International Journal of Innovative Computing, Information and Control
Volume **8**, Number **6**, June **2012**

ICIC International ©2012 ISSN 1349-4198
pp. 4111–4122

EXTENDED SUPPORT VECTOR REGRESSION BASED DATA RECONCILIATION AND ITS APPLICATION TO PLANT-WIDE MASS BALANCE

Hongren Zhan[1,2], Yu Miao[3] and Wei Wang[3]

[1]School of Materials and Metallurgy
Northeastern University
Shenyang 110004, P. R. China

[2]School of Mechanical Engineering
Shenyang Institute of Chemical Technology
No. 11 Street, Economic and Technological Development Zone, Shenyang 110142, P. R. China
z_hr555@163.com

[3]College of Control Science and Engineering
Dalian University of Technology
No. 2, Linggong Road, Ganjingzi District, Dalian 116024, P. R. China
{ miaoyu; wangwei }@dlut.edu.cn

Received October 2010; revised March 2011

Abstract. *Process data measurements are important for process monitoring, control and optimization. However, process data may be deteriorated by gross errors in measurements. Therefore, it is significant to detect and estimate gross errors with data reconciliation. Meanwhile, in any modern petrochemical plant, the plant-wide mass data derived from process data rendering the real conditions of manufacturing are the key to the operation managements such as production planning, production scheduling and performance analysis. In this paper, an extended support vector regression approach for data reconciliation and gross error detection is proposed and applied to deal with the plant-wide mass balance problem. The proposed approach could simultaneously detect and estimate gross errors like measurement bias and process leaks. Then the proposed approach is applied to address the plant-wide mass balance problem with measurement bias and mass movement information lost, because of its superior characteristic for the issue. Both simulation and application results in this paper demonstrate that the proposed approach is accurate and effective to address plant-wide mass balance.*

Keywords: Data reconciliation, Support vector regression, Parameter estimation, Gross error detection

1. **Introduction.** The function of plant-wide mass balance in a refinery is to help production manager to figure out reason of production loss, to calculate accurate product amounts inlet and outlet units, and to provide necessary supports to production performance analysis, production plan optimization and production scheduling optimization. However, it is hard to obtain plant-wide mass balance directly from raw process data, because process data are inevitably corrupted by errors during measurement and processing, as makes the process data deviated from their true values and do not obey the mass balance and other constraints.

Process data are usually corrupted by two types of errors – random errors and gross errors. Random errors cannot be completely eliminated and always present in any measurement. They cannot be predicted with certainty. The only possible way to characterize these random errors is to use probability distributions. On the other hand, gross errors are caused by nonrandom events and can be divided into two categories, namely measurement related such as mal-functioning sensors and process related such as process leaks that can be seen as mass movement information lost in plant-wide mass balance.

Since errors in process data could lead to significant deterioration in plant-wide mass balance, it is very important to minimize the effects of both random and gross errors.

Data reconciliation and gross error detection is a technique to improve the accuracy of process data by reducing the effect of random and gross errors in measurements and process model. Many methods have been developed to form data reconciliation and gross error detection. The most widely used methods for data reconciliation and gross error detection are the global test (GT) [1], the measurement test (MT) [2], the nodal test (NT) [1], the generalized likelihood ratio (GLR) [3], and the principal component test (PCT) [4]. Several strategies were developed to identify multiple gross errors, such as serial elimination [5], serial compensation [3], simultaneous or collective compensation [6]. Some new methods have been applied to real industrial processes, which range from statistical test methods to robust statistics methods [7,8], from sequential or combinatorial methods to simultaneous data reconciliation and gross error detection methods. For dynamic systems, the data reconciliation can be considered as a state estimation technique, and some approaches have been proposed to address the issue in dynamic systems [9-11]. Meanwhile, process data may be collected within different sampling times, consequently estimators for multi-rate sampled systems recently invoked a research interest [12]. However, few methods can deal with process related gross errors such as leaks [13,14].

As mentioned above, plant-wide mass balance results in a refinery are derived from the raw process data and production movement network, which should obey the materials balance constraints of the production movement network through abstracting actual products into some virtual products. As a result, the plant-wide mass balance can be addressed as a data reconciliation problem. Therefore, several strategies have also been applied to address plant-wide mass balance problem. Zhang et al. have discussed the application of data reconciliation and measurement related gross error removal to mass balance in a refinery [15]. Wang et al. have proposed a two-stepped mass balance strategy based on data reconciliation and gross error detection [16]. Wang et al. introduced Bayesian network to simplify mass balance model so as to enhance the feasibility of data reconciliation and gross error detection used for achieving plant-wide mass balance in a refinery [17]. At the same time, there are some sophisticated commercial pieces of software for practical plant-wide mass balance, such as the *Aspen Operations Reconciliation and Accounting* and the *Honeywell Production Balance*. However, in practice, sometimes production movement information would get lost by non-systematic errors. In this case, the production movement information can be considered as a model related error like process leaks. As mentioned above, few researches and applications have focused on process leaks, so mass movement information lost for plant-wide mass balance in refineries has not been addressed yet.

Recently, a support vector (SV) regression approach for data reconciliation and gross error or outlier detection has been proposed [18], which has also been applied to achieve plant-wide mass balance with only measurement biases [19]. The SV regression approach is firmly grounded in the framework of statistical learning theory, and minimizes regular risk with VC dimension instead of empirical risk. In the SV regression based data reconciliation, measurement bias is considered to be related to the complex of the model or VC dimension so that it is very efficient for detecting measurement related gross errors and outliers. Furthermore, as the SV regression approach takes the linear objective function with a variable to adjust sparseness of the optimization problem, it is suitable for large-scale problem such as plant-wide mass balance which is usually derived from a linear material balance network. However, the SV regression approach mentioned above still cannot address process related gross errors like leaks in processes. Therefore, it cannot deal with mass movement information lost when applied in plant-wide mass balance.

In this paper, the SV regression based data reconciliation approach is extended, where process model related gross errors like process leaks are also enclosed into the VC dimension. The main contribution of this paper is that, by minimizing regular risk with the extended VC dimension, the extended SV regression based data reconciliation approach proposed is able to achieve not only data reconciliation but also joint measurement bias and process leak detection and estimation simultaneously, and it has been applied in plant-wide mass balance in practice, as draw little attention in previous researches but is a significant aspect of data reconciliation in practice. Due to applying the proposed approach on plant-wide mass balance in refinery, both measurement bias and production movement information lost could be detected and estimated, which is the key feature of the approach for application.

A simulation study is provided to demonstrate the validity and accuracy of the approach proposed to simultaneously detect and estimate measurement bias and process leaks. Moreover, an application of the proposed approach to a practice case is introduced to illustrate that the extended SV based approach is effective and accurate to fulfill plant-wide mass balance with measurement bias and production movement information missing.

2. Support Vector Regression Approach for Data Reconciliation and Gross Error Detection.

2.1. Simultaneous data reconciliation and measurement related gross errors detection based on SV regression. The measurement model without gross errors can be written as,

$$\mathbf{x}^{\mathrm{M}} = \mathbf{x} + \boldsymbol{\varepsilon} \tag{1}$$

$$\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{c} = 0 \tag{2}$$

where $\mathbf{x}$ is the vector of true values of the variables, $\mathbf{x}^{\mathrm{M}}$ is the vector of measurements, $\boldsymbol{\varepsilon}$ is the vector of random errors which are assumed to follow a normal distribution $\mathrm{N}(0, \boldsymbol{\sigma})$, and $\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{c} = 0$ is the process model constraint such as mass balance constraints, $\mathbf{A}$ is the incidence matrix of variables, $\mathbf{c}$ is the constant vector of the process model, $\mathbf{B}$ is the incidence matrix of the constants.

If gross errors are present in process measurements, the measurement bias model for gross errors of unknown magnitude $\boldsymbol{\mu}$ is given by

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\varepsilon} + [b_1\mu_1, b_2\mu_2, \cdots, b_i\mu_i, \cdots, b_n\mu_n]^{\mathrm{T}} \tag{3}$$

where n is the number of measured variables, b is an indicator, which indicates the location of biases in the model, and it can be defined as

$$b_i = \begin{cases} 0 & \text{if no gross error presents in the } i\text{th measurement} \\ 1 & \text{if gross error presents in the } i\text{th measurement} \end{cases} \tag{4}$$

Then a regression problem with a larger degree of freedom is formulated, and the SV regression approach for data reconciliation and measurement related gross error detection aims to estimate the function

$$\begin{aligned} &f(\mathbf{x}) = \mathbf{x} + \boldsymbol{\varepsilon} + [b_1\mu_1, b_2\mu_2, \cdots, b_i\mu_i, \cdots, b_n\mu_n]^{\mathrm{T}} \\ &b_i = \begin{cases} 0 & \text{if no gross error presents in the } i\text{th measurement} \\ 1 & \text{if gross error presents in the } i\text{th measurement} \end{cases} \end{aligned} \tag{5}$$

based on the measurements $\mathbf{x}^{\mathrm{M}}$ and process constraints

$$\mathbf{Ax} + \mathbf{Bc} = 0 \tag{6}$$

As the function of (3) is linear, according to the theory of statistical learning [20], the VC dimension of (5) is the number of free parameters, which defines the complexity of the measurement bias model and can be written as

$$\sum_{i=1}^{n} b_i \tag{7}$$

The SV regression approach for data reconciliation and measurement related gross error detection takes the form as

$$\begin{array}{ll} \min & \sum\limits_{i=1}^{n} b_i + C \sum\limits_{i=1}^{n} \left(\upsilon\varepsilon + \dfrac{(\xi_i + \xi_i^*)}{\sigma_i} \right) \\ \text{s.t.} & \begin{array}{l} \mathbf{Ax} + \mathbf{Bc} = 0 \\ x_i + \mu_i - x_i^{\mathrm{M}} \le \varepsilon_i + \xi_i \\ x_i^{\mathrm{M}} - (x_i + \mu_i) \le \varepsilon_i + \xi_i^* \\ \mu_i - U_i b_i \le 0 \\ -\mu_i - U_i b_i \le 0 \\ 0 \le x_i \le \mathrm{X}_i \\ 0 \le \upsilon \le 1 \\ \xi_i^{(*)} \ge 0, \quad \varepsilon \ge 0 \\ b_i \in \text{binary} \end{array} \end{array} \tag{8}$$

where n is the number of measured variables, x_i is the reconciled value of the ith measurement, $\mathbf{x}_i^{\mathrm{M}}$ is the ith measurement of the ith variable, σ_i is the standard deviation of the ith measurement, μ_i is the magnitude of measurement related gross error in the ith variable, U_i is the upper bound on gross error in the ith variable, ε_i is the error tolerant for the ith measurement variable, $\xi^{(*)}$ are slack variables which are penalized in the objective function, C and υ are coefficients which are chosen a priori, and b_i is a binary variable denoting existence of bias in the ith measurement.

From the SV regression form above, it can be seen that the approach does not consider process related gross errors. Therefore, when there is any process related gross error such as mass movement information missing in plant-wide mass balance, the approach would provide deviated balanced result. In order to overcome this demerit, the original SV regression approach should be extended to simultaneously address measurement related and process related gross errors.

2.2. **Extended SV regression approach to simultaneously estimate measurement related and process related gross errors.** According to (1) and (2), if there are measurement related gross errors and process related gross errors at the same time, the corresponding model should be

$$\begin{array}{l} \mathbf{x}^{\mathrm{M}} = \mathbf{x} + \boldsymbol{\varepsilon} + [b_1\mu_1, b_2\mu_2, \cdots, b_i\mu_i, \cdots, b_n\mu_n]^{\mathrm{T}} \\ \mathbf{Ax} + \mathbf{Bc} = [l_1\mu_{\mathrm{L}1}, l_2\mu_{\mathrm{L}2}, \cdots, l_j\mu_{\mathrm{L}j}, \cdots, l_m\mu_{\mathrm{L}m}]^{\mathrm{T}} \end{array} \tag{9}$$

where n is the number of measured variables, m is the number of mass balance constraints, μ_i is the magnitude of measurement related gross error in the ith variable, $\mu_{\mathrm{L}j}$ is the magnitude of process related gross error in the jth mass balance constraint. b and l are indicators, which indicate the location of biases in the measurements and mass movement information missing, respectively. They can be defined as

$$b_i = \begin{cases} 0 & \text{if no gross error presents in the } i\text{th measurement} \\ 1 & \text{if a gross error presents in the } i\text{th measurement} \end{cases}$$
$$l_j = \begin{cases} 0 & \begin{array}{l}\text{if no mass movement information missing} \\ \text{presents in the } j\text{th mass balance constraint}\end{array} \\ 1 & \begin{array}{l}\text{if mass movement information missing} \\ \text{presents in the } j\text{th mass balance constraint}\end{array} \end{cases} \tag{10}$$

Then the problem is reformed to estimate the function

$$\begin{array}{l} f(\mathbf{x}) = \mathbf{x} + \boldsymbol{\varepsilon} + [b_1\mu_1, b_2\mu_2, \cdots, b_i\mu_i, \cdots, b_n\mu_n]^{\mathrm{T}} \\ b_i = \begin{cases} 0 & \text{if no gross error presents in the } i\text{th measurement} \\ 1 & \text{if gross error presents in the } i\text{th measurement} \end{cases} \end{array} \tag{11}$$

based on measurement $\mathbf{x}^{\mathrm{M}}$ and constraints

$$\mathbf{Ax} + \mathbf{Bc} = [l_1\mu_{\mathrm{L}1}, l_2\mu_{\mathrm{L}2}, \cdots, l_j\mu_{\mathrm{L}j}, \cdots, l_m\mu_{\mathrm{L}m}]^{\mathrm{T}} \tag{12}$$

According to (7), the VC dimension of the extended form, which considers both measurement related and process related gross errors, can be written as

$$\sum_{i=1}^{n} b_i + \sum_{j=1}^{m} l_j \tag{13}$$

Furthermore, according to (8), the extended SV regression approach for data reconciliation to simultaneously address measurement related gross errors and process related gross errors takes the form as

$$\begin{array}{ll} \min & \left(\sum\limits_{i=1}^{n} b_i + \sum\limits_{j=1}^{m} l_j\right) + C\sum\limits_{i=1}^{n}\left(\upsilon\varepsilon + \dfrac{(\xi_i + \xi_i^*)}{\sigma_i}\right) \\ \text{s.t.} & \begin{array}{l} \mathbf{Ax} + \mathbf{Bc} = \mathbf{L}\boldsymbol{\mu}_L \\ \mathbf{D}\boldsymbol{\mu}_L = 0 \\ x_i + \mu_i - x_i^{\mathrm{M}} \le \varepsilon_i + \xi_i \\ x_i^{\mathrm{M}} - (x_i + \mu_i) \le \varepsilon_i + \xi_i^* \\ -U_i b_i \le \mu_i \le U_i b_i \\ -U'_{\mathrm{Lj}} l_j \le \mu_{\mathrm{Lj}} \le U_{\mathrm{Lj}} l_j \\ 0 \le x_i \le \mathrm{X}_i \\ 0 < \upsilon \le 1 \\ \xi_i^{(*)} \ge 0, \quad \varepsilon \ge 0 \\ b_i, l_j \in \{0, 1\} \end{array} \end{array} \tag{14}$$

where $\mathbf{L}$ is a $m \times m$ matrix whose every diagonal element is l_j and non-diagonal elements are zeros, $\boldsymbol{\mu}_{\mathrm{L}}$ is the vector of process related gross errors, $\mathbf{D}\boldsymbol{\mu}_{\mathrm{L}} = 0$ are the mass movement constraints which provide potential mass movement information, and $\mathbf{D}$ is the incidence matrix of $\boldsymbol{\mu}_{\mathrm{L}}$. $U'_{\mathrm{L}j}$ and $U_{\mathrm{L}j}$ are the low and up bound of $\mu_{\mathrm{L}j}$, respectively.

As we know, for plant-wide mass balance problem, the mass balance constraints $\mathbf{Ax} + \mathbf{Bc} = 0$ are constant with mass movement information. Therefore, if there is any mass movement information missing to establish the mass balance constraints for the data reconciliation problem, the constraints do not represent the true production balance and biased production balance results would be obtained. To overcome this problem, the missing mass movement information should be modeled, as is done in (9), where μ_{Lj} represents product quantity of the missing mass movement and l_i indicates whether a mass movement missing happens in the ith constraint. To guarantee the estimates of missing mass movement be accurate and reasonable, the potential mass balance constraints about the missing mass movement, $\mathbf{D}\boldsymbol{\mu}_{\mathrm{L}} = 0$, are incorporated into the data reconciliation problem.

From (14) it can be seen that the proposed extended form of SV regression explicitly considers both measurement related gross errors and missing mass movement information, so that the extended approach can simultaneously estimate measurement biases and mass movement missing information for plant-wide mass balance.

3. Case Study.

3.1. A simulation case. The recycle system used for simulation study here consists of four units and seven streams, which is shown in Figure 1 and has also been studied with other approaches [5,13].

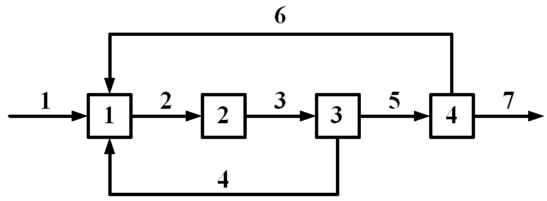

Figure 1. Flowsheet of recycling system

In this example the true flow rate values are $[5, 15, 15, 5, 10, 5, 5]$. The flow rate standard deviations (SD) are taken as 2.5% of the true flow rates. Measurement values for each simulation trial are taken as the average of ten random generated values. Each result is based on 100 simulation trials where the random errors are changed and the magnitudes of biases and leaks are fixed. The average number of type I errors (AVTI), the overall power (OP) and perfect identification (OPF) [13] are used as the criteria for judging the performance of the proposed strategy. Those are defined as

$$\mathrm{AVTI} = \frac{\#\text{ of unbiased variables wrongly identified}}{\#\text{ of simulation trials}} \tag{15}$$

$$\mathrm{OP} = \frac{\#\text{ of biased variables correctly identified}}{\#\text{ biased variables simulated}} \tag{16}$$

$$\mathrm{OPF} = \frac{\#\text{ of trials with perfect identification}}{\#\text{ of simulation trials}} \tag{17}$$

Table 1 and Table 2 indicate the performance of the extended SV regression based data reconciliation approach when both bias and leak are present. Fixed bias magnitudes of 7 standard deviations and 4 standard deviations were considered for the corresponding flow rates and leaks. As we expended, from Table 1, it can be seen that the extended SV regression approach achieved perfect joint identification of biases and leaks almost in every case. Table 2 shows that the result of the extended SV regression approach exhibits a smaller SD, because of accurate biases and leaks detection results. As a result, the extended SV regression based data reconciliation approach could achieve effective detection of both measurement bias and process leak, meanwhile, it can produce accurate estimates of biases and leaks.

Average solution times for the extended SV regression based data reconciliation and gross error detection approach are shown in Table 3. From the table, we can see that, as both of the objective function and constraints of the SV regression based approach are linear, computational load is reduced much more, as makes our approach efficient for application on linear systems even for online application.

TABLE 1. Performance results when two biases are introduced and leaks are present[a]

Leak & Gross error	AVTI	OP	OPF
L2, B4	0.00	1.00	1.00
L2, B5	0.00	1.00	0.99
L2, B6	0.00	1.00	1.00
L2, B7	0.00	1.00	1.00
L3, B2	0.00	1.00	0.99
L3, B6	0.00	1.00	1.00

[a]Ln means a leak in node n and Bn means a bias in stream n.

TABLE 2. Estimation results when two biases are introduced and leaks are present[a]

Leak & Gross error	Size	Estimates	SD
L2, B4	1.800	1.801	0.028
	0.625	0.629	0.015
L2, B5	1.800	1.784	0.018
	1.250	1.252	0.030
L2, B6	1.800	1.801	0.016
	0.625	0.626	0.026
L2, B7	1.800	1.800	0.025
	0.625	0.623	0.028
L3, B2	1.250	1.252	0.015
	1.875	1.869	0.036
L3, B6	1.250	1.251	0.015
	0.625	0.626	0.025

[a]Ln means a leak in node n and Bn means a bias in stream n.

TABLE 3. Average solution times (s)

Leak & Gross error	
L2, B4	0.06
L2, B5	0.05
L2, B6	0.07
L2, B7	0.07
L3, B2	0.07
L3, B6	0.07
Average	0.065

3.2. **A plant-wide mass balance application case.** Petroleum chemistry industry is a typical process industry, the petrochemical productions flow from one unit to others, as constructs a complex mass balance network. In this section, a part of a practical petrochemical process in a refinery is introduced to test plant-wide mass balance using the extended SV regression approach proposed. The practical petrochemical process is abstracted into a mass flow measurement network by merging specified pipes, units and nodes. The simplified mass flow measurement network includes 4 processing units (R_1 to R_4), 9 in-out nodes (IN_1 to IN_3 and S_1 to S_6), 12 tanks (T_0 to T_{11}) and 28 mass flows (F_1 to F_{28}). The plant-wide mass balance network is shown in Figure 2 [11], and the corresponding model data are listed in Table 4.

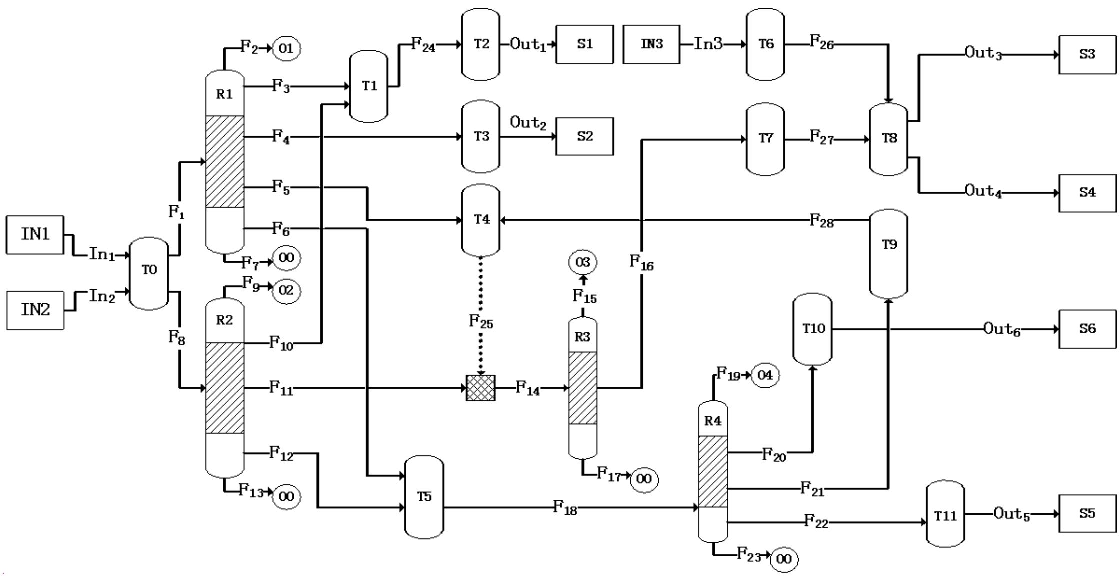

FIGURE 2. Plant-wide mass balance network

TABLE 4. Plant-wide mass balance information (ton)

Variable	Measurement	SD	Value	Constant	Measurement
F_1	10307.00	σ_1	20.64	In_1	12080
F_2	86.84	σ_2	1.29	In_2	9600
F_3	398.80	σ_3	0.79	In_3	500
F_4	1584.20[a]	σ_4	1.55	Out_1	100
F_5	4034.50	σ_5	20.23	Out_2	200
F_6	4970.20	σ_6	22.70	Out_3	1150
F_7	59.56	σ_7	0.60	Out_4	1400
F_8	9919.10	σ_8	19.86	Out_5	450
F_9	94.90	σ_9	1.42	Out_6	300
F_{10}	2045.90	σ_{10}	20.58	T_1	−1430
F_{11}	2641.70	σ_{11}	13.15	T_2	−1650
F_{12}	5075.00	σ_{12}	25.43	T_3	−703
F_{13}	59.54	σ_{13}	0.60	T_4	−573
F_{14}	6254.90	σ_{14}	12.54	T_5	−1131
F_{15}	292.09	σ_{15}	4.43	T_6	−5736
F_{16}	5022.40[a]	σ_{16}	11.83	T_7	−230
F_{17}	60.17	σ_{17}	0.60	T_8	−1915
F_{18}	4295.40	σ_{18}	8.62	T_9	−1720
F_{19}	209.00	σ_{19}	2.10	T_{10}	−406
F_{20}	2082.60	σ_{20}	20.80	T_{11}	−1780
F_{21}	1130.30	σ_{21}	2.26	T_{12}	−410
F_{22}	858.14	σ_{22}	4.30		
F_{23}	29.90	σ_{23}	0.30		
F_{24}	799.53	σ_{24}	4.01		
F_{25}	3632.40[b]	σ_{25}	18.20		
F_{26}	270.34	σ_{26}	1.35		
F_{27}	4006.50	σ_{27}	20.00		
F_{28}	717.70	σ_{28}	3.62		

[a]Measurement related gross error; [b]Mass movement information lost.

According to the practical situation in a refinery, the productions in tanks and in-let, out-let refineries are very important for operation management. Therefore, the measurements of tank volumes and in-let or out-let flows are quite accurate and usually without gross errors. Consequently, in this case study, the tank volume changes and in-let and out-let flows are considered as constants in the plant-wide mass balance model, and the variables to be reconciled are 28 mass flows.

The corresponding matrixes of plant-wide mass balance model shown in Figure 2 are as the followings,

$$\mathbf{x}^{\mathrm{M}} = \left[F_1^{\mathrm{M}}, F_2^{\mathrm{M}}, \cdots, F_{28}^{\mathrm{M}}\right]^{\mathrm{T}} \tag{18}$$

$$\mathbf{c} = \left[In_1, \cdots, In_3, Out_1, \cdots, Out_6, T_1, \cdots, T_{12}\right]^{\mathrm{T}} \tag{19}$$

$$\boldsymbol{\sigma} = \left[\sigma_1, \sigma_2, \cdots, \sigma_{28}\right]^{\mathrm{T}} \tag{20}$$

In order to investigate the effects of the SV regression approach and its extended approach to fulfill plant-wide mass balance, especially when measurement related gross errors and process related gross errors are present, two measurement related gross errors and one missing mass movement are introduced in the following two study cases.

Table 5. Plant-wide mass balance with only measurement biases

Variable	Measurement	True Value	SV	Extended SV
F_1	10307.00	10320.00	10324.66	10324.66
F_2	86.84	86.00	86.84	86.84
F_3	398.80	395.00	398.80	398.80
F_4	1584.20[a]	773.00	773.00	773.00
F_5	4034.50	4046.00	4036.26	4036.26
F_6	4970.20	4960.00	4970.20	4970.20
F_7	59.56	60.00	59.56	59.56
F_8	9919.10	9930.00	9925.34	9925.34
F_9	94.90	95.00	94.90	94.90
F_{10}	2045.90	2058.00	2054.20	2054.20
F_{11}	2641.70	2631.00	2641.70	2641.70
F_{12}	5075	5086.00	5075.00	5075.00
F_{13}	59.54	60.00	59.54	59.54
F_{14}	6254.90	6270.00	6271.26	6271.26
F_{15}	292.09	295.00	296.09	296.09
F_{16}	5022.40[a]	5915.00	5915.00	5915.00
F_{17}	60.17	60.00	60.17	60.17
F_{18}	4295.40	4310.00	4309.20	4309.20
F_{19}	209.00	210.00	209.00	209.00
F_{20}	2082.6	2080.00	2080.00	2080.00
F_{21}	1130.3	1130.00	1130.30	1130.30
F_{22}	858.14	860.00	860.00	860.00
F_{23}	29.9	30.00	29.90	29.90
F_{24}	799.53	803.00	803.00	803.00
F_{25}	3632.40	3639.00	3629.56	3629.56
F_{26}	270.34	270.00	270.00	270.00
F_{27}	4006.50	4000.00	4000.00	4000.00
F_{28}	717.70	724.00	724.30	724.30

[a]Measurement related gross error.

In the first study case, only two measurement related gross errors are used to investigate the ability of the proposed approach to detect and estimate measurement related gross errors. The measurements F_4 and F_{16} are contaminated by gross errors, respectively. The simulation result is shown in Table 5 taking the original SV regression approach as the comparison.

It can be seen from Table 5 that, when there are only measurement related gross errors in the plant-wide mass balance, both the original SV regression approach and the extended SV regression approach proposed can detect and estimate the gross errors accurately. Although the extended SV regression approach simultaneously considers process related gross errors and measurement related gross errors, it has the same effect to detect and estimate measurement related gross errors as the original SV regression approach, when there are only measurement related gross errors in the mass balance model.

In the second case, one more missing mass movement is introduced besides the two measurement related gross errors. The missing mass movement is the mass flow F_{25} from Tank4 to the junction, which is shown in Figure 2 by a dashed line. The original

Table 6. Plant-wide mass balance with measurement biases and production movement information missing

Variable	Measurement	True Value	SV	Extended SV
F_1	10307.00	10320.00	6695.10	10324.66
F_2	86.84	86.00	86.84	86.84
F_3	398.8	395.00	398.80	398.80
F_4	1584.20[a]	773.00	773.00	773.00
F_5	4034.50	4046.00	406.70	4036.26
F_6	4970.20	4960.00	4970.20	4970.20
F_7	59.56	60.00	59.56	59.56
F_8	9919.10	9930.00	13554.90	9925.34
F_9	94.90	95.00	94.90	94.90
F_{10}	2045.90	2058.00	2054.20	2054.20
F_{11}	2641.70	2631.00	6271.26	2641.70
F_{12}	5075.00	5086.00	5075.00	5075.00
F_{13}	59.54	60.00	59.54	59.54
F_{14}	6254.90	6270.00	6271.26	6271.26
F_{15}	292.09	295.00	296.09	296.09
F_{16}	5022.40[a]	5915.00	5915.00	5915.00
F_{17}	60.17	60.00	60.17	60.17
F_{18}	4295.40	4310.00	4309.20	4309.20
F_{19}	209.00	210.00	209.00	209.00
F_{20}	2082.60	2080.00	2080.00	2080.00
F_{21}	1130.30	1130.00	1130.30	1130.30
F_{22}	858.14	860.00	860.00	860.00
F_{23}	29.90	30.00	29.90	29.90
F_{24}	799.53	803.00	803.00	803.00
F_{25}	3632.40[b]	3639.00	–	3629.56
F_{26}	270.34	270.00	270.00	270.00
F_{27}	4006.50	4000.00	4000.00	4000.00
F_{28}	717.70	724.00	724.30	724.30

[a]Measurement related gross error; [b]Mass movement information lost.

SV regression approach is again taken as the comparison, and the simulation results are demonstrated in Table 6.

From the results shown in Table 6, it can be seen that when there are both measurement related gross errors and mass movement information missing in plant-wide mass balance, though the original SV regression approach can detect and estimate the measurement biases, it cannot address the missing mass movement. Therefore, the original SV regression approach provides biased estimations on mass flows F_1, F_5, F_8 and F_{11}. On the other hand, the extended SV regression approach can detect and estimate both the measurement biases and the missing mass movement. As a result, more accurate plant-wide mass balance is obtained. This is because the extended SV regression approach considers process related gross errors in the data reconciliation and gross error detection model, and it can simultaneously address measurement related gross errors and process related gross errors. For plant-wide mass balance, the extended SV regression can detect and estimate measurement biases and missing mass movement information accurately.

4. **Conclusions.** Plant-wide mass balance is very important for production planning, production scheduling and production performance analysis. Because the raw measurements used by plant-wide mass balance are usually deteriorated by random and gross errors, data reconciliation and gross error detection techniques have been widely applied on plant-wide mass balance. However, few data reconciliation and gross error detection approach can address process related gross errors, so that few focuses are drawn onto addressing the mass movement information missing in plant-wide mass balance.

In the paper, an extended support vector regression approach is proposed, which considers both measurement related gross errors and process related gross errors simultaneously. Therefore, the proposed approach is suitable for plant-wide mass balance especially when mass movement information missing presents. In the case study, a recycle system is first used to demonstrate the validity and accuracy of the approach proposed to simultaneously detect and estimate measurement bias and process leaks. Then, a plant-wide mass balance study case is introduced to demonstrate that the proposed approach is accurate and effective for detecting and estimating gross errors and production movement information missing.

REFERENCES

[1] P. M. Reilly and R. E. Carpani, Application of statistical theory of adjustment to material balances, *Proc. of the 13th Canada Chemical Engineering Conference*, Montreal, Quebec, 1963.

[2] R. S. H. Mah and A. C. Tamhane, Detection of gross errors in process data, *AIChE Journal*, vol.28, no.5, pp.828-830, 1982.

[3] S. Narasimhan and R. S. H. Mah, Generalized liklihood ratio method for gross error identification, *AIChE Journal*, vol.33, no.9, pp.1514-1521, 1987.

[4] H. W. Tong and C. M. Crowe, Detection of gross errors in data reconciliation by principal component analysis, *AIChE Journal*, vol.41, no.7, pp.1712-1722, 1995.

[5] J. Rosenberg, R. S. H. Mah and C. Iordache, Evaluation of schemes for detecting and identifying gross errors in process data, *Industrial & Engineering Chemistry Research*, vol.26, no.3, pp.555-564, 1987.

[6] I. W. Kim et al., Robust data reconciliation and gross error detection: The modified MIMT using NLP, *Computers & Chemical Engineering*, vol.21, no.7, pp.775-782, 1997.

[7] N. Arora and L. T. Biegler, Redescending estimators for data reconciliation and parameter estimation, *Computers & Chemical Engineering*, vol.25, no.11-12, pp.1585-1599, 2001.

[8] D. Wang and J. A. Romagnoli, Generalized T distribution and its applications to process data reconciliation and process monitoring, *Trans. of the Institute of Measurement and Control*, vol.27, no.5, pp.367-390, 2005.

[9] M. Ikenoue et al., On bias-compensated least-squares algorithm via prefiltering, *International Journal of Innovative Computing, Information and Control*, vol.1, no.3, pp.493-507, 2005.

[10] M. F. Hassan, M. Zribi and M. Tawfik, State estimation of constrained nonlinear discrete-timed dynamical systems, *International Journal of Innovative Computing, Information and Control*, vol.6, no.10, pp.4449-4470, 2010.
[11] M. Basin, A. Loukianov and M. Hernandez-Gonzalez, Mean-square joint state and noise intensity estimation for linear stochastic systems, *International Journal of Innovative Computing, Information and Control*, vol.7, no.1, pp.327-334, 2011.
[12] I. Mizumoto, Y. Fujimoto, N. Watanabe and Z. Iwai, Fast rate adaptive output feedback control of multi-rate sampled systems with an adaptive output estimator, *International Journal of Innovative Computing, Information and Control*, vol.7, no.7(B), pp.4377-4394, 2011.
[13] M. Sanchez et al., Simultaneous estimation of biases and leaks in process plants, *Computers & Chemical Engineering*, vol.23, no.7, pp.841-857, 1999.
[14] Z. H. Abu-el-zeet, V. M. Becerra and P. D. Roberts, Combined bias and outlier identification in dynamic data reconciliation, *Computers & Chemical Engineering*, vol.26, no.6, pp.921-935, 2002.
[15] P. Zhang, G. Rong and Y. Wang, A new method of redundancy analysis in data reconciliation and its application, *Computers & Chemical Engineering*, vol.25, no.7-8, pp.941-949, 2001.
[16] X. Wang and G. Rong, Data rectification based on refinery two-step mass balance, *Control and Instruments in Chemical Industry*, vol.16, no.4, pp.16-19, 2005.
[17] X. Wang and G. Rong, A method of data rectification based on Bayesian network, *Journal of Chemical Industry and Engineering (China)*, vol.57, no.6, pp.1385-1389, 2006.
[18] Y. Miao et al., Support vector regression approach for simultaneous data reconciliation and gross error or outlier detection, *Industrial & Engineering Chemistry Research*, vol.48, no.24, pp.10903-10911, 2009.
[19] L. Xiao, Y. Miao and H. Su, Production tracking and mass balance in a refinery based on manufacturing operations management and data reconciliation, *Proc. of the 1st International Conference on Information Science and Engineering*, Nanjing, China, 2009.
[20] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 1995.