Fault Diagnosis for Large Complex Petrochemical Plant

sandip k lahiri
CHEMCON, 2008.
Fault Diagnosis for Large Complex Petrochemical Plant

S.K. Lahiri
Department of Chemical Engineering, National Institute of Technology, Durgapur, India
sk_lahiri@hotmail.com

Nadeem Khalfe
Jubail United Petrochemical Company, SABIC.
khalfenm@united.sabic.com

K.C. Ghanta
Department of Chemical Engineering, National Institute of Technology, Durgapur, India
kcghanta@yahoo.com

Keywords: principal component analysis, fault diagnosis

ABSTRACT

Often it is time consuming to monitor the plant condition in modern complex process industries as there is abundance of instrumentation that measure thousands of process variables in every few seconds. This has caused a "data overload" and due to the lack of appropriate analyses very little is currently being done to utilize this wealth of information. Fortunately, in process, groups of variables often moves together because more than one variable may be measuring the same driving principle governing the behavior of the process. Multivariate statistical methods such as Principal Component Analysis (PCA) are capable of compressing the information down into low dimensional spaces which retain most of the information. Using this method of statistical data compression a multivariate monitoring procedure has been developed to efficiently monitor the performance of large processes and to rapidly detect and identify important process changes. A graphical interface was developed which will facilitate to plot the real time plant condition by a single point. The location of this point will give the idea whether whole plant is running in normal or abnormal zone and diagnosis online any fault in the process. This software was implemented in commercial Ethylene glycol plant and successfully diagnoses process upsets and abnormalities. Two implemented case studies were presented (namely compressor monitoring and ethylene oxide reactor monitoring) where an early detection of fault at its incipient stage can avoid money loss and process upsets. This generalized intelligent software can be useful and extend to any other process plant also.
1. INTRODUCTION:
Monitoring performance and detecting faults is an integral part of the successful operation of any process. Performance can be monitored by comparing the actual results to the predictions from a mechanistic model, or by using statistical process control (SPC) charts e.g. Shewart charts [5] to compare the current state of the process against "normal operating conditions". The challenge currently facing process monitoring is the enormous amount of correlated data being collected from a multitude of sensors every few seconds. This "data overload" and the lack of appropriate analytical tools have meant that very little is being done to utilize this wealth of information. The biggest drawback to using a mechanistic approach is the need for a detailed model. Even if a detailed mechanistic model is available its parameters are uncertain, and often they need to be updated in real-time. Using the established statistical control charting methods has the advantage that they need no model, but rather uses the operational data directly. The major drawback to these control charts is that they were developed for monitoring univariate problems or sets of independent variables, and the expansion to handle the case of many highly correlated variables is difficult [7]. These methods are still being used on a one variable at a time basis on multivariate processes, either formally using Shewart charts or informally by operators monitoring key variables. Generally, this approach has been adequate, although extremely inefficient; however, if the variables are correlated this approach can lead to erroneous results [2].

This paper presents a different and more efficient approach to process monitoring based on multivariable statistical methods of Principal Component Analysis (PCA). These methods are particularly suited to analyzing large sets of correlated data. In the general problem process measurements are collected from the plant and arranged in a (n x k) matrix X consisting of n observations on k variables. The objectives of a multivariate SPC scheme would be to monitor the process and product quality using these observations and detect process upsets, equipment malfunctions or other special events. The final step in the procedure would be to find and remove the assignable causes for these events, thus improving process performance. Given the current process control computer systems, on-stream analyzers, and automated quality control labs, it is not uncommon to measure hundreds of process variables on line every few seconds or minutes, and tens of product variables every few minutes or hours. Although a large number of variables may be measured, they are almost never independent; rather, they are usually very highly correlated with one another. The true dimension of the space in which the process moves is almost always much lower than the number of measurements.

In some situations this is due to underlying fundamental relationships among the variables. For example, in the hypothetical reaction of $A + B \rightarrow C$ where $A$ and $B$ are fed to the reactor in a specified ratio, although the concentrations of $A$, $B$ and $C$ are being measured (3 dimensional measurement space), the actual problem is univariate (the stoichiometric relationship and the fixed feed ratio each eliminate a degree of freedom). In other situations the placement of the measurements and the nature of the process make the measurements highly correlated. Consider a distillation column where only three variables change independently, reflux, reboil and feed composition. Originally, measurements are being made of the temperature profile using every fourth tray temperature. If the number of measurements is increased to every tray temperature to
obtain a more detailed temperature profile, the dimension of the measured variable space has been greatly increased, but the actual dimension of the problem has not changed. As a final illustration consider another situation where measurements on many different variables are made, but the nature of the process and the disturbances are such that they only allow the variables to move is a much lower dimensional space. For example, in the manufacture of synthetic fibers it is not uncommon to measure more than ten quality variables such as denier, elongation under different loads, breaking strength, dye depths, etc. The physical meaning of these measurements guarantees that the process is only capable of making fibers with certain combinations of properties, and disturbances to the process will affect many of these variables in a highly correlated manner. For example, fibers with very small deniers (weight / unit length) cannot be made with very high breaking strengths, and disturbances which lead to a reduction in denier lead to a reduction in breaking strength. As discussed, fortunately, in data sets with many variables, groups of variables often move together because more than one variable may be measuring the same driving principle governing the behavior of the system. In many petrochemical systems there are only a few such driving forces. But an abundance of instrumentation allows us to measure dozens of system variables. When this happens, we can take advantage of this redundancy of information. We can simplify our problem by replacing a group of variables with a single new variable. Principal components analysis is a quantitatively rigorous method for achieving this simplification. The method generates a new set of variables, called principal components. Each principal component is a linear combination of the original variables. All the principal components are orthogonal to each other so there is no redundant information. The principal components as a whole form an orthogonal basis for the space of the data. Principal Component Analysis was originally developed by [4] and is a standard multivariate statistical method described in many textbooks e.g. [1, 3, 6].

2. PRINCIPAL COMPONENT ANALYSIS:

2.1 The Basic:
Principal Component Analysis (PCA) is a data transformation method which rotates data such that the principal axis of the data is in the direction of maximum variation (fig.1). We can view the rotated data on the new Principal Axes. The ‘coordinates’ of the data in this new ‘coordinate system’ are known as Principal Component Scores. These are essentially projections of the data onto the Principal Axes. The Principal Axes (Components) are essentially vectors in the original variable space. These ‘vectors’ are known as Principal Component Loadings.

2.2 Why Do We Need To Rotate The Data?
Although data set may contain many variables (i.e. “n” dimensions), variation is often limited to a few key directions. By re-orienting our coordinate system to align with these key (multivariable) directions we can often find that some of the directions perpendicular to this space contain very limited variation and can therefore be ignored. Hence, by re-orienting the way we view the data we can squeeze as much information as possible into the 3 key Principal directions.
2.3 How Do We Generate Principal Components?

The principal components represent the selection of a new coordination system obtained by rotating the original variables and projecting them into the reduced space defined by the first few principal components, where the data are described adequately and in a simpler and more meaningful way. The principal components are ordered such that the first one describes the largest amount of variation in the data, the second one the second largest amount of variation, and so on. With highly correlated variables, one usually finds that only a few principal components are needed to explain most of the significant variation in the data.

Principal Components should have the following properties:

• They should be orthogonal (i.e. independent)
• The first principal component should point in the direction of maximum variance in the data.

In order to guarantee the above properties the columns of the original data must be orthonormal (normalized).

We can normalize a data set $X$ by the formula:

$$X_s = X_d \cdot D$$

(1)

Where the “mean centered” matrix $X_d$ can be obtained by subtracting the respective mean from each element of the matrix; that is,

$$X_d = X - \bar{X}$$

$D$ and $S_{xi}$ are defined as follows:

$$D = \begin{bmatrix}
\frac{1}{S_{x_1}} & 0 & 0 \\
0 & \frac{1}{S_{x_2}} & 0 \\
0 & 0 & \frac{1}{S_{x_3}}
\end{bmatrix}$$
Standard deviation \( S_{x_i} = \sqrt{\text{variance}} = \sqrt{\frac{\sum (X_i - \bar{X}_i)^2}{m-1}} \)

It can be shown that for a normalized data matrix \( X_s \) (i.e. \( X_s^T X_s = 1 \)) then the eigenvector of \( X_s^T X_s \) (the covariance matrix) with largest eigenvalue will point in the direction of maximum variance in the data. In fact, the variance of the data along each eigenvector is equal to the eigenvalue of that eigenvector.

The proof also shows that eigenvectors are, by definition, orthogonal to each other. Hence the **eigenvectors of the covariance matrix** are the **Principal Components** of the data.

### 2.4 Steps to Calculating Principal Components

- Consider a normalized data matrix \( X \) (each column has zero mean and unit standard deviation) of size \( n \) rows and \( m \) columns (\( n \times m \))
  
  The covariance of \( X = X^T X \)

- Calculate all eigenvectors and eigenvalues of \( X^T X \)

- Place the eigenvectors in order of decreasing eigenvalue. This matrix is called the **Loadings Matrix** \( P \) (\( m \times m \))

- To calculate the Principal Component Scores (i.e. the projections of original normalized data onto this new eigenvector basis) multiply the normalized matrix \( X \) by the Loadings Matrix \( P \).

  Scores Matrix will be: \( T = XP \) (\( n \times m \))

The loading vectors \( P \) are orthonormal and provide the directions with maximum variability. The \( T \) scores from the different principal components are the coordinates for the objects in the reduced space. They are uncorrelated and therefore are measuring different underlying "latent structures" in the data. By plotting the scores of one principal component vs. another, one can easily see which of the objects have similarities in their measurements and form clusters, and which are isolated from the others and therefore are unusual objects or outliers. The power of PCA arises from the fact that it provides a simpler and more parsimonious description of the data covariance structure than the original data.

### 3. CASE STUDY:

#### 3.1 Monitoring a reciprocating reclaim compressor:

In commercial ethylene glycol plant reciprocating compressor was used to pressurize reclaim process gas from 0.1 barg to 20.2 barg pressure for recovery of ethylene. The reclaim compressor is a 3-stage reciprocating compressor with intermediate cooler and separator at each stage. The loader valves used in this compressor are prone to damage (usual life is one year only) if slight moisture carryover or condensation takes place from process gas. While compressor is in running condition, it is very difficult to identify the loader valves performance deterioration in spite of its effect seen on inter stage discharge temperature. If the abnormalities of loader valves were not detected at its incipient stage, the sudden compressor shutdown will take place and causes substantiate process
upsets. The purpose of using multivariate process monitoring system on this compressor is to
diagnosis the problems of loader valves at its incipient stage so that their maintenance can be
planned and emergency compressor trip and subsequent process upset can be avoided.

To develop a monitoring system based on PCA 2162 hrs of historical data was collected from
process computers for all the process parameters (total 29 parameters like suction, discharge and
inter stage temperature and pressure, inter stage separators level and 10 vibrations indication
positioned at different location of compressor) available in DCS related to this compressor. The
projections of hourly average data (after gross outliers were removed) into the plane defined by the
first three principal component are shown in figure 2, 3 and 4 for this reclaim compressor. The data
points appear to cluster into 2 distinct regions which correspond to different operating conditions.
Simply examining the individual plots of the process variables would be confusing, time consuming
and would not reveal such information. To help diagnose the reasons for these shifts in process
operation, one can interrogate the underlying multivariate model and display the process variable
contributions to these shifts. After detail investigation of the process data it is concluded that green
cluster corresponds to normal operation of the compressor and red cluster corresponds to loader
valve malfunction and subsequent performance deterioration of compressor. After detecting the
different cluster, all the data corresponds to normal operation (green cluster) are separated and PCA
was done to identify the range of normal operating zone in first three principal component planes.

Figure 5 gives a historical trend plot of 1st principal component with time (i.e. record number).
Marked increase of 1st principal component represents the transition from normal operating zone to
abnormal zone due to loader valve malfunction. A “Scree” plot (refer figure 6) is a Pareto plot of
the percent variability explained by each principal component. We can see that the first three
principal components explain roughly two thirds of the total variability in the data. Thus it can be
concluded that instead of monitoring 29 parameters of compressor , one can only monitor first three
principal component and able to get enough information about compressor is running normal or not.

After establishing the normal and abnormal operating zone from historic data, we make the plot 2,
3, 4 and 5 online in an Exaquantum interface. In the computer screen only one point is plotted for
every minute on real time basis and from the location of that point one can identify easily whether
compressor is in green zone (normal) or red zone (abnormal). The locus of the point can give idea
that compressor is moving slowly outside normal operating zone when loader valves start
deteriorating. Thus it is possible to diagnosis the fault in the system at its incipient stage.
3.2 Monitoring of EO Reactor:
Oxidation of ethylene to produce ethylene oxide (EO) is an important reaction in the petrochemical industry for synthesis of glycol. Commercially EO is produced in shell and tube type EO reactor by reacting oxygen and ethylene at high temperature and pressure in presence of silver based catalyst. The oxidation of ethylene involves a main reaction producing EO and an undesirable side reaction producing carbon dioxide (CO2).

There are four process parameters (namely oxygen and ethylene concentration in cycle gas at reactor inlet, reactor temperature and inhibitor flow) has major impact on EO reaction. If oxygen and ethylene concentration in cycle gas at reactor inlet is maintained low, the selectivity drops rapidly and affects the economics of the process. Reactor temperature is important to maintain production rate. Optimum quantity of inhibitor flow is required to maximize the selectivity and production. Overdosing of inhibitor reduces the production rate by inhibiting the reaction. Under
dosing of inhibitor reduces the selectivity and produce less EO and more CO2. Early detection of abnormalities of any of the above four parameters is essential for large scale commercial Ethylene Glycol plant to maintain production target. Any delay to diagnosis the fault causes major production loss and subsequently reduce profit. One year hourly historical data was collected for 60 variables (e.g. reactor temperature, pressure, cycle gas flow, inhibitor flow, reactor inlet and outlet concentration of all components in cycle gas including oxygen and ethylene) to calculate the PCA for reactor. These data includes the abnormal operation, namely low oxygen and ethylene concentration at reactor inlet, low reactor temperature, overdosing and under dosing of inhibitor. Abnormal zone for each of cases were identified in PCA. After identifying different abnormal and normal zones we run the program on online Exaquantum interface where all the 60 input process parameters were collected on real time basis and plotted as a single point on the diagram. The location of this point will give idea about type of abnormality (if any) and thus enable the DCS operator and engineers to quickly identify the fault in the process.

5. Conclusion:
In this paper a multivariate PCA procedure has been proposed for monitoring EO reactor in EG plant. Multivariate statistical procedures are used to reduce the dimensionality of these large and highly correlated data sets down to a few principal components which contain most of the information about the process behavior under normal operating conditions. One implemented case study was presented where an early detection of fault at its incipient stage can avoid money loss and process upsets. The main advantage of this system is that one has to monitor only few principal components to detect abnormality or fault. The potential of this new monitoring system is enormous and its application is being tried to develop whole EG plant monitoring system. This generalized intelligent software was made in-house and can be useful and extend to any other process plant also.

References