DATA VALIDATION FOR REAL TIME OPTIMISATION CYCLE

H. W. Lee[*] and A. Ahmad

Laboratory of Process Control, Faculty of Chemical and Natural Resources Engineering, University Teknologi Malaysia, 81310 Skudai, Johor, Malaysia.
[*]Corresponding author. Phone: +607-5535877, Fax: +607-5581463
Email: leehw_78@hotmail.com

ABSTRACT

As a prerequisite for RTO implementation, data collected from a process plant must be treated by three important data validation stages, i.e., steady state detection, data reconciliation and gross error detection to eliminate errors and inconsistencies. This paper discusses the development of the data validation stage for real time optimization in a fatty acid fractionation process. To evaluate the proposed data validation system, single and multiple gross errors in the form of leakages in process stream were imposed to the process. The results proved the capability of the proposed tool in identifying and rectifying all the gross errors introduced to the process.

Keywords: Data Validation, Steady State Detection, Data Reconciliation, Gross Error Detection.

1 INRODUCTION

Process industries have been undergoing substantial changes in order to cope with new challenges resulting from high energy and manpower costs, strict safety and environmental regulations, stringent product specification and scarcity of reduced variation feedstock as well as stiff competition from new players. Due to these reasons, real time optimisation (RTO) has been brought forward for chemical industries as potential solutions to the increasingly intense production challenge.

Genichi *et al.* (1994) proposes that the second stage in the real time optimisation (RTO) cycle is data validation. Data validation is consisted of three major components, which are steady state detection, data reconciliation and gross error detection.

RTO moves processes from one steady state operating condition to another setting that are more profitable. Therefore, steady state detection is required to determine the condition of the process and the time required for an optimisation cycle. RTO is only implemented if the plant is at steady state and the plant must therefore be allowed to settle so that the desired condition is obtained. Darby and White (1988) proposed that the time period between two real time optimisation executions must be much longer than the process settling time to ensure that the process returns to steady state operation before optimisation is conducted again.

Measurements of process variables such as flow rates, pressures, levels, concentrations and temperatures in a chemical process are subject to error, both random and gross. These errors are often unavoidable and they may foul signals during measurements, processing and transmission stages. Random and gross errors can lead to deterioration in the performance of the control systems. At the limit, larger gross error can even nullify gains achievable through process optimisation. These erroneous data can even drive the process to uneconomic or unsafe operating regime. As a result, measures to eliminate these effects must therefore be in place to ensure the success of any optimisation tasks. Data reconciliation and gross error detection were required to accomplish these important tasks.

Data reconciliation is developed to improve the accuracy of measurements by reducing the effect of error in the data. Meanwhile, these measurements are adjusted to satisfy the conservation laws and the balance constraints (e.g. energy, mass, etc.). This was accompanied by a gross error detection mechanism to remove any gross error from the measurement.

2 DATA VALIDATION

2.1 STEADY STATE DETECTION METHOD

A number of methods have been proposed to certify that the required steady state condition is already achieved prior to the execution of the optimisation cycle. In this study, two methods were considered. The first was the composite statistical test proposed by Narasimhan *et al.* (1986) and the second was the mathematical theory of evidence (MTE) suggested by Narasimhan *et al.* (1987).

In this paper, the mathematical theory of evidence (MTE) is used. The MTE is defined in the following equation:

$$t_{1,i}^2 = \frac{N(\bar{x}_{2,i} - \bar{x}_{1,i})^2}{(S_{1,i}^2 + S_{2,i}^2)} \quad (1)$$

where $t_{1,i}^2$ is the random variables which obeying the Hotteling's T^2 distribution with numerator degrees of freedom 1, and denominator degrees of freedom $2N-2$. A level of significance, α is chosen, and $T_{1,2N-2}^2$ (α) is calculated to be the upper value of the T^2 distribution. If $t_{1,i}^2 \leq T_{1,2N-2}^2$, the plant is assumed to be at steady state with respect to the variable i. If $t_{1,i}^2 \geq T_{1,2N-2}^2$, the plant is taken as not being at a steady state condition.

2.2 DATA RECONCILIATION

Data reconciliation is developed to improve the accuracy of measurements by reducing the effect of error in the data. Meanwhile, these measurements are adjusted to satisfy the conservation laws and the balance constraints (e.g. energy, mass). Normally, data reconciliation is a constrained minimisation problem. The type of data reconciliation that is required depends on the problem and process units involved. When only process flowrates are reconciled, a linear data reconciliation problem is applied. On the other hand, when composition, temperature or pressure measurements are reconciled along with flowrates, a nonlinear data reconciliation scheme is required. In this case, nonlinear data reconciliation is required because it is dealing with process nonlinearity such as thermodynamic equilibrium relationship and complex correlation for the thermodynamic and other physical properties. The general nonlinear data reconciliation problem can be formulated as a weighted least square minimisation (Narasimhan, 2000):

$$\underset{x}{Min}\; w(y-x)^T \Sigma^{-1}(y-x) \quad (2)$$

Subject to

$$f(x) = 0$$

$$g(x) \leq 0$$

where f is m x 1 vector of equality constraints, g is q x 1 vector of inequality constraints, Σ is n x n variance-covariance matrix, x is n x 1 vector of measured variables, y is n x 1 vector of measured values of measurements of variables x and w is n x 1 vector of weighting factor of measurements of variables x

2.3 GROSS ERROR DETECTION

The most commonly used method for detecting gross error is statistical hypothesis testing that require selecting a statistic for the test with a known distribution and performance characteristics. A gross error is declared if the computed test statistic exceeds a critical value selected from a table of distribution. Among others, the statistical hypothesis tests include global test (GT), nodal or constraint test (NT), generalized likelihood ratio test (GLR), bounded generalized likelihood ratio (BGLR) method, measurements test (MT), iterative measurement test (IMT) and modified iterative measurement test (MIMT).

In this work, the Measurement Test (MT) method was used because it is suitable for RTO application. The test statistic is as given in Equation (3) below:

$$z_{e,i} = \left| \frac{e_i}{\sigma_i} \right| \tag{3}$$

Equation (3) also can be written as:

$$z_{e,i} = \left| \frac{y_i - x_i}{\sigma_i} \right| \tag{4}$$

Here, $z_{e,i}$ is the standardised measurement errors, e_i is the measurement errors, σ_i is the standard deviation of measurements, y_i is the reconciled data and x_i is the measurements.

3 PLANT SIMULATION

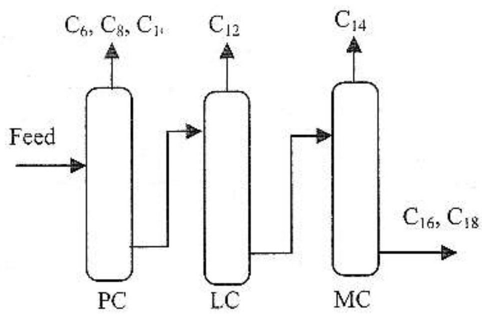

Figure 1. Schematic diagram of a fatty acid fractionation plant

The fatty acid fractionation (FAF) plant considered in this paper is consisted of three connecting packed distillation columns. The sequence of the columns is illustrated in Figure 1. Here, crude fatty acids are separated into its components through three columns, i.e., Pre-cut Column (PC), Light-cut Column (LC) and Middle-cut Column (MC). In the pre-cut column, light products (C_6, C_8 and C_{10} fatty acids) are recovered in the overhead. The bottom stream is then fed to the light-cut column where the C_{12} fatty acid is separated from the rest of the fatty acids. The bottom product of this column then enters the middle-cut column where the C_{14} fatty acid is recovered leaving the C_{16} and C_{18} fatty acids as bottom product.

In this study, the FAF plant was represented by a dynamic simulation model developed using HYSYS.Plant™ simulator. For the purpose of RTO implementation, the required steady state model was also developed using the same simulator. The former took the role of the actual plant whilst the latter provided the expected conditions if the plant was actually at a steady state. At the same time, the steady state model needed to provide systematic search for process optimisation especially in data reconciliation. Since the quality of both models plays pivotal role in producing the intended end-results, the models must therefore be accurate and robust.

To ensure the consistency of the model output, the steady state model must therefore be validated against the dynamic model. The results revealed that close agreement between the models with an overall difference of less than 3% was obtained and therefore considered adequate for this study.

4 EFFECT OF GROSS ERROR

In this section, gross errors were created by simulating process leak in pipelines. It was implemented by leaking the flowrate for the distillate stream at 60 minutes after the process in steady state condition. The gross error was detected as the distillate flowrate was one of the variables in the data reconciliation. The amounts of the leakage were highly dependent on the distillate flowrate. During implementation, care was taken not to let the distillate flowrate to dry out. The amounts of leakage were shown in table below:

Table 1. The amount of the leakage stream

Stream	Flowrate (Kg/h)
PC-Leakage	47.6579.
LC-Leakage	301.1780
MC-Leakage	124.3640

There were two ways of implementations. The first involved creating a single gross error by leaking the distillate stream either in PC, LC or MC column. The second approach introduced multiple gross errors by leaking all the distillate streams in the PC, LC and MC columns. The dynamic responses of these both cases were similar, either in single or multiple gross errors situation. Figure 2 illustrates the dynamic response of the process when multiple gross errors were introduced to the plant.

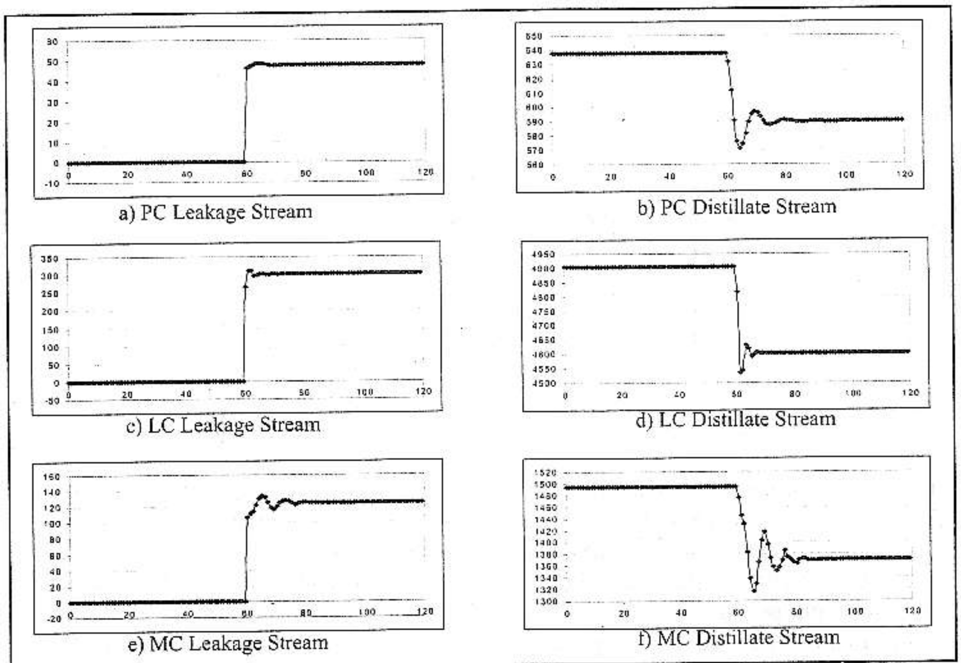

Figure 2. The dynamic response of leakage and distillate streams

4.1 STEADY STATE DETECTION

In this case, 6 key measurements were selected, i.e., the flowrate of the PC distillate stream, PC bottom stream, LC distillate stream, LC bottom stream, MC distillate stream and MC bottom stream. Each measurement was collected over a period of 30 minutes from the plant model (dynamic model). The mathematical theory of evidence method was used to carry out the intended task and the Hotteling T^2 was used to test the difference between the two means. The level of significance, α was specified at 0.05% and the $T^2_{12,2N-2}(\alpha)$ is 47.04 as determined from Equation 1. The value of $t^2_{1,i}$ for each measurement is lower than the $T^2_{1,2N-2}(\alpha)$. FAF process achieved steady state condition within the period of 30 minutes. Therefore, a 30 minutes or bigger cycle time can therefore useful for RTO.

4.2 DATA RECONCILIATION

It is important to reiterate here that the purpose of data reconciliation in this study is to support the RTO implementation where measured data must be reconciled to satisfy mass and energy balances. This is done only when the process is at a true steady state condition and any adjustment required to the process data would be made as small as possible to guarantee that the reconciled conditions are still accurately representing the true operating condition. Results of this implementation are shown in Table 2.

Table 2. The result of leakage study before and after data reconciliation

	Before Data Reconciliation	After Data Reconciliation
PC-Leakage	402.86	164.04
LC-Leakage	1184.30	1074.77
MC-Leakage	878.92	836.03
All-Leakage	1952.25	1865.10

The value of the objective function before data reconciliation was highly dependent to the amount of the distillate lost. It was proven from the table 2. The value of the objective function was large when all the distillate steams were leaked before data reconciliation is carried out. It is then followed by the leakage of the LC-distillate steam. It is due to the amount of the distillate lost were greater if compared to the lost of PC and MC distillate streams. The data reconciliation problem was solved using the Successive Quadratic Programming (SQP) algorithm provided by HYSYS optimiser. Data reconciliation was able to reduce these gross errors and the detection of the gross error was carried out by gross error detector.

4.3 GROSS ERROR DETECTION

The same 6 measurements that used in the steady sate detection were adopted in this study again. The plant data and the result of the data reconciliation were compared based on the critical value C. The critical value, C was determined based on the overall significant level, α which was specified as 0.05 (e.g. 95% of confidential interval), and the value was 2.8044 from the standard normal distribution with accumulated probability at 0.9989. If the value test statistic, $|z_{e,i}|$ exceeds the critical value C, then this measurement is said to contain gross error. Otherwise, this measurement is considered free of gross error. Measurements containing gross error are marked underline.

Flow Measurements (Kg/h)	**PC-Leakage**		**LC-Leakage**	
	Plant Data	***Reconciled Data***	***Plant Data***	***Reconciled Data***
PC Bottom	9400.83	9389.98	9400.83	9389.98
PC Distillate	589.93	617.53	589.93	617.53
LC Bottom	4483.59	4461.74	4483.59	4461.74
LC Distillate	4904.59	4890.82	4904.59	4890.82
MC Bottom	2982.89	2967.49	2982.89	2967.49
MC Distillate	1494.36	1474.77	1494.36	1474.77

Flow Measurements (Kg/h)	**MC-Leakage**		**All-Leakage**	
	Plant Data	***Reconciled Data***	***Plant Data***	***Reconciled Data***
PC Bottom	9400.83	9386.46	9400.83	9386.46
PC Distillate	637.59	622.39	637.59	622.39
LC Bottom	4483.49	4475.74	4483.49	4475.74
LC Distillate	4904.10	4826.07	4904.10	4826.07
MC Bottom	2983.02	2959.09	2983.02	2959.09
MC Distillate	1370.04	1471.62	1370.04	1471.62

In the PC-leakage case, one error, i.e., the PC distillate flowrate was detected. On the other hand, LC distillate flowrate was detected containing error in the LC-leakage case. Similarly, the MC distillate flowrate was identified in MC-leakage case. On the contrary, all the distillate streams were detected to contain gross error when these streams were subjected to leakage. By comparing to the three single gross error cases, the distillate streams after leakage were the same as the case of multiple gross error. All the gross errors in distillate streams were detected either in single or in multiple situations.

The distillate stream on each column was identified containing gross error when the leakage valve of the column was opened. As mentioned earlier, when the leakage valve was opened, the leakage stream started to flow causing the reduction in the distillate flowrate. The effect of gross error caused by the leakage was detected in the distillate stream.

5 CONCLUSIONS

Data validation is crucial to the success of the real time optimisation. Steady state detection had been applied to detect the process condition based on 6 keys measurements. When the process achieved its steady state condition, data reconciliation had been applied to adjust and reconcile the measurements in order to fulfil mass and energy balances. Gross error detection was used to identify the existing of the gross error in the measurements. As expected, gross error was detected in the distillate stream when the leakage of the distillation stream was occurred. These results exposed the ability of the data reconciliation and gross errors in the measurements.

ACKNOWLEDGMENTS

This project is funded by the Ministry of Science, Technology and the Environment through IRPA research grant. Our heartiest appreciations are for everybody who has directly or indirectly contribute to the success of this project.

REFERENCES

Darby, M. L., and White, D. C. (1988). "On–line Optimization of Complex Process Unit." Chemical Engineering Progress, 84(10), 51–59.

Genichi, E., Yodhinori, O., Hitoshi, M., Morimasa, O., Doughlas, B., Rosalyn, F. P., and John, S. A. (1994). "Integrated Advanced Control and Closed–Loop Real–Time Optimization of an Olefins Plant." Proceeding of IFAC Advanced Control of Chemical Processes, Kyoto, Japan, 95–100.

Narasimhan, S., Chen, S. K., and Mah, R. S. H. (1987). Detecting Changes of Steady State Using the Mathematical Theory of Evidence. *AICHE Journal.* 33(11): 1930–1932.

Narasimhan, S. and Harikumar, P. (1993). "A Method to Incorporate Bounds in Data Reconciliation and Gross Error Detection –I. The Bounded Data Reconciliation Problem." Computers and Chemical Engineering, 17(11), 1115–1120.

Narasimhan, S. and Jordache, C. (2000). Data reconciliation & Gross Error Detection: An Intelligent Use of Process Data. Gulf Publishing Company, Texas.