Monitoring chemical processes for early fault detection using multivariate methods

by Dr Frank Westad, Chief Scientific Officer, CAMO Software

Multivariate statistical methods can be used monitor process variables and predict final product quality at an early stage, while also providing deeper understanding of the process. This allows engineers and production managers to optimize their processes, thereby realizing significant cost and time savings. This white paper includes a background and explanation of the method, as well as a practical example of its application within a real company.

INTRODUCTION

Multivariate Statistical Process Monitoring (MSPM) has been established as a valuable tool for ensuring reliable product quality in the process industry. However,many organizations today are still not fully utilizing its potential to make significant improvements in their production environment. The MSPM approach to process monitoring involves the use of multivariate models to simultaneously capture the information from as few as two process variables, up to thousands. The methodology provides means for increased process understanding, fault detection and on-line prediction, all typical tasks for the process engineer and production manager.

With MSPM approaches, it is possible to not only control the final product quality data, but also all of the available process variable data in terms of the underlying systematic variations in the process. The variables measured in a process are often correlated to a certain extent, e.g. when several tempertures are measured in a distillation column. This means that the events or changes in a process can be visualized in a smaller subspace that may give a direct chemical or physical interpretation. If we want to keep such a process "in control", traditional univariate control charts – due to the covariance or interaction between variables - may not assure this efficiently. Figure 1 exemplifies a situation where two process variables are both inside their univariate control limits given as two standard deviation, but fails to detect that the general trend of correlation between these two variables is broken for the sample shown in red.

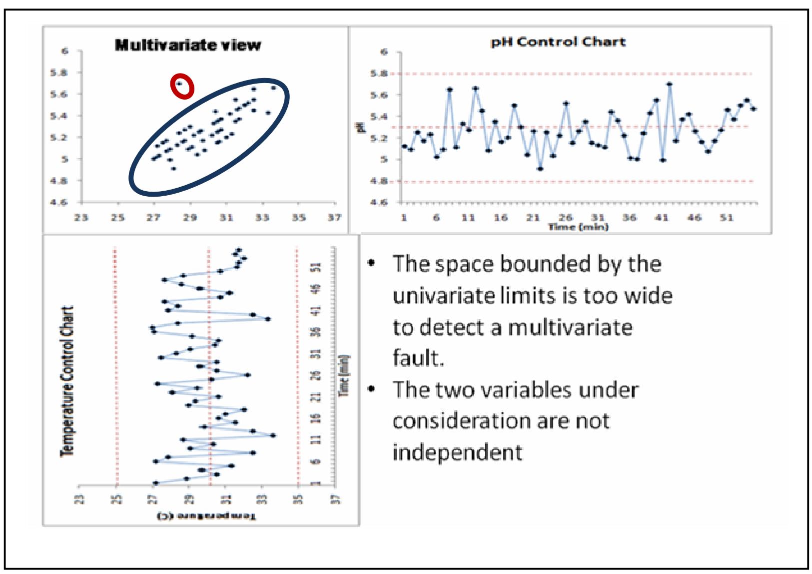


Figure 1 Illustration why multivariate methods are needed for process monitoring

The most frequently applied methods are Principal Component Analysis (PCA) and Partial Least Squares Regression (PLSR). The PCA answers the question "Is the process under control?" but does not provide a quantitiave model for the final product quality. Typical applications of PCA for this purpose are raw materal identification , adulteration in the food industry, counterfiet products, and on-line testing of product quality. The PLS regression provides in addition to the monitoring aspect, also quantitative prediction of the final product quality based on all or a subset of the process variables. One vital aspect in this context is to reduce the off-line laboratory work, both to have the prediction at an early stage as the product properties are not available on-line, and to reduce the labour-intensive work.

Critical statistical limits can be derived from the empirical data chosen to establish a model when the process is under control. One limit is based on the space defined by the model, the so-called Hotelling T^2 statistic. Thus, this statistic indicates if there is too high or too low concentration of the quality variable of interest. The other limit is based on the distance to the model, meaning there is something new e.g. there is a change in the raw material. We refer to Jackson (ref. 1) for a detailed description.

Multivariate statistical methods are also excellent tools to develop processes further. With these methods we can look *inside* the process to gain the necessary information for optimising them (PLSR).

BACKGROUND AND DATA FROM REAL-LIFE EXAMPLE

A paper mill monitors the quality of newsprint by applying ink to one side of the paper. By measuring the reflectance of light on the reverse side of the paper, a reliable, practical measure of how visible the ink is on the opposite side is obtained. This property, "Print through", is an important quality parameter. The paper is also analyzed with regard to several other product variables and raw material parameters.

The samples were collected from the production line over a considerable time interval, in the hope that the measurements would span the important variations in production. The data consists of 66 samples with 15 process and product attribute variables and the response variable, Print through. 16 of the samples were in this context test samples for prediction using the model based on the calibration data of 50 samples.

The process variables are given in Table 1.

Table 1: Variables for the paper production

X-var	Name	Description
X1	Weight	Weight / sq. m
X2	Ink	Amount of Ink
X3	Brightness	Brightness of the paper
X4	Scatter	Light scattering coefficient
X5	Opacity	Opacity of the paper
X6	Roughness	Surface roughness of the paper
X7	Permeability	Permeability of the paper
X8	Density	Density of the paper
X9	PPS	Parker Print Surf number
X10	Oil absorb	A measure of the paper's ability to absorb oil
X11	Ground wood	The % of ground wood pulp in the paper
X12	Thermo pulp	The % of thermomechanical pulp
X13	Waste paper	The % of recycled paper
X14	Magenf	The amount of additive
X15	Filler	The % of filler

The purpose was to establish a model that can be used for quality control and production management. The objectives were:

1. Predict quality from the process variables and other product variables.
2. Rationalize the quality control process by reducing the number of variables measured, i.e. build a model that includes as few variables as possible.

Using The Unscrambler X multivariate analysis software, a PLS regression model was run with 50 calibration samples and the 15 process and product variables with Print through as the response variable. As mentioned above one important aspect of multivariate modelling is that the dimensionality of the process typically is lower than the number of process variables measured, i.e. there is a redundancy among the observed variables. This is exemplified in the score plot in Figure 2 which summarises the model in the two underlying dimensions or "factors", or "latent variables" for the 15 original process variables. So instead of plots of the individual variables in one, two or three dimensions, the process can be visualized as a map of the samples in the latent variable space, the score plot.

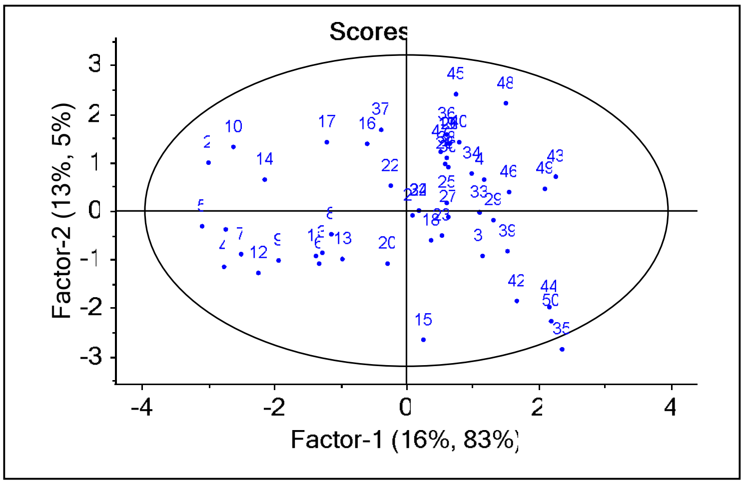


Figure 2. A score plot of the samples for factors 1 and 2.

Alternatively one may visualise the change in the process over time as a one-dimensional score plot if such a clear trend exists, Figure 3.

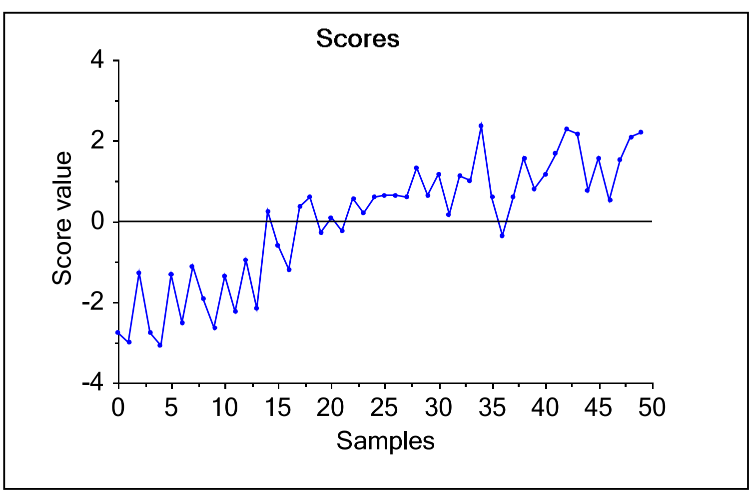


Figure 3 Line plot of score for factor 1 showing the change in the process over time

The overall importance of the process variables are most easily depicted in terms of the model coefficients as in Figure 4. While validating the model robustness a 95% confidence interval is estimated for each variable , thus indicating which are the important ones. This has an economical incentive in that if many variables describe the model in the same way, it is not necessary to measure all of them. That said, one may still monitor these variables but choose not to use them for prediction if the parsimonious (most simplistic version possible) model is better for that purpose.

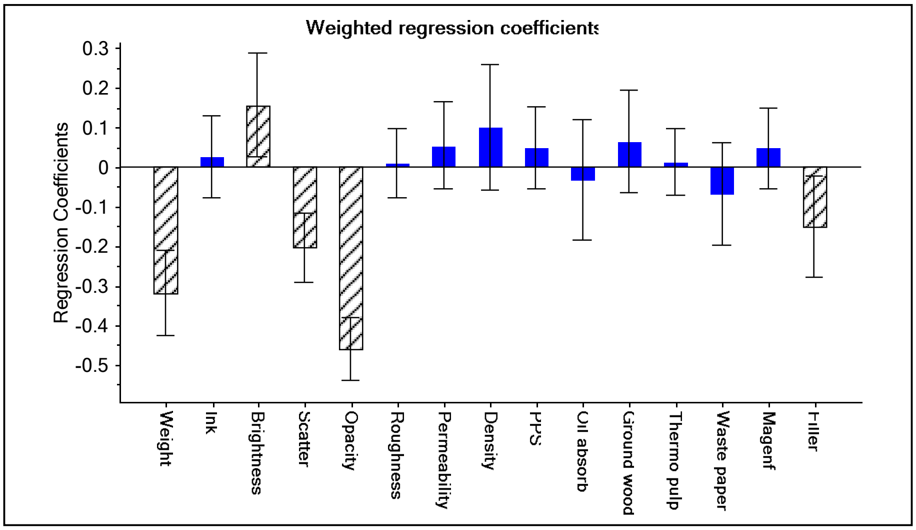


Figure 4. The model summarised in terms of the regression coefficients with significant variables marked.

From the results of the first model a reduced model with only five variables was chosen for on-line prediction. Interactive plots during the prediction stage give insight in any changes in the process. Upper and lower control limits for the print through are shown in real time as depicted in Figure 5 as is the Hotelling T^2 statistic in Figure 6.

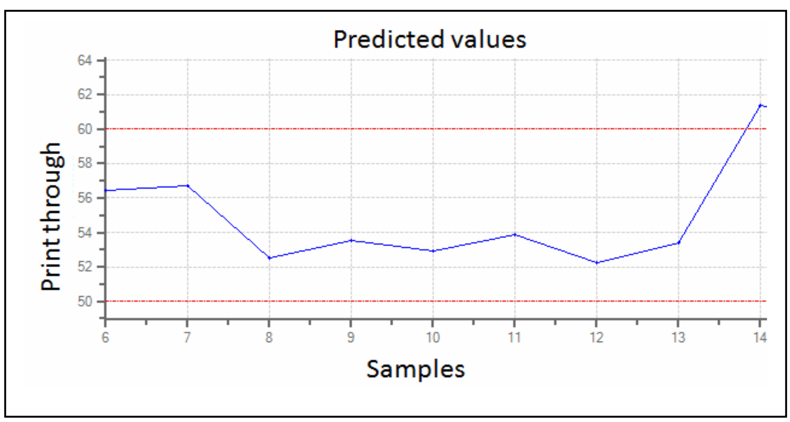


Figure 5. Predicted values with control limits.

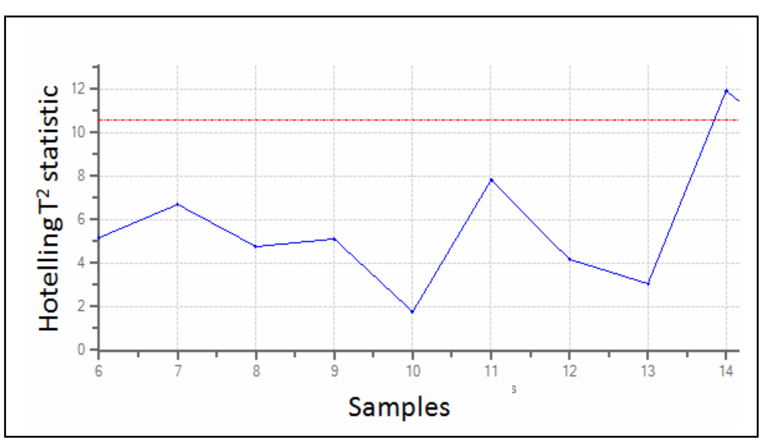


Figure 6. The Hotelling T^2 statistic with critical limits

When a new sample falls outside the critical limit a click in the plot produces Figure 7 thereby immediately showing which variable(s) are outside the limits as defined by the calibration.

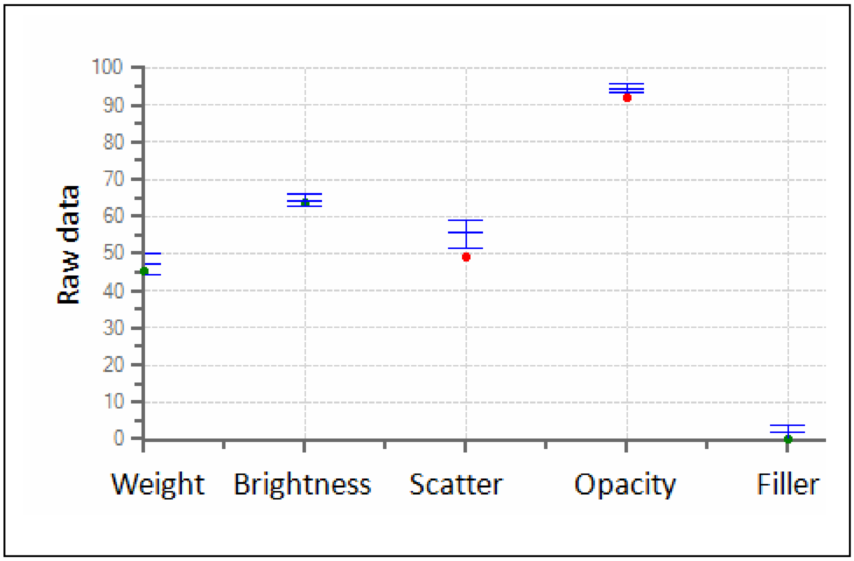
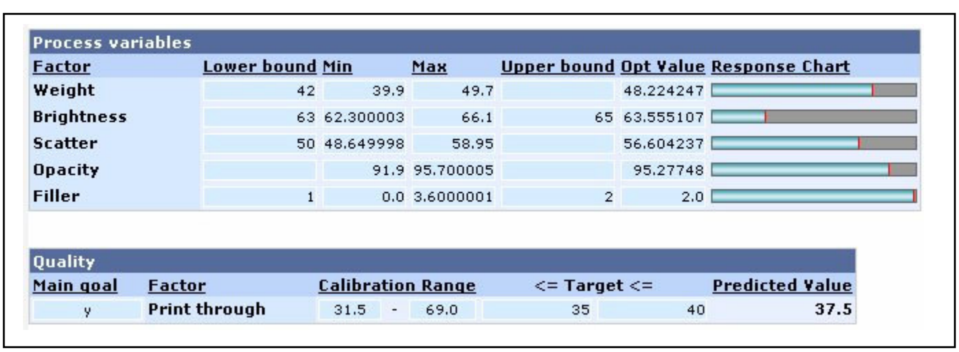


Figure 7. Plot of sample 14 showing Opacity and Scatter being outside the min and max from the calibration

OPTIMIZATION WITH CONSTRAINTS

Once a model has been established the next step may be optimization of the process with constraints on the process variables as well as the product quality.

In this case some constraints were given for the five important variables in the final model and at the same time the target range for the Print Through was set to be between 35 and 40. Figure 8 shows the result of the optimisation from the Unscrambler Optimizer software. It is also possible to add interaction and square terms but this was not pursued in detail as there were no indications that such additional model terms improved the model.



Process variables

Factor	Lower bound	Min	Max	Upper bound	Opt Value	Response Chart
Weight	42	39.9	49.7		48.224247	
Brightness	63	62.300003	66.1	65	63.555107	
Scatter	50	48.649998	58.95		56.604237	
Opacity		91.9	95.700005		95.27748	
Filler	1	0.0	3.6000001	2	2.0	

Quality

Main goal	Factor	Calibration Range			<= Target <=		Predicted Value
y	Print through	31.5	-	69.0	35	40	37.5

Figure 8. Optimization given constraints on the process variables and target range

SUMMARY

Multivariate methods have been shown to be an efficient tool in monitoring the process variables as well as for predicting final product quality at an early stage. It replaces univariate control charts with the more effective multivariate representation of the process' "sweet-spot". Disturbances in the process can easily be detected and the variables causing the upset can be interactively spotted in the on-line monitoring phase. Thereby, the process operators do not need to know the methodology behind the sytem as the plot of the original process variables is shown on screen. The concept of MSPM can also be extended to hierarchical models for classification and prediction of raw material quality in a complete production process quality system.

References

1. J.E.Jackson, *A Users Guide to Principal Components,* Wiley & Sons Inc., New York, 1991.

About CAMO Software

Founded in 1984, CAMO Software is a leader in multivariate data analysis and design of experiments software and solutions. Our flagship software, The Unscrambler® X, is recognized for its ease of use, outstanding visualization and powerful analytical tools. It is used by more than 25,000 people in over 3,000 organizations worldwide to analyze and make predictive models from complex data. Our in-house experts can advise on Process Analytical Technology (PAT) and Quality by Design (QbD) initiatives in addition to general data analysis across all industries. For more information please visit www.camo.com

camo.com

Bring data to life

NORWAY
Nedre Vollgate 8,
N-0158
Oslo
Tel: (+47) 223 963 00
Fax: (+47) 223 963 22

USA
One Woodbridge Center
Suite 319, Woodbridge
NJ 07095
Tel: (+1) 732 726 9200
Fax: (+1) 973 556 1229

INDIA
14 & 15, Krishna Reddy
Colony, Domlur Layout
Bangalore - 560 071
Tel: (+91) 80 4125 4242
Fax: (+91) 80 4125 4181

AUSTRALIA
PO Box 97
St Peters
NSW, 2044
Tel: (+61) 4 0888 2007

JAPAN
Shibuya 3-chome Square Bldg 2F
3-5-16 Shibuya Shibuya-ku
Tokyo, 150-0002
Tel: (+81) 3 6868 7669
Fax: (+81) 3 6730 9539