CHEMICAL ENGINEERING TRANSACTIONS

VOL. 39, 2014

A publication of
AIDIC
The Italian Association
of Chemical Engineering
www.aidic.it/cet

Guest Editors: Petar Sabev Varbanov, Jiří Jaromír Klemeš, Peng Yen Liew, Jun Yow Yong

ISBN 978-88-95608-30-3; **ISSN** 2283-9216

DOI: 10.3303/CET1439182

Data Reconciliation with Gross Error Detection using NLP for a Hot-Oil Heat Exchanger

Pathompong Kongchuay[a], Kitipat Siemanond*[a,b]

[a]The Petroleum and Petrochemical College, Chulalongkorn University, Chulalongkorn Soi 12, Phayathai Rd., Pathumwan, Bangkok 10330, Thailand
[b]Center of Excellence on Petrochemical and Materials Technology, Chulalongkorn Soi 12, Phayathai Rd., Pathumwan, Bangkok 10330, Thailand
kitipat.s@chula.ac.th

The measured data from instruments in process control activities usually consist of random and gross errors which reduce reliability of measurement. Data reconciliation (DR) technique is applied to improve the accuracy of measured data and satisfy the law of conservation. Moreover if data contains bias or gross errors in the system, DR is not as accurate as expected. This work performed DR with gross error detection (GED) technique to improve the data measurement of a simulated hot-oil heat exchanger. There are two kinds of GED; the conventional GED method and the traditional measurement test modified by using NLP. The gross errors or bias in some measured data, including volumetric flow rates, supply and target temperature of hot and cold process streams and overall heat transfer coefficient were generated. The DR with GED using NLP was done by commercial optimization software, GAMS, with a least-square objective function. The conventional GED and conventional gross error elimination applied statistical methods; basic global test and basic measurement test, respectively. The DR with GED technique produced more accurate estimates of process variables showing reductions in standard deviation. The other method, the modified measurement test, was studied for performance comparison. The performance of the modified measurement test using NLP was significantly better than the conventional method, in terms of the performance evaluation using the overall power (OP).

1. Introduction

The purpose to maximize production capacity using the minimum energy and minimum cost of operation is always presented as the primary motivation for process control system. The need for control of many petroleum and petrochemical processes is more fundamental because the manufacturing achievement would be dependent on maintaining process condition within certain boundaries and one important thing in the process control activities is the process measurements and its reliability of how closely the data satisfy mass balance, energy balance or other physical constraints of the process.

Process measurements in all petroleum and petrochemical processes are for the purpose of evaluating the process performance; however, not all variables are measured because of operating and maintenance costs or some technical infeasibility. Furthermore, the measurements often contain the random and gross errors from the changes in ambient condition, failure in network transmission or miscalibration, etc. (Narasimhan and Jordache, 2000), so the presence of random and gross errors in measurements leads to inaccurate process data. Moreover, the data also do not commonly satisfy the process constraints and the laws of conversion, thus, many process control activities are needed to correct this problem. The method of improving the accuracy of process data by adjusting the measured variables and estimating the unmeasured variable to be achieved in the process constraints and the laws of conversion, are known as "Data reconciliation" or DR. However, if the measurements are adjusted to satisfy the laws of conservation while in the presence of gross errors, all of the adjustments are often affected by such biases and would not generally be reliable. Thus gross errors must be detected and eliminated, by "Gross Error Detection" or GED.

Please cite this article as: Kongchuay P., Siemanond K., 2014, Data reconciliation with gross error detection using NLP for a hot-oil heat exchanger, Chemical Engineering Transactions, 39, 1087-1092 DOI:10.3303/CET1439182

The goal of this research is to perform the DR with GED technique to improve the data measurement of a simulated hot-oil heat exchanger from gas separation plant as a case study. DR problems use a least-square objective function and based on the assumption that random errors follow normal distribution with zero mean and known variance of each variable.

2. Physical system

A simulated hot-oil heat exchanger heating ethane consisted of simulated measurements for 8 variables, 3 equality constraints and 4 inequality constraints. The variables were the hot-oil inlet temperature, $T_{o,in}$, the hot-oil outlet temperature, $T_{o,out}$, the hot-oil volumetric flowrate, F_o, the cold-ethane volumetric flowrate, F_{et}, the cold-ethane inlet temperature, $T_{et,in}$, the cold-ethane outlet temperature, $T_{et,out}$, the overall heat transfer coefficient, U and heat duty, Q.
The three equality constraints are related to energy balance equation with process variables of heat duty of hot oil, ethane product, and heat exchanger as shown in Eq(1, 2 and 3), respectively. And the inequality constraints are shown in Eq(4, 5, 6 and 7), respectively.

$$Q = M_o\, C_{p,o}\, \Delta T_o \quad (1)$$
$$Q = M_{et}\, C_{p,et}\, \Delta T_{et} \quad (2)$$
$$Q = UA(LMTD) \quad (3)$$
$$T_{o,in} \geq T_{et,out} \quad (4)$$
$$T_{o,out} \geq T_{et,in} \quad (5)$$
$$T_{o,in} \geq T_{o,out} \quad (6)$$
$$T_{et,out} \geq T_{et,in} \quad (7)$$

Noted: $C_{p,o} = 0.0036T + 1.8089$ (8)

$C_{p,et} = -0.0068T + 2.58$ (9)

Where M_o is the hot-oil mass flowrate (kg/h), M_{et} is the cold-ethane mass flowrate (kg/h), $T_{o,in}$ is the hot-oil inlet temperature (°C), $T_{o,out}$ is the hot-oil outlet temperature (°C), $T_{et,in}$ is the cold-ethane inlet temperature (°C), $T_{et,out}$ is the cold-ethane outlet temperature (°C), U is the overall heat transfer coefficient (W/m^2 °C), A is the area of heat exchanger (m^2) and Q is the heat duty of heat exchanger (W).
Chen approximation (Chen, 1988) or LMTD value in Eq(3) is used to calculate log-mean temperature difference in this system as shown in Eq(10).

$$LMTD = \left[(T_{o,in} - T_{et,out}) \times (T_{o,out} - T_{et,in}) \times \frac{(T_{o,in}-T_{et,out})+(T_{o,out}-T_{et,in})}{2}\right]^{1/3} \quad (10)$$

The physical constants as shown in Table 1 are used for generating the simulated data of this model.

Table 1: Physical data for generating the simulated data of a hot-oil heat exchanger

Parameter	Value	Units
$C_{p,o}$	2.419	kJ/kg °C
$C_{p,et}$	2.473	kJ/kg °C
ρ_o	772.65	kg/m^3
ρ_{et}	1.334	kg/m^3
U	310.6	W/m^2 °C
A	46.10	m^2

Where $C_{p,o}$ is the heat capacity of hot-oil stream (kJ/kg °C), $C_{p,et}$ is the heat capacity of cold-ethane stream (kJ/kg °C), ρ_o is the density of hot oil (kg/m^3) and ρ_{et} is the density of cold ethane (kg/m^3)

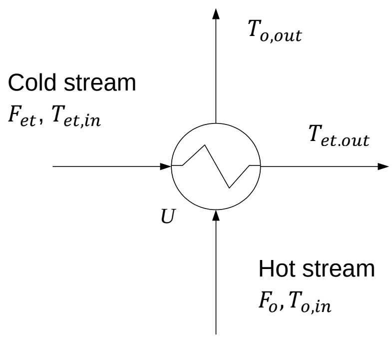

Figure 1: Measured variables of a simulated hot-oil heat exchanger

3. Simulating data

This task involves simulating true values and measured values. True value is the value which is assumed that the variable is directly measured without any errors. Measured value is the actual data which consist of some error. True values were simulated by fixing the values of volumetric flowrate of hot oil, inlet temperature of hot oil, volumetric flowrate of cold ethane, inlet temperature of cold ethane and heat overall coefficient and adjust the value of heat exchanger area closely to 46.10 m^2 by using goal seek technique and following the physical constant data in Table 1. Measured values were generated with two conditions of error; only random error and both random and gross errors. And the measured variables are shown in Figure 1.

3.1 Generating random errors

Measured value with random errors were obtained by adding Gaussian noise to true values using Gaussian Random Number Generator via random.org to generate 365 random numbers from a normal distribution and this measured values follow the assumptions where the standard deviations are 10 m^3/h for hot stream volumetric flowrate indicator and 10 °C for cold temperature indicator and the standard deviations are 25 m^3/h for cold stream volumetric flowrate indicator, 25 °C for hot temperature indicator and 25 W/m^2 °C for overall heat transfer coefficient indicator.

3.2 Generating gross errors

Measured values with gross errors were obtained by adding the randomness noise (comes from atmospheric noise) to measured values with random errors of a hot oil volumetric flowrate variable, F_o and a cold ethane inlet temperature, $T_{et,in}$ using Integer Random Number Generator via random.org to generate 365 values (days collected). The error magnitude was allowed to vary between approximately 1 to 300 % of the true values and the smallest measured values permitted were zero, corresponding to instrument failure or out of order. True values, measured values and some bias of variables above are shown in Figure 2 and Figure 3, respectively.

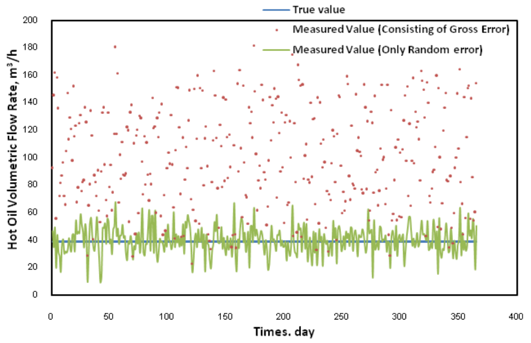

Figure 2: True values, measured values and bias of hot oil volumetric flow rate variable, F_o

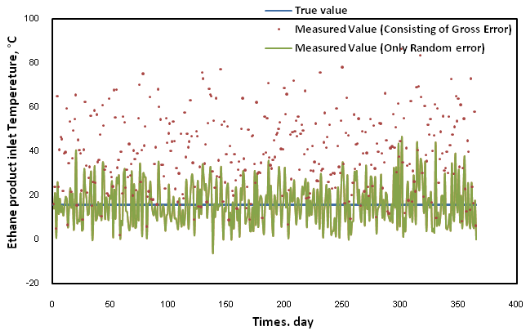

Figure 3: True values, measured values and bias of cold ethane inlet temperature variable, $T_{et,in}$

4. Application of data reconciliation with gross error detection

This research is divided into 3 parts; DR of measured values with only random error, DR of measured values with both random and gross errors and DR/GED of measured values with both random and gross errors.

4.1 Data reconciliation

Before doing the DR technique, Degree of freedom (*dof*) and degree of redundancy (*dor*) were done first. Degree of freedom (*dof*) is the difference between number of all variables (measured and unmeasured variables) and number of all equations. Degree of redundancy (*dor*) is the difference between number of measured variables and degree of freedom. In this case, degree of freedom is equal to 5 (*dof = 8 – 3 = 5*) and degree of redundancy is equal to 2 (*dor = 7 – 5 = 2*). To perform DR, DOR must be greater than or equal to 1 (*dor* ≥ *1*). GAMS program is used to perform DR by minimizing the objective function with the estimated values of flow rates, inlet and outlet temperatures of hot oil and cold ethane, and overall heat transfer coefficient suitable for constraints (Eq(1 to 7)). Objective function used to perform DR for this model, is shown in Eq(11).

$$Min\left(\frac{F_{oil}-Fr_{oil}}{\sigma_{Foil}}\right)^2+\left(\frac{F_{et}-Fr_{et}}{\sigma_{Fet}}\right)^2+\left(\frac{T_{o,in}-Tr_{o,in}}{\sigma_{To,in}}\right)^2+\left(\frac{T_{o,out}-Tr_{o,out}}{\sigma_{To,out}}\right)^2+\left(\frac{T_{et,in}-Tr_{et,in}}{\sigma_{Tet,in}}\right)^2 \\ +\left(\frac{T_{et,out}-Tr_{et,out}}{\sigma_{Tet,out}}\right)^2+\left(\frac{U-Ur}{\sigma_U}\right)^2 \tag{11}$$

Where subscript r represents the reconciled values of all measured variables and σ is the standard deviation.

4.2 Gross error detection

The gross error detection problem has been studied since 1960s. There are several review papers on this subject such as Mah (1987), Crowe (1994), and Yang (1992), etc.

4.2.1 The conventional gross error detection

Concept of the conventional gross error detection is to detect the systematic gross error by using some methods called Global Test (GT) and/or Measurement Test (MT), proposed by Mah and Tamhane (1982), using only the basic statistical concept of them. The GT is the statistical test to find the gross error occurring in the system but it cannot identify where gross errors are, many authors, Almasy and Szatno (1975), Madron et al. (1977) and Rip (1965) have suggested the use of a global chi-squared statistic constructed from the observed discrepancies in the constraints for GT method. This research also needs to use the MT technique to identify the position of gross errors, where the algorithm of this conventional technique follows the steps below

1. Perform DR and obtain the estimate values, $\hat{x}_i$ including the objective function value of them.
2. Find the Chi-squared value, $X^2_{1-\alpha,\mu}$ with degree of freedom, μ by using Chi-squared distribution table.
3. Compare the objective function value from Step 1. with Chi-squared value from Step 2., if the objective function value is less than $X^2_{1-\alpha,\mu}$, so gross errors are not expected in the system, then proceed to Step 7. Otherwise, continue
4. Compute the measurement adjustment, a_i for all measured variables.

   $$a_i = y_i - \hat{x}_i \qquad (12)$$

   Where y_i is the measured value and $\hat{x}_i$ is the reconciled value.
5. Find the confidence interval, $\pm Z\sigma_i$ of each measured variables where Z is the estimator value and σ_i is the standard deviation of measured variable.
6. Compare each of a_i from Step 4. with $\pm Z\sigma_i$ from Step 5., if the maximum value of a_i of which variable is larger than $\pm Z\sigma_i$, delete it and compute DR again and obtain $\hat{x}_{i,new}$ then return to Step 2. In contrast if a_i for all measured variables are less than $\pm Z\sigma_i$, then proceed to Step 7.
7. The measurements that were deleted from previous step are identified to contain gross error. The final reconciled values, $\hat{x}_{i,final}$ will be the results of the method.

4.2.2 The modified measurement test using NLP

Concept of the modified measurement test using NLP method is to detect the systematic gross errors or some bias, the purpose of this technique is the same as the conventional gross error detection but it was modified in some steps of them from the traditional MT, proposed by Mah and Tamhane (1982), to be able to use in non-linear (NL) processes. The modified MT using NLP adjusted two steps of DR. First, DR is performed by using NLP techniques in Step 1 and second, the new values of DR, $\hat{x}_{i,new}$ and the estimate values of variable containing gross error were computed by using NLP techniques in Step 5 and the algorithm of this technique follows the steps below.

1. Perform DR and obtain the estimate values, $\hat{x}_i$.
2. Compute the measurement adjustment, a_i for all measured variables by using Eq(12).
3. Compute the following statistical test.

   $$Z_{d\ or\ aj} = \frac{|d\ or\ a_j|}{\sqrt{W_{jj}}} \qquad (13)$$

   Where a_j is the measurement adjustment of each variable and W_{jj} is the covariance matrix of measurement adjustment.
4. Select the maximum $Z_{d\ or\ aj}$ from Step 3. and compare it with the statistical test criterion, $Z_{1-\beta/2}$. If $Z_{d\ or\ aj} < Z_{1-\beta/2}$ for all measured variables, then proceed to Step 6. Otherwise, continue.
5. Delete the variable which $Z_{d\ or\ aj} \geq Z_{1-\beta/2}$ and compute DR again and obtain $\hat{x}_{i,new}$ then return to Step 2.
6. The measurements that were deleted from previous step are identified to contain gross error. The final reconciled values, $\hat{x}_{i,final}$ will be the results of the method.

For method using NLP, lower and upper bounds of estimated values were set of 0.7 and 1.3 times corresponding to true values, respectively.

5. Results and Discussion

The results are shown in three case studies; case study with only random error in all measured variables (no gross error, NG), case study with random error in all measured variables and gross error in one measured variable (1 position, 1P), and case study with random error in all measured variables and gross error in two measured variables (2 positions, 2P).

Table 2: Objective function of data reconciliation compared to Chi-Squared distribution value (GT step)

Condition of gross error	Objective function values (1st DR / 2nd DR)	Chi-squared values (2 / 1 degree of freedom)
No Gross error (NG)	0.206	
1 Position (1P)	16.892 / 0.237	4.605 / 2.706
2 Positions (2P)	10.148 / 4.462	

Table 3: Measurement adjustment and confidence interval of a hot-oil heat exchanger for the conventional gross error detection (MT step).

Variable	Measurement adjustment		Confidence interval (1 Position / 2 Positions)
	1 Position	2 Positions	
F_o	-58.022	-61.082	±57.479 / ±57.479
F_{et}	-2.716	0.000	±41.408 / ±41.408
$T_{o,in}$	-6.871	-1.207	±40.902 / ±40.902
$T_{o,out}$	-2.311	-3.827	±40.061 / ±40.061
$T_{et,in}$	-0.918	-25.129	±16.070 / ±27.799
$T_{et,out}$	0.000	-0.478	±16.792 / ±16.792

Table 4: The following statistical test of a hot-oil heat exchanger for the modified MT using NLP.

Variable	The statistical test values, $Z_{d\ or\ a,j}$			Statistical test criterion, $Z_{1-\beta/2}$ (7 / 6 measured variables)
	No Gross error	1 Position (1st run / 2nd run)	2 Positions (1st run / 2nd run)	
F_o	1.4435	17.7215 / 1.3431	18.6960 / 0.6265	
F_{et}	0.0000	1.1938 / 0.0000	0.0000 / 0.0000	
$T_{o,in}$	0.2221	2.2952 / 0.3270	0.4416 / 0.3722	2.440 / 2.380
$T_{o,out}$	1.2857	0.7882 / 2.0505	1.4296 / 2.4582	
$T_{et,in}$	1.8007	0.7805 / 2.0247	13.5277 / 12.8131	
$T_{et,out}$	0.0480	0.0000 / 0.5611	0.2573 / 0.6537	

Noted: Used confidence interval, Chi-Squared values and statistical test criterions at 10% level of significant (For statistical criterion, 7 measured variables: β = 1.493 x 10-2 and 6 measured variables: β = 1.740 x 10-2)

The conventional gross error detection for three cases described above were tested and their results given in Table 2 and 3. From Table 2, the 1st reconciled objective function value of NG was 0.206, which is less than the criterion value (Chi-squared value with 2 degree of freedom, 4.605) but cases of 1P and 2P (16.892 and 10.148, respectively) are higher, so the gross errors were detected, the MT was applied to detect the position of gross error resulting in Table 3, the measurement adjustment of hot oil volumetric flowrate for case of 1P and 2P (-58.022 and -61.082, respectively) fall outside the confidence interval (±57.479) meaning the gross error occurring. When the gross error eliminations were completed, the 2nd reconciled objective function value of 1P was 0.237, which is less than the criterion but cases of 2P (4.462) are higher, so the gross error still existed in the case of 2P, but case of 1P was completed because its objective function value is less than the criterion value (Chi-squared value with 1 degree of freedom, 2.706).

The modified measurement test using NLP for three cases described above were tested and results were given in Table 4. Comparing the results with the statistical test criterion, the results show the same way as previous technique and the magnitudes of statistical test values, $Z_{d\ or\ a,j}$ of 1st run at hot-oil volumetric flowrate for cases of 1P and 2P (17.7215 and 18.6960, respectively) are most far from statistical test criterion (2.440) when compared to the others, so these variables can be identified to contain gross errors. When the gross error eliminations were completed, the $Z_{d\ or\ a,j}$ of 2nd run at ethane product inlet temperature for case of 2P was 12.8131 which is most far from the criterion (2.380) when compared to the others, so gross errors still existed in the system at this variable but DR cannot perform for this case because DOR will be less than 1, in contrast, for case of 1P, there were no the statistical test values exceeding the test criterion for all variables, meaning no gross errors existing in the system and the process was completed since the statistical test values for all measured variables are less than the criterion value (Statistical test criterion for 6 measured variables, 2.380). After the process was completed, the performance evaluation for 1P was used and the results are shown in Table 5.

Table 5: Performance evaluation of 3 conditions of data reconciliation with gross error detection

	DR: Only Random Error	DR: Random/Gross Error	DR/GED: Random/Gross Error
SD in Measurement Error	60.347	60.347	60.347
SD in Reconciled Error	5.159	7.820	7.336
% SD Reduction	91.15 %	87.04 %	87.84 %

Figure 4: SD reduction percentage of each *case*

Table 6: The GED performance evaluation

Techniques	Overall Power	
	1 Position	2 Positions
Conventional GED	1	0.5
Modified MT using NLP	1	1

**Overall power (OP) is equal to no. of gross error correctly identified divided by no. of gross error simulated.*

The GED performances between 2 techniques were calculated in term of overall power (OP)*, proposed by Narasimhan and Mah (1987) and the performance results are shown in Table 6.
From Table 5 and Figure 4 show the performance evaluation of each case, the standard deviation reduction percentage in case of DR combined with GED give higher value than case of only DR and Table 6 shows the OP of the modified MT using NLP is higher than the conventional GED technique for case of 2P.

6. Conclusion

The DR combined with GED technique gives more accurate estimates of process variables consistent with the constraints when compared to DR without GED, showing reductions in standard deviation. The GED techniques, the conventional method and the modified MT using NLP, were studied. The performance of the modified MT was significantly better than the conventional method, in terms of OP evaluation.

Acknowledgement

Authors would like to express our gratitude to the Petroleum and Petrochemical College, Chulalongkorn University, National Center of Excellence for Petroleum, Petrochemicals and Advanced Materials and Ratchadaphiseksomphot Endowment Fund for financial support.

References

Kim W., Kani S., Park W., Edgar T.F., 1997, Robust Data Reconciliation and Gross Error Detection: the Modified MIMT using NLP, Computers & Chemical Engineering, 21(7), 775-782.

Singhmaneeskulchai P., Angsutorn N., and Siemanond K., 2013, Dynamic Data Reconciliation in a Hot-oil Heat Exchanger for Validating Energy Consumption, Chemical Engineering Transactions, 35, 493-498, DOI: 10.3303/CET1335082.