Data Reconciliation in Gas Pipeline Systems

Miguel J. Bagajewicz* and Enmanuel Cabrera

School of Chemical Engineering & Materials Science, University of Oklahoma, 100 E. Boyd Street, T-335, Norman, Oklahoma 73019-1004

In natural gas pipeline systems, all product transactions are based on flow rate measurements. Thus, it is essential to have good estimates of these variables. Currently, these variables are estimated using data reconciliation based exclusively on material balances. To improve these estimations, an approximate methodology that includes material balances as well as mechanical energy balances is presented. This method is based on iterations between a simplified model and a rigorous simulation model developed in SIMSCI's PRO/II 5.55 system. A new method of performing observability analysis is also developed to take into account the nonlinear mechanical energy equations and pressure measurements. Finally, because temperature measurements are used to define density, the influence of temperature measurement errors on the accuracy of the estimators is analyzed. Several examples are presented to illustrate the effectiveness of the methodology. The proposed technique proves to be very effective computationally and generates better estimates than techniques in which only material balances are used.

Introduction

In pipelines systems, accurate flow rate data are essential for the proper calculation of transactions. Imbalances in these systems generate revenue losses to companies every year. Biased instrumentation can also affect the estimators, thus reducing the efficiency of operations. Because readings from instruments do not satisfy basic conservation laws, data reconciliation considers the adjustment of measured values so that unique estimates from all of the conflicting readings are obtained and biases and leaks are detected and estimated.

Data reconciliation is usually performed by minimizing a least-squares objective function, subject to model equations, that comes from maximum likelihood formulations and the assumption of normal error distributions. Such model equations range from simple material, component, and energy balances to full models involving all system variables and parameters. Even at the commercial level, a variety of software systems exist that are able to perform data reconciliation in full nonlinear systems (DATACON, SIMSCI, etc.), including gross error handling. However, most commercial data reconciliation packages limit themselves to material component and energy balances. One exception is Chemplant Technology, s.r.o., who developed a software package that also performs hydraulic calculations for pipes. This software uses an isothermal Bernoulli equation and provides means of reconciling pressures. Nevertheless, the details of the procedure and its accuracy are not published. A complete background on these techniques and their variants can be obtained from three recent books.[1−3]

The governing equations for pipeline systems are the material balances, the mechanical energy balances, and the heat balance equation. Of these, only the first two can be used to link redundant variables. The last involves heat exchanges with the environment, and therefore, it can be used only to estimate the heat exchanged, which is of lesser practical importance. Thus, temperature measurements can be used only to obtain more accurate values of gas density close to the flowmeters, but cannot be made redundant. This is not the case for pressure measurements, which can be made redundant through the mechanical energy balances and, therefore, can improve the accuracy of the estimators of mass flows.

Very little is available for systems where the mechanical energy balance is also used to improve the material balance data reconciliation. Only Coelho and Medeiros[4] have presented an analysis of data reconciliation and leak detection in pipelines with incompressible fluids.

The numerical difficulties associated with such problems are many. When performing data reconciliation in large pipeline networks handling compressible fluids (and possibly two phases), the nonlinear model can become too large and computationally expensive. Therefore, a method is needed to take these nonlinearities into account in an efficient and computationally fast manner. In this paper, the use of simple models as pieces of a successive approximation scheme is proposed, increasing the calculation speed. The method uses simple expressions for the mechanical energy balances, whose parameters are updated using rigorous simulations on each piping section in an iterative manner. The end result is then consistent with rigorous models. Finally, the influence of temperature measurements errors on the results is discussed.

Data Reconciliation with Rigorous Models

The data reconciliation model is

$$\min\{[\tilde{\mathbf{G}}_r - \mathbf{G}_r^+]^T \cdot \mathbf{S}_G^{-1} \cdot [\tilde{\mathbf{G}}_r - \mathbf{G}_r^+] + [\tilde{\mathbf{P}}_r - \mathbf{P}_r^+]^T \cdot \mathbf{S}_P^{-1} \cdot [\tilde{\mathbf{P}}_r - \mathbf{P}_r^+]\} \quad (1)$$

s.t.

$$f(\tilde{\mathbf{G}}_r, \tilde{\mathbf{P}}_r) = 0$$

* To whom correspondence should be addressed. E-mail: bagajewicz@ou.edu. Tel.: (405) 325-5458. Address: 100 E. Boyd St., T-335, Norman, OK 73019.

10.1021/ie020774j CCC: $25.00
Published on Web 10/02/2003

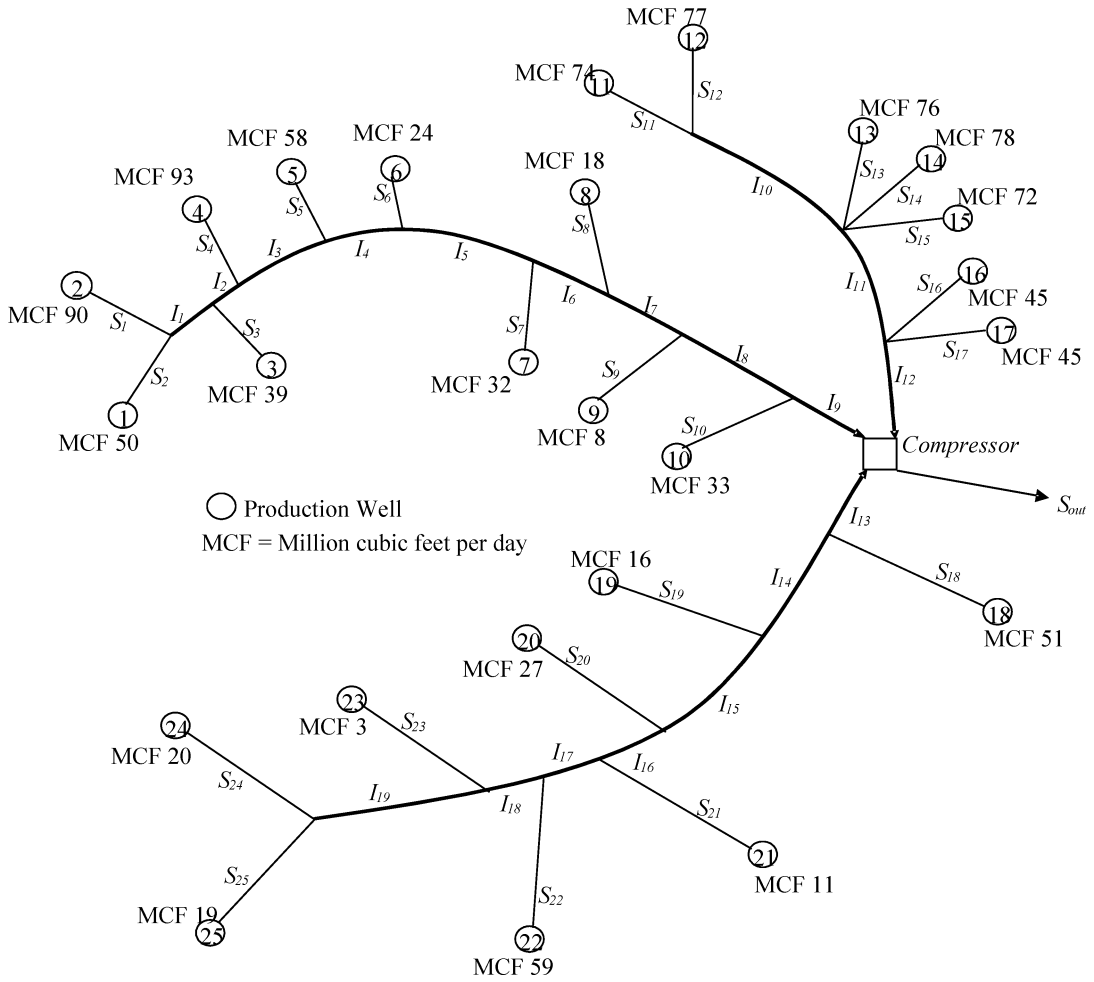


Figure 1. Cerro Fortunoso gas production and gathering field.

where $\mathbf{S_G}$ and $\mathbf{S_P}$ are the variance matrices of mass flow ($\mathbf{G}^+$) and pressure ($\mathbf{P}^+$) measurements, respectively, and $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{P}}$ represent the corresponding estimators. In turn, $f(\tilde{\mathbf{G}}_\mathbf{r},\tilde{\mathbf{P}}_\mathbf{r})$ represents material and rigorous mechanical energy balances. Temperatures are not reconciled because the total energy equation is not used.

Because the model should use compressible flow expressions, these are usually integrated using numerical algorithms along the length of each pipeline section. Considering that these sections can be very long, the computational time one expects is large. This is illustrated next.

Motivating Example. Consider the flowsheet of Figure 1. It represents a section of the Cerro Fortunoso gas production and gathering field located in the southern region of the province of Mendoza, in Argentina.[5] The system is composed of 25 different oil and gas production wells connected to a main pipeline, which transports the gas to processing plants.

Data reconciliation was performed for the Cerro Fortunoso network assuming the pipe dimensions given in Table 1. The flow rates presented in the figure are at normal conditions (0 °C, 1 atm). The feeds are considered to be pure methane at 25 °C. The flows of all inlet and outlet streams are measured, as well as the pressures of the inlet streams of the system. This data reconciliation problem was set up in SIMSCI's PRO/II 5.55 process simulator using an optimizer. The pressure drop in each section of piping was modeled using the Beggs−Brill−Moody equation,[6] the default method for PRO/II recommended for most systems, especially single-phase systems. Considering only the first branch of the network, streams S_1-S_{10}, the simulator takes 9 min and 38.03 s to solve the data reconciliation on a Pentium III PC, 850 MHz, 128 MB RAM computer.

Table 1. Pipe Dimensions for Cerro Fortunoso Network

stream	length (m)	internal diameter (mm)	stream	length (m)	internal diameter (mm)
S_1	515	90	S_{24}	657	63
S_2	615	102	S_{25}	606	63
S_3	781	90	S_{out}	1041	381
S_4	821	128	I_1	1132	154
S_5	884	102	I_2	1262	203
S_6	989	90	I_3	1239	203
S_7	982	90	I_4	1412	255
S_8	569	63	I_5	1134	255
S_9	723	53	I_6	1100	255
S_{10}	732	90	I_7	1170	303
S_{11}	778	102	I_8	1442	333
S_{12}	606	102	I_9	1217	303
S_{13}	706	102	I_{10}	1154	154
S_{14}	924	102	I_{11}	1186	255
S_{15}	839	102	I_{12}	1193	255
S_{16}	695	90	I_{13}	1234	203
S_{17}	616	90	I_{14}	1139	154
S_{18}	848	90	I_{15}	1299	154
S_{19}	980	63	I_{16}	1289	154
S_{20}	654	78	I_{17}	1124	128
S_{21}	755	53	I_{18}	1016	90
S_{22}	874	102	I_{19}	1029	90
S_{23}	798	35			

Considering that pipeline networks can have hundreds of streams, the computational time of data reconciliation can become unmanageable, especially when information about leaks and biases is desired quickly. Indeed, considering only the 25 oil and gas wells of the Cerro Fortunoso Network, with a total of 45 process streams, the data reconciliation problem is solved in 24 min and 34.92 s using PRO/II.

Using approximate models, such as the one presented in this paper, the computational time to solve the same Cerro Fortunoso Field problem of Figure 1 is about 40 s, both for the entire 25 oil and gas wells and for the first 10 wells alone. Therefore, the use of approximate models for the mechanical energy equation implies important reductions of the computational time of the data reconciliation problem.

Proposed Reconciliation Model

Instead of using a rigorous model, an approximation is proposed. Assuming incompressible flow and neglecting acceleration terms, one obtains the following expression of the mechanical energy balance

$$\rho_1 g(h_1 - h_2) + (p_1 - p_2) = \frac{1}{2\rho_1} f_f \left(\frac{G}{A}\right)^2 \frac{L}{D} \tag{2}$$

A discussion of this and other models is given in the Appendix. To account for the error of this model, a term Δ is introduced into the equation as follows

$$\rho_1 g(h_1 - h_2) + (p_1 - p_2) - \frac{1}{2\rho_1} f_f \left(\frac{G}{A}\right)^2 \frac{L}{D} = \Delta \tag{3}$$

We rewrite eq 3 as

$$aG^2 + (p_1 - p_2) + c = \Delta \tag{4}$$

where

$$a = -\frac{8}{\pi^2} \frac{f_f}{D^4} \frac{1}{\rho_1} \frac{L}{D}$$
$$c = \rho_1 g(h_2 - h_1) \tag{5}$$

The iterative scheme consists of assuming certain values for the model error Δ and running the data reconciliation using eq 2. Next, the process estimates are updated, and new pressure-drop estimates are obtained with rigorous models. The correction term Δ is then calculated by comparing these last pressure-drop values with the predictions of the simplified model obtained using the current variable estimates. The process is repeated until convergence is achieved.

Observability Analysis

Observability analysis is needed to classify the variables and perform data reconciliation using the correct set of equations and variables to avoid singularities. Such analysis can be made using approximate models because the objective is only to classify the process variables and not to obtain the value of the variables themselves. In other words, approximate models are as good at identifying redundant and observable variables as rigorous models.

In general, the system of balance equations can be written as

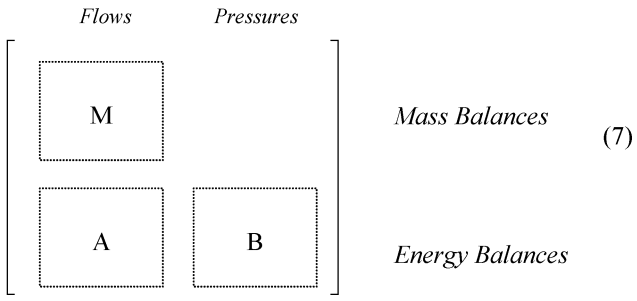
$$\begin{aligned} &\mathbf{M}\cdot\mathbf{G} = 0 && \text{mass balances} \\ &\mathbf{A}\cdot(\mathbf{G} \otimes \mathbf{G}) + \mathbf{B}\cdot\mathbf{P} + \mathbf{c} = 0 && \text{mechanical energy balances} \end{aligned} \tag{6}$$

where **M** is the occurrence matrix of the system, **A** is a diagonal matrix containing the a_i coefficients, **B** is the matrix of pressure coefficients (which are 1 or -1), and **c** is the vector of c_i coefficients. The term $\mathbf{G} \otimes \mathbf{G}$ is a vector such that $(G \otimes G)_i = G_i^2$. Pressure drops due to fittings, especially tees, are incorporated as head losses in the piping behind or ahead of the corresponding fittings.

The procedure for observability analysis has three main steps. The first step consists of finding what observability and redundancy information can be obtained from the mass balance equations. The second step involves the energy balance equations and aims at determining the observability and redundancy information that can be obtained from these equations. Finally, the third step combines the mass and energy balance equations.

The only nonlinear term present in eq 6 is quadratic in single variables (G_i^2). Therefore, in the observability analysis involving the mechanical energy balance equation, one can treat G_i^2 as a new variable and perform an observability analysis the same way as in linear systems. Details of this procedure follow:

1. Construction of the Occurrence Matrix. Mass balance equations are located in the upper side of the matrix, and energy balance equations are below them. Each row represents an equation, and each column represents a variable. The matrix is constructed by filling position (i,j) with the coefficient of the variable j in the equation i or leaving it blank if that variable does not appear in that particular equation. For quadratic terms (G_i^2), the coefficients are a_i, located on the columns of their corresponding flow as follows



(7)

The procedure is illustrated using the system of Figure 2, which is a straightforward gas gathering and transportation system with nine streams, three splitters, and one mixer. The rectangles represent the pipeline section to which the energy balances are applied. The measured variables in this example are G_1, G_2, G_8, G_9, P_1, P_2, P_8, and P_9, indicated by a star (★) in the figure.

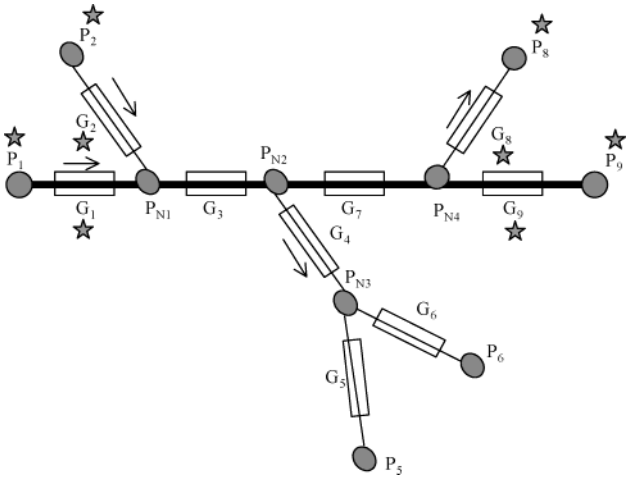
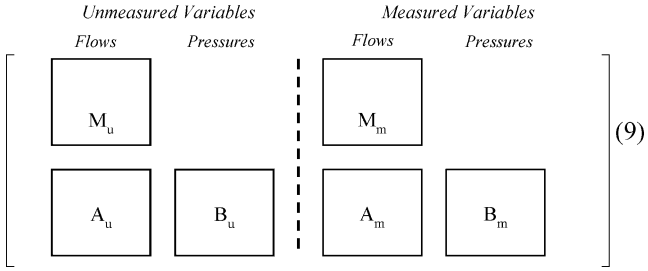


Figure 2. Gas gathering and transportation system.

For this system, the occurrence matrix is as shown in eq 8.

$$\begin{array}{cccccccccccccccccc}
G_1 & G_2 & G_3 & G_4 & G_5 & G_6 & G_7 & G_8 & G_9 & P_1 & P_2 & P_{N1} & P_{N2} & P_{N3} & P_{N4} & P_5 & P_6 & P_8 & P_9 \\
\hline
1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
a_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & a_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & a_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & a_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & a_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & a_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & a_7 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & a_8 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_9 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1
\end{array} \quad (8)$$

2. Rearrangement of the Occurrence Matrix. In this step, measured variables are separated from unmeasured ones. The unmeasured flows are located first, followed by the unmeasured pressures. Measured flows are in the first columns, trailed by measured pressures as shown in eq 9.

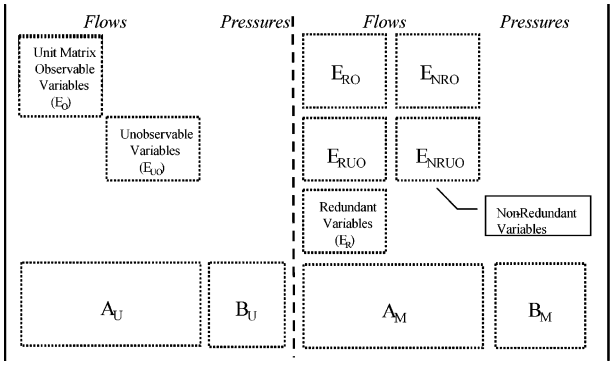


(9)

After the separation of the measured variables is performed on the example of Figure 2, the occurrence matrix is as given in eq 10.

$$\begin{array}{cccccccccc|cccccccc}
\multicolumn{10}{c|}{\text{Unmeasured Variables}} & \multicolumn{8}{c}{\text{Measured Variables}} \\
G_3 & G_4 & G_5 & G_6 & G_7 & P_{N1} & P_{N3} & P_5 & P_6 & P_{N4} & P_{N2} & G_1 & G_2 & G_8 & G_9 & P_1 & P_2 & P_8 & P_9 \\
\hline
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & a_1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & a_2 & 0 & 0 & 0 & 1 & 0 & 0 \\
a_3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & a_4 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & a_5 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & a_6 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & a_7 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & a_8 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & a_9 & 0 & 0 & 0 & -1
\end{array} \quad (10)$$

3. Canonical Representation of the Mass Balance Subset. Following the procedure proposed by Madron,[1] Gauss−Jordan factorization is used to obtain the largest possible identity square matrix in the upper left corner of the subset. The scheme of the occurrence matrix with the mass balance section in its canonical form is as shown in eq 11.



(11)

The matrix in eq 12 is the canonical representation of the mass balance subset for the system of Figure 2.

$$\begin{array}{cccccccccc|cccccccc}
G_3 & G_7 & G_4 & G_5 & G_6 & P_{N1} & P_{N3} & P_5 & P_6 & P_{N4} & P_{N2} & G_1 & G_2 & G_8 & G_9 & P_1 & P_2 & P_8 & P_9 \\
\hline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\
1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & a_1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & a_2 & 0 & 0 & 0 & 1 & 0 & 0 \\
a_3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & a_4 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & a_5 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & a_6 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & a_7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & a_8 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & a_9 & 0 & 0 & 0 & -1
\end{array} \quad (12)$$

For this example, G_3 and G_7 become observable using mass balances only.

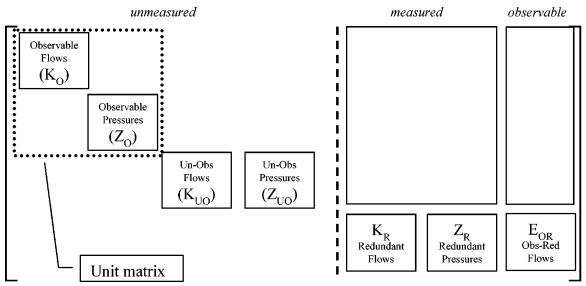
4. Separation of Observable Variables. The columns corresponding to the flows determined observable using only mass balance equations are put together with the measured variables. The set of measured and observable flows is now called the set of "known variables". Finally, the rows corresponding to redundant flows are moved to the bottom. The structure of the resulting incidence matrix is presented in eq 13.

$$\begin{array}{cc|ccc|c}
\multicolumn{2}{c}{\text{Pressures}} & & & & \\
\hline
0 & 0 & \boldsymbol{E_{RO}} & \boldsymbol{E_{NRO}} & 0 & \boldsymbol{E_O} \\
\boldsymbol{E_{UO}} & 0 & \boldsymbol{E_{RUO}} & \boldsymbol{E_{NRUO}} & 0 & 0 \\
\boldsymbol{A_{UO}} & \boldsymbol{B_U} & \boldsymbol{A_R} & \boldsymbol{A_{NR}} & \boldsymbol{B_M} & \boldsymbol{A_O} \\
0 & 0 & \boldsymbol{E_R} & 0 & 0 & 0
\end{array} \quad (13)$$

For the example, take the columns corresponding to G_3 and G_7 to the right side of the occurrence matrix, as indicated next in eq 14.

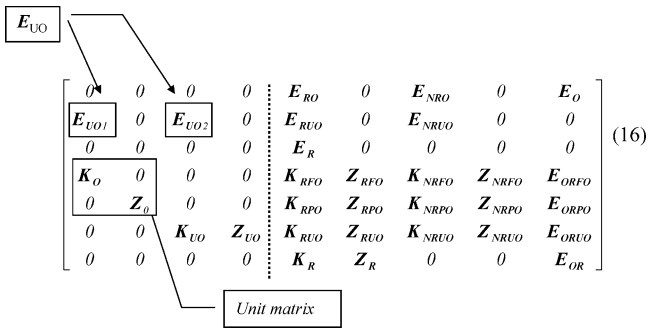
$$\begin{array}{cccccccccc|cccccccc|cc}
G_4 & G_5 & G_6 & P_{N1} & P_{N3} & P_5 & P_6 & P_{N4} & P_{N2} & G_1 & G_2 & G_8 & G_9 & P_1 & P_2 & P_8 & P_9 & G_3 & G_7 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\
1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & a_1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & a_2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_3 & 0 \\
a_4 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & a_5 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & a_6 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_7 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & a_8 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & a_9 & 0 & 0 & 0 & -1 & 0 & 0
\end{array} \quad (14)$$

5. Observability Analysis on the Energy Balance Equations. Gauss−Jordan factorization is used next to obtain the canonical of the energy balances. The resulting structure is



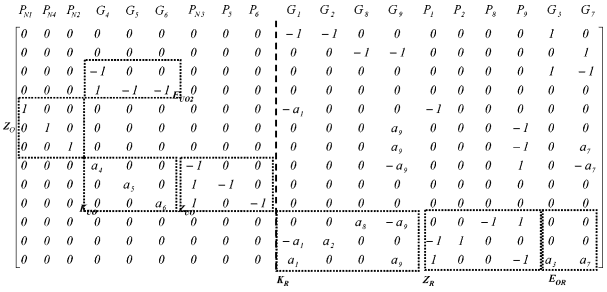
(15)

The whole occurrence matrix after this step is



(16)

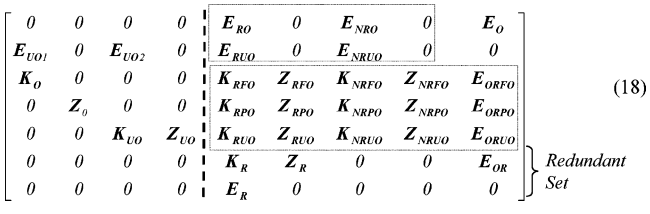
For the example, the canonical form equivalent to eq 16 is



(17)

Three different pressures become observable using energy balances. Notice that only pressures become observable by using energy balances in this particular example.

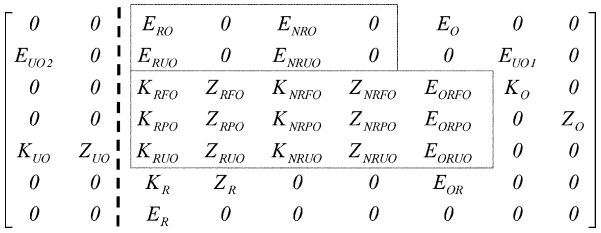
Next, the row with the submatrix $\mathbf{E_R}$ is taken to the bottom of the occurrence matrix to join the rest of the redundant equations as follows



(18)

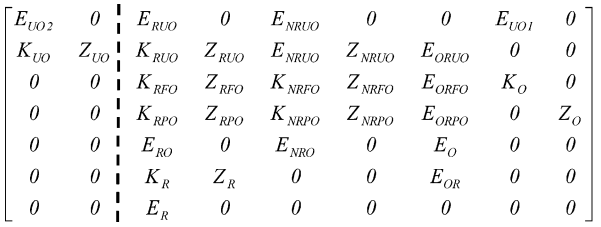
At this point, the maximum information about the observability and redundancy of the system has been obtained from the energy balances and mass balances independently. The task now is to determine whether some of the remaining unobservable variables can be made observable by combining mass and energy balances.

6. Combination of Balances. The first two columns in the matrix in eq 18 contain observable variables and are therefore moved to the known variables set



(19)

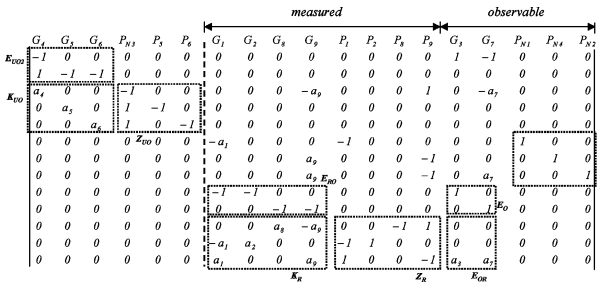
which can be rearranged to give the following structure



(20)

The rows and columns corresponding to the matrices $\mathbf{E_{UO2}}$, $\mathbf{K_{UO}}$, and $\mathbf{Z_{UO}}$ represent the combined subset that is analyzed in the next step of the procedure.

In the example, the matrix corresponding to eq 20 is



(21)

The procedure to solve the remaining nonlinear system is presented next.

7. Solution of the Nonlinear System. A technique based on the work of Steward[7,8] is used to classify the variables in the remaining nonlinear system. The method consists of performing a partitioning of the columns containing unmeasured variables of the occurrence matrix. It classifies unmeasured variables into observable and unobservable and measured variables into redundant and nonredundant.

Steward's Algorithm−**Acyclic Precedence Ordering.**

(a) Convert the combined subset into the occurrence matrix to be solved by filling position *ij* with an X if variable *j* appears in equation *i* and leaving it blank otherwise.

(b) Find a row containing only one entry. This entry represents a removable 1 × 1 subset. If this subset is allowable, remove it by deleting the row and column in which occurs and place it in the first open row and column in the reordered matrix. Enclose that entry in a box to indicate that this equation is to be solved for this flow. Repeat step 2 until no more entries can be made in the reordered matrix.

(c) Find all undeleted rows with only two entries. If two rows have their entries in the same two columns, a removable 2 × 2 subset has been found. If that subset is allowable, remove it from the occurrence matrix, place

it in the next two open rows and columns of the reordered matrix, and enclose it in a box. Repeat steps 2 and 3 until no more entries can be made in the reordered matrix.

(d) Repeat step c replacing two by three, that is, find all undeleted rows with only three entries check whether they form a removable 3 × 3 subset. Repeat replacing three by four, etc., until the rank of the largest allowable subset is reached.

For the example of Figure 2, the converted occurrence matrix is

$$\begin{array}{c|cccccc} & G_4 & G_5 & G_6 & P_{N3} & P_5 & P_6 \\ \hline & X & 0 & 0 & 0 & 0 & 0 \\ & X & X & X & 0 & 0 & 0 \\ & X & 0 & 0 & X & 0 & 0 \\ & 0 & X & 0 & X & X & 0 \\ & 0 & 0 & X & X & 0 & X \end{array} \quad (22)$$

Finally, the reordered matrix for the example is

$$\begin{array}{c|cccccc} & G_4 & P_{N3} & G_5 & G_6 & P_5 & P_6 \\ \hline & X & 0 & 0 & 0 & 0 & 0 \\ & X & X & 0 & 0 & 0 & 0 \\ & X & 0 & X & X & 0 & 0 \\ & 0 & X & X & 0 & X & 0 \\ & 0 & X & 0 & X & 0 & X \end{array} \quad (23)$$

Hence, G_4 and P_{N3} become observable, whereas G_5, G_6, P_5, and P_6 remain unobservable. These are the only variables that remain unobservable after the application of the entire proposed procedure. Using only mass balances, just two variables would have become observable, G_3 and G_7. This illustrates the advantage of using mechanical energy balances as part of the model equations.

Redundant Set of Equations

The redundant set is composed of the equations that include only measured variables after the occurrence matrix is in its canonical form. In the case of mass balances, the redundant equations can be expressed as

$$\mathbf{E_R}\cdot\mathbf{G_r} = 0 \quad (24)$$

where $\mathbf{G_r}$ is a vector composed only of measured and redundant flow rates. Another set of redundant equations comes from the subsystem of mechanical energy balances. This set can be written as

$$\mathbf{K_R}\cdot(\mathbf{G_r}\otimes\mathbf{G_r}) + \mathbf{Z_R}\cdot\mathbf{P_r} + \mathbf{E_{OR}}\cdot(\mathbf{G_O}\otimes\mathbf{G_O}) = \mathbf{c}^* \quad (25)$$

where $\mathbf{P_r}$ is composed of only measured and redundant pressures. $\mathbf{G_O}$ is the vector of observable flows from mass balances, and $\mathbf{c}^*$ is the original vector $\mathbf{c}$ when modified through all of the Gauss−Jordan factorization operations. However, $\mathbf{G_O}$ is expressed as a function of measured flows

$$\mathbf{E_O G_O} = -\mathbf{E_{RO}G_r} - \mathbf{E_{NRO}G_{nr}} \quad (26)$$

where $\mathbf{G_{nr}}$ is a vector composed of only nonredundant measured flows.

Material and Energy Balance Data Reconciliation

As described above, this is the data reconciliation problem expressed in terms of the approximate model

$$\min\{[\tilde{\mathbf{G}}_\mathbf{r} - \mathbf{G_r}^+]^T\cdot\mathbf{S_f}^{-1}\cdot[\tilde{\mathbf{G}}_\mathbf{r} - \mathbf{G_r}^+] + [\tilde{\mathbf{P}}_\mathbf{r} - \mathbf{P_r}^+]^T\cdot S_p^{-1}\cdot[\tilde{\mathbf{P}}_\mathbf{r} - \mathbf{P_r}^+]\} \quad (27)$$

s.t.

$$\mathbf{E_R}\cdot\tilde{\mathbf{G}}_\mathbf{r} = 0$$

$$\mathbf{K_R}\cdot(\tilde{\mathbf{G}}_\mathbf{r}\otimes\tilde{\mathbf{G}}_\mathbf{r}) + \mathbf{Z_R}\cdot\mathbf{P_r} + \mathbf{E_{OR}}\cdot(\mathbf{G_O}\otimes\mathbf{G_O}) - \mathbf{c}^* = \mathbf{\Delta}$$

$$\mathbf{E_O G_O} = -\mathbf{E_{RO}}\tilde{\mathbf{G}}_\mathbf{r} - \mathbf{E_{NRO}G_{nr}}$$

The objective is to find the proper value of the vector $\mathbf{\Delta}$, the model error factor, so that the equation above is satisfied for the estimates obtained by data reconciliation. The value of $\mathbf{\Delta}$ is assumed to be equal to zero for the first iteration, and the constants a and c are not updated in the iterative process. They are calculated with the measurements and by using the process simulator to obtain the pipe and flow characteristics. For the rest of the iterations, the following formula is used

$$\Delta_{i+1} = \Delta\tilde{P}_i - \Delta P_i^* + \Delta_i \quad (28)$$

where $\Delta\tilde{P}_i$ is evaluated using the estimates obtained from the data reconciliation and ΔP_i^* using the process simulator. However, using the approximation $(P_i^* - P_{out}^*) \approx (\tilde{P}_{in} - P_{out}^*)$, eq 28 becomes

$$\Delta_{i+1} = (P_{out}^*)_i - (\tilde{P}_{out})_i + \Delta_i \quad (29)$$

This scheme is similar to, but more efficient than successive linearization, a very well-known technique in optimization theory. Nevertheless, because the constraints are not convex, global optimality is not guaranteed, either in the scheme using rigorous expression directly or in the case where the approximate model is used iteratively. This issue cannot be resolved easily and is left outside the scope of this paper.

The algorithm was implemented in MS Visual Basic. The GAMS-MS Visual Basic Interface was used to solve the nonlinear optimization and to find the estimates in each iteration. No interface between PRO/II 5.55 and MS Visual Basic exists that allows input values to be entered into the flowsheet from Visual Basic. This was done by hand. As a consequence, the comparison between the rigorous models and the approximate method proposed in this paper was based on the calculation time for each stage of the optimization process and not on the interaction time among them.

The iterative process is terminated when the change in Δ becomes very small, that is, when $(\Delta_i - \Delta_{i+1}) \leq \epsilon$, with ϵ being a given small number. In the examples presented in this paper, we used $\epsilon = 0.01\%$. Two different examples are presented next to illustrate the use of the algorithm.

Example 1. Consider the flowsheet of Figure 2. Pure methane at $T = 25$ °C and $P = 25$ atm is fed into the

Table 2. Pipeline Dimensions and Constant Values for Example 1

pipe	diameter (mm)	length (m)	a_i
1	477.82	10 000	−0.238
2	477.82	7500	−0.178
3	574.65	5000	−0.047
4	303.23	750	−0.203
5	202.72	300	−0.686
6	202.72	300	−0.686
7	381.00	2500	−0.208
8	202.72	600	−1.378
9	254.51	500	−0.350

Table 3. Measurements and Statistical Data for Example 1

	measurement (kPa, kg/s)	standard dev (kPa, kg/s)	variance (kPa^2, kg^2/s^2)
P_1	2498.38	74.95	5618
P_2	2576.55	77.30	5975
P_8	2141.84	64.26	4129
P_9	2139.06	64.17	4118
G_1	19.66	0.59	0.348
G_2	24.78	0.74	0.553
G_8	7.46	0.22	0.050
G_9	15.07	0.45	0.204

pipes. The measured variables are the same as before (G_1, G_2, G_8, G_9, P_1, P_2, P_8, and P_9). There is no elevation change on the pipes. After the observability analysis is performed, the redundant set of the system is

$$G_3 = G_1 + G_2$$
$$G_7 = G_8 + G_9$$
$$-a_1 G_1^{\,2} + a_2 G_2^{\,2} - P_1 + P_2 = 0 \qquad (30)$$
$$a_8 G_8^{\,2} - a_9 G_9^{\,2} - P_8 + P_9 = 0$$
$$a_1 G_1^{\,2} + a_9 G_9^{\,2} + P_1 - P_9 + a_7 G_7^{\,2} + a_3 G_3^{\,2} = 0$$

There are no elevation changes in the pipeline sections, that is, $c_i = 0$, $\forall\, i$. Therefore, the data reconciliation problem for each i iteration is

$$P_i = \min\{[\tilde{G}_{r,i} - G_r^{+}]^{T} \cdot S_G^{-1} \cdot [\tilde{G}_{r,i} - G_r^{+}] + [\tilde{P}_{r,i} - P_r^{+}]^{T} \cdot S_P^{-1} \cdot [\tilde{P}_{r,i} - P_r^{+}]\} \qquad (31)$$

s.t.

$$\tilde{G}_{3,i} = \tilde{G}_{1,i} + \tilde{G}_{2,i}$$
$$\tilde{G}_{7,i} = \tilde{G}_8 + \tilde{G}_{9,i}$$
$$-a_1 \tilde{G}_{1,i}^{\,2} + a_2 \tilde{G}_{2,i}^{\,2} - \tilde{P}_{1,i} + \tilde{P}_{2,i} = \Delta_{2,i} - \Delta_{1,i}$$
$$a_8 \tilde{G}_{8,i}^{\,2} - a_9 \tilde{G}_{9,i}^{\,2} - \tilde{P}_{8,i} + \tilde{P}_{9,i} = \Delta_{8,i} - \Delta_{9,i}$$
$$a_1 \tilde{G}_{1,i}^{\,2} + a_9 \tilde{G}_{9,i}^{\,2} + \tilde{P}_{1,i} - \tilde{P}_{9,i} + a_7 \tilde{G}_{7,i}^{\,2} + a_3 \tilde{G}_{3,i}^{\,2} = \Delta_{9,i} + \Delta_{7,i} + \Delta_{1,i} + \Delta_{3,i}$$

Pipe dimensions, pipe lengths, and internal diameters are presented in Table 2.

The constants c_i are equal to zero because there are no elevation changes on the pipes. Measurement values and the statistical data from the measurements are given in Table 3. The algorithm required four iterations to meet the termination criterion of $|\Delta_i - \Delta_{i+1}| < 0.01\%$. The execution time on a Pentium III PC, 850 GHz, 128 Gb RAM computer was about 10 s. Iteration information and results of the problem are summarized in Table 4, with values in bold representing the final results of the example.

Table 4. Results of Example 1[a]

	iteration number				
	0	1	2	3	
P_1 (kPa)	2470.591	2470.054	2470.047	**2470.045**	
P_2 (kPa)	2488.523	2487.849	2487.792	**2487.794**	
P_8 (kPa)	2183.518	2184.557	2184.618	**2184.618**	
P_9 (kPa)	2178.549	2178.371	2178.353	**2178.353**	
G_1 (kg/s)	19.705	19.705	19.705	**19.705**	
G_2 (kg/s)	24.898	24.899	24.899	**24.899**	
G_8 (kg/s)	7.459	7.459	7.459	**7.459**	
G_9 (kg/s)	15.130	15.131	15.131	**15.131**	
ΔP_1 (GAMS)	92.29	94.78	94.80	94.80	
ΔP_2 (GAMS)	110.23	112.34	112.37	112.38	
ΔP_8 (GAMS)	76.68	75.28	75.29	75.29	
ΔP_9 (GAMS)	80.18	78.94	78.99	78.99	
ΔP_1 (PRO/II)	94.78	94.80	94.80	94.80	
ΔP_2 (PRO/II)	112.33	112.37	112.38	112.38	
ΔP_8 (PRO/II)	75.28	75.29	75.29	75.29	
ΔP_9 (PRO/II)	78.93	78.99	78.99	78.99	
Δ_1 (kPa)	0	2.489	2.512	2.512	**2.512**
Δ_2 (kPa)	0	2.105	2.137	2.140	**2.140**
Δ_8 (kPa)	0	−1.398	−1.385	−1.388	**−1.388**
Δ_9 (kPa)	0	−1.256	−1.201	−1.202	**−1.202**

[a] Values in bold indicate the final results of the example.

Notice that, using only mass balances, there are no redundant variables, and the estimates of the flows are equal to the measured values. This leaves the system without bias and leak detection capabilities.

Example 2: Cerro Fortunoso Field Case. Consider the Cerro Fortunoso Field presented earlier in Figure 1.[5] Inlets to the system are assumed to be pure methane at $T = 25\ °C$. The flows of all inlet and outlet streams are measured, as are the pressures of the inlet streams of the system. There is no elevation change on the pipes. The dimensions of the pipes assumed for the data reconciliation are shown in Table 1, based on the stream i.d.'s presented in Figure 1. Inlet pressures and flow rate measurements are presented in Table 5.

The redundant set in this problem consists of 44 equations and is omitted here for simplicity. The algorithm required five iterations to meet the termination criterion of $|\Delta_i - \Delta_{i+1}| < 0.01\%$. The execution time on a Pentium III PC, 850 GHz, 128 Gb RAM computer was about 40 s.

The results of the problem are summarized in Table 6. In this table, we compare the values obtained using material and mechanical energy balances with those obtained using material balances only. In this example, S_{out}, perhaps the most important flow, is corrected to give a reconciled flow that is 0.31% larger than the measurement, whereas the combined material balance and mechanical energy data reconciliation gives a value that is 0.37% lower. The difference between one estimator and the other is 0.68%. This is a significant difference that has a sizable economical impact in production accounting. Finally, the errors of the model for each iteration are shown in Table 7. All flows are redundant in both cases. This example takes a total of 24 min and 34.92 s to be solved using the SIMSCI PRO/II process simulator. Twenty-five iterations of approximately 59 s each are performed by PRO/II. Note that, although the pressure drop in Table 6 has large deviations, these deviations result from the difference of two pressures,

Table 5. Pipe Dimensions and Measurements for Example #2

	measurement (kPa)	standard dev (kPa)	variance (kPa^2)		measurement (kg/s)	standard dev (kg/s)	variance (kg^2/s^2)
P_1	3028	76	5731	G_1	1.1860	0.0297	0.000 879
P_2	3039	76	5773	G_2	2.1960	0.0549	0.003 014
P_3	3021	76	5702	G_3	0.8680	0.0217	0.000 471
P_4	3060	76	5852	G_4	2.2270	0.0557	0.003 100
P_5	3018	75	5692	G_5	1.4250	0.0356	0.001 269
P_6	3005	75	5643	G_6	0.5720	0.0143	0.000 204
P_7	3019	75	5695	G_7	0.7610	0.0190	0.000 362
P_8	3005	75	5643	G_8	0.3700	0.0093	0.000 086
P_9	2992	75	5596	G_9	0.1930	0.0048	0.000 023
P_{10}	3010	75	5661	G_{10}	0.7500	0.0188	0.000 352
P_{11}	3124	78	6100	G_{11}	1.7089	0.0427	0.001 825
P_{12}	3549	89	7872	G_{12}	1.7767	0.0444	0.001 973
P_{13}	3224	81	6497	G_{13}	1.7545	0.0439	0.001 924
P_{14}	3209	80	6437	G_{14}	1.8525	0.0463	0.002 145
P_{15}	3164	79	6256	G_{15}	1.6707	0.0418	0.001 745
P_{16}	3089	77	5963	G_{16}	1.0817	0.0270	0.000 731
P_{17}	2981	75	5553	G_{17}	1.0364	0.0259	0.000 671
P_{18}	2985	75	5567	G_{18}	1.2048	0.0301	0.000 907
P_{19}	3184	80	6337	G_{19}	0.3828	0.0096	0.000 092
P_{20}	3456	86	7466	G_{20}	0.6353	0.0159	0.000 252
P_{21}	3377	84	7128	G_{21}	0.2566	0.0064	0.000 041
P_{22}	3534	88	7808	G_{22}	1.3777	0.0344	0.001 186
P_{23}	3487	87	7599	G_{23}	0.0727	0.0018	0.000 003
P_{24}	3410	85	7270	G_{24}	0.4614	0.0115	0.000 133
P_{25}	3029	76	5734	G_{25}	0.4549	0.0114	0.000 129
				G_{out}	26.1910	0.6569	0.431 507

Table 6. Results of Example 2

	pressure (kPa)		flow rate (kg/s)		
stream	measurement	estimate	measurement	estimators	estimators (material balance only)
1	3076	2812	1.186	1.168	1.186
2	3440	2810	2.196	2.133	2.195
3	3084	2820	0.868	0.858	0.868
4	3466	2836	2.227	2.167	2.226
5	3219	2871	1.425	1.403	1.425
6	2910	2930	0.572	0.569	0.572
7	3001	2997	0.761	0.757	0.761
8	2904	3078	0.370	0.369	0.370
9	2772	3166	0.193	0.193	0.193
10	3011	3256	0.750	0.749	0.750
11	3114	3161	1.709	1.708	1.709
12	3479	3165	1.777	1.775	1.776
13	3233	3171	1.754	1.754	1.754
14	3145	3171	1.853	1.852	1.852
15	3194	3171	1.671	1.670	1.670
16	3062	3246	1.082	1.082	1.082
17	2963	3245	1.036	1.037	1.036
18	2928	3357	1.205	1.206	1.205
19	3204	3371	0.383	0.383	0.383
20	3540	3373	0.635	0.636	0.635
21	3331	3368	0.257	0.257	0.257
22	3558	3363	1.378	1.379	1.377
23	3464	3356	0.073	0.073	0.073
24	3371	3336	0.461	0.462	0.461
25	2963	3333	0.455	0.455	0.455
out	- -	- -	26.191	26.094	26.271

estimators that are, in turn, well within the standard deviation of the measurements.

As the problem size increases, the number of iterations also increases for both rigorous and approximate approaches. Nevertheless, the time per iteration does not increase much in the case of the approximate models because of the low level of nonlinearity of the problem. Indeed, when only the first branch of the Fortunoso Field Case is taken into account (streams S_1-S_{10}, see Figure 1), the data reconciliation problem (with the same characteristics as example 2) takes approximately 40 s to converge with the proposed methodology. In contrast, it takes almost 10 min for the process simulator to solve the optimization problem (same PC as before). Therefore, in larger cases, the rigorous model approach is still less effective than the one proposed here.

Precision of Estimates

Once the data reconciliation algorithm is completed, it is desirable to know the precision of the estimators obtained. Because the relationship between estimators and measurements is nonlinear (through the iterative model), the classical relationship used for linear systems no longer applies.

One way to find the precision of estimates is by linearization. Assuming that the measurements (x)

Table 7. Model Errors for Example 2

model error (kPa)	iteration number					
	0	1	2	3	4	5
Δ_1	0	3.414	3.381	3.382	3.367	3.367
Δ_2	0	4.482	4.412	4.421	4.400	4.401
Δ_3	0	2.591	2.571	2.576	2.559	2.560
Δ_4	0	6.042	5.901	5.915	5.886	5.887
Δ_5	0	1.756	1.749	1.754	1.739	1.739
Δ_6	0	0.518	0.522	0.525	0.516	0.516
Δ_7	0	0.228	0.238	0.242	0.227	0.227
Δ_8	0	−0.515	−0.505	−0.503	−0.511	−0.511
Δ_9	0	−0.412	−0.409	−0.408	−0.411	−0.411
Δ_{10}	0	−2.005	−1.983	−1.980	−1.990	−1.990
Δ_{11}	0	−0.517	−0.499	−0.493	−0.512	−0.511
Δ_{12}	0	3.837	3.804	3.809	3.793	3.794
Δ_{13}	0	0.669	0.671	0.677	0.659	0.659
Δ_{14}	0	0.647	0.652	0.659	0.633	0.634
Δ_{15}	0	−0.109	−0.095	−0.090	−0.109	−0.108
Δ_{16}	0	−2.481	−2.432	−2.426	−2.447	−2.446
Δ_{17}	0	−3.511	−3.459	−3.454	−3.471	−3.471
Δ_{18}	0	−9.054	−8.845	−8.841	−8.871	−8.870
Δ_{19}	0	−2.036	−2.006	−2.002	−2.017	−2.016
Δ_{20}	0	0.345	0.347	0.349	0.343	0.343
Δ_{21}	0	0.024	0.027	0.029	0.024	0.024
Δ_{22}	0	1.391	1.386	1.390	1.378	1.378
Δ_{23}	0	0.123	0.124	0.124	0.122	0.123
Δ_{24}	0	0.176	0.178	0.179	0.175	0.175
Δ_{25}	0	18.846	27.802	−3.255	−2.993	−2.995
Δ_{out}	0	−1.315	−1.295	−1.297	−1.292	−1.292

Table 8. Parameters for Measurement Samples

	mean	σ
P_1 (kPa)	2533.13	75.99
P_2 (kPa)	2533.13	75.99
P_8 (kPa)	2143.58	64.31
P_9 (kPa)	2142.17	64.26
G_1 (kg/s)	20.00	0.600
G_2 (kg/s)	25.00	0.750
G_8 (kg/s)	7.50	0.225
G_9 (kg/s)	15.00	0.450

are related to the estimates ($\mathbf{z}$) by some nonlinear function $\mathbf{z} = \mathbf{f}(x)$, then linearization gives $\mathbf{z} \approx \mathbf{f}(x_0) + \mathbf{f}'(\mathbf{x_0})(x - \mathbf{x_0})$. Then, the standard linear approach can be applied to the linearized $\mathbf{z}$ to obtain the precision of the estimates. Nevertheless, this method involves the error of truncating the Taylor series. Thus, instead of using linearization, a sufficient number of measurement samples is taken and used to obtain the corresponding group of estimators from which the precision and variance is obtained. An example illustrating the process is presented next.

Example 3. Consider example 1 with the same characteristics and requirements. Two groups of measurement samples were generated, one with 10 samples and one with 15 samples. The samples were obtained with a normal distribution random number generator assuming 3% error in the measurements and a mean equal to the "true value" of the variable. The mean values and standard deviations of variables are presented in Table 8.

The results for both sample groups are presented in Table 9. First, it can be verified that the estimates (<2.7%) are more precise than the measurements (3%), as predicted by statistics theory. Second, when going from 10 to 15 samples, the results are similar, although still fluctuating. Hence, although a large number of measurements is needed to obtain more stable values for the precision of the estimators, one can obtain a good idea by using a relatively small number of samples. In addition, because the precision of the instruments does not change in time, for a given system, such a study has to be done only periodically (when the precision of the instruments change), but not associated with each data reconciliation.

Table 9. Precision of Estimates for Different Sample Group Sizes

	10 samples		15 samples	
	average	σ (%)	average	σ (%)
P_1 (kPa)	2470.88	1.371	2482.85	1.400
P_2 (kPa)	2492.05	1.458	2502.83	1.482
P_8 (kPa)	2188.09	1.628	2194.58	1.503
P_9 (kPa)	2179.11	1.650	2187.12	1.566
G_1 (kg/s)	19.67	2.527	19.87	2.665
G_2 (kg/s)	25.27	2.118	25.33	2.701
G_8 (kg/s)	7.51	2.138	7.52	1.985
G_9 (kg/s)	14.97	2.616	15.07	2.518

Use of Temperature Measurements

Even though temperature cannot be made redundant using the proposed model, there are many temperature measurements along the pipelines, and those measurements have certain errors, which affect the accuracy of the estimators. Errors in temperature measurements have a weak effect on the data reconciliation of pipeline systems. Indeed, consider eq 4. Temperature has an effect on parameters a and c through the density. Assume that we have an error of two standard deviations in the measured temperature. This allows us to rewrite eq 4 as

$$a(T^*+\epsilon_T)G_T^2 + (p_{T,1} - p_{T,2}) + c(T^*+\epsilon_T) = \Delta_T \quad (32)$$

where T^* is the true value of temperature, ϵ_T is the error in the temperature measurement, $a(T^*+\epsilon_T)$ and $c(T^*+\epsilon_T)$ are the corresponding model parameters a and c calculated at $T = T^* + \epsilon_T$, G_T and p_T are the corresponding reconciled values for a different temperature, and Δ_T is the corresponding error. Expanding the terms $a(T^*+\epsilon_T)$ and $c(T^*+\epsilon_T)$ using Taylor series, we rewrite eq 32 as

$$a(T^*)G_T^2 + (p_{T,1} - p_{T,2}) + c(T^*) \approx \Delta_T - \left[\frac{\partial a}{\partial T^*}G_T^2 + \frac{\partial c}{\partial T^*}\right]\epsilon_T \quad (33)$$

Thus

$$\Delta T = \Delta + \frac{\partial \Delta'}{\partial T^*}\epsilon_T \quad (34)$$

where $\Delta'(T^*) = a(T^*)G_T^2 + c(T^*)$. We notice that, in the case where $(\partial\Delta'/\partial T^*)\epsilon_T \ll \Delta$, that is, when the correction to the equation due to changes in temperature is negligible, we can also expect $G_T \approx G$ and $p_T \approx p$. However, when $(\partial\Delta'/\partial T^*)\epsilon_T$ is not negligible compared to Δ, then an effect of temperature is seen. Whether this will affect the flows more than the pressures or both equally depends on the relative weights in the objective function (precision of measurements). It would be nice to have better tests that would indicate whether temperatures are an issue. Nevertheless, in practice, after reconciliation is completed, one can simply check the effect of temperatures by repeating the run for other temperatures, as in the example below, which illustrates a case in which temperature errors have a negligible effect in reconciled flows.

Table 10. Temperature (°C) Measurements for Example 4

stream	original case	2.5% std dev	7.5% std dev
S_1	25.00	25.67	23.07
S_2	25.00	24.43	23.12
S_8	25.00	25.58	26.75
S_9	25.00	24.40	26.82

Table 11. Comparison of the Outcome of the Algorithm for Different Temperature Measurement Errors

	temperature standard deviation		
	0%	2.5%	7.5%
P_1 (kPa)	2470.045	2469.526	2469.925
P_2 (kPa)	2487.794	2487.601	2487.607
P_8 (kPa)	2184.618	2185.127	2184.841
P_9 (kPa)	2178.353	2178.360	2178.348
G_1 (kg/s)	19.705	19.706	19.706
G_2 (kg/s)	24.899	24.899	24.899
G_8 (kg/s)	7.459	7.459	7.459
G_9 (kg/s)	15.131	15.131	15.131

Table 12. Effect of Temperature Measurements on Model Errors

	$\partial\Delta/\partial T$	ϵ_T (°C)		$(\partial\Delta/\partial_T)*\epsilon_T$ (kPa)		Δ
pipe	(kPa/°C)	2.5%	7.5%	2.5%	7.5%	(kPa)
1	−0.377	0.670	−1.930	−0.253	0.728	2.512
2	−0.434	−0.570	−1.880	0.247	0.816	2.14
8	−0.317	0.580	1.750	−0.184	−0.554	−1.388
9	−0.274	−0.600	1.820	0.164	−0.499	−1.202

Example 4. Consider example 1 with the same characteristics and requirements. Two different cases are studied here, one with a 2.5% standard deviation in temperature measurements and one with a 7.5% standard deviation. The temperature measurements used for the numerical experiment are presented in Table 10.

For this case, the effect of gross errors in temperature measurements is summarized in Table 11. These results indicate that temperature measurements have a moderate effect on the model errors, as indicated by Table 12. However, in this example, the reconciled values of flows obtained with the proposed iteration procedure are not significantly altered by gross errors in temperature measurements, as discussed below in detail. We leave a more detailed investigation of this issue for future work.

The results presented in Table 11 show that gross errors in temperature measurements as large as 7.5% do not significantly distress the outcome of the proposed iterative process. The changes in the reconciled values are no larger than 0.02% for any of the redundant variables. It can also be noticed that there is no significant increase in the inaccuracy of the model when going from a 2.5 to a 7.5% deviation in the temperature measurements; it is equally irrelevant in both cases. Thus, very large gross errors in temperature measurements are needed to generate noticeable effects in the results of the data reconciliation scheme proposed.

Conclusions

In this paper, mechanical energy balances in an approximate form are used as model equations in addition to mass balances for data reconciliation in pipeline systems. An observability analysis based on the use of both material balances and mechanical energy equations is proposed. It is shown that the use of energy equations increases the amount of redundant and observable variables. An iterative process to solve the data reconciliation problem is proposed to account for the error of the approximate model. It is revealed that using approximate models for the mechanical energy equations reduces the execution time for the data reconciliation problem. Finally, the effect of gross errors in temperature on the data reconciliation process is studied, revealing that this effect is insignificant for gross errors of 10% or lower.

Acknowledgment

Financial support from the NSF though Grant CTS-0075814 is gratefully acknowledged.

Nomenclature

a = flow rate coefficient in the model
A = internal pipe area
c = independent term coefficient in the model
D, d = pipe internal diameter
ϵ = termination criterioncriteria for iteration process
f_f = friction factor
δF = friction energy loss differential
g = gravity constant (=9.8 m/s^{2}m^2/sec)
G = flow rate
G_T, P_T = reconciled values of flow rate and pressure, respectively, for temperature T
$\tilde{G}$, $\tilde{P}$ = flow rate estimate, pressure estimate
G^+, P^+ = flow rate measurement, pressure measurement
h = pipe elevation of the pipe
L = pipe length
M_i = measurement sample
P, p = fluid pressure
R = universal gas constant [=8.314 51 J/(gmol K)]
Re = Reynolds number
$\mathbf{S_G}$ = variance matrix of flow rates
$\mathbf{S_P}$ = variance matrix of pressures
T = fluid temperature
T^* = true fluid temperature
V = fluid velocity
δW = work differential
x = vector of measurements
z = gas compressibility factor
$\mathbf{z}$ = vector of estimators
ϵ_T = gross error in temperature measurements
κ_R = nondimensional threshold value for the error resilience
ρ = fluid density
Δ = model error
Δ_T = model error when temperature measurements are included

Appendix: Mechanical Energy Balance Equation

The mechanical energy balance on a differential pipe length is

$$g\,\mathrm{d}h + \frac{\mathrm{d}p}{\rho} + V\,\mathrm{d}V + \delta W + \delta F = 0 \qquad \text{(A1)}$$

No work is done by the pipeline against the surroundings or by the surroundings against the pipeline; thus, the work term can be disregarded. Multiplying both sides of eq A1 by the density squared gives

$$\rho^2 g\,\mathrm{d}h + \rho\,\mathrm{d}p + \rho^2 V\,\mathrm{d}V + \rho^2\,\delta F = 0 \qquad \text{(A2)}$$

However $V = G/\rho A$ and

$$\delta F = -f_f \frac{V^2}{2}\frac{dL}{D} = -\frac{1}{2}f_f\left[\frac{G}{\rho A}\right]^2\frac{dL}{D} \quad (A3)$$

Therefore, eq A2 becomes

$$\rho^2 g\, dh + \rho\, dp + \left(\frac{G}{A}\right)^2 \rho\, d\left(\frac{1}{\rho}\right) - \frac{1}{2}f_f\left(\frac{G}{A}\right)^2\frac{dL}{D} = 0 \quad (A4)$$

Using the gas law $p = z\rho RT$, one obtains

$$\rho^2 g\, dh + \frac{p}{zRT}\, dp + \left(\frac{G}{A}\right)^2 \rho\, d\left(\frac{zRT}{p}\right) - \frac{1}{2}f_f\left(\frac{G}{A}\right)^2\frac{dL}{D} = 0 \quad (A5)$$

Equation A5 can be integrated along a defined section of a pipe to obtain the integral form of the mechanical energy balance. A different expression is found depending on the type of assumptions made. The particular case of natural gas pipelines is a compressible gas flowing at velocities mostly below 50 m/s, which are considered low velocities. For low-velocity compressible gas flow, the pressure loss causes a slow density change. In addition, it can also be shown that, in the absence of large heat losses (that is, assuming that the flow is adiabatic), this type of flow is almost isothermal. Therefore, one approximation is to assume that the temperature is constant. Next, elevation changes are neglected. Thus, one obtains

$$(p_1^2 - p_2^2) - 2\left(\frac{G}{A}\right)^2 p_1 \ln\left(\frac{p_1}{p_2}\right) - f_f\left(\frac{G}{A}\right)^2 p_1 \frac{L}{D} = 0 \quad (A6)$$

This expression is implicit for p_2, and therefore, if p_1 is known, then numerical methods are needed to solve for p_2.

Simpler expressions of similar accuracy can be obtained. For instance, the error caused by neglecting the kinetic energy change due to expansion ($V\, dV$) is always lower than 1% for pressure drops lower than 40% of the inlet pressure, i.e., $\Delta P/P_{in} < 0.4$.[9] Indeed, for natural gas at a temperature of 20 °C and a velocity of 50 m/s, the error is about 1.2%. For superheated steam at 400 °C and 50 m/s, it is 0.83%. Then, the integration of A5 gives

$$(p_1^2 - p_2^2) = f_f\left(\frac{G}{A}\right)^2 p_1 \frac{L}{D} \quad (A7)$$

which is explicit.

When the elevation change in the pipeline is considerable, gravity effects must be taken into account. Then, if the acceleration term $V\, dV$ is neglected, from eq A5, one obtains (after some omitted intermediate steps)

$$\left(\frac{p_2}{p_1}\right)^2 = \left[1 - \frac{f_f}{2Dg\sin\alpha}\left(\frac{G}{\rho_1 A}\right)^2\right] \times \exp\left(-\frac{2L\rho_1 g\sin\alpha}{p_1}\right) - \frac{f_f}{2Dg\sin\alpha}\left(\frac{G}{\rho_1 A}\right)^2 \quad (A8)$$

where α is the angle of elevation of the pipe.

The error generated by using eq A7 or A8 is less than 1% for pressure drops lower than 40% of the inlet pressure, i.e., for $\Delta P/P_{in} < 0.4$. The problem with these formulations is their level of nonlinearity.

Finally, if one assumes incompressible flow and neglects the kinetic energy change due to expansion, the integration of eq A5 gives

$$\rho_1 g(h_1 - h_2) + (p_1 - p_2) = \frac{1}{2\rho_1}f_f\left(\frac{G}{A}\right)^2\frac{L}{D} \quad (A9)$$

Equation A9 generates errors smaller than 10% for $\Delta P/P_{in} < 0.3$, which is worse than those obtained using eq A7, but the nonlinearity in pressures is eliminated.

To perform the data reconciliation, eq A9 is used. To account for the error of this model, a term Δ is introduced into the equation

$$\rho_1 g(h_1 - h_2) + (p_1 - p_2) - \frac{1}{2\rho_1}f_f\left(\frac{G}{A}\right)^2\frac{L}{D} = \Delta \quad (A10)$$

The correction term Δ is calculated through an iterative process during the data reconciliation scheme, which is explained in detail in the Material and Energy Balance Reconciliation section of this paper.

We rewrite eq A10 as follows

$$aG^2 + (P_1 - P_2) + c = \Delta \quad (A11)$$

where

$$a = -\frac{8}{\pi^2}\frac{f_f}{D^4}\frac{1}{\rho_1}\frac{L}{D} \quad (A12)$$

and

$$c = \rho_1 g(h_2 - h_1) \quad (A13)$$

Flow rates and pressures are considered to be the only variables of eq A11. All other variables are considered to be known and constant during each data reconciliation step. To calculate the friction factor, f_f, a variety of expressions can be used, e.g., Prandtl formula.[6]

Literature Cited

(1) Madron, F. *Process Plant Performance. Measurement and Data Processing for Optimization and Retrofits*; Ellis Horwood: West Sussex, U.K., 1992.

(2) Romagnoli, J.; Sanchez, M. *Data Processing and Reconciliation for Chemical Processes Operations*; Academic Press: New York, 1999.

(3) Narasimhan, S.; Jordache, C. *Data Reconciliation & Gross Error Detection. An Intelligent Use of Process Data*; Gulf Publishing Company: Houston, TX, 2000.

(4) Coelho, M. R.; Medeiros, J. L. Reconciliação de Dados e Detecção de Vazamentos em Redes Estacionárias de escoamento Incompressivel. Presented at Empromer 99, Florianópolis, Brazil, Aug 30−Sep 2, 1999.

(5) Galliano, G. R.; Digiorgis, G. L, Lopez, M. L.; Fornes, A. Modeling and Optimization Oil and Gas Production and Gathering Systems: Cerro Fortunoso Field Case. *SPE Prod. Facil.* **2000**, *15*, 3.

(6) Bird, R. B.; Stewart, W. E.; Edwin N. L. *Transport Phenomena*, 2nd ed.; John Wiley & Sons: New York, 2001.

(7) Steward, D. V. *Soc. Ind. Appl. Math. Rev.* **1962**, *4*, 321.

(8) Steward, D. V. *J. Soc. Appl. Math. Num. Anal.* **1965**, *B2*, 345.

(9) Bobok, E. *Fluid Mechanics for Petroleum Engineers*; Elsevier Science Publishers B.V.: Amsterdam, The Netherlands, 1993.

Received for review September 27, 2002
Revised manuscript received August 15, 2003
Accepted August 18, 2003

IE020774J