Proceedings of the 12th KACC, October 1997

통계적 분석 기법을 이용한 공정 운전 향상의 방법

Process Operation Improvement Methodology based on Statistical Data Analysis

Dae-Hee Hwang, Tae-Jin Ahn*, Chonghun Han
Automation Research Center and Dept. of Chemical Eng., POSTECH
Technology & Engineering Dept., POSCO*

Abstract
With dissemination of Distributed Control Systems(DCS), the huge amounts of process operation data could have been available and led to figure out process behaviors better on the statistical basis. Until now, the statistical modeling technology has been usually applied to process monitoring and fault diagnosis. However, it has been also thought that these process information, extracted from statistical analysis, might serve a great opportunity for process operation improvements and process improvements. This paper proposed a general methodology for process operation improvements including data analysis, backing up the result of analysis based on the methodology and the mapping physical phenomena to the Principal Components(PC) which is the most distinguished feature in the methodology from traditional statistical analyses.

The application of the proposed methodology to the Blast Furnace(BF) process has been presented for details. The BF process is one of the complicated processes, due to the highly nonlinear and correlated behaviors, and so the analysis of the process based on the mathematical modeling has been very difficult. So, the statistical analysis has come forward as a alternative way for the useful analysis. Using the proposed methodology, we could interpret the complicated process, the BF, better than any other mathematical methods and find the direction for process operation improvement. The direction of process operation improvement, in the BF case, is to increase the fludization and the permeability, while decreasing the effect of tapping operation. These guide directions, with those physical meanings, could save fuel cost and process operators' pressure for proper actions, the better set point changes, in addition to the assistance with the better knowledge of the process. Open to set point change, the BF has a variety of steady state modes. In usual, almost chemical processes are under the same situation with the BF in the point of multimode steady states. The proposed methodology focused on the application to the multimode steady state process such as the BF, consequently can be applied to any chemical processes set point changing whether operator intervened or not.

Keywords Process operation improvement, PCA, Data analysis

1. Introduction

Due to the increasing global competition among process industry, the productivity enhancement of process has become an essential task to survive and succeed in the world trade market. Process engineers have tried to find the ways to improve the productivity and reduce production cost. A variety of problem has not been possible to be overcome in any data analysis approaches based on mathematical modeling. Therefore, we propose a general process operation improvement methodology based on the statistical data analysis of operation data with multimode steady states.

The proposed methodology is illustrated using an application to the Blast Furnace in order to determine the desirable operation conditions for the increase of PCI without sacrificing the product quality and yield. The process operation improvement, the increase of the amount of PCI, leads to the production cost reduction due to saving the amount of expensive cokes.

As a result, we obtained a desirable operation condition saving process operators' pressure for the better operations and production cost. In addition, it is also, as a guideline, much easier to understand and more meaningful to operation engineers since the results were obtained based on the physical meanings mapped into PCs.

In conclusion, we presented other obtainable information from the analysis based on the methodology, such as process dynamics, process flow, and the possible contribution to other application fields such as fault diagnosis and design a statistical design. Finally, we also discussed process improvement such as discovering fouling and blocking. It might be extended to revamping and debottlenecking.

The proposed methodology serves a great opportunity for an alternate way of process analysis.

2. Process operation improvement methodology

Traditional statistical analysis[2-4] has usually not gone trough data analysis, so there should be uncertainty in the analysis. For an assurance, the proposed methodology introduces data collection, data selection, and data quality analysis procedures[1].

2.1 Data collection

Although much data, for a number of variables, became available with the installation of DCS, these data for all variables are not needed for modeling, rather than surplus variables can accumulate the same covariance structure and make the computation time longer[5]. So we must determine which variables are needed to explain the process fully enough, namely, enough to make a full covariance structure for the process, not to accumulate the covariance structure. It is called relevant variables search and can be accomplished by data reconciliation techniques such as matrix projection and graph theory.

Then, we need to consider for good data sampling interval, prefiltering, experimental duration, slow disturbance consideration, and the minimum negative impact on daily production so on[1].

Since sampling interval is strongly related to signal to noise ratio of data, it need to be determined prudently. It should be not too large not to describe high frequency dynamics and not too small to be sensitive to high frequency disturbances. In practice, it is usually chosen 10% of maximum value of either process time constant or process time delay. Prefiltering will be necessary, when high frequency disturbances are contaminating process data seriously and lead to low signal to noise ratio. The same effect to prefiltering can be achieved with a large sampling interval.

Experimental duration must cover long enough to have several operation modes including the desired operation range so that we can make a comparison. Also, if there were slow disturbances, it must be considered to get rid of this and we have to try to have as small negative effect of data collection on the daily operation as possible.

2.2 Data selection

The collected data might be contaminated by some causes during data acquisition. They must be removed before constructing statistical models. The main causes[5] for data contamination are follows.

- Regular maintenance(shutdown and start up)
- Process equipment, such as sensors or computers, malfunction
- Instability of process, due to disturbances or wrong operations
- Some events impacting on the process such as input belt malfunction

A normal operation data selection procedure is proposed as follows, based on what we done for this.

- Search for the major variables which represent the process state most
- Define the ranges of major variables for normal operation
- Search for the subranges which satisfy their range restriction for all major variables
- Steady state refinement in a steady state mode[5]
  Removal or replacement of physically unrealistic signal
  Removal or replacement of signals with physically unrealistic change
  Removal or replacement of signal contaminated by sensor calibration and equipment malfunction such as computer and sensor malfunction

Removal or replacement of signals, on statistical ground, outside the window of plus and minus 2 times the standard deviation

Removal of the remaining unstable operation data and the impacted data by some events

With any trial operations, they should be considered over.

2.3 Data quality analysis

Now, we need to confirm whether the measured input signals really have the sufficient degree of persistent exciting and the measured output signals have high signal to noise ratio actually[1]. The sufficient degree of persistent exiting for input signals implies the input signals excite the process sufficiently through our interested frequency range. High signal to noise ratio for output signals means the magnitude of the process output signal is much larger than that of disturbances such as process disturbances or measurement disturbances. So it is natural that the results of any analysis, dealing with these high signal to noise ratio data, have a strong certainty[1]. It is the purpose of these data treatments that acquire or check good data quality and get an assurance of the analysis based on the statistical modeling dealing with the process data.

2.3.1 Input signal

The forms of input signals are a important factor to be considered to figure out the degree of persistent exiting. The degree of persistent exciting can be known by means of the spectral density. The one point not zero represent the one degree of persistent exciting[1].

The amplitude is also considered important, because the lower amplitude leads to the lower signal to noise ratio[1].

2.3.2 Output signal

High signal to noise ratio[1] is the only thing to be considered important in output signals. Signal to noise ratio can be known by the variability[5] which is defined as the ratio of the standard deviation and the mean value of each variable over the whole period. Standard deviation is associated with noise and mean value with process signal.

2.4. Investigation of the most desirable operation range

The strategy of process operation improvement is to compare the characteristics of the most desirable operation mode with those of the rest of normal operation modes and to identify the difference between them and then to set it a direction for process improvement. What we have to do first is to set a criterion for a desirable range by using measured variables and search for the desired range, according to the criterion.

2.5 Division the whole dataset into several subsets

For comparison, we need to classify the operation modes. In usual, it would be better to classify the dataset into four subsets which comprise the most desirable, desirable, less desirable and least desirable one, considering the level of desirability of subranges, according to the criterion.

2.6 Statistical modeling(Principal Component Analysis modeling)

After dividing a dataset into several subranges, we construct a PCA model for each operation mode.

2.6.1 PCA introduction

PCA[2-4] is a multivariate Statistical Process Control(SPC) tool developed from the traditional univariate SPC tool, the Shewart chart. It has been a powerful analysis tool, specially for the complicated processes with highly correlated variables, by constructing a statistical linearized model called a PCA model. It has numerous applications such as process monitoring, fault diagnosis, data reconciliation, and controller design so on.

2.6.2 PCA theory

PCA[2-4] is to define new coordinates(PCs) orthogonal to one another from the largest variability one by one in a covariance structure, until a new coming coordinate has meaningless. It reduces variable dimension. The orthogonal coordinates are called Principal Components(PCs) or loading vectors and the coordinate values for observations score vectors(t), which are equivalent to projected values of observations on the PCs. A reduced hyper space spanned by the PCs is the very PCA model and there is always the model misfit, because the dimension has been reduced. Since the analysis is difficult in original dimension due to complexity and correlation, it is the strong power of PCA that makes the analysis possible in the reduced dimension.

For PCA modeling, a reference set(X) must be constructed with observations. Then, the orthogonal decomposition, along directions determined by the eigenvectors of the covariance matrix of X, is performed as follows.

$$X = t_1 p_1^T + t_2 p_2^T + \ldots + t_a p_a^T + E_n = \sum_{i=1}^{a} t_i p_i^T + E_n = T_a P_a^T + E_n$$

where t is score vector, p^T is loading vector and a is the reduced dimension.

2.6.3 PCA application

PCA has been applied to a lot of industrial area specially such as process monitoring and fault diagnosis[2-4]. This paper propose another application for process operation improvement, which includes the mapping of physical meanings, such as physical phenomena or physical properties, to PCs. It will be explained next in detail.

2.7 Mapping the physical meanings(phenomena) to PCs

Until now, it has been thought PCs have no physical meanings[3]. The only concepts used for the analysis in PCA are the distance(T^2 value) from the target value to new observation in the reduced dimension and the discrepancy(Squared Prediction Error) between the real value and the model value. If the PCs have physical meanings like original variables such as pressure or temperature, we can analyze a movement in score plot. The mapping physical meanings to PCs, if we can, must raise the analyzing power of PCA by interpreting the behavior in score plot with the mapped meanings, just as we do with original variables having physical meanings.

What we have done to map physical meanings to PCs is 2 step dimension reduction. First, we define major physical meanings such as physical phenomena or properties and then map original variables into the major physical meanings. This is the fist dimension reduction. We propose a procedure for the first dimension reduction as follows.

1. Define the major physical meanings that are considered important in the process or are thought to let the process have the large variation in the data space.
2. Map original variables to the physical meanings with the following methods.

i) Control system developed by expert system

If control system constructed by expert system are available, we may infer how the original variables relate to the major physical meanings by searching the rules within the control system; otherwise, we can generate some rules needed for the mapping by cluster analysis, sensitivity analysis or dimension analysis.

ii) Qualitative reasoning for process dynamics

Process engineer may be possible to reason physical relationships, based on basic physical knowledge such as thermodynamic relations, between the original variables and the physical meanings. More systemically, it can be done by a process centered technique in qualitative reasoning. For the more rigorous reasoning, we can apply qualitative simulation, a constraint centered technique.

iii) Simple mathematical model

This method can be said to give the most rigorous mapping because we can simulate the relationships with a simple model[6]. We usually use the simplest model enough to get a feel of a process. Rather than a qualitative model or qualitative constrains, it gives the more rigorous quantitative relations of a process, for based on a mathematical model.

After we obtained the relationship of mapping original variables into major physical meanings, we need to map the major physical meanings to PCs, which is the second dimension reduction. In this stage, we use loading contribution plots, obtained as a result of PCA, which represent the contribution of original variables to PCs, which means the relationships between original variables and PCs. Then, since we had already found the relationship between original variables and physical meanings, we can map physical meanings to PCs. Original variables play the role of bridge between physical meanings and PCs.

The PCs can be seen the pseudo independent physical meanings which is actually the linear combination of the major physical meanings. If the one physical meaning contributes to the first PC more dominantly than other physical meanings, we can use the dominant physical meaning, as if it were equivalent to the first PC, in every analysis such as fault detection, in addition to process operation improvement. We confirmed this fact from the BF case study.

2.8 Identification of the direction for process operation improvement

In the same way for the other PCs in the subrange, we can also identify the dominant physical meanings. Repeating the same procedure to the other subranges, we can make a table for the dominant physical meanings in each subrange. With comparison based on the results in table, we can set the direction for process operation improvement.

2.9 Validation

We need to validate the direction found for the process operation improvement. It can be done the score plot analysis in the subranges. If there are some observations distinguished from the other observations in the score plot in a subrange, the observation can be explained how they are different form others in terms of the dominant phenomena mapped into the PCs. The explanation will give us the hypothesized action we can take to improve the process after the observations. And the action really taken by process operators after the observation can be known in the action data. If they coincides with each other, it can concluded the direction is pretty much valid.

3. Overview of the Blast Furnace

A blast furnace[6] is a large reactor where many reactions occur in complicated ways, simultaneously with physical phase transitions, heat, and mass transfer. Iron ore is charged down from the reactor top and Pulverized Coals(PC) is charged up from the waist level near the bottom with oxygen. The higher level of PC injection(PCI) is, the more fuel will be saved. The study of increasing the PCI level, based on a variety of modeling techniques, has been conducted by numerous researchers for the past decade in order to save the fuel cost.

Although many mathematical models, such as thermal analysis model, operation diagram, kinetic model and control model, have been available and give a great amount of insight on the internal behavior of the BF and operation conditions to a extent, they are not accurate enough to determine optimal operation conditions when we want to change operating conditions such as the increase of Pulverized Coal Injection (PCI). This is due to the unknown parameters of the complex mathematical models, the unknown input information such as the quality change of ore and cokes, how iron ores and cokes are injected.

Hence, using the proposed methodology for process operation improvement based on statistical data analysis, we defied to find the answer of the unsolved problem on any mathematical analysis, the optimal operation condition for the increase of the PCI level.

5. Process operation improvement

5.1 Data collection

We have collected the last six month operation data from the fourth blast furnace of POSCO, a iron making company, at pohang in Korea. The measured variables are composed of 329 temperatures through the whole reactor, 31 pressures in shaft region, the upper part of the reactor, and 4 gas concentrations in the top. The relevant search for the appropriate variables among the measured variables has not been done, because any on line computations such as prediction are not involved in our application. Since a measured value was a average for 1 hour measurements, sampling interval was 1 hour. It was appropriate enough to include high frequency dynamics and not to be sensitive to high frequency disturbances. One hour averaging plays the same role with a prefiltering, a lowpass filtering, so the prefiltering has not been considered. For the duration of data acquisition, we collected operation data for six month when is long enough to include specially the desired operation, the 2150-2506 observations in this case, and a variety of operation modes.

5.2 Data selection procedure

We could set four major variables as K value, the blast pressure(V_b) action, PCI action, and O_2. The defined normal operation ranges for them are K<2.7, V_b action<30 Nm/min, PCI action<2 ton/hr, and constant level of O_2. Figure 1, 2, 3 show the results of normal operation selection procedure.

5.3 Data quality analysis

We checked the degree of persistent exiting in input data using the power spectral density for input data and how high the signal to noise ratio in output data using the variability. Figure 4 shows the power spectral density for the blast pressure(V_b), a typical input, in which we can see it is possible to develop a model for the BF using this input even if the BF is a infinite order process. So the quality of input data has turned out very good.

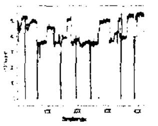
Figure 1 Original data(PCI)

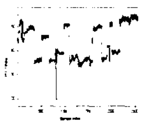
Figure 2 Reconstructed data

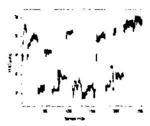
Figure 3 Refined data

Figure 5 also shows the variability of all outputs where we can see almost variables have higher than 10 and the average is 26. The fact the least signal to noise ratio is 3 and the average 26 means that signals have much larger magnitude and play the much more powerful role than the existing disturbances inside the BF. So the signal to noise ratio of output data is high enough to make it possible the useful data analysis.

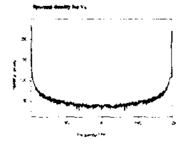
Figure 4 Spectral density of V_b

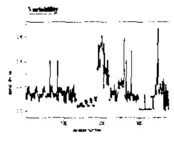
Figure 5 Variability for all variables

5.4 Investigation of the desired operation range

We set the criterion for the desired operation, low K value and large amount of PCI. K value is defined as a index including the pressure difference between the bottom and the top, so it is a index indicating the state of the gas flow within the BF and low K value means the operation is stable. Also the large amount of PCI means saving much fuel cost. The desired operation range, searched according to the criterion, is the operation range corresponding to 2150-2506 observations.

5.5 Division of the whole range into several subranges

For comparison, we divided the dataset into four subranges. The result of division is shown in table 1.

Range	Observations	Desirability	Level of PCI	Level of K value
1st range	1-760	less desirable	middle	middle
2nd range	761-1545	least desirable	low	low
3rd range	1546-2506	desirable	high	low
4th range	2150-2506	most desirable	high	low

Table 1 Division of the dataset into four subsets

5.6 Statistical modeling(PCA modeling)

We developed the PCA models for all subranges. Figure 6 shows the loading contribution plot for each subrange from which we can see the contribution of some original variables to the first PC are getting larger, while some variables contribute to the second PC less and less, as desirability is getting better with time. Also there are a lot of similarity between the PCs for the third range and those of the forth range.

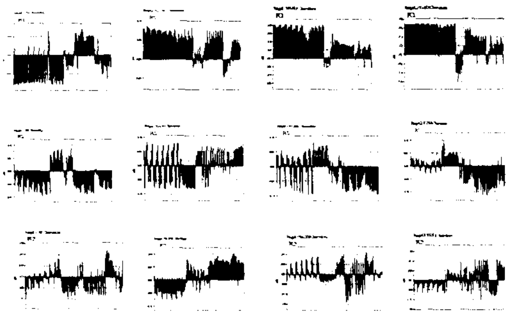
Figure 6 Loading contribution plot

5.7 Mapping the physical phenomena to PCs

We defined major physical phenomena as the fludization, the permeability and tapping operation, which are considered very important to infer the process situation by operator. For the first dimension reduction, we used the second proposed method using the experienced operators' qualitative reasoning. The fludization is the physical phenomena meaning the flow of pig iron in the bottom of the BF. If we change it by some actions, the temperatures in the bottom of the BF will be changed directly with time lags smaller than sampling interval It means the only temperatures in the bottom should be mapped into the fludization. Tapping operation is to tap out the melt pig iron through the holes around the reactor trunk near the bottom. The temperature variables existing in the vicinity of the tapping holes should be affected by tapping operation, so they should be mapped into tapping operation. The permeability means the state of gas flow within the BF. The temperatures and pressures in the upper part of the BF are strongly related to this physical phenomenon.

We have not thought over yet whether the physical phenomena are good or bad. It must be considered. So we have six major physical phenomenon, two for each one, when it is good or bad.

The second dimension reduction is to map the six major physical phenomena into the PCs obtained from the PCA modelings. With the loading contribution plot in Figure 6, we found the major phenomena by which the PCs are linearly combined. For example, let see the first PC of the first range. The variables(1-187), temperature in the bottom, contribute to the first PC and they were mapped into the good fludization, so the good fludization is one of the components taking part in the linear combination. Another component in the linear combination is the good permeability into which the variables(244-299) were mapped. Therefore, the first PC in the first range is a independent physical phenomenon which is a linear combination composed of two major physical phenomena, the good fludization and the good permeability. The same procedure has applied to all PCs and the table 2 shows the mapping results for all PCs with the subranges. The bold characters mean the dominant physical phenomena composing the independent pseudo physical phenomena and G and B imply 'Good' and 'Bad' respectively.

PCs	1st range	2nd range	3rd range	4th range
1	**F_G(-) P_G(+)**	**F_G(+) P_G(+)**	**F_G(+) P_G(+)**	**F_G(+) P_G(+)**
2	F_B(+) **T(-)** P_G(-)	T(-) P(+)	F_B(+) **P_G(-)**	F_B(+) **P(-)**
3	F_B(-) P_B(+)	T **P(+)**	P_B	P

Table 2 The mapping results

Table 2 shows that the dominant physical phenomenon of the first PC in all four range is the fludization and, when looking at Figure 7, the extent in dominance is getting clearer, as moving to the forth range. It means the fludization is the most important phenomenon explaining the variability of the data space best in all ranges and the importance of the fludization is getting larger as moving to the forth range. For the second PCs, tapping operation is certainly the dominant physical phenomena in the first range, but its contribution is coming to be faint and disappeared with time. It means taping operation is in the first range the important independent phenomena, next to the fludization, but its importance is getting weaker with time. The third PCs are mostly related to the permeability, but, in the forth range, the permeability come up to be the dominant phenomena mapped into the second PC, which means the importance of the permeability is increased, while tapping operation come to be meaningless there. Comparing these results, we can find how we can reach the desired operation range, the forth range.

5.8 Identification of the direction for process operation improvement

As a result of comparison in table 2, it can be concluded that we need to perform the operation increasing fliudization and permeability, while keeping the effect of tapping as low as possible, in order to increase the amount of PCI, the process operation improvement.

It is very worth that we found the priority from the fludization, tapping operation and to the permeability among the major physical phenomena. The fludization must be ensured first, tapping operation next and the permeability last in the order.

5.9 Validation

In the case of the first range, there is a group of observations 26 to 78 separated from the other groups, which means they have a different covariance structure from those of others, in Figure 7. They are separated by the negative scores of the first and second PCs. The driving force of the isolation in the first PC is the good fludization giving a major negative effect. For the second PC, the observations are affected by the good permeability giving a negative effect. As a result of the score plot analysis, the observation 26 to 78 are corresponding to the operation with the good fludization and good permeability. According to the found direction, we would increase the amount of PCI, because we confirmed the good fludization and the permeability. By the way, the amount of PCI after these observations was really increased by process operators. The hypothesized action based on the direction coincides with the action in the real operations. Accordingly, it can be concluded the direction is pretty much valid. For the other ranges, we performed the same validation processes and drew the same conclusion.

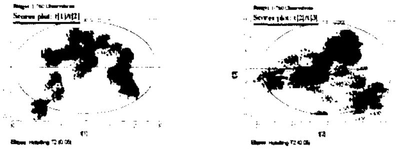

Figure 7 Score plot

6. Conclusion

For process operation improvement using the proposed methodology, we have done data analysis first to affirm data quality, discovered physical meanings of PCs, found a direction of process operation improvement, and then validate the direction. In the long run, what we have done was to present another PCA application with a new concept, the mapping of physical meanings to PCs.

The mapping results can help the analysis of the movement in the score plot. The driving force of the movement can be explained by physical phenomena mapped into PCs and then the situation of the process can be interpreted much more in detail. Furthermore, the analysis based on the physical meanings may help the interpretation of dynamics, stream flow pattern, fault diagnosis, and design a statistical controller. In the BF case, we figured out the bad fludization unstablized the process during the observations 284 to 446 and knew how the fluid flows in the bottom of the BF. And when we detected a fault and obtained for the contribution plot, if the variables are highly correlated, we could not find the cause of the fault, because too many variables are appeared in the contribution plot. The diagnosis based on the physical phenomena mapped into the PCs can present a solution about the cause of the fault in terms of physical meanings. For the design of a statistical controller, since it has been thought there are no physical meanings in the PCs, we have to solve a optimization problem with original variables to control the movement in the score plot. With physical meanings, however, we can design a more meaningful statistical controller with the variables of the physical meanings.

Also we might improve the process. But, in this time, we have to map something that can be considered important for the process improvement, such as Renolds number and schmidt number, so that we can recognize something wrong with the process, such as fouling and blocking. The analysis based on the mapping of those indexes into PCs might help achieve process improvement such as revamping or debottlenecking.

7. Acknowledgment

The authors would like to thank Korea Automatic Control Conference(KACC) for financial support through Automation Research Center at Pohang University of Science and Technology (POSTECH).

8. Reference

1. Soderstorm and Stoica, "System Identification," chap. 5 and 12, prentice hall, (1988)
2. Kourti, T. and J.F. MacGregor, "Process Analysis, Monitoring and Diagnosis using Multivariate Projection Method," Chemometrics and Intelligent Laboratory Systems, 28,3-21,(1996)
3. B.M. Wise, N.L. Ricker, D.F. Veltkamp and B.R. Kowalski, "A Theoretical Basis for the Use of Principal Component Analysis for Monitoring Multivariate Processes," Process Control and Quality, 1, pp. 41-51, (1990)
4. G. Chen and T. McAvoy, "Process Control Utilizing Data Based Multivariate Statistical Models," The Canadian J. of Chem. Eng., 74, Decem., (1996)
5. A. Santen, G.L.M. Koot and L.C. Zullo, "Statistical Data Analysis of a Chemical Plant," PSE'97-ESCAPE-7, Suppl., 21, pp. S1123-1128, (1997)
6. Y. Omori, "Blast Furnace Phenomena and Modeling," Elsevier Applied Science, London and New York, (1987)