

# A Systematic Decision Procedure to Identify Vital Few Causes of Variation in Large-scale Petrochemical Process

Minjin Kim<sup>†, ‡</sup>, Young-Hak Lee<sup>‡</sup>, and Chonghun Han<sup>‡, \*</sup>

<sup>†</sup> Department of Chemical Engineering, Pohang University of Science and Technology, San 31, Hyoja-dong, Nam-gu, Pohang, Gyeongbuk 790-784, South Korea

<sup>‡</sup> School of Chemical Engineering and Institute of Chemical Processes, Seoul National University, Shillim-dong, Kwanak-gu, Seoul 151-742, South Korea

**Abstract** We propose a systematic decision procedure which helps us identified vital few causes (VFCs) of variation in an industrial large-scale chemical process. The previous process improvement strategies require significant time of analysis when applied to a large-scale chemical process with complex causal relationships because they are just provide various tools or simple heuristics to identify VFCs. On the other hand, the conventional variable selection methods can not guarantee their performance to identify the reasons of the empirical correlation. The proposed stepwise decision procedure using systematic decision methods enables to reduce analysis time since the focus is systematically shifted to the significant causes for variation. Additionally, accuracy of the identified VFCs is improved using a causal analysis based on technical information as well as a correlation analysis based on a statistical hypothesis test. This procedure has been successfully applied to purified terephthalic acid (PTA) process to discover VFCs. The identified VFCs have been validated, and they have greatly contributed to improve the degraded process.

**Keywords** Vital Few Causes; Process Improvement; Systematic Procedure; Large-scale; Petrochemical Process

## 1. Introduction

In large-scale chemical process, the quality of the final product is not right, but the causes remain unknown, frequently. Variation in quality can be caused by various operational variations and disturbances, but the manufacturer needs VFCs that can be easily used at improvement of process. Generally, design of experiments (DOE) can be considered in order to evaluate the factors that affect the quality and to find optimal operating condition prior to improvement of process. DOE force the process to operate in certain predefined conditions according to a properly designed set of experiments. If we do not know vital causes and have too many candidates, we must accept a heavy cost of experiments and take a risk of obtaining a low-quality product or threaten the security of

---

\* To whom all correspondence should be addressed. Tel: +82-2-880-1887. Fax: +82-2-888-7295. E-mail: chhan@snu.ac.kr.

the process during experiments. Therefore, to consistently guarantee product quality regardless of operational variations and disturbances, it is important to identify VFCs that make major impact on the quality variation. The selected VFCs allow us to focus our efforts on removing their contributions to the defects of quality

The previous process improvement strategies, such as Six Sigma, which have been applied in practice and have proven to be successful<sup>1</sup> provide various supporting tools to identify VFCs: the qualitative ones such as fishbone diagrams, process mapping, and pareto charts and the quantitative ones such as control chart, ANOVA, multivariate chart, correlation study, and histogram.<sup>2-5</sup> The chemical process has many potential causes of variations and their causal relationships are complex due to recycle loop, buffering, and control system. These strategies are so limited to provide usable tools and/or simple heuristics for identifying vital causes<sup>6</sup> that significant amount of analysis time is required due to their ambiguous decision procedure based on 'intelligent guess' when they applied to the chemical process. It's successful identification of VFCs depends on the empirical knowledge of professionals.

There are many systematic solutions to the problem of variable selection, however most of methods may be applied to chemometric data and structure-activity relationship data rather than process data. Currently, Lazraq compare the effectiveness of several variable selection methods in the context of PLS regression, on several real datasets of chemical manufacturing processes.<sup>7,8</sup> These statistical methods are based on the "observed variability" of the process. In the analysis of observed data with these statistical methods, historical quantitative data are so distorted due to time delay, operational noise, and sensor error that it is difficult to guarantee accuracy of extracted variables. Furthermore, the variables significantly correlated with the final quality do not necessarily cause the variation of quality. It is impossible to identify the reasons of the observed correlation, unless empirical models are supplemented with technical information of the process.<sup>9</sup>

In this study, we newly propose a systematic decision procedure which helps us identify VFCs used for improving a large-scale chemical process. The proposed methodology allows us to handle the complexity of large-scale chemical process by deciding the potential causes and the VFCs, stepwise. At the first step, the potential causes are selected using systematic decision methods such as HOQ and 20:80 rule based on empirical knowledge without requiring a subjective decision by a professional. At the second step, VFCs are screened by analyzing causality and correlation based on technical information which enable VFCs to be robust for noise of observation. The proposed approaches have been validated through the industrial application to the PTA manufacturing process.

## **2. Proposed Methodology**

The proposed methodology consists of two steps such as identification of potential

causes and screening VFCs.

**2.1. Identify potential causes of variation.** Prior to identify VFCs, the potential causes for variation of a critical to quality (CTQ) which is the key to the quality improvement should be firstly identified to reduce candidates of VFCs. Generally in the chemical process, the several sub-processes have several tens of processing systems that affect the CTQ due to hundreds of thousands of variation sources of them such as production load, feed composition, catalyst, operating condition, recycle, equipment trouble, operator variation, and so on. The significant amount of time is needed to analyze their complex causality and correlation. Therefore it is not a good choice to identify the VFCs among all the candidates though all influence of them on the CTQ are not significant. Screening potential causes contributes to reduce analysis time and to improve an accuracy of analysis by handling complexity of the large-scale process.

First of all, brainstorming is performed to select important processing systems affect the CTQ, mainly. And then, to identify potential causes of variation, the qualitative data about possible sources, which may affect the CTQ, are collected using questionnaire based on the proposed standard candidates of variation causes. A good approach for the collection of all the possible causes is provided as 5M1E (Man, Materials, Manufacture, Machine, Measurements, and Environment) by Ashton.<sup>5</sup> However, the 5M1E give only categories on a higher level so that it is difficult to investigate all the specific sources in the large-scale chemical process. The proposed standard candidates helps the respondents to find out all the specific sources which can occur in the target process without missing ones. The questionnaire ask several professional engineers and operators to select one of the 4 levels of strong, medium, weak and unclear according to the correlations between CTQ and standard candidates in the selected major processing systems. This operational knowledge based analysis is understandable as well as acceptable to the field.

For the first time, HOQ is introduced to the proposed methodology in order to plays a role to qualify and order influences of causes on the CTQ using the large amount of qualitative information obtained from the questionnaire. Thereby we can systematically find out what are the potential causes for the variation of the CTQ. HOQ originated as a systematic technique for identifying those product features which contribute strongly to product quality, and where engineering effort is needed.<sup>11</sup> The matrix rows represent the sources of causes, and the columns the respondents. The central part of the matrix is used to show the relationships between the causes and the CTQ determined by each respondent. Each selected level is scored by exchanging from strong, medium, weak and unclear into 9, 3, 1 and -1, respectively. The right hand side of the central matrix is used to record the weight which represent relative importance. The weight is calculated according to the variation of opinions from respondents as like eq. 1. Finally, the related score is decided by multiplying the average score by the weight.

$$\bar{s}_i = \sum_{j=1}^n s_{ij} / n \quad (1)$$

$$w_i = \frac{1}{\sqrt{\sum_{j=1}^n (s_{ij} - \bar{s}_i)^2 + 1}} \quad (2)$$

$$s_i^r = \bar{s}_i w_i \quad (3)$$

where,  $n$  is the number of respondents,  $s_{ij}$  is the score of  $i$ -th cause from  $j$ -th respondent,  $\bar{s}_i$  is the average score of  $i$ -th cause,  $w_i$  is the weight of  $i$ -th cause, and  $s_i^r$  is the related score of  $i$ -th cause. The causes are ordered in an ascending series for the relative score using pareto chart. We select the possible sources whose cumulative score takes more than 80 percent of the summation of all the related scores. And then, the potential causes are specified from the possible sources through brainstorming. These potential causes can be classified into an operational variable that represent quantitative data and a conditional variable that indicate qualitative information. Some of the conditional variables might be the binding constraints that are on the edge of normal operational bound thereby cannot freely adjust.

**2.2. Screen vital few causes.** Prior to final selection of VFCs, the critical candidates are preliminarily screened based on causal analysis. The potential causes might be supposed to have cause and effect relationship within them. We connect CTQ and potential causes according to their causal relationship such that an arrow is linked from one cause to its origin for all cause as shown in Figure 1. It is called “causal line” that a series of potential causes linked by from the CTQ to a final node (a cause). For example, there are 4 causal lines such as C<sub>1</sub>-C<sub>2</sub>, C<sub>1</sub>-C<sub>3</sub>-C<sub>4</sub>, C<sub>5</sub>-C<sub>6</sub>-C<sub>7</sub>, and C<sub>5</sub>-C<sub>8</sub>-C<sub>9</sub> in Figure 1. Among them, the improvable causal lines that do not include any binding constraint are selected because binding constraint prevents intended implementation. After all, the nearest operational variable is preliminarily screened as critical candidate of VFC, for example, C<sub>2</sub> and C<sub>3</sub> in Figure 1. Other operational variables, which are in the same line, are usually used as a manipulated variable to control the nearest operational variable.

Final VFCs are selected based on statistical hypothesis test of correlation between the CTQ and the pre-screened VFC. The null hypothesis is eq. 4 and the test statistic is calculated by eq. 5. The null hypothesis is rejected when test statistic is in the critical region,  $|T| > t_{(\alpha/2; n-2)}$  with level of significance,  $\alpha$ .

$$H_0 : \rho = 0 \quad (4)$$

$$T = \frac{(\sqrt{n-2})r}{\sqrt{1-r^2}} \quad (5)$$

where,  $\rho$  and  $r$  is correlation coefficient and sample correlation coefficient. Pre-

screened VFCs that have statistically significant correlation with the CTQ are finally selected as VFCs.

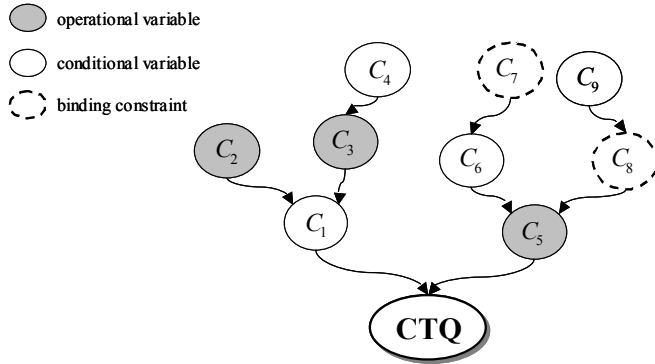


Figure 1. Relations diagram which represent causal relationship among the CTQ and the potential causes.

### 3. Case Study

**3.1. Process Description.** Purified terephthalic acid (PTA) is a monomer used to manufacture polymer such as polyethylene terephthalate (PET) and polyester which then is formed into films, textiles, bottles, and plastic molds. Since the demand for PTA has been growing steadily and the market is large, the manufacturing process is receiving increased attention. Several types of commercial processes each licensed by Amoco, Estman-Kodak, or Mitsubishi are being operated to manufacture PTA in the world<sup>10</sup>. Process flow diagram of this large-scale process consists of the major 6 sub-processes: oxidation, centrifuging, digestion, filtering and drying, catalyst purification, and solvent separation. In these sub-processes, there are about 40 equipments and about 3000 variables.

**3.2. Problem Define.** From customer surveys and market analysis, it is known that the 4-carboxybenzaldehyde (4-CBA) concentration contained in PTA product is the main quality that degrade quality of polymer product made from PTA product. In a TPA manufacturing process, PX is partially oxidized with air to PTA, and the 4-CBA is inevitably formed as an undesirable impurity during the oxidations. Since a great quantity of the 4-CBA cause unstable polymerization in down stream, the formation of the 4-CBA should be within the desired value in the operating range of interest so that quality of polymer can be maintained uniformly during a certain period of time. The target value of the 4-CBA concentration is 150 ppm and the target performance of the 4-CBA concentration has to be at least 3.5 sigma. The performance of the 4-CBA concentration in PTA is calculated using the historical 4-CBA concentration data during 1 month before implementation. The average of the 4-CBA concentration is about 155 ppm and its performance is 2.0 sigma. We can know that the average is quite larger than the target and its performance is much worse than the target. Quality of polymer product has degraded due to the poor

performance of the 4-CBA concentration and dissatisfaction of customers has grown about that gradually. Therefore our goal specification is to improve the 4-CBA up to the target performance and to maintain improved 4-CBA, consistently.

**3.3. Identify potential causes of variation.** Oxidizer (OXD), digester (DIG), and centrifuge (CF) are selected as major processing systems from brainstorming. The 4-CBA is formed during the reactions in oxidizer and digester and then it is reduced throughout separators such as centrifuge and filter. Therefore, the 4-CBA has strong relationship with 3 selected processing systems. The questionnaire has been done with total fourteen operators and engineers. HOQ assigns a total of 88 candidates within the major processing systems for the related score, respectively. Among them, a total of 21 candidates are selected as the possible sources which highly influence on the 4-CBA. Finally, a total of 25 potential causes for the variation of the 4-CBA are identified by specifying the selected candidates. They are classified into operational variables, conditional variables, and binding constraints.

**3.4. Screen vital few causes.** A total of 5 VFCs are preliminarily screened based on causal analysis. In order to see the gradual causality, the diagram has been focused in two kinds of effect, 4-CBA in CTA and 4-CBA in PTA. The causes of the significant amount of 4-CBA generated during oxidation are related to the 4-CBA in CTA. The other causes of following changes of the 4-CBA arising from separation and second oxidation are related to the 4-CBA in PTA. There are 11 lines, but only 6 lines do not be affected by binding constraints. We preliminarily select the operational variable located most close to the corresponding 4-CBA as VFC for each line. Catalyst into oxidizer is VFC for two lines. The pre-screened VFCs are catalyst into oxidizer, impurity into oxidizer, oxidizer temperature, catalyst into digester, and air into digester.

A total of 4 VFCs are finally screened based on statistical hypothesis test for correlation between the 4-CBA in PTA and the pre-screened VFCs. The data matrices are collected with the historical data measured during the last 1 month operation for the 11 process variables of the pre-screened VFCs and 1 quality variable of the 4-CBA. The data matrix is rearranged to resolve the problem not only that the quality variables are measured much less frequently than the process variables but also that each equipment have different time delay from resident time. And then statistical outliers that might be caused by measurement errors or abnormal operations were removed from the data set on the basis of principal component analysis<sup>12</sup>. The critical region,  $t_{(0.025;304)}$  of the hypothesis test is 1.96 when significance level,  $\alpha$  is 0.05 and the number of observations,  $n$  is 306. The final VFCs are catalyst into oxidizer, oxidizer temperature, catalyst into digester, and air into digester. They have the variables such as Br, Co, Mn and Ir concentration into oxidizer, oxidizer temperature, Br, Co and Mn concentration into digester, and air flowrate.

#### **4. Validation**

In order to validate the influence of identified VFCs upon the 4-CBA in PTA, a stabilization simulation of the PLS model for the 4-CBA in PTA is used to avoid the risk of real plant test. The variation in the 4-CBA is obtained at the assumed scenario in which the VFCs are respectively stabilized such that the standard deviations of these are halved. The PLS model is build with the operational variables nearest to the 4-CBA in all lines and it has explained variance of 70%. The inputs of the model are catalyst concentration into oxidizer, impurity, PX feed, air flowrate, oxidizer temperature, catalyst concentration into digester, CTA feed, and air flowrate under the assumptions that these variables can be independently and freely adjustable. Then, the effect of the stabilization of each VFC can be roughly estimated by comparing the variation of the 4-CBA before and after stabilizing the VFC. The standard deviation of the 4-CBA is decreases after the stabilization of each VFC. The reduction of variation is expected to greatly improve the sigma level of the 4-CBA from 2.0 to 3.7. Therefore, it is confirmed that all the identified VFCs significantly affect on the 4-CBA. Note that this simulation approach gives an useful appraisal for the stabilization of the VFCs even though the estimated variations reduced by the stabilization may differ from the actual values since the simulation results are obtained from the PLS models in which considerable amounts of the variations in product quality are treated as noises.

#### **5. Conclusions**

To consistently guarantee product quality regardless of various variations and disturbances, it is important to find out vital few causes that make major impact on the quality variable. Identification of these VFCs allows us to focus our efforts on their contributions to the variations in quality. However previous process improvement strategies are limited to just provide usable tools and simple heuristics to identify VFCs, they require significant time of analysis when applied to a large-scale continuous chemical process which includes complex causal relationships due to recycle, buffer, control system and time delay. On the other hand, the conventional variable selection methods are impossible to identify the reasons of the observed correlation, thereby we can not guarantee accuracy of results. A systematic decision procedure which helps us identified VFCs of variation in an industrial large-scale chemical process is proposed. It's stepwise zooming decision allows us to greatly reduce analysis time since the focus is systematically shifted to the significant factors for variation. At the first step, the potential causes of variation are systematically determined using house of quality (HOQ) based on empirical qualitative data. At the second step, accurate VFCs are screened using a causal analysis based on technical information as well as a correlation analysis based on a statistical hypothesis test. The systematic decision procedure has been successfully applied to PTA process to discover VFCs of quality variation. Three vital causes to CTQ

are systematically identified and the accuracy of identification has been statistically validated. We believe the proposed methodology can be widely applied to many large-scale chemical processes to identify vital few variables for CTQ variable.

### **Acknowledgements**

The authors would like to acknowledge the support for the fulfillment of this work by IMT2000 project (ID: 00015993) fund of Ministry of Information and Communication.

### **Literature Cited**

- (1) Hahn, G. J.; Doganaksoy, N. The Evolution of Six Sigma. *Quality Engineering*, **2000**, 12(3), 317-326.
- (2) Trivedi, Y.B. Applying 6 Sigma. *Chemical Engineering Progress*, **2002**, 98(7), 76-81.
- (3) Wheeler, J.M. Getting Started: Six-Sigma Control of Chemical Operations. *Chemical Engineering Progress*, **2002**, 98(6), 76-81.
- (4) Deshpande, P.B.; Makker, S.L.; Goldstein, M. Boost Competitiveness via Six Sigma. *Chemical Engineering Progress*, **1999**, 95(9), 65-70.
- (5) Ashton, H.C. A Statistical Quality-driven Approach to In-polymer Additive Analysis. *Angewandte Makromolekulare Chemie*, **1998**, 262, 9-23.
- (6) Mast, J.D. Quality Improvement from the Viewpoint of Statistical Method. *Quality and Reliability Engineering International*, **2003**, 19, 255-264.
- (7) Lazraq, A.; Cleroux, R; Gauchi, J.-P. Selecting both Latent and Explanatory Variables in the PLS1 Regression Model. *Chemometrics and Intelligent Laboratory Systems*, **2003**, 66(2), 117-126.
- (8) Gauchi, J.-P.; Changnon P. Comparison of Selection Methods of Explanatory Variables in PLS Regression with Application to Manufacturing Process Data. *Chemometrics and Intelligent Laboratory Systems*, **2001**, 58(2), 171-193.
- (9) Zarzo, M.; Ferrer, A. Batch Process Diagnosis: PLS with Variable Selection versus Block-wise PCR. *Chemometrics and Intelligent Laboratory Systems*, **2004**, in press.
- (10) Kroschwitz, J.I. Encyclopedia of Chemical Technology. *John Wiley & Sons*, New York, USA **1991**.
- (11) House of Quality reference
- (12) Wold, S.; Esbensen, K. and Geladi, P. Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, **1987**, 2, 37-52.