

유틸리티 플랜트에 대한 최적 총계 오차 판별 방법의 선택

김민수 · 한중훈^{†*}

포항공과대학교 화학공학과, *아이시스텍(주)
(1999년 6월 29일 접수, 2000년 4월 6일 채택)

Selection of Optimal Gross Error Detection Method for Utility Plant

Minsu Kim and Chonghun Han^{†*}

Dept. of Chem. Eng., POSTECH and *ISYSSTECH, Inc.
(Received 29 June 1999; accepted 6 April 2000)

요 약

공정에서 여러 장치에 의해 측정된 데이터들은 서로 일관성을 갖기 힘들며 물질, 에너지 수지를 만족시키지 않는다. 이러한 현상을 극복하여 모델에 일치하는 값을 얻기 위해 데이터 보정이 쓰인다. 측정 장치의 잘못된 영점 조정, 라인의 누출, 기기의 오동작 등에 의한 총계 오차는 데이터 보정 방법을 쓰기 위해서는 반드시 제거되어야 한다. 이 논문에서는 현재까지의 주요 총계 오차 판별 방법들을 유틸리티 플랜트에서 발생할 수 있는 여러 가지 경우에 대해 적용하여 최적 방법을 선별하고 이 과정을 통하여 각 테스트들의 장단점을 파악하였다. 사용한 방법은 Chi Square 테스트, 측정 값 테스트, 제약 조건 테스트, Maximum Power 테스트, Principal Component 테스트이며 각 테스트별 최적 공정을 제시하였다.

Abstract – The measured data in a utility plant have little consistency in both mass and energy balances due to random and gross errors, *Data Reconciliation* tries to resolve this inconsistency in measured data. *Gross Errors* that include leak, miscalibration of measurement instrumentation, and sensor malfunction should be detected and eliminated from the measured data. Several methods for gross error detection such as Chi-square test, Measurement test, Constraint test, Maximum power test, and Principal component test are compared and analyzed based on several cases in a utility plant. This article summarized the distinguishing features in every tests and proposed optimal processes for each tests.

Key words: Data Reconciliation, Gross Error Detection

1. 서 론

석유 화학 공장은 보일러와 반응기, 증류탑, 터빈 등의 주요 구성 단위들이 복잡한 네트워크로 연결되어 있다. 센서로부터 얻어지는 flow나 온도 그리고 각 성분의 농도의 측정 값은 정상 상태에 있어서 일반적으로 물질 수지나 에너지 수지를 만족하지 않는데 이는 데이터 내에 무작위 오차(random error)나 총계 오차(gross error)가 존재하기 때문이다. 총계 오차는 잘못된 영점 조정, 측정 기계의 오동작, 예기치 못한 누출로 인하여 발생하며 반드시 제거되어야 한다.

재정적인 부담이나 기술적인 한계로 인하여 모든 변수에 대한 측정 값을 구하는 것은 힘들다. 그러므로 측정 값은 조절하고 비측정 값은 계산해 내어 물질, 에너지 수지 제약 조건을 맞추는데 이를 데이터 보정이라고 한다. 총계 오차의 존재는 잔차의 최소자승의 합을 구하는 데이터 보정의 통계적인 근간을 변화시킨다. 따라서 총계 오차의 존재를 판별해 내어 측정값에서 제거해야 하는데 최근 수 십년 동안 이러한 총계 오차를 찾는 통계적 테스트들이 계속 발전되어 왔다.

Reilly와 Carpani[1]에 의해서 데이터 보정에 처음 쓰인 방법은 Chi-

square테스트인데 데이터 보정에서는 수학적 모델에서의 목적함수의 최적 값과 테이블화된 chi-square 값을 비교한다. 그들은 또한 공정 제약 조건의 잔차를 가지고 테스트를 하는 제약 조건 테스트를 제안하였다. 측정값 테스트는 측정 값 각각에 대한 조정 값을 대상으로 하는데 Mah와 Tamhane, Crowe 등[2]이 제안하였다. Almasy와 Sztanof[3]는 측정 값에 단지 하나의 총계 오차가 있을 때 가장 뛰어난 능력을 가지는 Maximum Power(MP) 테스트를 제안하였고 Crowe[4]는 MP 제약 조건 테스트를 제안하였다. Tong과 Crowe[5]는 상관 관계가 있는 변수를 상관 관계가 없는 변수로 변환시키는 방법인 Principal component analysis (PCA)를 이용하여 총계오차를 판별하는 Principal component(PC) 테스트를 제안하였다.

이 논문에서는 현재까지 제시된 주요 총계 오차 판별 방법들을 유틸리티 공장에서 발생할 수 있는 여러 경우에 적용하여 각 경우에 따른 최적 테스트를 제시하고 각각의 장단점을 파악하고자 하였다. 사용한 방법들은 Chi-square 테스트, 측정값 테스트, 제약 조건 테스트, MP 제약 조건 테스트, MP flow 테스트, PC 제약조건 테스트, PC flow 테스트인데 총계 오차의 개수와 크기에 따라 테스트마다 다른 판별 특성을 보인다. 이에 따라 각 테스트별로 성능을 가장 잘 발휘할 수 있는 공정을 제시하였다.

[†] E-mail: chan@postech.ac.kr

2. 이론적 배경

Flow, 온도, 성분의 농도의 측정 값들은 오차를 필연적으로 포함하고 있기 때문에 물질, 에너지 수지를 만족시키지 못한다. 정상 상태에서의 데이터 보정은 측정 값들의 최적 조정 값을 찾아 물질수지나 에너지 수지에 해당하는 제약 조건을 만족시키는 값을 구하는 것이 목적이다. 결국 선형 제약 조건에 대한 잔차의 가중 최소 자승(Weighted Least-square)을 최소화하는 문제가 되며 여기서 쓰이는 가장 행렬은 주로 측정 값의 분산-공분산 행렬이 된다.

Crowe 등[6]은 행렬 투영법(matrix projection)을 이용하여 정상상태 공정에서의 비측정 변수를 제거하고 측정 변수로만 구성되어 축약되고 중복된 제약 조건을 얻어내었다. 이 논문에서는 행렬 투영법을 거쳐 변환된 제약조건을 다루었다.

일반적으로 데이터 보정 문제의 목적 함수는 식 (1), (2)와 같이 나타낼 수 있다.

$$\text{Min}_x (x - \hat{x})^T \Sigma^{-1} (x - \hat{x}) \quad (1)$$

$$\text{subject to } Bx = 0 \quad (2)$$

\hat{x} 는 측정 값, x 는 보정 값, 그리고 Σ 는 측정 값의 분산-공분산 행렬이다. B 는 수식 행렬로써 행은 node를 나타내고 열은 flow에 해당한다. Node는 입력이면 1, 출력이면 -1, 아니면 0인 원소로 구성된다. Lagrange multiplier방법을 이용하여 답을 구하면 식 (3)과 같다.

$$\hat{x} = \bar{x} - \Sigma B^T (B \Sigma B^T)^{-1} B \bar{x} \quad (3)$$

만일 Σ 대신에 다른 행렬을 사용한다면 답의 Σ 자리에도 다른 행렬이 위치한다.

모든 측정 값에는 오차가 포함되어 있으므로 보존 법칙이나 제약 조건을 만족하지 않게 된다. 따라서 선형 정상 상태의 공정에서 축약된 제약조건의 잔차를 e 라고 할 때 식 (4)와 같이 나타낼 수 있다.

$$e = B \bar{x} \quad (4)$$

e 는 총계 오차 분석에 있어서 기본적인 벡터이다.

측정 변수 \bar{x} 는 기대 값이 x 이고 분산행렬 Σ 인 분포를 따른다.

$$\bar{x} \sim (x, \Sigma) \quad \text{such that } Bx = 0 \quad (5)$$

이때 잔차 e 는 평균이 0이고 공분산이 H 인 분포를 따르게 된다.

$$e \sim (0, H) \quad (6)$$

$$H = B \Sigma B^T \quad (7)$$

H 에는 B 의 형태로 공정 구조에 대한 정보가 들어있다.

데이터 보정을 수행하기 전에 비측정 변수가 제약 조건 내의 다른 측정 변수로부터 유일하게 결정되는지 여부를 아는 것은 중요하다. 공정을 디자인하는 사람은 각각의 변수가 측정되거나 유일하게 결정되도록 측정 장치를 적절한 위치에 달아야 한다.

Redundant한 측정 변수는 실제로 측정을 하는 변수지만, 만약 측정을 하지 않았다면 observable한 변수이기 때문에 소거하더라도 대상 공정의 observability에 영향을 주지 않는 변수이다. 반면 Non-redundant 변수는 다른 측정 변수들에 의해 유일하게 결정될 수 없으므로 측정된 값 자체가 그대로 사용되어 데이터 보정 효과를 살릴 수 없다[7].

비측정 변수 값들 중 다른 측정 값들로부터 유일한 값으로 결정될 수 있는 변수 값은 observable하다고 하고 자유도가 높아 유일하게 결정되지 않는 변수 값은 unobservable하다고 한다. Unobservable한 값들이 있다면 데이터 보정시 0의 값으로 설정이 전체 수지에 문제를 일으킬 수 있으므로 반드시 제거시켜야 한다.

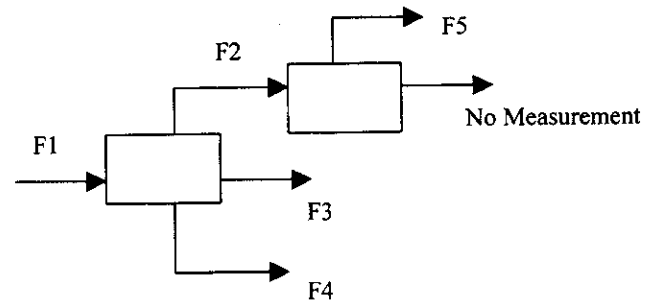


Fig. 1. Redundancy & Non-Redundancy.

(F1, F2, F3, F4: redundant flow; F5: non-redundant flow)

가령 Fig. 1과 같은 공정을 생각해 본다면 F1, F2, F3, F4는 Redundant한 stream이므로 데이터 보정으로부터 $\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4$ 가 얻어지고 F5는 Non-Redundant한 stream이므로 $\hat{x}_5 = y_5$ 가 되므로 유량계기가 없는 stream의 flow는 $\hat{x}_2 - y_5$ 로 계산된다.

3. 최적 총계 오차 판별 방법의 선택

3-1. 다양한 총계 오차 판별 방법들의 해석

정상상태에서 잘못된 영점 조정, 측정 기계의 오동작, 예기치 못한 누출로 인하여 발생하는 총계 오차를 판별해 내는 문제는 많이 연구되어져 왔다. 현재까지 여러 가지 통계적인 방법이 제시되어 왔는데 대표적인 방법들을 소개하면 다음과 같다.

3-1-1. Chi-square 테스트

Reilly와 Carpani[1]는 총체적 측정법으로 Chi-square test를 이용하였다. 이 방법은 모든 잔차들을 한번에 고려한 것으로 m 개의 자유도에 대하여 식 (8)과 같이 나타낼 수 있다.

$$\chi_m^2 = e^T H_e^{-1} e \quad (8)$$

e 는 잔차 벡터이고 H_e 는 잔차의 공분산이다. 잔차의 제곱 합(residual square sum)의 값들이 원하는 범위 수준의 Chi-square의 분포를 따른다는 가정을 이용하는 방법으로 값이 유의 수준의 Chi-square 테스트 값보다 큰 경우 총계 오차가 존재한다고 판단한다. 이 방법은 총계 오차의 유무만을 확인할 뿐 오차의 위치에 대한 정보는 줄 수 없다.

3-1-2. 제약 조건 테스트

Reilly와 Capani[1]는 각각의 제약 조건의 잔차를 그 제약 조건의 잔차의 표준편차로 나누어 제약 조건의 단위 정상 분포를 정의하였다.

$$Z_{ek} = \frac{e_k}{\sqrt{H_{kk}}} \sim N(0, 1) \quad (9)$$

e_k 는 k 번째 제약조건의 잔차이고 H_{kk} 는 잔차의 공분산 행렬의 대각 원소이다. 측정 값들은 정상 분포를 이룬다는 가정 하에 제약 조건의 단위 분포 값을 구해 원하는 유의 수준의 단위 정상 분포 값과 비교하여 벗어날 경우 총계 오차가 있는 것으로 판단한다. 상관 관계에 대한 정보가 없고 문제가 있는 제약 조건만을 판별하나 식이 간단하고 H 가 nonsingular하므로 적용하기가 쉬워 실제 데이터를 다룰 때 자주 쓰이는 방법이다.

3-1-3. 측정값 테스트

Mah와 Tamhane[2]는 제약 조건 테스트와 유사하게 각 변수의 측정 조정 값 $a(x - \hat{x})$ (measurement adjustment)를 측정 보정치 a 의 표준편차로 나누어 주는 것으로 측정 값 테스트를 정의하였다.

$$Z_{aj} = \frac{a_j}{\sqrt{Q_{jj}}} \sim N(0, 1) \quad (10)$$

a_j 는 j 번째 제약조건의 측정 조정값이며 Q_{jj} 는 측정 조정 값의 공분산의 대각 원소이다. 이 방법은 측정 변수 각각에 대해 판단할 수 있는 특징이 있으나 데이터의 상관 관계에 대한 정보가 없고 실제 데이터에서는 Q 가 singular일 수 있으므로 실제 데이터에 적용할 때 문제가 생긴다.

3-1-4. Maximum Power 테스트

Almasy와 Sztano[3]는 하나의 총계 오차를 찾아내는데 있어 탁월한 기능을 가지는 Maximum Power 테스트를 정의하였다.

$$Z_{aj}^* = \frac{\sum^{-1} a_{ij}}{\sqrt{\sum^{-1} Q \sum^{-1} I_{jj}}} \sim N(0, 1) \quad (11)$$

a 는 측정 조정 값이고 \sum 와 Q 는 측정 값과 보정 치의 공분산이다. \sum^{-1} 는 데이터의 상관 관계에 대한 정보를 포함한다.

Crowe[4]는 위의 식에 대응하여 각각의 제약조건에 적용할 수 있는 Maximum Power 제약조건 테스트를 정의하였다.

$$Z_{e,i}^* = \frac{(H_e^{-1} e)_i}{\sqrt{(H_e^{-1})_{ii}}} \sim N(0, 1) \quad (12)$$

e 는 잔차이고 H_e 는 잔차의 공분산이다. H_e^{-1} 에 상관 관계에 대한 정보가 있어 식 (11)과 마찬가지로 하나의 총계 오차를 발견하는데 있어 최대 성능을 가지나 두 개 이상의 총계 오차에서는 이런 공분산의 역 행렬들에 의해 상관관계 구조가 정상에서 크게 벗어나게 되어 부정확한 위치에서 판별되는 단점이 있다.

3-1-5. Principal component (PC)테스트

Crowe[5]는 PCA방법을 이용하여 상관관계가 존재하는 변수들을 상관관계가 없는 변수들로 전환하여 총계 오차의 정확한 위치와 크기 판별이 가능하도록 하였다.

$$y_{e,i} = (W_e^T e)_i \sim N(0, 1) \quad (13)$$

벡터 y_e 는 principal component(PC)들로 구성되며 각각의 원소의 값 $y_{e,i}$ 는 PC score값들을 의미한다. 선형 결합 계수를 나타내는 W_e^T 는 PC에서의 loading에 해당하며 H_e 의 eigenvector들로 구성되고 상관 관계에 대한 정보를 가지고 있다. 식 (13)은 상관된 데이터인 e 에서 서로 상관되지 않는 $y_{e,i}$ 로 변환시킨 것이다. 만일 i 번째 score값이 유의 수준을 벗어났다면 식 (14)의 기여도(contribution)식에 의해 몇 번째 제약 조건에서 이상이 있는지 알아낼 수 있다.

$$g_j = (w_{e,i})_j e_j \quad j=1, \dots, m \quad (14)$$

$g=(g_1, \dots, g_m)^T$ 라고 정의되며 m 은 기여자(contributor)의 수이다. 만일 유의 수준을 벗어난 score값에서 k 번째 제약 조건에 의한 기여도가 상대적으로 매우 크다면 그 제약 조건에서 총계 오차가 있다고 판별한다. PC를 식 (15)와 같이 잡으면 각각의 flow에 의한 판별이 가능하다.

$$y_{a,j} = (W_a^T a)_j \sim N(0, 1) \quad (15)$$

이 식에 대응하는 기여도식은 식 (16)과 같다.

$$g_j \sim (w_{a,i})_j a_j \quad j=1, \dots, n \quad (16)$$

PC 테스트는 상관 관계가 없는 PC에 의한 분석을 하므로 총계 오차의 정확한 위치를 판별해 낼 수가 있다.

3-2. 최적의 판별 기준

테스트가 최적으로 판별한다는 의미는 총계 오차를 정확한 위치에서 좋은 감도로 판별해 낼 수 의미한다. 결과에서처럼 flow 테스트라도 같은 제약조건 내의 다른 flow들에도 총계 오차가 존재한다고 판별하기

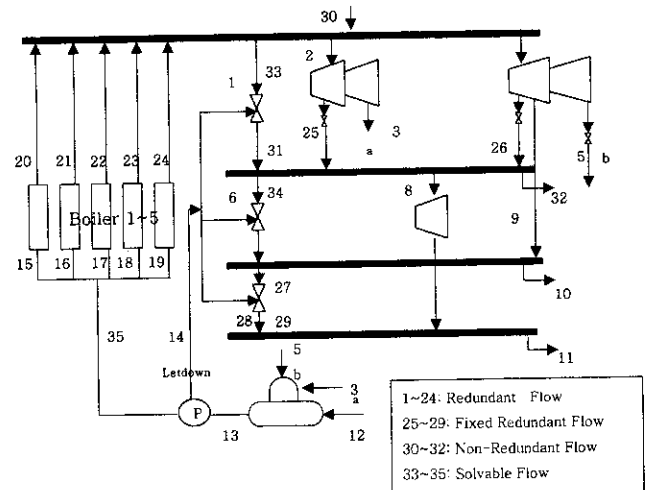


Fig. 2. A schematic diagram for a utility plant.

도 하는데 실제 총계 오차가 있는 flow 역시 존재한다고 판별하므로 잘못 잡아냈다고 볼 수 없다. 하지만 다른 제약 조건 내의 flow에 총계 오차가 있다고 판별해 낸다면 명백한 잘못이다. 3-1에 제시된 방법들은 총계 오차의 개수와 크기, 위치에 따라서 감도와 정확성에 있어서 차이를 보인다.

4. 사례연구

4-1. 대상 유틸리티 플랜트

Fig. 2의 유틸리티 공장은 스팀을 생산하는 보일러와 전기를 생산하는 터빈, 탈기기와 헤더로 구성된다. 공정에서 사용되는 스팀은 압력에 따라 XP, HP, MP, LP로 구분되는데 밸브로 조정되어 각 단위에 알맞은 스팀으로 변화되어 공급된다. 공정의 flow는 모두 35개이고 29개의 redundant flow와 3개의 non-redundant flow, 그리고 3개의 solvable flow로 구성되며 redundant flow중 5개는 고정 값을 갖는다. Non-redundant flow도 고정 값으로 다루어 지므로 전체 고정 값을 갖는 flow는 모두 8개이다. Flow의 번호는 이런 구분별로 순서대로 매겼다. Fig. 2의 공정도에 redundant flow, non-redundant flow, solvable한 flow를 구분지어 제시하였다.

4-2. 총계 오차 판별을 위한 실험 설계

4-2-1. 데이터 전처리 구성

기본적으로 제약 조건 테스트가 flow 테스트보다 실제적으로 많이 쓰이는 이유는 flow 테스트에 쓰이는 a 의 공분산인 Q 가 singular일 경우가 많기 때문이다. 이런 현상을 피하기 위해서 실험을 구성하였는데 실험 데이터는 시뮬레이션 데이터를 가지고 대략 5%의 분산을 가지도록 난수 60개를 주어 데이터 집합을 만들었다. 이 데이터 집합은 무작위 오차만을 포함한다.

측정 값들은 서로 독립이 아니어서 상관관계가 존재한다. 만일 상관관계가 없는 데이터를 가지고 실험한다면 측정 값 테스트와 MP 테스트와 PC 테스트는 같은 결과가 나온다. 데이터 보정에서 최적 값을 얻어내기 위하여 SQP(Successive Quadratic Programming)를 이용하였으며 제약 조건은 각 node에 대한 물질 수지이므로 선형 제약 조건이다. 즉, 정상 상태에서의 선형 제약 조건을 다루는 데이터 보정 문제를 다루었다.

4-2-2. 행렬 투영법(Matrix Projection Method) 적용

실험에서 사용된 테스트 중 제약 조건 테스트, MP 테스트, PC 제약조건 테스트는 Crowe[6]의 행렬 투영법을 기본으로 한다. 수식 행렬 B 는 $[B_0 \ B_1 \ B_2]$ 와 같이 구분될 수가 있는데 B_0 는 데이터 보정의 대상인

고정 값을 가지지 않는 redundant flow에 해당하는 부 행렬(sub-matrix)로써 flow 1-24번째에 해당한다. B_1 은 flow로는 25-32번째에 해당하는 고정된 값을 가지는 redundant flow와 non-redundant flow로 구성된 부 행렬이다. B_2 는 측정 값이 없는 33-35번째 flow로 구성되며 보정 값이나 non-redundant 값을 통하여 구해진다.

$$B_0(\hat{x}+a)+B_1c+B_2u=0 \quad (17)$$

a 는 조정 값이고 c 는 고정된 flow, u 는 비 측정 flow에 해당한다. 비 측정 flow에 해당하는 열을 제거시키기 위한 Y^T 를 먼저 구하고 제약 조건에 곱하면 식 (18)이 얻어진다.

$$Y^T[B_0(\hat{x}+a)+B_1c]=0 \quad (18)$$

이 식은 식 (19), (20)으로 나타내어질 수 있다.

$$Y^T[B_0x+B_1c]=e \quad (19)$$

$$Y^T[B_0B_1]\begin{bmatrix} x \\ c \end{bmatrix}=e \quad (20)$$

$Y^T[B_0B_1]$ 행렬의 열수는 측정된 flow의 수와 같고 행렬의 행의 수는 잔차 e 의 수와 같다. 행의 수는 15개이므로 e 의 개수도 15개가 되며 여기서 구해진 잔차 e 값으로 테스트가 행해지게 된다.

4.3. 사례별 해석

각각의 테스트의 감도와 위치 판별력의 차이를 밝히는 목적으로 총계 오차를 3개까지 임의로 생성시켜 테스트별 정확성을 비교하였고 각 총계 오차의 크기에 따른 테스트별 감도 비교도 할 수 있는 최소한의 실험을 설계하였는데 Case 1에서는 작은 총계 오차로 테스트의 감도 비교를, Case 2에서는 적당한 크기인 총계 오차가 하나일 때의 일반적인 경우를 나타내었다. Case 3에서는 센서 고장의 경우처럼 매우 큰 크기의 총계 오차가 존재하는 특이한 경우, Case 4는 두 개의 총계 오차가 다른 제약 조건에 존재할 경우를 다루었으며 Case 5에서는 세 개의 총계 오차를 같은 제약 조건에 둔 경우를 나타내었다. 여기서 ‘작은’, ‘적당한’의 표현은 주관적인 것으로 MP 테스트만이 판별해 낼 수 있을 정도의 크기를 작은 총계 오차라고 하였고 총계 오차가 하나일 때 모든 테스트가 올바르게 판별해 내는 크기의 데이터를 ‘적당한’ 총계 오차라고 명명하였다.

4.3-1. 작은 총계 오차 1개가 포함될 경우

처음 데이터를 보정했을 때의 데이터 보정의 목적함수 값은 21.0849이다. Fig. 2에서 보일러 1의 출력 flow인 20번 flow의 크기를 원래 데이터 90에서 100으로 바꾸어 작은 총계 오차를 주었다(Fig. 3). 유의수준을 $\alpha=0.01$ 이라 하여 오차 판별의 기준으로 삼고 데이터 보정을 하고 총계 오차를 판별해 냈을 때 Table 1과 같은 결과 값을 얻었고 이 경우의 목적 함수 값은 28.6322였다. Table 1의 오른쪽 열의 숫자들은 판별된 총계 오차의 위치(stream number)를 나타낸다. Flow 관련 테스트들은 판별된 flow의 위치를 나타내고 제약조건에 관련된 테스트는 판별된 제약조건을 나타낸다. Chi-square 테스트는 총계 오차의 유무만을 따지므로 존재, 비존재로 나타내었다.

제약조건 테스트와 MP 제약조건 테스트는 11번 제약조건에서 총계 오차가 있다고 판별했는데 이 제약조건을 구성하는 flow에서 총계 오차가 있다고 보면 된다. 11번 제약조건은 보일러 1에서의 입력과 출력을 나타내는 15, 20번 flow로 구성되므로 정확하게 판별해 냈다.

Flow 테스트들 중에 MP flow 테스트만이 15와 20에서 총계 오차가 있음을 판별해내어 가장 좋은 감도를 갖는다. Flow 테스트들의 감도를 비교한 그래프는 Fig. 4과 같다. $\alpha=0.01$ 인 경우는 2.3263이 총계 오차의 기준 값이 되므로 MP 테스트만이 판별되는 것이다. MP 테스트는 자체 식 내의 Σ^{-1} , H^{-1} 에 상관 관계에 관한 정보를 갖는다. 이

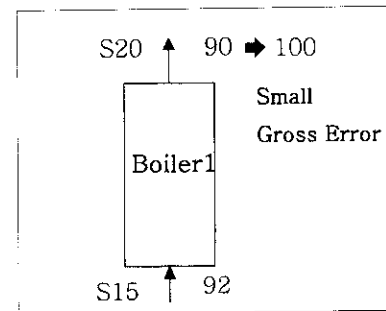


Fig. 3. Case 1: One small gross error in a constraint.
(11th constraint; S15, S20: 15th, 20th stream)

Table 1. Case 1: One small gross error in a constraint

Chi-square method	Exist
Measurement test	No gross error
Constraint test	11
MP flow test	15 20
MP constraint test	11
PC flow test	No gross error
PC constraint test	No gross error

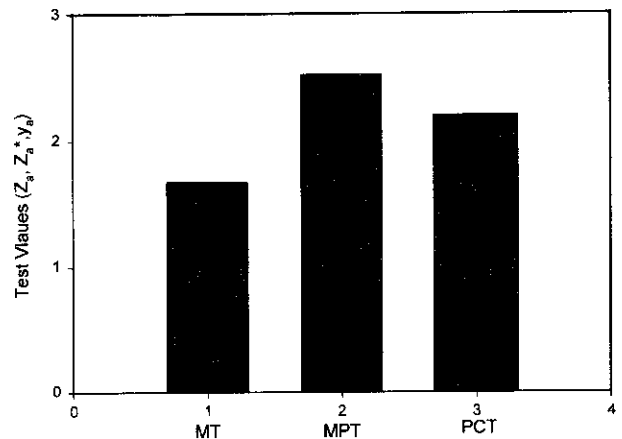


Fig. 4. Comparison of the sensitivity of flow tests.

는 총계 오차가 하나 있을 경우 측정값 테스트나 PC 테스트보다 감도가 뛰어나게 되어 판별력에 있어 최대 성능(maximum power)을 갖게 한다. 따라서 총계 오차가 기껏해야 하나 이하로 예상되는 소규모 공정에서는 MP 테스트를 사용하는 것이 가장 좋다.

MP flow 테스트는 15, 20번 flow에서 총계 오차가 있다고 판별해 냈는데 모두 11번째 제약 조건을 구성하는 flow이다. 따라서 측정값 테스트와 마찬가지로, MP 테스트 결과 값을 통해 총계 오차의 정확한 위치를 알아내려면 총계 오차가 있다고 판별된 flow가 속한 제약 조건내의 모든 flow를 의심해 봐야 한다.

4.3-2. 적당한 크기의 1개의 총계 오차가 포함될 경우

Fig. 2에서 보일러 1의 출력 flow인 20번 flow의 크기를 원래 데이터 90에서 110으로 바꾸어 총계 오차를 주었다(Fig. 5). 유의수준을 $\alpha=0.01$ 이라 하여 오차 판별의 기준으로 삼고 데이터 보정을 한 뒤 총계 오차를 판별해 냈을 때 Table 2와 같은 결과 값을 얻었다. 목적 함수 값은 63.9447였다.

이 경우 모든 테스트가 올바르게 판별하고 있음을 보여준다. 3가지 flow 테스트 모두 15와 20에서 판별되는 것으로 보아 어떤 flow가 총

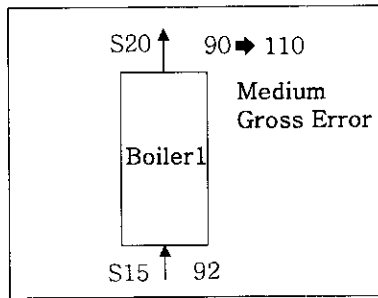


Fig. 5. Case 2: One medium gross error in a constraint.
(11th constraint; S15, S20: 15th, 20th stream)

Table 2. Case 2: One medium gross error in a constraint

Chi-square method	Exist
Measurement test	15 20
Constraint test	11
MP flow test	15 20
MP constraint test	11
PC flow test	15 20
PC constraint test	11

제 오차가 있다고 판별이 되면 그 flow가 포함된 제약 조건 내의 모든 flow를 의심해 보아야 한다.

4-3-3. 매우 큰 1개의 총계 오차가 포함될 경우

보일러 1의 출력 flow인 20번 flow를 90에서 0으로 바꾸었다. 이 경우는 센서 고장의 특이한 경우로 flow가 있는데도 불구하고 0의 값을 센서가 내보내는 경우이다(Fig. 6). 유의수준은 $\alpha=0.01$ 이고 데이터 보정 후 총계 오차 판별 결과는 Table 3과 같다. 목적 함수 값은 1529.547이다.

Flow가 있는데도 불구하고 센서가 0의 값을 내보내는 경우 이를 판별하지 않고 단순히 데이터 보정을 수행하면 제약 조건 하에서 전체적으로 잔차의 합을 최소화하는 것이 데이터 보정이므로 고장난 센서가

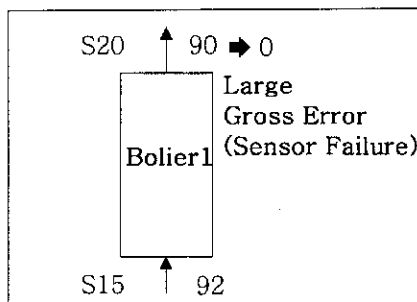


Fig. 6. Case 3: One large gross error in a constraint.
(11th constraint; S15, S20: 15th, 20th stream)

Table 3. Case 3: One large gross error in a constraint

Chi-square method	Exist
Measurement test	2,3,4,5,13,15,16,17,18,19,20,21,22,23,24
Constraint test	1 11
MP flow test	1,3,4,6,7,9,10,12,13,15,16,19,20,23
MP constraint test	2,3,4,5,7,9,10,11,12,13,14,15
PC flow test	15 20
PC constraint test	1 11

있는 node의 불균형이 전체적으로 퍼지는 결과를 초래하여 정확한 측정 값조차 부정확한 보정 값을 나타내는 심각한 결과를 나타낸다. 따라서 수식 내에 측정 값과 보정 값의 차이를 직접적으로 다루는 측정 값 테스트나 MP flow 테스트는 이런 경우에 쓸 수가 없다. 두 테스트는 정상 상태일 경우 이상치가 있다면 상태 전체를 바꾸어 제약조건을 만족시키려 하는 반면 PC flow 테스트는 정상 상태를 유지하면서 이러한 이상치들의 영향을 최소화하면서 제약 조건에 맞는 목적 함수 값을 구할 수가 있으므로 위의 결과처럼 정확한 위치에서 총계 오차를 판별해 내었다.

Flow 테스트들과는 반대로 잔차를 다루는 제약조건 테스트들은 총계 오차를 잘 잡아내었다. 이 테스트들은 잔차만을 다루기 때문에 20번 flow가 포함된 1번째, 11번째에서 이상이 있음을 알 수가 있는 것이다.

그런데 MP 제약조건 테스트는 데이터의 상관관계에 대한 구조를 포함하는 잔차의 공분산 H^{-1} 을 사용하는데 이러한 값(0)이 큰 영향을 주어 상관 관계 구조가 정상에서 크게 벗어나게 되므로 잔차를 다루더라도 테스트의 기능을 수행할 수가 없다.

4-3-4. 적당한 크기의 총계 오차 2개가 각각 다른 제약조건에 존재할 경우

탈기기의 12번 flow를 1.5에서 3.5로 바꾸고 보일러 1에서 20번 flow를 24.6에서 29로 변화시켰다(Fig. 7). 12번 flow는 3, 5, 12, 13으로 구성된 9번째 제약조건에 속하고 20번 flow는 15와 20의 flow로 구성된 11번째 제약조건에 속한다. 데이터 보정 결과 목적 함수 값은 170.8143이고 Table 4를 얻었다.

총계 오차가 2개 존재할 때는 다른 테스트들은 이상을 보이지 않으나 MP 테스트가 잘못된 위치에서 총계 오차를 판별하였다. 그 이유는 MP 테스트는 데이터의 상관관계의 정보를 Σ^{-1} , H^{-1} 에 포함하는데 총계 오차가 2개 이상인 상관 관계의 구조가 정상 상태에서 크게 벗어나 각 flow나 제약 조건에 대한 테스트 값을 매우 크게 만들어 잘못된 위치에서 총계 오차가 판별되는 단점을 갖기 때문이다. 즉 판별력에 있어 최대 성능을 가지는 경우는 단지 총계 오차가 하나일 때이다.

측정값 테스트는 이러한 상관관계에 대한 정보를 가지고 있지 않으며 PC 테스트는 이러한 상관관계의 구조가 식에서 eigenvector에 들어 있기 때문에 이러한 현상을 보이지 않는다.

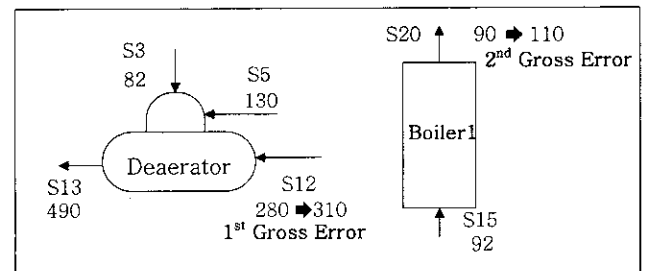


Fig. 7. Case 4: Two gross errors in two constraints.

(9th constraint; S3, S5, S13, S12: 3rd, 5th, 12th, 13th stream 11th constraint; S15, S20: 15th, 20th stream)

Table 4. Case 4: Two gross errors in two constraints

Chi-square method	Exist
Measurement test	12 15 20
Constraint test	9 11
MP flow test	9 12 14 15 17
MP constraint test	1 3 6 7 9 11 12 13 14 15
PC flow test	12 15 20
PC constraint test	9 11

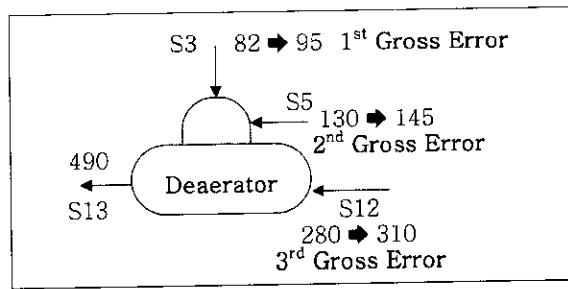


Fig. 8. Case 5: Three gross errors in a constraint.
(9th constraint; S3, S5, S12, S13: 3rd, 5th, 12th, 13th stream)

Table 5. Case 5: Three gross errors in a constraint

Chi-square method	Exist
Measurement test	2 3 5 12 21
Constraint test	6 7 9
MP flow test	1 3 5 7 12 14 17 23
MP constraint test	3 9 11 13 14
PC flow test	3 5 12
PC constraint test	9(very big contribution value)

4-3-5. 적당한 크기의 총계 오차 3개가 같은 제약조건 내에 존재할 경우 3, 5, 12, 13의 flow로 구성되는 9번째 제약 조건인 탈기기의 제약조건 내의 3번(82→95), 5번(130→145), 12번째 flow(280→300)에 총계 오차가 존재(Fig. 8)할 때, 데이터 보정의 목적함수 값은 200.7이고 Table 5와 같은 결과 값을 얻을 수 있었다.

측정값 테스트와 제약조건 테스트는 식 자체에 데이터의 상관 관계의 정보를 가지고 있지 않기 때문에 데이터의 상관 관계 정도가 크다면 판별 정확성에 문제가 생길 수 있다. 2번 flow는 flow 2, 3, 25로 구성된 6번째 제약조건내의 flow이고 21번은 잘못 판정한 값이다. 따라서 다변량 데이터는 다변량 기준에 입각한 테스트로써 판단되어야 한다. 그런데 MP 테스트의 Σ^{-1} , H^{-1} 나 PC 테스트의 W^T 등과 같이 상관 관계에 해당하는 정보를 얻을 필요가 없기 때문에 가장 간단하게 적용할 수가 있는 방법들이다. 따라서 많은 수의 총계 오차가 예상되지 않는 간단한 공정의 상관 관계가 적은 데이터를 다룰 경우에는 이 두 개의 테스트들이 최적이다.

PC 테스트는 이 경우에 가장 좋은 결과 값을 보이고 있다. 정확한 위치를 판별하였으며 PC 제약조건 테스트의 경우에는 9번째 기여자(contributor)가 매우 큰 값을 보여서 이 곳에 총계 오차가 있음을 알 수 있는데, 정상 상태를 유지하면서 이러한 이상치들의 영향을 최소화하면서 제약조건에 맞는 목적함수 값을 구하기 때문에 정확한 위치에서 총계 오차를 판별해 낼 수 있는 것이다.

PC 테스트는 측정 값의 확률 분포가 알려진다면 어느 경우에 대하여도 적용 가능하다. 정상 분포가 가정되지 않았다면 중심 극한 정리에 의하여 독립적으로 얻어진 데이터의 수가 증가할수록 측정 값들은 정상 분포에 가까워지게 된다. 그러나 측정 값들은 일반적으로 시간에 따라 상관 관계를 가지고 있으므로 이러한 중심 극한 정리가 성립되지 않는다. 분포가 알려져 있지 않다면 기준 분포를 얻기 위하여 충분한 데이터가 얻어져야만 한다. PC 테스트는 eigenvalue와 eigenvector를 계산하는데 있어서 좀더 많은 계산이 이루어지며 기여도 그래프에 관련된 분석이 부가적으로 이루어지기 때문에 비교적 시간이 걸린다. 또한 조업 조건이 변하면 다시 많은 데이터를 수집해야 하는 번거로움을 가진다. 따라서 데이터 수집이 많이 이루어지고 조업조건이 자주 바뀌지 않으며 측정 값의 상관 관계 정도가 큰 대규모 공정에서는 PC 테스트가 최적이라고 할 수 있다.

5. 결 론

이 논문은 주요 총계 오차 판별 방법들을 유틸리티 플랜트의 여러 경우에 적용하여 각 경우에 따른 최적 테스트를 밝혀내고 이 과정에 따른 테스트별 장단점을 파악하고 각 테스트별 최적 공정을 제시하였다.

공정데이터를 가지고 총계 오차 판별법을 적용할 때는 공정의 복잡한 정도를 살펴봐야 하고 데이터의 상관관계 정도도 알아야 하며 데이터의 수집 정도를 살펴봐야 한다. 작은 수의 총계 오차가 기대되는 간단한 공정의 상관관계가 작은 데이터를 다룰 경우에는 측정 값 테스트와 제약 조건 테스트가 적당하다. 그리고 총계 오차가 기껏해야 하나 이하로 예상되는 소규모 공정에는 MP 테스트를 사용하는 것이 가장 좋으며 만일 데이터 수집이 많이 이루어지고 조업조건이 자주 바뀌지 않는 측정치 간의 상관 관계 정도가 큰 대규모 공정에서는 PC 테스트가 최적이라고 할 수 있다.

감 사

본 연구는 한국과학재단지정 우수 연구 센터인 공정 산업의 지능화 연구 센터, 한국과학재단 목적 기초 연구(98-0502-05-01-3) 및 2000년도 두 뇌 한국 21 사업의 지원에 의하여 수행된 연구 결과로써 지원에 감사드립니다.

사용기호

- a : vector of adjustments to measured component flow rates
- B : balance matrix of process($m \times n$)
- B_0 : columns of B corresponding to measured flow rates
- B_1 : columns of B corresponding to fixed flow rates
- B_2 : columns of B corresponding to unmeasured flow rates
- \bar{x} : vector of flow measurement
- \hat{x} : vector of reconciled value
- e : vector of imbalance
- H : covariance of e
- z, y : unit normal variate, scalar or vector
- Q : covariance of adjustment
- $w_{e,i}$: ith eigenvector of H_e

그리스 문자

- Σ : covariance of \bar{x}

참고문헌

1. Reilly, P. M. and Carpani, R. E.: "Application of Statistic Theory of Adjustments to Material Balances," 13th Can. Chem. Eng. Conf. Montreal Que. Oct., (1963).
2. Mah, R. S. H. and Tamhane, A. C.: *AIChE J.*, **28**, 828(1982).
3. Almsy, G. A. and Sztano, T.: *Probl. Control Inf. Theory*, **4**, 57(1975).
4. Crowe, C. M.: *Can. J. Chem. Eng.*, **70**, 1030(1992).
5. Tong, H. and Crowe, C. M.: *AIChE J.*, **41**, 1712(1995).
6. Crowe, C. M., Garcia Campos, Y. A. and Hrymak, A.: *AIChE J.*, **31**, 1187(1985).
7. Crowe, C. M.: *Chem. Eng. Sci.*, **44**, 2909(1989).
8. Ham, P. G., Cleaves, G. W. and Lawlor, J. K.: "Operation Data Reconciliation: An Aid to Improved Plant Performance," Proc. 10th World Petrol. Congr. Bucarest., **10**, 281(1979).

9. Hodouin, D., Alliot, N. and Flament, F.: *Int. J. Mineral Processing.*, **32**, 213(1991).
10. Holly, W., Cook, R. and Crowe, C. M.: *Can. J. Chem. Eng.*, **67**, 595(1989).
11. Mah, R. S. H.: "Chemical Process Structures and Information Flows," Butterworths, Stoneham, MA(1990).
12. Van der Heijden, R. T. J. M., Heijnen, J. J., Hellinga, C., Romein, B. and Luyben, K. Ch. A. M.: *Biotech. Bioeng.*, **43**, 3(1994).