# Data Analysis Technology in Plant Control

MATSUI Tetsuro* MURAKAMI Kenya* SUZUKI Satoshi*

**ABSTRACT**

For safe and optimized control of factory equipment and plants, various control technologies has been developed. Fuji Electric is developing technologies for analyzing massive data measured in plants and making effective use of them. Fault diagnosis technology by means of multivariate statistical process control can be used for recognizing faults in manufacturing processes accurately to improve the operation rate. In addition, quality estimation technology based on the partial least squares method using manufacturing process information leads to improved quality and yield. Furthermore, log analysis technology by pattern mining can be used for analyzing past data and making tacit knowledge into explicit knowledge, thereby assisting operators.



## 1. Introduction

Diverse technologies are being developed for the purpose of ensuring the safe and optimized control of factory equipment and plants. These technologies include prediction technologies, which collect and analyze measured data to predict the future behavior of plants, fault diagnosis technologies intended to discover faults at an early stage, technologies for optimizing the energy efficiency of energy supply plants, and control technologies that enable plants to operate in a stable manner.

Fuji Electric Co., Ltd. has developed the following technologies: Original neural network technology as core technology for prediction and fault diagnosis; metaheuristics optimization technology, such as particle swarm optimization (PSO) suitable for mathematical programming and nonlinear large-scale optimization problems; PID control technology; model prediction control technology for multivariable systems; and control performance monitoring technology that monitors deterioration of control performance resulting from changes in the characteristics of the object controlled[1].

With the sophistication of these technologies, an increasing number of plants in various fields are introducing automatic control. On the other hand, the startup and shutdown of plants and response to unusual states, such as faults, still rely on the judgment of experienced operators in many cases. Drawing attention as an approach to advanced automation of plants is data analysis technology, which analyzes massive volumes of data measured at plants and utilizes them for improving operations, improving manufacturing quality and responding to unusual states.

This paper presents an overview of trends in plant control technology and describes what Fuji Electric's data analysis technology is like.

## 2. Trends in Plant Control Technology

### 2.1 Demands for data analysis technology in view of plant control

To maintain the quality of products manufactured at plants, upper and lower limits are set on process variables that will affect the quality of products to determine whether there is any problem with the quality. However, since multiple process values are in correlation with one another, it is difficult to conduct correct diagnosis by monitoring any single variable. In addition, for a process in a normal condition with great variation in process values, proper lower and upper limits are difficult to set. For this reason, technology that enables us to conduct diagnosis with consideration given to the correlation between multiple variables is required.

In connection with the improvement of the yield and quality of products, the quality of finished products needs to be evaluated, and if any quality-related problem is found, such measures as equipment adjustment must be taken. However, this approach will give rise to defective lots because it evaluates finished products. What is needed as a solution to this issue is technology capable of preventing defective products by estimating the quality of products prior to completion and taking measures against expected problems.

Startup and shutdown of plants and response to unusual states must also be performed by field personnel. However, it takes field personnel time to respond to faults because they have less experience in faults and the number of experienced operators is on the decrease. Plant operations are carried out according to manuals to a certain degree, but some manual

* Corporate R&D Headquarters, Fuji Electric Co., Ltd.

operations are performed in different ways depending on the operator. Therefore, technology capable of realizing optimum operations in response to different situations is required. In case of a fault, operators will be swamped with alarms, which will be displayed on the history monitoring screen of the distributed control system (DCS), and they will become unable to determine each of them, making it difficult to identify the fault. To deal with these issues, it is desired today that measurement trend data, alarm data, operation log data and log data on manual operations collected at plants be analyzed and utilized to assist plant operations.

### 2.2 Trends in related technologies

When it comes to activities of societies engaged in plant data analysis technologies, the 143rd Committee on Process Systems Engineering of the Japan Society for the Promotion of Science conducts workshops organized by universities, user companies and vendor companies, such as "Alarm Management (2008 to 2010)" and "Soft Sensors (2010 to 2012)," compiled study results on issues confronting user companies and the latest technological trends. In the Institute of Electrical Engineers of Japan, the Electronics, Information and Systems Society has the "Technical Committee on Technological Survey of Use of Big Data (2012 to 2014)" and the "Technical Committee on Survey of Data-based Adaptive Smart Systems (2012 to 2014)," which are actively committed to surveys and research of technologies for analysis and use of data. The Society of Instrument and Control Engineers presented a feature article titled "Data-driven Control-New Approaches and New Horizons" in its journal "Measurement and Control" in 2013. As described thus far, many organizations have been actively devoting their energy to data analysis technologies for the last several years.

## 3. Fuji Electric's Commitment to Data Analysis Technology

### 3.1 Approaches to use of data analysis technology

It is important to enhance productivity by having the stable operation of a plant and improving manufacturing quality and yield. Thus, Fuji Electric takes the following three main approaches to the development of data analysis technology:

(a) improving the operation rate through accurate fault diagnosis of the manufacturing process;
(b) estimating product quality based on information obtained during manufacture and thereby improving quality and yield; and
(c) assisting operators in running the plant by analyzing past data and turning tacit knowledge into explicit knowledge.

We are developing fault diagnosis technology backed by multivariate statistical process control (MSPC) for (a), quality estimation technology adopting the partial least squares for (b), and log analysis technology based on pattern mining for (c) (See Fig. 1).

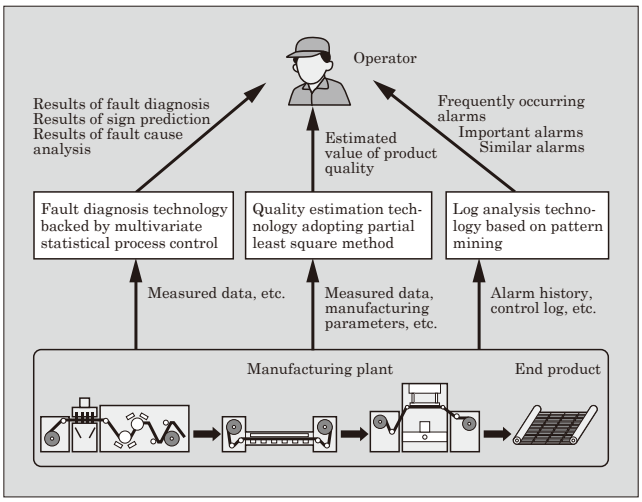


Fig.1 Approaches to use of data analysis technology

### 3.2 Fault diagnosis technology backed by multivariate statistical process control

Statistical process control means monitoring the operating state of the production process using a statistical method. It is a technique for preventing products that are nonconforming to specifications and thereby improving productivity.

Univariate statistical process control (USPC), which has been widely applied, is a technique for diagnosing faults by setting upper and lower limits, or control limits, on each process variable that significantly affects quality. However, if process values considerably vary, it is difficult to properly control upper and lower limits on variables even if the production process concerned is normal because the method for controlling these upper and lower limits has the precondition that the production process in normal state is working in a stable manner. MSPC, as opposed to USPC, is a technique capable of effectively monitoring a large number of variables, which affect one another in a complicated manner, by considering the correlation between these variables (see Fig. 2).

Next, we describe principal component analysis (PCA)*1, which is a technique of MSPC. Following steps are taken when applying PCA to fault diagnosis:

(a) Taking normal sample from a collected sample.
(b) Performing PCA with normal data to model the characteristics of the normal data.
(c) Setting a threshold on two indices used to evaluate the degree of divergence of the measured

*1: PCA: Principal component analysis. It is the method of taking characteristics from a massive amount of data by compiling the amount of information into orthogonal principal components based the correlation between many variables.

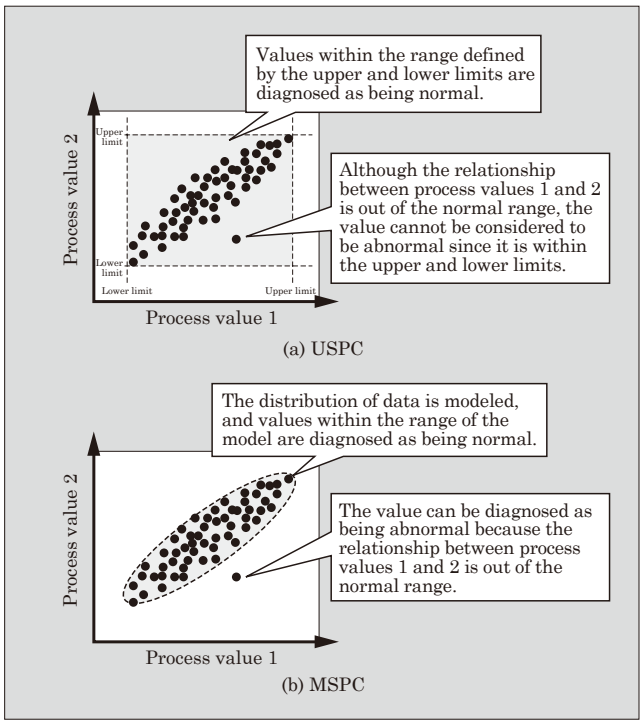


Fig.2 Comparison of fault diagnosis between USPC and MSPC

data from the normal model, $Q$ statistic*2 and $T^2$ statistic*3.

The $Q$ statistic and $T^2$ statistic are calculated using the measured data, and the values are diagnosed as being abnormal if they exceed the thresholds. The $Q$ statistic and the $T^2$ statistic are calculated by Equation (1) and Equation (2), respectively.

$$Q = \|x - \hat{x}\|^2 = \sum_{n=1}^{N} (x_n - \hat{x}_n)^2 \quad \cdots\cdots (1)$$

$Q$: $Q$ statistic
$\hat{x}$: Approximate value of input variable $x$ on PCA model
$N$: Number of variables

The $Q$ statistic is the index for evaluating the deviation from the correlation between the modeled data variables had and it can detect abnormality in the correlation between the variables.

$$T^2 = \sum_{m=1}^{M} \frac{t_m{}^2}{\sigma_{t_m}{}^2} \quad \cdots\cdots (2)$$

$T^2$: $T^2$ statistic
$t$: Principal component score
$\sigma_{t_m}$: Standard deviation of the $m$th principal component score
$M$: Number of principal components

*2: $Q$ statistic: Degree of deviation of the correlation between variables from the normal state.

*3: $T^2$ statistic: Degree of deviation of amplitude from the normal state although the correlation between variables is maintained.

The $T^2$ statistic corresponds to the distance from the average to each sample within the principal component space obtained by compressing original variables, and represents the degree of divergence from the average of the model. Consequently, even if the correlation between variables is maintained, abnormalities attributable to the fact that values themselves are large can be detected. An example of the application of PCA to fault diagnosis is shown in Section 4.1.

### 3.3 Quality estimation technology adopting partial least square method

Quality estimation is a technology that, with the aid of a multivariate analysis method, models the correlation between the manufacturing process and product quality based on the operating state of the process and the set values of manufacturing conditions and consequently estimates product quality in the middle of manufacturing. If the quality of the end product can be estimated in the middle of manufacturing, measures, such as parameter or system adjustments, can be taken to prevent the occurrence of defective products. Therefore, quality estimation technology can make a contribution to the achievement of stable quality and improved yield.

Even if there are many input variables, partial least squares (PLS*4) makes it possible to easily create a model using all input variables as they are, unlike a multi-regression analysis model, which requires input variables to be narrowed down in preliminary analysis. PLS thus helps significantly reduce the time required to create a model. PLS is very effective for processes with particularly many input variables and processes with multicollinearity.

Following steps are taken when applying PLS to quality estimation.

(a) Collecting production conditions, process values and quality values.
(b) Carrying out PLS using the collected values and creating a PLS model based on the relationship between quality and the production conditions and the process values.
(c) Performing a simulation with different production conditions to find conditions that improve quality.
(d) During diagnosis, sequentially estimating product from the production conditions and the pro-

*4: PLS: Partial Least Squares. PLS is a kind of modeling technique developed in the field of economics. A proper model can be obtained even if there is multicollinearity because for input variables that are correlated with one another, they are aggregated into intermediate variables and then output variables are expressed. Multicollinearity means that input variables cannot be properly modeled by general multi-regression analysis when input variables are strongly correlated with one another.

cess values.

An estimated value of quality based on a PLS model can be calculated from Equation (3).

$$\hat{y} = Q\ (W^{\mathrm{T}} P)^{-1} W^{\mathrm{T}} x \quad \cdots\cdots (3)$$

$\hat{y}$: Estimated value of quality
$x$: Input variable
$W$: Weighting matrix
$P$, $Q$: Coefficient matrices relating to input variable and output variable

An example of the application of PLS to quality estimation is shown in Section 4.2.

### 3.4 Log analysis technology based on pattern mining

Pattern mining means extracting "a pattern of a phenomenon fulfilling a constraint such as events, motion, operation or alarm," which exists frequently in a database. Recently, many monitoring control systems have been designed to give operators messages and, at the same time, save event logs as log data whenever such events as alarms, operations by operators, and actions by automatic control occur. Log data, in many cases, includes the times of occurrence and types (alarm, operation, control action) of events and the contents of messages and is accumulated as text data listed in time sequence.

By applying pattern mining to this log data, characteristic patterns corresponding to various situations can be extracted. Using these patterns enables us to embody improvements as shown below toward the achievement of optimum operation.

(1) Reducing unnecessary alarms

By extracting consecutive alarm patterns, in which only specific alarms occur, or patterns of the occurrence of notification alarms that do not need any operation, unnecessary alarms can be reduced through, for example, the aggregation of alarms.

(2) Automating routine operations

By extracting patterns of combinations of operations relative to the occurrence of alarms, operations can be automated as routine operations.

(3) Turning operation know-how into explicit knowledge

By extracting the operation patterns of experienced operators, operation know-how can be turned into explicit knowledge or automated. In addition, when multiple groups of operators working in shifts operate the plant, they can share operation know-how different among the groups by, for example, extracting the operation pattern of each group against the same alarm.

(4) Fault factor analysis

By extracting the patterns of the occurrence of events, including the occurrence of faults that should be noted, it is possible to analyze the factors and process of the occurrence of faults.

The log analysis flow of a control system adopting pattern mining is shown in Fig. 3. Below are the main items of log analysis.

(a) Data selection: Selecting the subject range from all data
(b) Preprocessing: Dealing with missing data
(c) Data conversion: Coding and aggregating data
(d) Serial pattern analysis: Carrying out pattern mining
(e) Filtering: Filtering the analysis result
(f) Pattern extraction: Extracting the analysis pattern

## 4. Application Examples

### 4.1 Fault diagnosis by PCA

What is particularly important to maintain the performance of a thermal power plant is to discover faults with the turbine shaft. This section cites an example of the application of PCA to turbine shaft fault diagnosis.

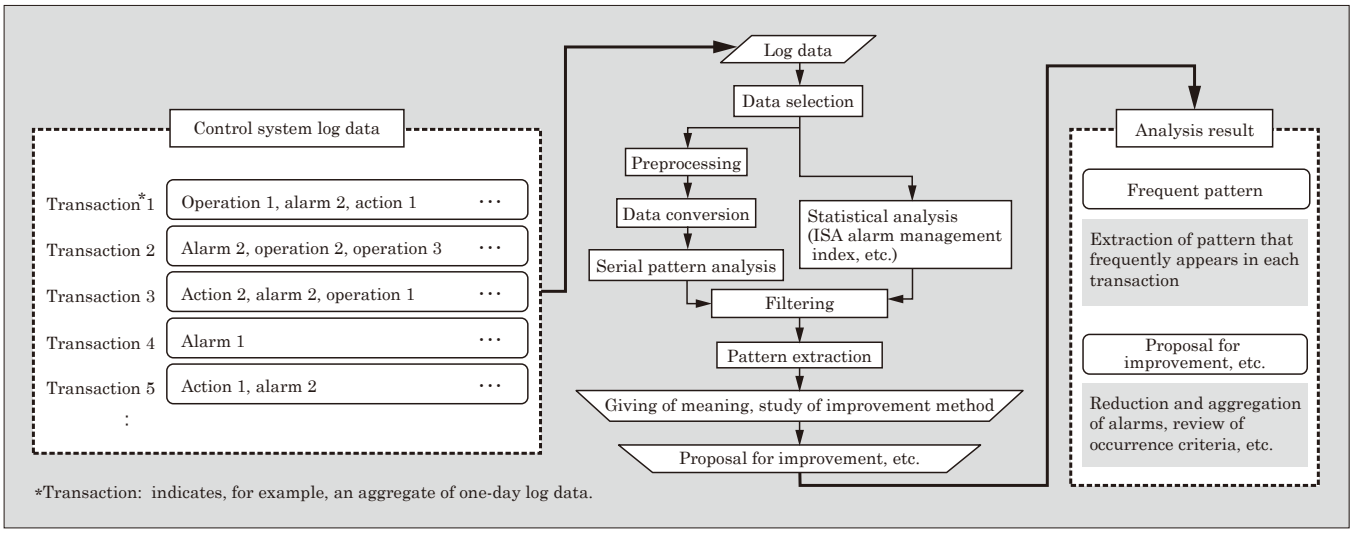


Fig.3 Log analysis flow of control system

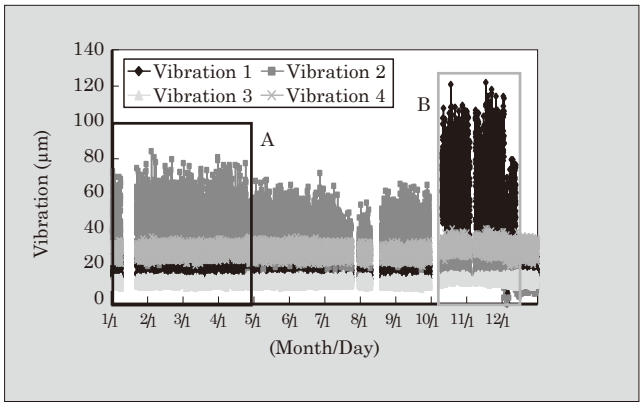


Fig.4 Vibration data on turbine shaft

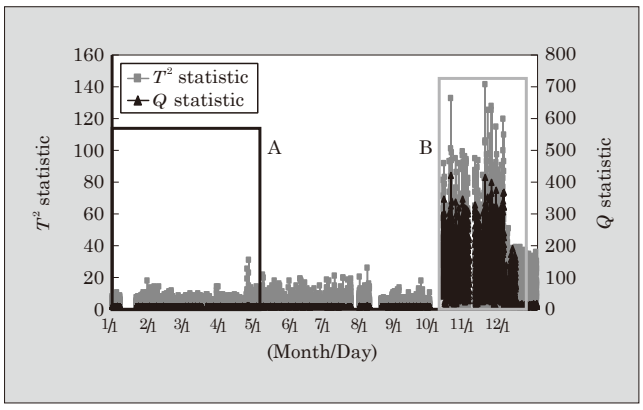


Fig.5 Analysis results of vibration data by PCA

In conventional control based on upper and lower limits, vibrations in excess of 120 μm are considered abnormal, but there is a possibility that faults cannot be detected when the value is 120 μm or lower. An example of vibration data is shown in Fig. 4. Area B is the period showing abnormal measured data because of sensor fault. Most of the data in this period was at or below the fault judgment threshold, 120 μm, and it is difficult for conventional upper/lower limit checks to detect abnormal data. PCA evaluation results of the $Q$ and $T^2$ statistics by creating a PCA model based on normal data (see area A in Fig. 4) are shown in Fig. 5. The $Q$ and $T^2$ statistics in area B were much larger than those in the other areas and helped detect abnormal vibrations.

### 4.2 Quality estimation by PLS

This section presents an example of quality estimation by PLS in a film formation process for thin-film photovoltaic cells. Based on the results of experiments performed by changing many manufacturing conditions, the relationship between the parameters of the manufacturing conditions and the conversion efficiency representing product quality was modeled by PLS.

The photovoltaic cell consists of about ten layers. The film formation process (see Fig. 6) uses about ten parameters of manufacturing conditions for the formation of each film, including temperature, pressure, film

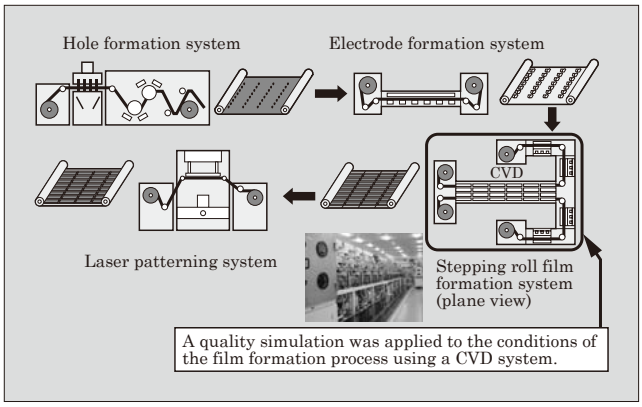


Fig.6 Film formation process for thin-film photovoltaic cell

formation time, film formation speed, film thickness, and doping level, and about 100 parameters in total for the formation of all films. These parameters are correlated with one another, and whenever any manufacturing condition is changed, the other ones correlated with the changed condition also need to be changed accordingly.

There are about 100 parameters in total for the formation of all films. These parameters are correlated with one another, and whenever any manufacturing condition is changed, the other ones correlated with the changed condition also need to be changed accordingly.

We conducted manufacturing experiments by gradually changing 30 of these about 100 parameters, which greatly affects product quality. A PLS model was constructed using data on the 115 obtained samples. The relationship between two main variables and conversion efficiency is shown in Fig. 7. In the figure, the dots are measured data, and the plane is the quality characteristic plane based on the PLS model, indicating that the distribution of the actual data can be approximated. The figure shows that higher efficiency is desirable, and that efficiency can be improved by changing the parameters of the manufacturing condi-

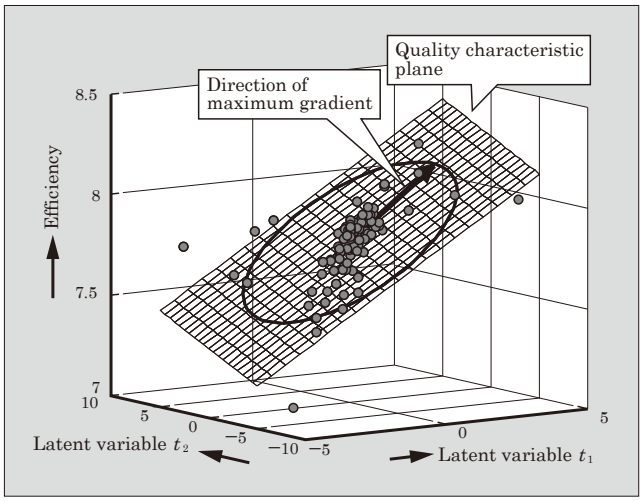


Fig.7 Example of quality estimation

tions in the steepest gradient direction (maximum gradient direction) of the plane.

## 5. Postscript

This paper described fault diagnosis by multivariate statistical process control, quality estimation technology adopting the partial least squares, and log analysis technology based on pattern mining as data analysis technologies for plant control. Of these technologies, log analysis technology based on pattern mining is in the process of being developed toward early practical use.

We are determined to establish more advanced, sophisticated technologies through application to actual plants and consequently make a contribution to the realization of a society offering a higher level of safety and efficiency.

### Reference

(1) Kurotani, K., Todaka, Y. Current Status and Prospects of Measurement and Control Technologies. FUJI ELECTRIC REVIEW. 2012, vol.58, no.1, p.2-8.

Fuji Electric

*Innovating Energy Technology*