None	None	None	None
UpLim	64.647	807	6626	3841	10.940	19.349	19.325	0.76541	91.508	1.47897	95.217	79.150	91.286	67.682
Actual	50.628	276	2406	1828	7.180	19.133	19.133	X	X	X	X	X	X	X
LoLim	20.728	100	2453	779	6.769	18.391	18654	0.36494	89.103	1.29229	94.572	78.103	89.100	66.557

Figure 9 Operating advice to correct alarms or optimise the process is highlighted in blue. Moving the indicated variables to their blue-dot values will cause the green envelope to change in shape to the blue one. Advice can be used for open-loop or closed-loop control.

It is also comforting to recognise that the basis of geometry and an interior point algorithm means that this is the only intrinsically safe control algorithm in that it can only ever generate inwards movement advice into the Design Envelope.

The same algorithm is used when no alarms are present to find moves that will maximise and minimise variables previously identified for optimisation of secondary objectives thus delivering RealTime Optimisation (RTO) without the expense of developing a first-principles model.

Bibliography

1. Plumb P., Rowe R. and York P. Graphical Representation of Formulation Data for Analysis and Optimisation. Pharmaceutical Technology Europe, February 2007
2. Brooks R., Thorpe R., and Wilson J. Geometry Unifies Process Control, Production Control and Alarm Management.. IEE Computing and Control Engineering. March 2004
3. Inselberg, A. and Dimsdale, B. Parallel Coordinates – A Tool for Visualising Multivariate Relations. Human-Machine Interactive Systems edited by A. Klinger. Plenum Publishing Corporation 1991.