Solving Quality Problems with Multivariate Data Analysis: Basics and Applications

Manabu Kano
manabu@cheme.kyoto-u.ac.jp
Kyoto University
Nishikyo-ku, Kyoto 615-8510, Japan
Phone: +81 -75-383-2687 Fax: +81-75-383-2657

To realize target quality in manufacturing processes, we have used correlation analysis, multiple regression analysis, and other statistical techniques including even super-advanced methods. However, our processes are too complicated and factors or parameters to be investigated are too many; it seems difficult to achieve good results. How can we solve such problems? In this lecture, the know-how of using simple, useful, suitable data analysis techniques is explained with various industrial applications.

First of all, it is beneficial for us to recognize that manufacturing processes in different industries are quite different in appearance but they have very similar problems from the viewpoint of quality issue.

In the chemical industry, for example, product quality is not usually measured in real time due to high investment and maintenance cost of on-line analyzers. To meet the customers' specifications, inferential control has been widely used. The key idea is to develop a soft-sensor that can estimate product quality from measured process variables and to use the estimates for feedback control.

In the steel industry, it is crucial to minimize defects in steel products and improve product yield, but the problem is not easy because various products are manufactured through various equipments such as a blast furnace, a steel converter, a continuous-casting machine, and hot and cold rolling mills. Engineers have great difficulty in finding the key operating condition to reduce the defects.

In the pharmaceutical industry, Quality by Design (QbD) has become a keyword very recently. QbD is an emerging concept that the productivity can be greatly improved by realizing quality assurance through real-time monitoring instead of sample-based inspections. Process Analytical Technology (PAT) plays an important role in the framework of QbD.

The above-mentioned problems in various industries are similar to the problems in the semiconductor industry: how to estimate product quality, how to improve product quality and yield, and how to detect process faults. Therefore, it is beneficial for you to know the technologies that have succeeded in solving problems in other industries.

In this lecture, three kinds of technologies are explained with industrial applications: quality estimation, quality improvement, and fault detection. All technologies are based on multivariate data analysis, because it is quite common that first-principle models (mechanistic models) are not available but a huge amount of data is available instead. In fact, a major problem in various companies seems how to use enormous data stored in expensive databases for process innovation.

To estimate product quality, it is very important to understand the basics of linear regression analysis. Of course, many user-friendly software tools enable you to execute advanced modeling methods. However, such tools assume that you use them correctly on your own responsibility. Are you confident that you can build a linear regression model correctly even when colinearity occurs? In this lecture, after the basics of multivariate data analysis are given, applications of soft-sensors in the chemical industry are introduced. Partial Least Squares (PLS) is a well-known powerful method for building a regression model[1]. Here, an application of PLS to an industrial distillation process[2] is illustrated. To improve the estimation accuracy under the existence of unknown disturbances, two-stage subspace identification (TS-SSID) is useful[3]. In addition, to cope with changes in process characteristics, that is, to adapt a soft-sensor to process characteristics changing with time, Correlation-based Just-In-Time (CoJIT) modeling works well[4].

To improve product quality and yield, a model that relates product quality with operating conditions needs to be developed. Regression analysis can be used for this purpose. However, product quality cannot be always measured quantitatively; sometimes only judgment on the quality, i.e. good or bad, is available. Discriminant analysis is useful in such a case. Here, a successful application of multivariate data analysis in the steel industry[5] is introduced. The objective was to find key variables and operating conditions for reducing product defects and consequently improving product yield. A method combining principal component analysis (PCA) and linear discriminant analysis (LDA) was developed for relating product yield and operating conditions in various equipments, finding key variables, and optimizing operating conditions. As a result, drastic improvement in product yield was achieved.

To detect faults, multivariate statistical process control (MSPC) has been widely used in various industries. The key concept of MSPC is to check correlation among process variables for improving the fault detection performance. Although many methods have been proposed, the most important method is PCA-based MSPC[6]. Multiway PCA can be used for monitoring batch processes[7]. A good application of both PCA-based MSPC and Multiway PCA was reported[8], and it is introduced here. MSPC was used to prevent a serious trouble, i.e. breakout, in the steel industry. To make MSPC more practicable, it is crucial to distinguish faults from normal changes in operating conditions and process characteristics. External analysis is useful for this purpose[9]. Among various advancements in this field, independent component analysis is introduced to further improve the monitoring performance[10].

Through collaboration with various companies in various industries, it becomes clear that there remain major challenges. Model maintenance or model adaptation is very important for coping with changes in process characteristics. This should be done automatically. In addition, integration of a statistical model and a first-principle model is necessary to improve the model accuracy. Furthermore, efficient data preprocessing is crucial in practice because it is dominant over all the other modeling activities.

Here is a conclusion. Simple multivariate data analysis will solve your quality problems even though they seem quite complicated. You need to understand basic methods correctly and to use them appropriately for your purpose. It is really important to clarify the objective of your project: what you have to achieve. Otherwise, you cannot choose an appropriate method for solving the problem. Be patient! Think logically!

Acknowledgement

The author would like to acknowledge the technical and financial support of various companies.

References

[1] M. Kano, K. Miyazaki, S. Hasebe, I. Hashimoto: Inferential Control System of Distillation Compositions Using Dynamic Partial Least Squares Regression. *Journal of Process Control*, 10, 157-166 (2000)

[2] H. Kamohara, A. Takinami, M. Takeda, M. Kano, S. Hasebe, I. Hashimoto: Product Quality Estimation and Operating Condition Monitoring for Industrial Ethylene Fractionator. *Journal of Chemical Engineering of Japan*, 37, 422-428 (2004)

[3] M. Kano, S. Lee, S. Hasebe: Two-Stage Subspace Identification for Softsensor Design and Disturbance Estimation. *Journal of Process Control*, 19, 179-186 (2009)

[4] K. Fujiwara, M. Kano, S. Hasebe, A. Takinami: Soft-Sensor Development Using Correlation-Based Just-In-Time Modeling. *AIChE Journal*, 55, 1754-1765 (2009)

[5] M. Kano, Y. Nakagawa: Data-Based Process Monitoring, Process Control, and Quality Improvement: Recent Developments and Applications in Steel Industry. *Computers & Chemical Engineering*, 32, 12-24 (2008)

[6] T. Kourti, J.F. MacGregor: Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods. *Chemometrics and Intelligent Laboratory Systems*, 28, 3-21 (1995)

[7] P. Nomikos, J. MacGregor: Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics*, 37, 41-59 (1995)

[8] Y. Zhang, M.S. Dudzic: Online Monitoring of Steel Casting Processes Using Multivariate Statistical Technologies: From Continuous to Transitional Operations. *Journal of Process Control*, 16, 819-829 (2006)

[9] M. Kano, S. Hasebe, I. Hashimoto, H. Ohno: Evolution of Multivariate Statistical Process Control: Application of Independent Component Analysis and External Analysis. *Computers & Chemical Engineering*, 28, 1157-1166 (2004)

[10] M. Kano, S. Tanaka, S. Hasebe, I. Hashimoto, H. Ohno: Monitoring Independent Components for Fault Detection. *AIChE Journal*, 49, 969-976 (2003)