Journal of the Society of Korea Industrial and Systems Engineering
Vol, 31, No. 3, pp.125－130, September 2008.

자동차 차체 조립공장에서 주성분 분석의 응용 : 사례 연구

이명득* · 임익성**† · 김은정***

*미국 포드자동차 연구혁신센터
**남서울대학교 산업경영공학과
***한국정보문화진흥원 조사연구팀

Application of Principal Component Analysis in Automobile Body Assembly : Case Study

Myung D. Lee* · Ik Sung Lim**† · Eunjung Kim***

*Research and Innovation Center, Ford Motor Co. MI, U.S.A.
**Dept. of Industrial and Management Engineering, Namseoul University, Korea
***Research and Policy Development Team, Korea Agency for Digital Opportunity and Promotion, Korea

이 논문은 자동차 차체 조립과정에서, 품질관리의 일환으로써, 비접촉 자동측정시스템을 이용하여 검사해야 하는 수많은 비독립적인 검사점을 다변량분산분석과 주성분분석을 이용하여 효율적으로 검사점을 감소시키는 방법을 설명하고 있다. 이 연구의 목적은 다변량분산분석, 주성분 분석의 개념과 이러한 기법들을 산업체 제조분야에서 응용하는 방법을 설명하여 독자의 사례 응용 이해를 돕는데 있으며, 또한 특히 주성분분석을 이용하여 수 많은 비독립적인 검사점을 어떻게 유효하게 줄여나가는지를 보여주고자 한다. 독자의 이해를 돕기 위하여 위와 같은 절차를 순서대로 설명하였으며, 실제 자동차 조립공장에서 발생하는 사례를 수치 예를 들어 설명하였다.

Keywords : Data reduction, Control Points, OCMM, PCA, MANOVA

1. Introduction

Multivariate statistical methods have been successfully applied to biology, computer science, sociology and quality control areas. Principal Component Analysis (PCA) was applied on the standardized full near-infrared (NIR) spectral data for vegetable oils to totally capture the NIR spectral pattern [18].

The application of Multivariate Analysis of Variance (MANOVA) in the quality control area has increased and it appears that it will be used frequently in the future for data analysis. Shewart's control charts presented the first unified treatment of the subject of quality control [19]. Shewart's process control charts are based on the assumption that only one critical check point is of interest. The quality of body-in-white in the automotive assembly process, though, is determined by several critical check points. When more than one check point is of interest and each is controlled using separate univariate Shewhart process control charts, the probability of type I error or the probability of a false alarm, becomes highly distorted.

The application of multivariate statistical analysis can overcome these difficulties. Multivariate statistical techniques have

논문접수일 : 2008년 08월 27일　　논문수정일 : 2008년 09월 17일　　게재확정일 2008년 09월 18일
† 교신저자

been developed by Hotelling [9, 10] and applied to quality control plan. The multivariate analogue of the Shewart control chart is the T^2 control chart. Hotelling's paper combined the characteristics mean, variance and covariance into one statistic. Also, he discussed the relevant small sample distributions and illustrated these techniques by actual case studies involving the test of bombsights.

Also, few real world cases have been published in applying multivariate statistical analysis to the manufacturing industry. Since manufacturing processes are, in many ways, different from other fields. These methods should be modified. The dimensional measurement systems in the automotive industry take multiple dimensional measurements for each part. Each check point on a part behaves as a random variable. Therefore, a new methodology is required to analyze data that is dependent on each other. With the availiability of new sensor technology, an OCMM (Optical Coordinate Measurement Machine) that consists of a large rigid frame supporting a large number of sensors can measure the dimensions of every unit produced, resulting in 100% measurement data. This set-up provides massive amount of process data that are multivariate in nature [11].

2. Principal componenet analysis

2.1 concept between correlation and covariance

The covariance is an absolute measure of the joint variation of two variables. By definition, the covariance of the elements Xi and Xj of X as the product moment of those variates about their respective means.

$$Cov(X_{i.}X_j) = E[X_i - E(X_i)][X_j - E(Xj)] \quad (2.1)$$

and $E(X_i)$ and $E(X_j)$, i, j = 1, 2, ⋯, p, are the marginal means. When i=j, the covariance becomes the marginal variance. If there is a relationship between X_i and X_j, then

$$X_j = \alpha_{ij} + \beta_{ij} + \epsilon_j \quad (2.2)$$

In many multivariate procedures, the linear relationship is led naturally by combining (2.1) and (2.2).

$$Cov(X+X) = E[X_i - E(X_i)][\alpha_{ij} + \beta_{ij} + \epsilon_j - E(\alpha_{ij} + \beta_{ij} + \epsilon_j)] \quad (2.3)$$

$$= E\beta_{ij}[X_i - E(X_i)]^2 \quad (2.4)$$

$$= \beta_{ij} Var(X_i) \quad (2.5)$$

or equivalently,

$$\beta_{ij} = \frac{Cov(X_i,\ X_j)}{Var(X_i)} \quad (2.6)$$

The value of β_{ij}, which is the slope of X_i and X_j, is a very important factor because this can represent the magnitude of a linear relationship. However, the correlation coefficient provides the amount of linear association between the random variables X_i and X_j. There are two kinds of PCA. One is based on the correlation matrix and the other is based on the covariance matrix. Based on the Yang's paper [21], the PCA on covariance matrix should be used in dimensional quality control and measurement system analysis because it will provide a magnitude of variable relationships and a clear geometrical interpretation of the results.

2.2 Concept of principal component analysis

Principal Component Analysis (PCA) is a statistical technique that linearly transforms a given set of variables into a new set of composite variables. These new variables are orthogonal to each other and capture most of the information in the original variables. A small set of uncorrelated variables is much easier to understand and use in further analysis than a larger set of correlated variables. It is for these reasons that principal components are often used in exploratory data analysis and data reduction. The concept was originally conceived by Pearson [16] and independently developed by Hotelling [9].

Principal components, $Z_1, Z_2, \cdots, Z_p$, are particular linear combinations of the variables, $X_1, X_2, \cdots, X_p$ with n measurements. After the PCA is conducted, $Z_1, Z_2, \cdots, Z_p$ are uncorrelated to each other, which can be expressed as :

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{p1} \end{bmatrix} Z_1 + \cdots + \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{pi} \end{bmatrix} Z_i + \cdots + \begin{bmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{pp} \end{bmatrix} Z_p \quad (2.7)$$

and $Cov(Z_i, Z_j) = 0$ for all $i \neq j$ which indicates that all geometrical variation modes are independent of each other. The key idea from matrix algebra related to the method of principal components is that a p x p symmetric, nonsingular matrix, such as the covariance matrix Σ, may be reduced to a diagonal matrix **D** by premultiplying and postmultiplying by a particular orthonormal matrix **A** such that

$$A\Sigma A' = D \tag{2.8}$$

The diagonal elements of **D**, $\lambda_1, \lambda_2 \cdots, \lambda_{10}$ are called the eigenvalues of Σ. The columns of **A**, $a_1, a_2, \cdots, a_p$ are called eigenvectors of Σ. The eigenvalues may be obtained from the following determinant equation

$$|\lambda I - \Sigma| = 0 \tag{2.9}$$

where **I** is the identity matrix. This produces a p^{th} degree polynomial in λ from which the values $\lambda_1, \lambda_2, \cdots, \lambda_p$ are obtained. Therefore :

$$Var(Z) = \begin{bmatrix} Var(Z_1) & 0 & 0 & 0 \\ 0 & Var(Z_2) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & Var(Z_p) \end{bmatrix} =$$

$$A\Sigma A' = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_p \end{bmatrix} \tag{2.10}$$

Clearly, $\lambda_i = Var(Zi)$, since λ is obtained by letting $\det(\lambda I-\Sigma) = 0$, so $\lambda_1 + \cdots + \lambda_p = Var(X_1) + \cdots + Var(X_p) = \sum_{i=1}^{p}\sigma_i^2$. It shows clearly that the percentage of total variance of original variables for each principal component can be explained by its corresponding eigenvalue.

3. Case study : data reduction of the automotive body assembly

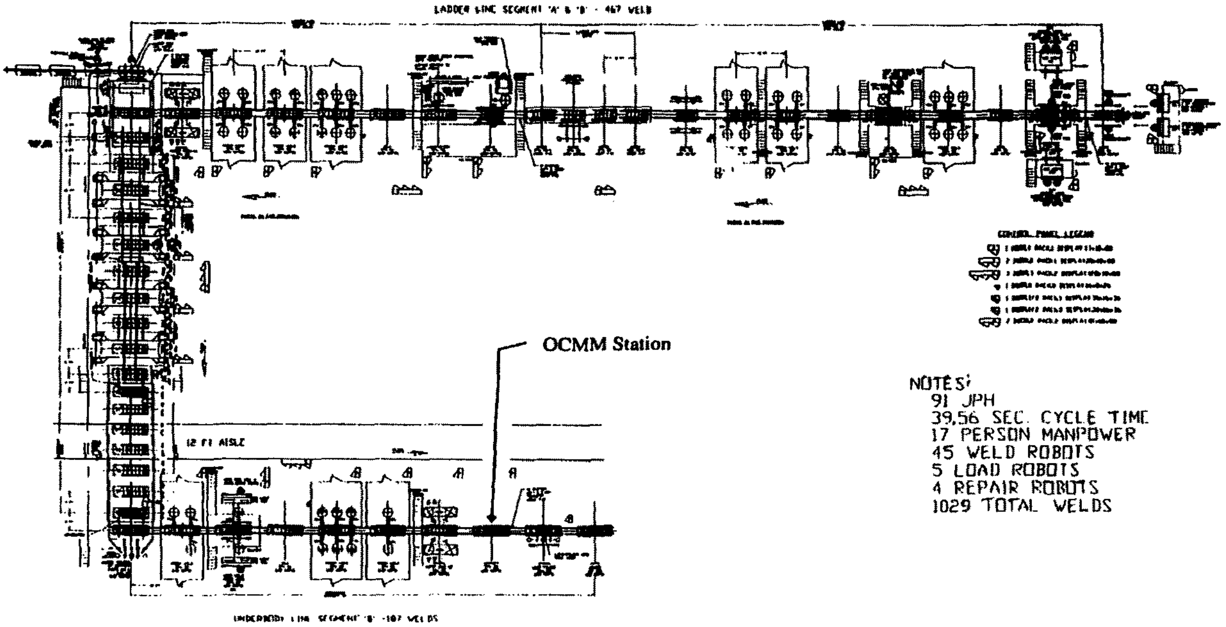
As there are an infinite number of possible points on a body-in-white that can be monitored, the body shop area requires a practical system to track and report variation of body assembly. A minimum number of check points must be selected in order to permit the body-in-white to identify, monitor and control the assembly process of these critical build characteristics. In the conventional method, the engineer correctly believes all variables have been tried, but only a fraction of the total possible combinations have been tried. Additionally, the undisciplined manner in which the trials are typically made provides data of questionable validity and creates more confusion than new knowledge. The engineer also has no information about the combinations of variables, which are important because, generally variables are not independent of each other. Thus, changing only one variable at a time, the true effect of a variable may go undiscovered, and the problem will reamin unsolved. The question that will arise will be how to select and analyze each measurement point individually and concurrently. This is complicated because it requires expertise in different stations, at several different points, and at the assembly process. It is also a slow and time consuming process when dealing with the huge amounts of data and the many characteristics of an assembly process.

As a case study for data reduction, the underbody assembly is presented. The underbody assembly is a very complicated multi-station process in which many manufacturing variables are contained through each work station. The assembly work stations include material handling, sealer and welding stations. Typical welding operations are performed by robots, hard tooling, and are press or manual in operation. The dimensional integrity of the underbody can be affected by the sequence of sub-assembly and the sequence of spot-welding. The clamping position, force and sequence can also impact dimensional stability. Since there are so many manufacturing variables and so many dimensional control stations, selecting a minimum number of checking points is not an easy task.

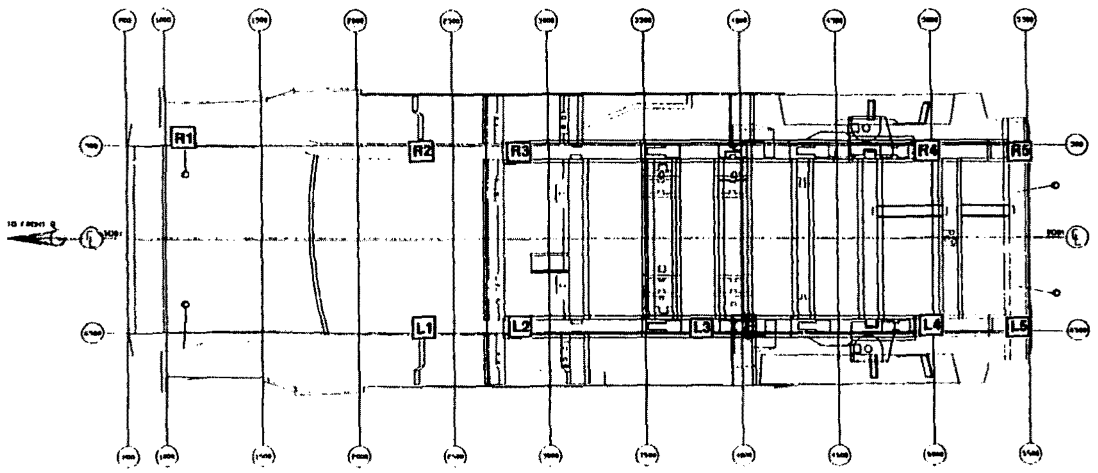
The master control points on an underbody are measured by an OCMM that locates at the end of assembly tooling line, as shown in <Figure 1>. In this case study, the Hi/Lo coordinate variation of the underbody is of interest and only the dimensional deviations from the nominal dimensions are recorded, <Figure 2>. <Table 1> shows the results of the PCA conducted on the covariance matrix for the Hi/Lo measurements on these ten master control points. The statistics of the PCA are the ten variances $\lambda_1, \cdots, \lambda_{10}$ ordered by size and the associated variable weight vectors $a_1, \cdots, a_{10}$. the first column are the weights associated with the linear composite that has maximum variance. The linear composite is

$$(L1, L2, L3, L4, L5, R1, R2, R3, R4, R5)^T = (0.267, 0.321, 0.300, 0.355, 0.329, -0.269, -0.331, -0.322, -0.309, -0.348)^T Z_1 \tag{3.1}$$

The first five weights are associated with left hand side and positive variables and the last five weights are associated with right hand side and negative variables. It is clearly indicated that the left hand side moved to the same direction together,



<Figure 1> Underbody Process Layout



<Figure 2> Underbody Assembly Master Control Points (H/L)

and the right hand side moved together in the opposite direction. This twisting deformation pattern is the largest variation mode which accounts for 88.7% of total variation of the underbody Hi/Lo position. The second principal component is the linear composite

$$(L1, L2, L3, L4, L5, R1, R2, R3, R4, R5)^T = (-0.529, -0.400, -0.225, -0.212, -0.150, -0.298, -0.245, 0.351, -0.334, -0.242)^T Z_1 \quad (3.2)$$

The weights are about equal so that each variable is equally represented in the linear composite. Accordingly, the second principal component which is 4.2% of total variation could be interpreted as a whole underbody moves up and down which represents translation movement. To select a subset of variables from a larger set of variables, PCA is used. Applying the PCA to the positive group, <Table 2> that the first and second principal component have a eigenvalue of 10.66 and accounts for 95.7% of the total variance. The linear composite is shown as follows :

$$\begin{bmatrix} L_1 \\ L_2 \\ L_3 \\ L_4 \\ L_5 \end{bmatrix} \cong \begin{bmatrix} -0.393 \\ -0.495 \\ -0.424 \\ -0.497 \\ -0.456 \end{bmatrix} Z_1 + \begin{bmatrix} -0.210 \\ 0.185 \\ 0.441 \\ 0.338 \\ -0.783 \end{bmatrix} Z_2 \quad (3.3)$$

<Table 1> PCA Data

Var.	PC1	PC2	PC3	PC4	PC5
L1	0.267	-0.529	-0.099	-0.399	-0.174
L2	0.321	-0.400	0.062	0.447	-0.457
L3	0.300	-0.225	0.357	0.060	0.318
L4	0.355	-0.212	0.279	-0.116	0.236
L5	0.329	-0.150	-0.504	0.029	0.466
R1	-0.269	-0.298	0.195	-0.156	-0.398
R2	-0.331	-0.245	0.034	-0.259	0.046
R3	-0.322	-0.351	-0.404	-0.280	0.136
R4	-0.309	-0.334	-0.253	0.665	0.109
R5	-0.348	-0.242	0.510	0.103	0.442

Var.	PC6	PC7	PC8	PC9	PC10
L1	0.128	-0.449	0.123	0.246	-0.395
L2	0.153	0.174	-0.362	-0.373	-0.007
L3	-0.207	0.413	-0.250	0.592	-0/032
L4	0.123	0.203	0.659	-0.349	0.265
L5	-0.396	-0.252	-0.245	-0.225	0.257
R1	-0.435	-0.149	0.029	0.134	0.623
R2	-0.493	0.359	-0.013	-0.398	-0.480
R3	0.497	0.404	-0.156	0.069	0.273
R4	-0.092	-0.038	0.555	0.231	-0.122
R5	0.240	-0.423	-0.273	-0.207	0.005

Eigenanalysis of the Covariance Matrix						
Eigenvalue	19.985	0.956	0.677	0.345	0.204	0.135
Proportion	0.887	0.042	0.030	0.015	0.009	0.006
Cumulative	0.887	0.929	0.959	0.974	0.983	0.989

Eigenvalue	0.111	0.067	0.035	0.028
Proportion	0.005	0.003	0.002	0.001
Cumulative	0.994	0.997	0.999	1.000

The weights for the first principal component are about equal so that each master control point of the left hand side is about equally represented in the linear composite. This concludes that PCA searches for a few uncorrelated linear combinations of the original variables that capture most of the information in the original variables. This case study shows that four principal components account for most of the variance in the original ten master control points which combines left and right hand sides <Table 2>.

The structure of the data set is simplified considerably. If univariate analysis such as statistical process control is used for the same problem, it would require ten control point charts at the same time to understand and analyze how the underbody system varies together.

Reducing the number of control points that are supposed to be checked will save time for collecting and analyzing data. Tooling costs for measurement systems will be reduced significantly by deleting mechanical units or measurement sensors. Most of all, the availability of a data set which contains most of the original value will provide the opportunity to improve the dimensional quality of the bodies in a short time.

<Table 2> PCA for Left/Right Side

Principal Component Analysis for Left Side					
Var.	PC1	PC2	PC4	PC4	PC5
L1	-0.393	-0.210	0.724	0.493	0.183
L2	-0.459	0.185	0.367	-0.777	-0.132
L3	-0.424	0.441	-0.334	0.077	0.713
L4	-0.497	0.338	-0.254	0.367	-0.663
L5	-0.456	-0.783	-0.406	-0.114	0.035
Eigenanalysis fof the Covariance Matrix					
Eigenvalue	10.380	0.376	0.242	0.180	0.058
Proportion	0.924	0.033	0.022	0.016	0.005
Cumulative	0.924	0.957	0.979	0.995	1.000

Principal Component Analysis for Right Side					
Var.	PC1	PC2	PC3	PC4	PC5
L1	-0.382	-0.244	0.262	0.606	0.599
L2	-0.464	-0.037	0.395	0.261	-0.748
L3	-0.456	0.588	0.367	-0.485	0.278
L4	-0.438	0.385	-0.765	0.266	-0.058
L5	-0.489	-0.667	-0.236	-0.509	0.034
Eigenanalysis fof the Covariance Matrix					
Eigenvalue	10.449	0.410	0.247	0.145	0.058
Proportion	0.924	0.036	0.022	0.013	0.005
Cumulative	0.924	0.960	0.982	0.995	1.000

4. Summary

Multivariate analysis is a rapidly expanding approach to data analysis. One specific technique in multivariate analysis is Principal Component Analysis (PCA). PCA is a statistical technique that linearly transform a given set of variables into a new set of composite variables. These new variables are orthogonal to each other and capture most of the information in the original variables. PCA is used to reduce the number of control points to be checked by measurement system. Therefore, the structure of the data set is simplified significantly. It is also shown that eigenvectors obtained by conducting principal component analysis on the basis of the covariance matrix can be used to physically interpret the pattern of relative deformation for the points. This case study reveals the twisting deformation pattern of the underbody which is the largest mode of the total variation.

References

[1] Baron, J.; "Dimensional Analysis and Process Control of Body In While processes," Ph.D. Dissertation, University of Michigan, Ann Arbor, MI, U. S. A., 1992.

[2] Bennich, P., Cvesto, P. C., Soons, H., and Trapet, E.; "Calibration of Co ordinate Measuring Machines," WELL Technical Guidelines, draft version, U.S.A. October., 1992.

[3] Burdick, R. K., Borror C. M., and Montgomery, D. C.; "A review of methods for measurement systems capability analysis," *Journal of Quality Technology*, 35(2), 342-354, 2003.

[4] Chen, J., Bandoni, J. A., Romagnoli, J. A.; "Robust PCA and Normal Region in Multivariate Statistical Process Monitoring," *AICHE Journal*, 42(12). 3563-3566, 1996.

[5] Folan, P. and Browne, J.; "A review of performance measurement : Towards performance management," *Computers in Industry*, 56, 663-680, 2005.

[6] Gimmi, K. J.; Measurement Uncertainty, ASQC Quality Congress Transactions, Boston, 916-921, 1993.

[7] Greer, D.; "On line Machine Vision Sensor Measurements in a Coordinate System," SME paper # IO, 88-289, 1988.

[8] Holmes, D. S. and Mergen, A. E.; "Improving the Performance of the T2 Control Chart," *Quality Engineering*, 5(1), 619-625, 1993.

[9] Hotelling, H.; "Analysis of a Complex of Statistical variables into Principal Components," *Journal of Educational Psychology*, 24(1), 498-520, 1993.

[10] Hotelling, H.; Multivariate Quality Control Techniques of Statistical Analysis, Eisenhart, Hastay and Wallis, Editors, McGraw-Hill, 111-184, 1947.

[11] Hu, S. and Wu, S. M.; "Impact of 100% In-Process Measurement on Statistical Process Control in Automobile Body Assembly," *ASMEW inter Annual Meeting*, Dallas, Texas, Nov., U. S. A. 355-359, 1990.

[12] Knowles, G., Vickers, G. and Anthony, J.; "Implementing Evaluation of the Measurement Process in an Automotive Manufacturer : a Case Study," *Quality and Reliability Engineering International*, 19, 397-410, 2003.

[13] Lee, Myung Duk; "Methodology for Measurement Systems Analysis and Dimensional Control Process of Automotive Body Manufacturing," Ph.D. Dissertation, Wayne State University, Detroit, MI, U. S. A., 1996.

[14] Mast, J. and Trip, A.; "Gauge R&R studies for destructive measurements," *Journal of Quality Technology*, 37, 40-49, 2005.

[15] Morris, A. S.; Measurement and Calibration for Quality Assurance, Prentice Hall International (UK) Ltd. 1991

[16] Pearson, K.; "On Line and Planes of Closet Fit to Systems of Points in Space," *Philosophical Magazine*, Sep. 6(2), 559-572, 1991.

[17] Proud, P. and Ermer, D. S.; "A Geometrical Analysis of Measurement system Variations," ASQC Quality Congress Transactions, Oct., p. 929-935, 1993.

[18] Sato, T.; "Application of Principal Component Analysis on Near-Infrared Spectroscopic Data of Vegetable Oils for Their Classification," *JAOCS*, 71(3), March, 1994.

[19] Shewhart, W. A.; Economic Controlof Quality of Manufactured Product, Van Nostrand Co., Inc. NewYork, 1931.

[20] Wu, S. K.; "A Methodology for Optimal Door Fit in Automobile Body Manufacturing," Ph. D. Dissertation, The University of Michigan, Ann Arbor, U. S. A., 1991.

[21] Yang, K.; "Improving Automotive Dimensional Quality by Using Principal Component Analysis," *Quality and Reliability Engineering International*, 12, 401-409, 1996.

[22] Yang, P. W., Liu, F.; "Application Comparison of PCA and PPCA in Chemical Separation Process Monitoring," *Control and Instruments in Chemical Industry*, 34(6), 7-11, 2007.