Cumhuriyet Üniversitesi Fen Fakültesi
Fen Bilimleri Dergisi (CFD), Cilt:36, No: 3 Özel Sayı (2015)
ISSN: 1300-1949

Cumhuriyet University Faculty of Science
Science Journal (CSJ), Vol. 36, No: 3 Special Issue (2015)
ISSN: 1300-1949

Data reconciliation and gross error detection: application in chemical processes

Mohammad Hosein EGHBAL AHMADİ

Research Assistant, Process Development and Equipment Technology Division, Research Institute of Petroleum Industry (RIPI), P.O. Box 18745-4163, Tehran, Iran

Received: 01.02.2015; Accepted: 05.05.2015

Abstract. Measured data are normally corrupted by different kinds of errors in many chemical processes. In this work, a brief overview in data reconciliation and gross error detection believed as the most efficient technique in reducing the measurement errors and obtaining accurate information about the process is presented. In addition to defining the basic problem and a survey of recent developments in this area that is categorized in "Real Time Optimization" field, we will describe about advanced optimization methods in nonlinear cases. At the end, implementation of data reconciliation is illustrated on a challenging process of Claus as a case study and as a result, a modified and consistent model with regard to measured data is presented by simultaneous estimation of key model parameters. In our case study, automation capability of ASPEN HYSYS is used to provide interface environment to reach global optimum.

Keywords: Data Reconciliation, Gross Error Detection, Optimization, Modeling, Claus Process.

1. INTRODUCTION

Today, with the advent of modern desktop computers, hundreds or thousands of process measurements are simultaneously measured and stored in massive storage media for the continuous monitoring of the process behavior and for performing process studies, such as development of robust process models and the online optimization of the process operation. Hence, Industrial plants may provide very detailed and rich data sets for development of fundamental modeling studies. Yet, measured data are normally corrupted by different kinds of errors. As a consequence, the collected data generally do not satisfy the process constraints, including mass and energy balances as well as other relationships between process variables. For this reason, implementation of data rectification procedures is essential for obtainment of accurate and reliable information about the process behavior. The data rectification procedure comprises different steps, namely, variable classification, data reconciliation (DR) and gross error detection (GED). Variable classification is required in order to determine if the available information is sufficient for solving the proposed reconciliation problem and to identify sets of observable and non-observable process variables. Different solutions have been proposed for the variable classification problem during the last four decades and detailed description of published procedures can be found in Romagnoli and Sanchez [1] or Narasimhan and Jordache [2].

Data reconciliation may be defined as the adjustment of process measurements and parameters in order to satisfy the process constraints, while minimizing some sort of objective function that formulated as deviations between corrected and observed plant values [3]. This procedure is used in different applications, such as process control and optimization, process monitoring, plant safety, plant operational efficiency improvement, fault detection, among others [4]. Consequently, data reconciliation represents an important step for many engineering

*Corresponding author. *Email address: eghbalahmadimh@ripi.ir*

Special Issue: The Second National Conference on Applied Research in Science and Technology

http://dergi.cumhuriyet.edu.tr/cumuscij

activities in industrial processes. It is necessary to mention that the principal difference between data reconciliation and other filtering techniques is that data reconciliation explicitly makes use of process model constraints, so that adjusted and estimated measurements satisfy the constraints.

Unfortunately, if we adjust the measurements to conform to the conservation laws but in the presence of biases, all of the adjustments are greatly affected by such biases and would not be reliable indicators of the state of the process. Thus biases, or generally gross errors, must be both detected and identified, then the measurements be corrected or discarded. Commonly, statistical tests for the treatment of gross errors are adopted [5].

In chemical engineering, Kuehn and Davidson [6] were the first to publish an analysis of data reconciliation in the early sixties, presenting the general solution when all flows are measured. Later other researchers focus on more complex problems. There are various numbers of techniques and strategies which have been adopted for tackling the problem of data reconciliation as well as gross error detection (and identification) which are described in the following sections.

2. DATA RECONCILIATION – DR

Generally, data reconciliation is a constrained optimization problem. The constraints may be linear in the simplest case but are in general nonlinear. The objective function is some sort of quadratic form in the adjustments to the measurements that is a general weighted sum of squares. Data reconciliation procedure is based on measurement redundancy. This concept may be classified in two types of sensor and topological (spatial) redundancy. The former arising from multiple sensors of the same quantity at the same time at the same place and the latter related to model information where a single variable can be estimated in several independent ways, from separate sets of measurements with regard to plant structure. Data reconciliation uses information redundancy and conservation laws to correct measurements and convert them into accurate and reliable information. As a result, the reconciled values exhibit a lower variance compared to original raw measurements.

The classical general data reconciliation problem deals with a weighted least-squares minimization of the measurement adjustments subject to the model constraints:

$$Ax = 0 \qquad A \in R^{n,v}, \quad x \in R^{v} \tag{1}$$

Where x is the state of the process and A describes the constraints with rank (A) = n according to the fact that the constraints are independent. The measurement devices give the information

$$\hat{x} = x + \varepsilon \qquad \varepsilon : N(0, V) \tag{2}$$

Where ε is a vector of random errors characterized by a normal probability density function (pdf) with a diagonal variance matrix V. Solution procedure to minimize a classical least-squares objective function (eq.(3)), subject to process constraints (steady state or dynamic constraints), gives the accurate estimate of x.

Obj. Fun:

$$\min_{x} f = \frac{1}{2} \sum_{i=1}^{v} \left(\frac{x_i - \hat{x}_i}{\sigma_i} \right)^2 \tag{3}$$

Problems related to DR could be categorizes in linear, bilinear and nonlinear systems. If we wish to take into consideration nonlinear relationships such as thermodynamic equilibrium ones, nonlinear data reconciliation (NDR) must be used. In addition if we impose bounds on the estimate of variables or other feasibility constraints, constrained data reconciliation solution techniques must be used.

In chemical engineering, Kuehn and Davidson [6] were the first to publish an analysis of data reconciliation. Since then, many articles on this subject have appeared in the literature. General reviews of data reconciliation have been published by Hlavacek [7], Tamhane and Mah [8], Mah [9], Madron [10], Crowe [5] and Bagajewicz [11]. The minimization of objective function subject to equality or inequality constraints can be achieved by using optimization techniques. Method of Lagrange multiplier [12], using Newton–Raphson iterative method which is based on a quasi-Newton linearization of the nonlinear model [13], method of successive linear data reconciliation [14], Nonlinear Programming method (NLP) such as Sequential Quadratic Programming (SQP) [15&16], Generalized Reduced Gradient (GRG) method [17&18] among many others are solution methods which are reported in literature.

3. OPTIMIZATION ALGORITHMS FOR DATA RECONCILIATION

Data reconciliation problems are inherently difficult to solve with conventional optimization methods especially in nonlinear cases, because of the likely existence of multiple local minima [20]. Heuristic optimization methods, such as Particle Swarm Optimization (PSO) [21], Simulated Annealing (SA) and Genetic Algorithm (GA) can be used to overcome these difficulties. The first study regarding the use of heuristic approaches to solve DR problems was reported by Parsopoulos et al. [22]. These methods also present some additional advantages, such as the global character of the search (which avoid local minima), the unnecessary computation of derivatives and the simplicity of the implementation, although they are usually characterized by the high number of objective function evaluations, which may require more CPU time than conventional methods. However, the computed values of the objective function can be used for rigorous statistical analyses of the confidence regions of parameter estimates, which can also constitute an important benefit of these algorithms [21].

4. GROSS ERROR

Commonly, the technique of classical data reconciliation depends on the assumption that only random errors are present in the data and systematic or gross errors (instrument biases and leaks) either in the measurements or the model equations are not present [2]. Detection of gross errors is of fundamental importance for adequate model building and interpretation of process data, as poor input data generally lead to very poor model responses.

Any comprehensive gross error treatment strategy should preferably possess the capabilities of detecting the presence of one or more gross errors in data, identifying the type and location of the gross error, locating and identifying multiple gross errors which may be present simultaneously in the data and finally (but not necessary) ability to estimate the magnitude of the gross errors. There are two main approaches for treatment of gross errors. In the first approach, prior to data reconciliation operations, gross errors have to be detected and eliminated (or compensated). Another approach is to treat gross errors simultaneously with data reconciliation problem. It could be performed by utilizing robust objective functions which are less sensitive to large errors or adding terms in objective function, and as a result, some sort of penalty is imposed for increasing the number of gross errors. In this approach, the

measurements containing a gross error are not removed prior to solving the data reconciliation problem [23].

The most common test used for detection of gross errors is statistical hypothesis testing, which requires selection of a proper statistics for the test. A gross error is declared when the computed test statistics exceeds a critical value. Many traditional methods, such as the Global Test (GT) [24], the Measurement Test (MT) [25], the Nodal Test (NT) [24], the Generalized Likelihood Ratio Test (GLR) [26], the Maximum Power Test (MP) [27], the Bonferroni Test [28], and Principal Component Test [29], among others, have been developed to handle the problem of detecting single or multiple gross errors in raw measured data. Combinations among these statistical tests have also been proposed [30&31].

In order to deal simultaneously with the DR and multiple gross error detection (MGED) problems, alternative approaches based on cluster analysis [19] and robust estimators [23&32] have been proposed in the literature. The use of robust estimators for data analysis is becoming increasingly popular, including real industrial applications, as some of examples are illustrated in Table 1. These studies have shown the many advantages of using robust statistics to perform DR in the presence of gross errors, such as the low sensitivity to possible model mismatches and atypical measurements.

Table 1. Examples of robust estimators used for DR and GED

References	Estimator
Johnson and Kramer [33]	Contaminated normal and Lorentzian
Zhou et al [34]	Huber
Arora and Biegler [35]	Fair and Hampel
Lid and Skogestad [36]	Contaminated Normal
Wang and Romagnoli [37]	Generalized T (GT)
Siyi Jin et al [38]	New target function

The benefits derived from data reconciliation and gross error detection in chemical processes are numerous. They include improvement of measurement layout, fewer routine analyses, reduced frequency of sensor calibration (only faulty sensors need to be calibrated), removal of systematic measurement errors, systematic improvement of process data, a clear picture of plant operating conditions, and reduced measurement noise for key variables. Moreover, monitoring through data reconciliation leads to early detection of sensor deviation and equipment performance degradation, actual plant balances for accounting and performance follow-up, safe operation closer to the process limits and improved quality and performance at the process level.

5. INDUSTRIAL APPLICATIONS OF DATA RECONCILIATION AND GROSS ERROR DETECTION

The number of works that report the use of DR and GED techniques in real industrial applications is small. This is particularly true when one considers problems with the simultaneous estimation of model parameters [3]. Some of these were based on actual data in real processes and some on artificial data. Relatively fewer applications have been published on raw plant data. Table 2 presents some of works which have been applied for industrial plants.

Table 2. Real industrial applications of DR and GED techniques.

References	Industrial plant
Sanchez et al. [39]	Ethylene plant
Dempf and List [40]	Vinyl acetate and ketene plant
Bagajewicz and Cabrera [41]	Gas pipeline systems
Al-Arfaj [42]	Methyl-terc-butyl-ether plant
Hu and Shao [43]	Coking plant
Shunji. H et al [44]	Nuclear plan

6. CASE STUDY

As a case study, in this section application of data reconciliation using a new approach by applying automation capability of a simulator on Claus process is illustrated.

The Claus process that is a part of sulfur recovery unit (SRU) is a chemical process for converting gaseous hydrogen sulfide (H_2S) found in sour gases derived from refining petroleum crude oil and other industrial facilities into elemental sulfur. In the straight through process, up to 70% of sulfur conversion occurs in the reaction furnace (RF) and waste heat boiler (WHB) that are referred as thermal stage in Claus process. The cooled gases are sent to catalytic stages for further sulfur recovery [45].

Fig.1 depicts typical thermal stage in Claus process.

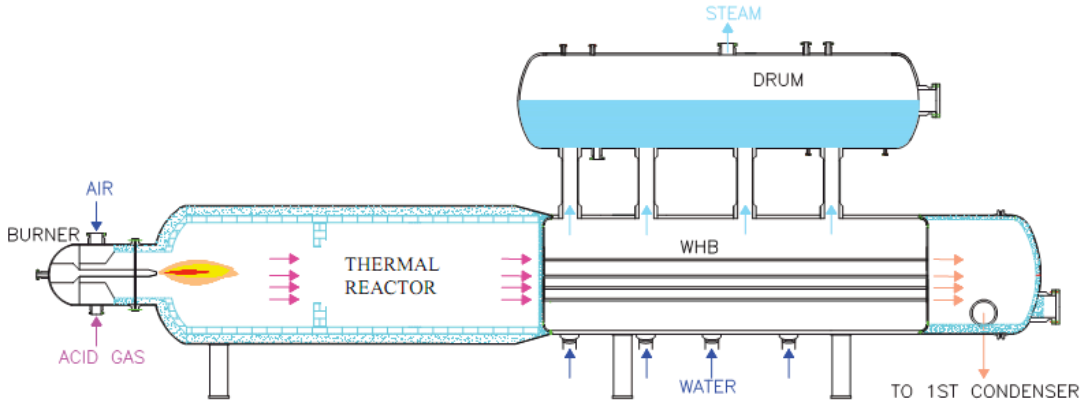

Figure 1. Thermal stage (RF +WHB) of Claus process [46].

Numerous known and unknown reactions occur in the RF. Both methods of using equilibrium and empirical correlations fail to accurately predict the behavior of RF. It is due to not considering the kinetic limitations and complexity of reactions encountered in the actual RF as well. Kinetic modeling was proved as the best approach to model reaction furnace [47&48]. However, there is not a comprehensive and accurate kinetic model data in open literature. The main reactions in RF are presented in table 3.

Table 3. Main reactions in RF.

1) $CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$	7) $2H_2S + SO_2 \leftrightarrow 1.5S_2 + 2H_2O$
2) $CH_4 + 1.5O_2 \rightarrow CO + 2H_2O$	8) $H_2S \leftrightarrow 0.5S_2 + H_2$
3) $C_2H_6 + 3.5O_2 \rightarrow 2CO_2 + 3H_2O$	9) $2CO + S_2 \leftrightarrow 2COS$
4) $3H_2S + O_2 \rightarrow 2H_2S + SO_2 + H_2O$	10) $CH_4 + 2S_2 \rightarrow CS_2 + 2H_2S$
5) $H_2S + 0.5O_2 \rightarrow 0.5S_2 + H_2O$	
6) $H_2S + O_2 \rightarrow H_2 + SO_2$	

Four categories of measurements were taken at 4 different times for entering and leaving streams of RF which are assumed to be free of gross error are considered as measured data. In order to properly predict the behavior of RF, by using data reconciliation technique, the accuracy of measurements were improved and kinetic parameters were estimated simultaneously. As a result, modified model of RF which is consistent to measurement data is presented. To increase the accuracy and reliability of optimum results, ASPEN HYSYS was applied as simulator and its connectivity capability with optimization algorithm was utilized.

For simulating the RF, combustion reactions are defined in Gibbs reactors (reactions 1 to 6), while the kinetic reactor was chosen for other reactions (reactions 7 to 10). The best kinetic models of these reactions taken from literature [49&50] are presented in following.

reaction 7

$$r_{H_2S} = k_1 \exp(-E_1 / RT) P_{H_2S} P_{SO_2}{}^{0.5} - k_2 \exp(-E_2 / RT) P_{H_2O} P_{S_2}{}^{0.75}$$

$k_1 = 488.69 \quad (mol.m^{-3}.s^{-1}.\mathrm{Pa}^{-1.5}) \qquad E_1 = 208580 \quad (\mathrm{J}.mol^{-1})$

$k_2 = 0.879 \quad (mol.m^{-3}.s^{-1}.\mathrm{Pa}^{-1.75}) \qquad E_2 = 187682 \quad (\mathrm{J}.mol^{-1})$

recation 8

$$-r_{H_2} = k_3 \exp(-E_3 / RT) C_{H_2} C_{S_2}, \qquad -r_{H_2S} = k_4 \exp(-E_4 / RT) C_{H_2S} C_{S_2}{}^{0.5}$$

$k_3 = 3.46 \times 10^6 \quad (m^3.mol^{-1}.s^{-1}) \qquad E_3 = 131300 \quad (J / mol^{-1})$

$k_4 = 2.26 \times 10^9 \quad (m^{1.5}.mol^{-0.5}.s^{-1}) \qquad E_4 = 216600 \quad (J / mol^{-1})$

reaction 9

$$r_{COS} = k_5 \exp(-E_5 / RT) C_{CO} C_{S_2} - 2k_6 \exp(-E_6 / RT) C_{COS}^{2(n)}$$

$k_5 = 318 \quad (m^3.mol^{-1}s^{-1}) \qquad E_5 = 55703.8 \quad (J.mol^{-1})$

$k_6 = 2.18 \times 10^6 \quad (m^3.mol^{-1}s^{-1}) \qquad E_6 = 179831.8 \quad (J.mol^{-1})$

reaction 10

$$r_{CS_2} = k_7 \exp(-E_7 / RT) C_{CH_4} C_{S_2}$$

$k_7 = 5.53 \times 10^7 \quad (m^3.mol^{-1}.s^{-1}) \qquad E_7 = 160630 \quad (J.mole^{-1})$

Between above kinetic parameters, some are less precise than others and were candidate for being re-optimized through data reconciliation procedure.

Optimum values for kinetic parameters as well as reconciled data were obtained for RF (Fig.2) are presented in table 4. It should be mentioned that averaged data during 4 different times were reported in table 4. SQP algorithm was applied to achieve optimal solution.

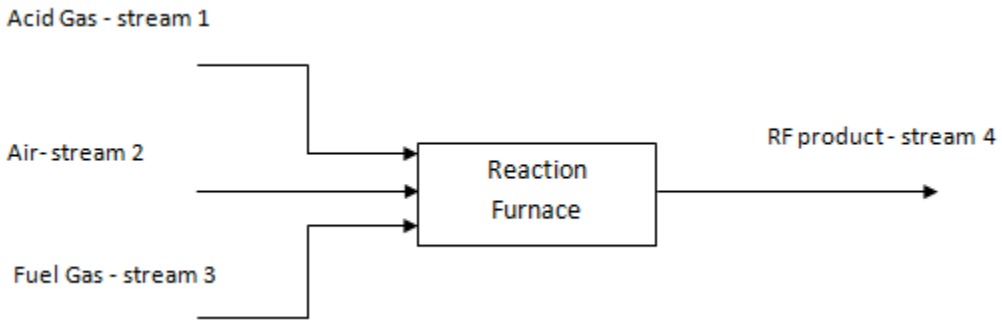

Figure 2. RF schematic view.

Table 4. Reconciliation solution for RF.

Measurements									
Var	Name	Before reconciliation				After reconciliation			
		Date 1	Date 2	Date 3	Date 4	Date 1	Date 2	Date 3	Date 4
Flow/ kgmol.s^{-1}	Stream 1	0.00200	0.00194	0.00193	0.00187	0.00186	0.00186	0.00193	0.00193
	Stream 2	0.00347	0.00371	0.00374	0.00360	0.00365	0.00365	0.00351	0.00351
	Stream 3	0.00022	0.00021	0.00022	0.00022	0.00021	0.00021	0.00022	0.00022
	Stream 4	0.00573	0.00573	0.00583	0.00578	0.00576	0.000576	0.00576	0.00576
Temperature/□ C	Stream 1	50.19	50.08	51.45	48.13	50.51	50.53	50.48	50.45
	Stream 2	57.59	59.81	60 .46	59.60	58.61	58.65	58.54	58.52
	Stream 3	25.58	25.84	26.27	24.97	25.25	25.26	25.25	25.24
	Stream 4	1214.19	1238.26	1144.01	1129.40	1188.24	1188.24	1188.24	1188.24

Parameters					
name	Optimum value	name	Optimum value	name	Optimum value
k_1	493.58	k_4	2280955772	k_7	55853000
k_2	0.888	k_6	2201800		
k_3	3494600	n	2.02		

7. CONCLUSIONS

Application of data reconciliation and gross error detection in chemical engineering and some recent developments were reviewed. It was discussed that the more complex problems arising from nonlinear cases need the more robust algorithms such as using heuristic optimization methods and new strategies for detection of gross errors. Afterward, some industrial case studies reported in literature were presented and it was demonstrated that number of works that report the use of DR and GED techniques in real industrial applications is small particularly in cases of simultaneous estimation of model parameters. Finally application of data reconciliation on a reaction furnace (RF) of Claus process has investigated. Having utilized the automation capability of ASPEN HYSYS as a simulator, key kinetic parameters of RF have been estimated so that a consistent and up to dated model was presented considering reconciled measured data.

REFERENCES

[1] Romagnoli, J. A., & Sanchez, M. C., Academic Press, (1999).

[2] Narasimhan, S., & Jordache, C, Gulf Professional Publishing, (1999).
[3] Prata, D. M., Schwaab, M., Lima, E. L., & Pinto, J. C., Chem. Eng. Sci., 64, 3953, (2009).
[4] Martinez Prata, D., Schwaab, M., Luis Lima, E., & Carlos Pinto, J., Chem. Eng. Sci., 65, 4943, (2010).
[5] Crowe, C. M., J. Process. Contr, 6, 89, (1996).
[6] Kuehn, D. R., & Davidson, H., Chem. Eng. Prog, 57, 44, (1961).
[7] Hlaváček, V., Comput. Chem. Eng., 1, 75, (1977).
[8] Tamhane, A. C., & Mah, R. S., Technometrics, 27, 409, (1985).
[9] Mah, R. S., & Mah, R. S. Butterworths, Boston, 1990.
[10] Madron, F., Ellis Horwood, Chichester, 1992.
[11] Bagajewicz, M. J., Latin. Am. Appl. Res, 30 (4) (2000) 335-342.
[12] Britt, H. I., & Luecke, R. H., Technometrics, 15, 233, (1973).
[13] Stephenson, G. R., & Shewchuk, C. F., AICHE. J, 32, 247, (1986).
[14] Serth, R. W., Valero, C. M., & Heenan, W. A., Chem. Eng. Commun, 51, 89, (1987).
[15] Han, S., J. Optimiz. Theory. App. 22, 297, (1977).
[16] Chen, H. S., & Stadtherr, M. A., Comput. Chem. Eng, 8, 229, (1984).
[17] J. Abadie, Sijthoff and Noordhoff, Holland, (1978).
[18] Lasdon, L. S., & Waren, A. D., Sijthoff and Noordhoff, Holland, (1978).
[19] Chen, J., & Romagnoli, J. A., Comput. Chem. Eng, 22, 559, (1998).
[20] Wongrat, W., Srinophakun, T., & Srinophakun, P., Comput. Chem. Eng, 29, 1059, (2005).
[21] Schwaab, M., Biscaia Jr, E. C., Monteiro, J. L., & Pinto, J. C., Chem. Eng. Sci, 63, 1542, (2008).
[22] Parsopoulos, K. E., Laskari, E. C., & Vrahatis, M. N. Proceedings of the Conference on Artificial Intelligence and Applications, USA, (2001).
[23] Özyurt, D. B., & Pike, R. W., Comput. Chem. Eng, 28, 381, (2004).
[24] Reilly, P. M., & Carpani, R. E, Proceedings of the 13th Canadian Chemical Engineering Conference, Montreal, Quebec, 1963.
[25] Mah, R. S. H., & Tamhane, A. C., AICHE. J, 28, 828, (1982).
[26] Narasimhan, S., & Mah, R. S. H., AICHE. J, 33, 1514, (1987).
[27] Crowe, C. M., AICHE. J, 35, 563, (1985).
[28] Rollins, D. K., & Davis, J. F., AICHE. J, 38, 563, (1992).
[29] H. Tong, C.M. Crowe., AICHE. J, 41, 1712, (1995).
[30] Yang, Y., Ten, R., & Jao, L. Comput. Chem. Eng, 19, 217, (1995).
[31] Wang, F., Jia, X. P., Zheng, S. Q., & Yue, J. C. Comput. Chem. Eng, 28, 2189, (2004).
[32] Prata, D. M., Pinto, J. C., & Lima, E. L., Comp. Aid. Ch, 25, 501, (2008).
[33] Johnston, L. P., & Kramer, M. A., AICHE. J, 41, 2415, (1995).
[34] L. Zhou, H. Su, J. Chu, Chinese. J. Chem. Eng, 14, 357, (2006).
[35] Arora, N., & Biegler, L. T., Comput. Chem. Eng, 25, 1585, (2001).
[36] Lid, T., & Skogestad, S. J. Process. Contr, 18, 320, (2008).
[37] Wang, D., & Romagnoli, J. A., Ind. Eng. Chem. Res, 42, 3075, (2003).
[38] Jin, S., Li, X., Huang, Z., & Liu, M., Ind. Eng. Chem. Res, 51, 10220, (2012).
[39] Sanchez, M., Bandoni, A., & Romagnoli, J., Comput. Chem. Eng, 16, 499, (1992).
[40] Dempf, D., & List, T. Comput. Chem. Eng, 22, S1023, (1998).
[41] Bagajewicz, M. J., & Cabrera, E. Ind. Eng. Chem. Res, 42, 5596, (2003).

[42] Al Arfaj, M. A. AICHE. J, 52, 414, (2006).
[43] Hu, M., & Shao, H., Ind. Eng. Chem. Res, 45, 8973, (2006).
[44] Homma, S., Watanabe, M., Koga, J., & Matsumoto, S., J. Nucl. Sci. Technol, 48, 1285, (2011).
[45] K.A. Hawbolt, Ph.D. thesis, The University of Calgary, Canada, (1998).
[46] Manenti, G., Papasidero, D., Manenti, F., Bozzano, G., & Pierucci, S., Procedia. Eng., 42, 376, (2012).
[47] Karan, K., Mehrotra, A. K., & Behie, L. A. Ind. Eng. Chem. Res, 37, 4609, (1998).
[48] Monnery, W. D., Hawboldt, K. A., Pollock, A. E., & Svrcek, W. Y., Ind. Eng. Chem. Res, 40, 144, (2001).
[49] Monnery, W. D., Hawboldt, K. A., Pollock, A., & Svrcek, W. Y. Chem. Eng. Sci, 55, 5141, (2000).
[50] Hawboldt, K. A., Monnery, W. D., & Svrcek, W. Y., Chem. Eng. Sci, 55, 957, (2000).