

주성분 분석을 이용한 효과적인 화학공정의 이상진단 모델 개발

박재연 · 이창준[†]

부경대학교 안전공학과

(2014. 1. 15. 접수 / 2014. 5. 1. 채택)

Principal Component Analysis Based Method for Effective Fault Diagnosis

Jae Yeon Park · Chang Jun Lee[†]

Department of Safety Engineering, Pukyong National University

(Received January 15, 2014 / Accepted May 1, 2014)

Abstract : In the field of fault diagnosis, the deviations from normal operating conditions are monitored to identify the type of faults and find their root causes. One of the most representative methods is the statistical approaches, due to a large amount of advantages. However, ambiguous diagnosis results can be generated according to fault magnitudes, even if the same fault occurs. To tackle this issue, this work proposes principal component analysis (PCA) based method with qualitative information. The PCA model is constructed under normal operation data and the residuals from faulty conditions are calculated. The significant changes of these residuals are recorded to make the information for identifying the types of fault. This model can be employed easily and the tasks for building are smaller than these of other common approaches. The efficacy of the proposed model is illustrated in Tennessee Eastman process.

Key Words : fault diagnosis, principal component analysis, tennessee eastman process

1. 서론

화학 산업은 고도의 정밀 기술이 요구되는 장치산업이 주를 이루고 있다. 21세기에 들어 에너지·화학 산업이라는 기존의 틀에서 벗어나 나노·무기소재, 정밀 유기화학, 전기화학 등 다양한 형태의 공정들이 개발되고 있으며 이에 따라 공정도 매우 복잡해지고 있다. 최적의 공정조건에서 원하고자 하는 최종 생성물을 얻기 위해서는 환경적 영향, 안전성, 생산량, 제품 질 등 여러 사항을 고려해야 한다. 또한, 생산량 혹은 제품 질의 저하를 야기할 수 있는 안전과 관련된 문제들도 매우 중요하게 다뤄져야 한다.

보다 안전성이 높은 공정을 확보하기 위해서는 이상의 원인을 조기에 감지하거나 미연에 방지하는 것이 매우 중요하다¹⁾. 여기서 말하는 이상은 물리적 고장, 교란, 작업자의 오류 등의 원인으로 공정의 운전조건에서 정상상태에서 벗어나는 경우를 의미한다²⁾. 예상치 못한 센서·제어기 고장, 파울링(fouling), 막힘(blockage), 누출(leaking), 원료의 변화, 유틸리티의 변

화 등이 이에 해당한다.

화학공정이 가지는 특징으로서 복잡성을 들 수 있다. 공정 안에 많은 순환 흐름과 제어 루프가 존재하기 때문에 이상을 감지하는데 많은 어려움이 있다. 화학공정의 이상을 진단하는 방법론은 공정지식을 얻는 방법에 따라 모델기반과 데이터기반 방법으로 나누어지며 각각의 지식 형태에 따라 정량적, 정성적 방법으로 나뉘진다¹⁻²⁾. 모델 기반 방법론은 공정을 에너지·물질 수지 식을 이용하여 모델링을 한 후 실제 데이터와 예측 결과를 비교하여 이상의 원인을 진단하는 방법이다. 하지만 이러한 방법은 공정을 모델링 하는데 많은 어려움 및 노력이 필요하다는 단점이 있다. 최근에는 이를 대체하면서 비교적 간단한 데이터 기반 방법론이 각광 받고 있다. 이러한 방법론들은 공정에 아무런 이상이 발생하지 않는 경우인 정상 상태 데이터를 이용하여 정상 상태의 고유 패턴을 설정한다. 그리고 발생 가능한 모든 이상에 대해 정상 상태의 고유 패턴과 구분이 되는 패턴을 설정하게 된다. 공정에서 실제 이상이 발생했을 경우 데이터의 패턴을 계산하여 기존의

[†] Corresponding Author : Chang Jun Lee, Tel : +82-51-629-6465, E-mail : changjunlee@pknu.ac.kr

Department of Safety Engineering, Pukyong National University, 45, Yongso-ro, Nam-gu, Busan 608-737, Korea

패턴과 비교함으로써 이상의 발생 유무 및 그 종류를 판별하게 된다. 이를 위해 PCA(principal component analysis), HMM(hidden Markov model) 등의 통계적 방법들을 활용하여 데이터의 패턴을 파악하여 이를 이상 진단에 활용하고 있다¹⁻³⁾. 하지만, 기존의 방법론들은 이상의 세기가 틀려짐에 따라 그 패턴 또한 같은 종류의 이상이라도 틀려진다는 단점이 있다¹⁾. 즉 기존의 통계를 이용한 방법론은 발생 가능한 모든 이상에 대한 패턴을 미리 계산하지만, 같은 종류의 이상이 발생할 경우라도 그 발생 세기에 따라 데이터의 특성이 틀려지게 되며 이에 따라 패턴 또한 틀려지게 된다. 이러한 특성으로 인하여 공정의 이상진단 시 오 진단을 하게 될 가능성이 매우 커지게 된다.

본 논문에서는 PCA를 이용하여 이상의 세기가 틀려짐에 따라 발생할 수 있는 문제점을 해결하는 새로운 방법론을 제시하고자 한다. 정량적인 방법과 정성적인 방법 두 가지를 혼용하여 이상의 세기가 변함에 따라 발생할 수 있는 통계적 방법의 단점을 극복하고자 한다.

먼저 2장에서는 PCA에 대해 소개하고, 3장에서는 PCA를 이용한 이상진단 방법을 사례연구를 통하여 제안하였다. 4장에서는 본 연구에서 제안한 모델을 이용한 진단 결과에 대해 설명하였으며 마지막으로 연구 결론에 대하여 서술하였다.

2. PCA (principal component analysis)

PCA(Principal Component Analysis)는 변수들의 선형 결합으로 표시되는 새로운 주성분(Principal Component (PC)) 을 찾아서 이를 이용하여 차수가 큰 원형 데이터를 적은 차수의 데이터로 데이터를 축약하는 방법이다. 이를 통하여 복잡한 데이터의 요약 및 용이한 해석에 활용한다¹⁾.

PCA의 기본원리는 원형 데이터들의 선형 조합을 통해 주성분으로 불리는 새로운 특성 변수(feature variable)들을 만들어 내는 것으로 이러한 특성 변수들은 분산(variance)이 최대가 되도록 설정이 된다³⁾. 이러한 과정에서 작은 분산 값을 갖는 주성분들을 제거하게 되면 결국 변수들 간의 상관관계에 대한 정보를 크게 잃지 않으면서 원형 변수들의 개수보다 작은 수의 특성변수를 구할 수 있게 되어 고차원 원형 변수 집합에서 저차원의 특성 변수 집합으로의 차원감소를 이룰 수 있게 되어 고차원을 다루는 어려운 문제를 해결할 수 있게 된다⁴⁾.

화학공정의 수많은 계측장치로부터 얻어지는 변수들의 측정값에는 측정오차를 비롯한 공정의 노이즈가

포함되어있는 것이 대부분이고 이러한 노이즈들은 공정 변수들의 실제 값에 비해 상대적으로 작은 분산을 갖는다고 볼 때 작은 분산 값을 갖는 특성변수 즉, PC들을 제거하는 것이 타당하며 이는 결국 노이즈 필터링의 기능을 부가적으로 수행한다고 볼 수 있다.

원형 데이터의 조합인 $X = X_1, X_2, \dots, X_p$ 의 주성분 $Y = Y_1, Y_2, \dots, Y_a$ 는 다음과 같은 식을 이루고 있다.

$$\begin{aligned} Y_1 &= l'_1 X = l_{11} X_1 + l_{12} X_2 + \dots + l_{1p} X_p \\ Y_2 &= l'_2 X = l_{21} X_1 + l_{22} X_2 + \dots + l_{2p} X_p \\ &\vdots \\ Y_a &= l'_p X = l_{p1} X_1 + l_{p2} X_2 + \dots + l_{pp} X_p \end{aligned} \quad (1)$$

여기서 $Var(Y_i) = l'_i \Sigma l_i = \lambda_i$ 이다. 첫 번째 주성분 Y_1 이 최대분산을 갖도록 계수 l'_1 를 유도한다. 그리고 두 번째 주성분 Y_2 는 Y_1 과는 독립이며, 최대분산을 갖도록 유도된다. 이러한 과정을 통해 Y_1, Y_2, \dots, Y_a 를 차례로 구한다. 이를 수식으로 표현하면 다음과 같다.

$$Y = L' X \quad (2)$$

여기서 L' 은 Loading 벡터이며 Y 는 score 벡터이다. X 는 차수가 p 이며 데이터 수가 n 개인 $p \times n$ 인 데이터이며 L' 은 $a \times p$ 인 데이터이다. 식 (2)를 통해서 Y 는 $a \times n$ 의 데이터를 갖게 된다.

Y 에 의해서 설명되는 원래 데이터의 정보의 정도는 다음의 식으로 표시된다.

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_a}{\lambda_1 + \lambda_2 + \dots + \lambda_p}, \quad (a < p) \quad (3)$$

만약 (식 3)의 값이 80~90%이상 된다면 전체 데이터내용의 대부분을 a 개의 주성분으로 설명할 수 있다는 결론을 내리게 된다. 그러므로 큰 정보의 손실이 없이 원래 데이터 차수보다 적은 a 개의 변수(주성분)로 대체가 가능하다⁴⁾.

본 연구에서는 정상 데이터를 이용하여 PCA를 이용하여 L' 을 계산한 후 이상이 발생했을 시 데이터를 계산하여 정상상태와 비교하고자 한다. 각각의 주성분을 비교함으로써 다양한 이상을 정확하게 진단하는 모델을 구현하고자 한다.

3. 사례연구

3.1. TE(Tennessee Eastman) Process

공정제어기법이나 최적화, 이상진단 등의 평가를 위

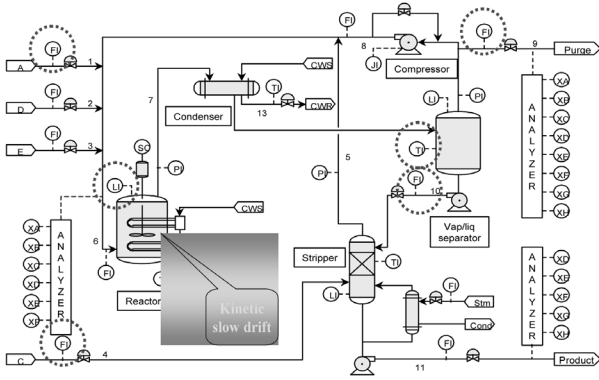
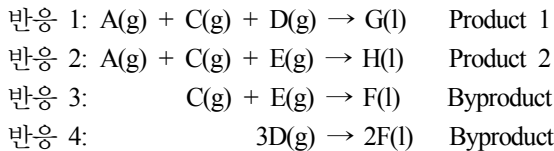


Fig. 1. Tennessee Eastman process.

하여 Downs와 Vogel(1993)⁵⁾에 의하여 제시된 것으로, 성분, 반응식 및 운전 조건들을 수정하였으나 각종 제어 기술을 실제 공정에 대하여 적용한 것과 동일한 효과를 나타내도록 Eastman 사의 TE 공정을 실제적으로 모사한 것이다. 이를 이용하여 공정에 대한 많은 이상진단, 제어, 최적화에 대한 연구가 수행되어 왔다¹⁻⁵⁾.

Fig. 1에서 보는 바와 같이 TE 공정은 반응기, 스트리퍼(stripper), 기액 분리기(vapor/liquid separator), 압축기 및 환류가 있는 공정으로 41개의 측정변수와 12개의 조작변수가 있으며 4가지 반응물 A, C, D, E로부터 2 가지 G, H가 생성된다. 반응기에서는 액상에 녹아있는 촉매에 의하여 다음과 같은 비가역 발열반응들이 일어난다⁵⁾.



발생하는 반응열을 제거하기 위하여 냉각수 코일이 반응기 내부에 설치되어 있으며, 생성물은 미반응물과 함께 증기 상태로 반응기 상부로 빠져 나오며, 응축기를 거치면서 응축이 되어 기액 분리기로 보내진다. 기액 분리기에서는 응축되지 않은 성분들을 분리하여 분리기 상부로 내보내며 압축기에 의해 반응기로 환류된다. 기액 분리기 상부로부터 나오는 비 응축 성분 중 일부는 외부로 방출되어 비활성 성분 B와 부산물 F를 공정으로부터 제거한다. 응축된 성분들은 기액 분리기 하부의 펌프에 의해 스트리퍼로 보내며 스트리퍼에서는 잔여 미 반응물을 4번 흐름으로 스트리핑하여 상부로 분리해낸다. 생산물은 스트리퍼의 하부를 통해 나오며 펌프를 통해 다른 공정으로 이송된다⁵⁾.

Down과 Vogel(1993)⁵⁾은 41개의 측정변수와 11개의 조작변수가 포함된 공정에 대하여 반응기로 공급되는

냉각수의 계단 변화에 따른 측정변수들의 동적응답에 대한 연구를 하였다. 공정 모델 수립 시 Down과 Vogel이 사용한 가정은 다음과 같다.

- ① 증기는 이상기체법칙을 따른다.
- ② 상들은 잘 섞여 있다.
- ③ 반응은 비가역 발열반응이며, 반응속도의 온도의존성은 Arrhenius 법칙을 따른다.
- ④ 모든 반응은 증기 상에서 일어나며, 반응속도는 반응기내의 증기의 부피에 비례한다.
- ⑤ 반응속도는 각 반응물의 분압에 비례한다.
- ⑥ A, B, C 성분은 모든 액상에 녹지 않는다.
- ⑦ 증기와 액상이 같이 존재하면 평형상태라고 가정한다.
- ⑧ 1, 2, 3번 피드(feed)는 순수한 A, D, E가 공급되며, 4번 피드는 A, C 및 소량의 B가 섞여있으며 이들의 조성도 바뀐다.

데이터로 사용한 15개의 데이터 Set은 52개의 변수로 이루어져 있다. 11개는 조작변수이며 측정변수는 22개, 조성변수는 19개이다. 측정변수는 1분마다, 조성변수는 6분내지 9분마다 한 번씩 측정된 것이다. 이 공정의 시뮬레이션 코드는 포트란이며 모든 공정 변수들은 가우시안 noise를 포함하고 있다. Table 1에서 보는 바와 같이 총 15개의 이상 데이터를 이용하여 본 연구를 수행하였다.

3.2. PCA 분석

먼저, 이상이 발생하지 않은 정상 상태의 데이터를 이용하여 PCA 모델을 계산하였다. Loading vector인 L' 을 계산한 후 이상이 발생했을 시 데이터를 계산하여 주성분 행렬 Y 를 계산하였다. 그리고 정상상태의 주성분 행렬 데이터와 이상이 발생했을 때의 주성분 데이터 행렬의 차이인 잔차를 이용하여 각 잔차가 주성분마다 어떻게 변화하는지를 조사하였다. 잔차는 다음과 같은 식으로 정의할 수 있다.

$$r = Y' - Y \quad (4)$$

각 주성분 마다 경향을 정성적으로 비교하여 각 이상별로 주성분의 어떠한 형태로 변동하는 지 조사하였다. 일반적으로 PCA를 이용하여 모델을 만들고 이를 이용하여 이상을 판별할 때 Hotelling's T-squared distribution 모델이나 Q-statistic 방법을 이용하여 정상상태의 데이터와 이상 상태의 데이터를 비교한다⁶⁾. 하지만, 같은 종류의 이상이라도 이상 발생 정도가 틀리면 이러한 방법

Table 1. The descriptions of fault types

Fault	Description
F1	A/C Feed Ratio, B Composition Constant (Stream 4)
F2	B Composition, A/C Ratio Constant (Stream 4)
F3	D Feed Temperature (Stream 2)
F4	Reactor Cooling Water Inlet Temperature
F5	Condenser Cooling Water Inlet Temperature
F6	A Feed Loss (Stream 1)
F7	C Header Pressure Loss-Reduced Availability (Stream 4)
F8	A, B, C Feed Composition (Stream 4)
F9	D Feed Temperature (Stream 2)
F10	C Feed Temperature (Stream 4)
F11	Reactor Cooling Water Inlet Temperature
F12	Condenser Cooling Water Inlet Temperature
F13	Reaction Kinetics
F14	Reactor Cooling Water Valve
F15	Condenser Cooling Water Valve

들은 다른 크기의 값들을 산정하기 때문에 이상 판별에 큰 혼란을 가져올 수 있다. 따라서 본 연구에서는 주성분 행렬들의 잔차의 변화 경향을 이용하여 이를 이상 판별에 활용 하였다. 이상이 확산됨에 따라 잔차의 경향도 복잡해지기 때문에 초기에 잔차가 어떻게 변하는지를 조사하였다. Table 2는 1번 이상부터 8번 이상까지 주성분에 따라 그래프 유형이 초기에 어떻게 변동하는지를 보여주고 있다. 이러한 각 주성분의 변화를 활용

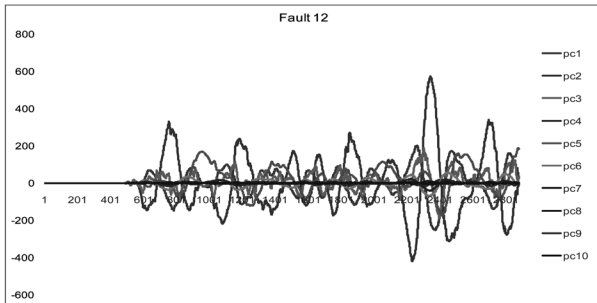


Fig. 2. The changes of principal components for Fault 12.

Table 2. The changes of PCs for assumed faults (fault 1-8)

Fault	pc1	pc2	pc3	pc4	pc5	pc6	pc7	pc8	pc9	pc10
F1	-/+	-/+	-/-	+/+	+/-	-/-	-/-	-/-	-/+	-/+
F2	-/-	+/+	-/+	+/-	+/+	-/+	-/-	-/+	-/-	-/-
F3	-/+	-/+	-/+	+/-	+/-	-/+	-/+	+/-	-/+	-/-
F4	-/+	-/+	-/+	+/-	-/+	-/+	+/-	+/-	+/-	-/-
F5	+/-	+/-	+/-	-/+	-/-	-/+	+/+	+/-	-/+	-/-
F6	-/+	+/+	-/+	+/-	+/-	+/+	+/+	+/+	-/-	-/+
F7	+/-	+/-	+/-	-/+	-/+	+/-	+/-	-/+	+/-	+/-
F8	+/-	+/-	+/-	-/+	-/+	+/-	+/-	+/-	+/-	+/-

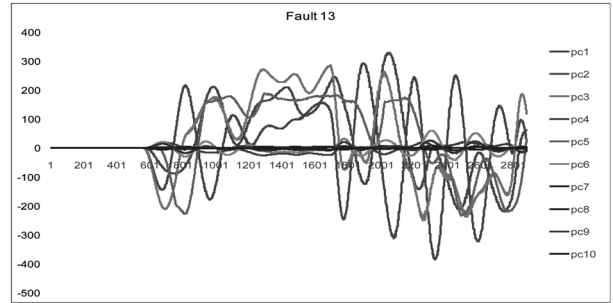


Fig. 3. The changes of principal components for Fault 13.

하면 이상을 효과적으로 판별할 수 있다.

Fig. 2와 3은 12, 13번째 이상이 발생하는 경우 잔차의 변화가 어떤 형태로 변화하는지를 보여주고 있다. 여기서 X축은 시간을 의미하며 Y축은 상대적인 주성분의 크기를 나타내며 단위는 무차원이다.

4. 진단 결과

모든 이상의 초기 변화를 이용하여 Fig. 4와 같은 이상 판별 알고리즘을 도출하였다. 이를 이용하면 TE공

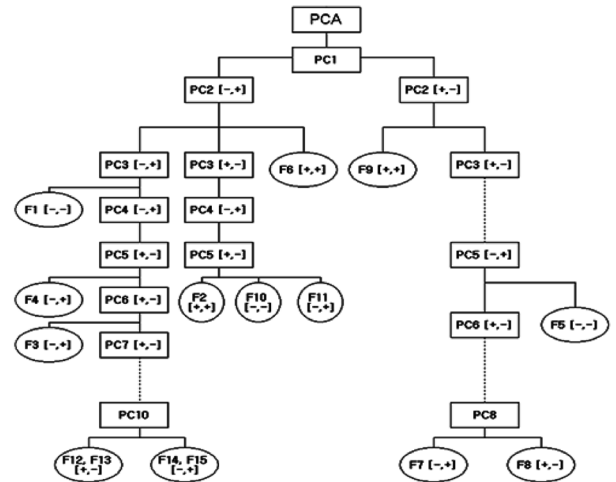


Fig. 4. The algorithm of fault diagnosis based on PCA for TE process.

Table 3. The results of fault diagnosis with unknown data

Unknown data	Fault type
Data1	F1
Data2	F13
Data3	F10
Data4	F5
Data5	F12
Data6	F2

정에서 발생하는 모든 이상에 대한 판별이 가능하다. Fig. 4에서 보는 바와 같이 주성분의 초기 잔차의 변화를 이용하면 이상의 판별이 가능하다. 최소 10개의 주성분을 이용해야 모든 이상의 판별이 가능하다. 원래 데이터의 변수 수가 52개임을 고려하면 적은 차수의 데이터만 이용해도 충분히 가정된 이상의 분류가 가능함을 보여준다. 이는 PCA에서 주성분이 공정의 중요한 변수를 포함하는데 따른 것이다. Table 3은 이상의 원인이 무엇인지 모르는 데이터를 이용하여 판별한 결과를 보여주고 있다. 6개 데이터 모두 정확하게 이상 진단을 수행함을 확인할 수 있다.

5. 결론

본 연구에서는 화학 공정상에서 발생할 수 있는 이상을 효과적으로 진단하기 위해 정량적인 기법인 PCA를 이용하였다. 많은 변수가 존재하는 복잡한 화학 공정의 특성 상 많은 변수들을 모니터링 하는 것은 불가능하다. 따라서, PCA를 이용하여 정보의 손실이 최대한 적도록 하면서 데이터 차수를 효과적으로 줄일 수 있으며, 정상 데이터와 이상 데이터 간의 차이를 생성한 후 각 주성분의 변화를 정성적인 형태로 관찰하여 이상을 판별하는 알고리즘을 제안하였다.

기존의 통계를 이용한 진단 기법들은 같은 종류의 이상이 발생하더라도 그 발생 세기에 따라 다른 값을 가질 수 있으며 이는 이상진단에 큰 문제를 일으킬 가능성을 가지고 있다. 본 연구에서 제안한 정량적인 기법과 정성적인 방법을 혼용하면 이상의 세기가 틀려지더라도 효과적으로 이상을 판별할 수 있다. 본 연구의 경우 15개의 데이터에 대해 진단을 수행하는 알고리즘을 구현하였으며 모든 이상의 판별이 가능하였다. 6개의 이상의 원인 및 그 세기를 모르는 임의의 데이터를 이용하여 테스트를 하였으며 모두 정확하게 판별함을 확인할 수 있었다.

하지만 본 논문에서 제안한 모델은 사전에 알고 있는 이상의 정성적인 패턴을 이용하여 진단하기 때문에 새롭게 발생한 이상의 원인에 대해서는 진단이 불가능하다. 따라서 앞으로 다양한 기법들을 결합하여 보다 더 효과적인 이상을 진단할 수 있는 방법론에 대한 연구가 필요할 것이며 이를 통해 공정의 전반적인 능력 향상을 도모할 필요가 있다.

감사의 글: 이 논문은 2013학년도 부경대학교의 지원을 받아 수행된 연구임 (PK-2013-C-D-2013-0478).

References

- 1) C. J. Lee, G. Lee and J. M. Lee, "A Fault Magnitude Based Strategy for Effective Fault Classification", *Chemical Engineering Research and Design*, Vol. 91, No. 3, pp. 530-541, 2012.
- 2) C. J. Lee, S. O. Song and E. S. Yoon, "The Monitoring of Chemical Process using The Support Vector Machine", *Korean Chemical Engineering Research*, Vol. 42, No. 5, pp. 538-544, 2004.
- 3) M. Kano, S. Hasebe, I. Hashimoto and H. Ohno, "A New Multivariate Statistical Process Monitoring Method using Principal Component Analysis", *Computers and Chemical Engineering*, Vol. 25, pp. 1103-1113, 2001.
- 4) L. H. Chiang, E. L. Russell and R. D. Braatz, "Fault Detection and Diagnosis in Industrial Systems", Springer-Verlag, New York, 2001.
- 5) J. J. Downs and E. F. Vogel, "A Plant-Wide Industrial Process Control", *Computers and Chemical Engineering*, Vol. 17, No. 3, pp. 245-255, 1993.
- 6) V. Venkatasubramanian, R. Rengaswamy, K. Yin S. N. Kavuri, "A Review of Process Fault Detection and Diagnosis Part III: Process History Based Methods", *Computers and Chemical Engineering*, Vol. 27, No. 3, pp. 327-346, 2003.