

Clasificación de variables y reconciliación de datos en ingeniería de procesos

Miguel A. Lozano y Jesús A. Remiro

Departamento de Ingeniería Mecánica
Universidad de Zaragoza, Centro Politécnico Superior
María de Luna 3, 50018 Zaragoza, España
Tel.: 34-976-76 20 39, Fax: 34-976-76 26 16
e-mail: mlozano@posta.unizar.es

Resumen

Las medidas de proceso de las plantas térmicas se utilizan para tareas de control, evaluación de rendimientos, optimización de la operación, etc. Frecuentemente se dispone de más medidas que las necesarias. La reconciliación consiste en ajustar las medidas redundantes de modo que obedezcan las leyes de conservación y cualquier otra restricción que incorpore el modelo matemático de la planta. Como resultado: (i) se detectan y eliminan los errores sistemáticos de medida; (ii) se obtiene un conjunto consistente de medidas ajustadas; (iii) se estiman las variables no medidas por el método de máxima verosimilitud y (iv) se obtienen intervalos de confianza para los resultados. En este trabajo se propone el método de la descomposición QR para analizar, descomponer y resolver los problemas de reconciliación lineal. Los problemas de reconciliación no lineal pueden resolverse iterativamente linealizando las ecuaciones del modelo matemático. Se presenta un ejemplo para ilustrar la aplicación de los procedimientos de cálculo propuestos.

Palabras clave:

reconciliación de datos, clasificación de variables, sistemas energéticos, rendimiento.

VARIABLE CLASSIFICATION AND DATA RECONCILIATION IN PROCESS ENGINEERING

Summary

Process data are used in thermal plants for the purpose of control, performance evaluation, optimization, etc. Usually more data is measured than necessary. Reconciliation is based in statistical adjustment of redundant process data to obey the conservation laws and any other constraint imposed by the mathematical model of the plant. As a result: (i) detection and elimination of gross measurement errors are possible; (ii) new consistent set of data is obtained; (iii) unmeasured variables of the model are estimated on the maximum likelihood principle and (iv) confident intervals for results are obtained. In this work, the application of QR decomposition to analyze, decompose and solve the linear reconciliation problem is presented. Non-linear reconciliation problems can be also solved by linearizing the equations of the mathematical model. The calculation procedure is demonstrated by an illustrative example.

Key words:

data reconciliation, variable classification, energy systems, performance.

INTRODUCCIÓN

Obtener la mayor cantidad de información sobre la operación de la planta analizada y conocer su calidad son cometidos importantes del ingeniero de procesos. En este artículo se explica y aplica una formulación matemática para resolver los problemas de reconciliación lineal y no lineal aprovechando la redundancia de medidas. Su contribución al diagnóstico de procesos resulta relativamente sencilla con las herramientas de cálculo computacional de que hoy se dispone.

Por razones de coste, conveniencia o factibilidad técnica no se miden todas las variables de un proceso. Sin embargo, todavía podemos ser capaces de estimar su valor a partir del de otras variables medidas. Esto dependerá de cuál sea la estructura física de la planta analizada y del tipo, número y localización de los instrumentos de medida.

Sea $\mathbf{f}(\mathbf{z}) = \mathbf{0}$ el conjunto de M ecuaciones independientes a satisfacer por el conjunto de variables $\mathbf{z}[N \times 1]$ que describen la operación del proceso analizado. Sea $\mathbf{x}[I \times 1]$ el subconjunto de variables medidas e $\mathbf{y}[J \times 1]$ su complementario que incluye a las variables no medidas. A partir de las ecuaciones podremos calcular algunas de estas últimas a las que calificaremos de observables o calculables. Por otro lado, si alguna de las variables medidas dejara de estarlo, podría ocurrir que fuese calculable a partir del resto de las variables medidas. Calificaremos a dichas medidas de redundantes. La presencia de medidas redundantes hace posible la detección de errores groseros de medida y mejora la calidad de los resultados. Parece claro el interés tanto de que el mayor número de variables medidas sean redundantes como el de que el mayor número de variables no medidas sean calculables.

Según sea el subconjunto de variables medidas pueden presentarse distintos casos. Uno frecuentemente preconizado en la normativa sobre pruebas de rendimiento para la aceptación de equipos y sistemas es el de medir un conjunto de $I = N - M$ variables tales que el resto de las variables sean calculables a partir de las ecuaciones. Este caso, que podemos denominar como convencional, se analiza frecuentemente en la literatura y se caracteriza porque todas las variables medidas son no redundantes. En este artículo estudiaremos el caso general, que es el que se presenta con mayor frecuencia en el diagnóstico de la operación en tiempo real de equipos y sistemas.

Los avances recientes tanto en los aspectos teóricos de la clasificación de variables y reconciliación de datos como en la capacidad de cálculo están permitiendo incorporar en la práctica procedimientos de diagnóstico robustos y eficientes de los procesos químicos y energéticos que operan en estado estacionario.¹ Una buena muestra del grado de madurez alcanzado es la publicación reciente de libros dedicados al tema específico que aquí nos ocupa.²⁻⁵ En ellos se destaca la conveniencia de realizar una aproximación integral para resolver los distintos problemas del diagnóstico: (i) definición del número, tipo y localización de los instrumentos de medida, (ii) detección de los errores de medida, (iii) determinación de las variables de procesos calculables y (iv) evaluación de la calidad de los resultados obtenidos.

En este trabajo se propone el método de la descomposición QR para analizar, descomponer y resolver los problemas de reconciliación lineal. Los problemas de reconciliación no lineal pueden resolverse iterativamente linealizando las ecuaciones del modelo matemático. Se presenta un ejemplo para ilustrar la aplicación de los procedimientos de cálculo propuestos.

TÉCNICAS DE RECONCILIACIÓN

Reconciliación lineal

El problema matemático de reconciliación lineal cuando se dispone de medidas para todas las variables puede expresarse como

$$\begin{aligned} \text{minimizar } F &= (\mathbf{x}_m - \mathbf{x})^T \mathbf{Q}_m^{-1} (\mathbf{x}_m - \mathbf{x}) \\ \text{sujeto a } \mathbf{A}(\mathbf{x}_m - \mathbf{v}) &= \mathbf{a} \end{aligned} \quad (1)$$

donde $\mathbf{x}_m[N \times 1]$ son las medidas, $\mathbf{x}[N \times 1]$ son los valores reconciliados, $\mathbf{v} \equiv \mathbf{x}_m - \mathbf{x}$ son los ajustes de las medidas, $\mathbf{A}[M \times N]$ es la matriz de incidencia de las ecuaciones, $\mathbf{a}[M \times 1]$ es el término independiente de las ecuaciones y $\mathbf{Q}_m[N \times N]$ es la matriz de covarianza de las medidas.

Formulando la función lagrangiana

$$L(\mathbf{v}, \lambda) \equiv \mathbf{v}^T \mathbf{Q}_m^{-1} \mathbf{v} - 2\lambda^T (\mathbf{A}\mathbf{x}_m - \mathbf{A}\mathbf{v} - \mathbf{a}) \quad (2)$$

donde $\lambda[M \times 1]$ es el vector de multiplicadores de Lagrange y derivan respecto de las variables \mathbf{v} y λ , se obtienen las dos condiciones necesarias y suficientes de optimalidad

$$\frac{\partial L}{\partial \mathbf{v}} = 2\mathbf{Q}_m^{-1} \mathbf{v} + 2\mathbf{A}^T \lambda = \mathbf{0} \quad (3)$$

$$\frac{\partial L}{\partial \lambda} = \mathbf{A}\mathbf{x}_m - \mathbf{A}\mathbf{v} - \mathbf{a} = \mathbf{0} \quad (4)$$

De la primera

$$\mathbf{v} = -\mathbf{Q}_m \mathbf{A}^T \lambda \quad (5)$$

Sustituyendo en la segunda y despejando

$$\lambda = -(\mathbf{A}\mathbf{Q}_m \mathbf{A}^T)^{-1} \mathbf{r} \quad (6)$$

siendo $\mathbf{r}[M \times 1]$ los residuos de las ecuaciones

$$\mathbf{r} \equiv \mathbf{A}\mathbf{x}_m - \mathbf{a} \quad (7)$$

Para el cálculo de los ajustes \mathbf{v} se procederá resolviendo en orden inverso las tres últimas ecuaciones, tras lo cual podrán calcularse los valores reconciliados \mathbf{x} . La incertidumbre de las medidas producirá resultados también inciertos. Puede demostrarse que las matrices de covarianza de las variables calculadas son

$$\mathbf{Q}_r = \mathbf{A}\mathbf{Q}_m \mathbf{A}^T \quad (8)$$

$$\mathbf{Q}_v = \mathbf{Q}_m \mathbf{A}^T \mathbf{Q}_r^{-1} \mathbf{A}\mathbf{Q}_m \quad (9)$$

$$\mathbf{Q}_x = \mathbf{P}\mathbf{Q}_m \mathbf{P}^T \quad (10)$$

con

$$\mathbf{P} = \mathbf{I} - \mathbf{Q}_m \mathbf{A}^T \mathbf{Q}_r^{-1} \mathbf{A} \quad (11)$$

Clasificación de variables

Sea $\mathbf{z}[N \times 1]$ el vector de variables que deben satisfacer el sistema de M ecuaciones lineales siguiente

$$\mathbf{Cz} = \mathbf{c} \quad (12)$$

Si el rango de la matriz \mathbf{C} es igual a M , entonces todas las ecuaciones son independientes. Si además $M = N$, el problema tendrá solución sin necesidad de disponer de ninguna medida

$$\mathbf{z} = \mathbf{C}^{-1}\mathbf{c} \quad (13)$$

En general sólo dispondremos de medidas para algunas variables. También puede ocurrir que no puedan determinarse los valores de algunas de las variables no medidas a partir de las ecuaciones.

Si $\mathbf{x}[I \times 1]$ es el vector de variables medidas e $\mathbf{y}[J \times 1]$ el vector de variables no medidas, podemos realizar la partición del vector \mathbf{z} como

$$\mathbf{z} = \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} \quad (14)$$

Reordenando las columnas de \mathbf{C} para mantener coherente el sistema de ecuaciones resulta

$$\mathbf{C} = [\mathbf{B} \quad \mathbf{A}] \quad (15)$$

con $\mathbf{B}[M \times J]$ y $\mathbf{A}[M \times I]$. El sistema de ecuaciones podrá escribirse ahora como

$$\mathbf{B}\mathbf{y} + \mathbf{A}\mathbf{x} = \mathbf{c} \quad (16)$$

y el problema de reconciliación resulta

$$\begin{aligned} &\text{minimizar } F = (\mathbf{x}_m - \mathbf{x})^T \mathbf{Q}_m^{-1} (\mathbf{x}_m - \mathbf{x}) \\ &\text{sujeto a } \mathbf{B}\mathbf{y} + \mathbf{A}\mathbf{x} = \mathbf{c} \end{aligned} \quad (17)$$

Se han propuesto distintos métodos para simplificar la resolución de este problema a la par que se clasifican las variables. Los más utilizados son el de eliminación de Gauss-Jordan^{3,6} y el de la matriz de proyección.¹ Aquí describiremos una versión de este último propuesta por Sánchez y Romagnoli⁷ que utiliza la descomposición QR de la matriz \mathbf{B} . Como veremos, este método permite responder a las preguntas clave del problema de clasificación: (i) cuáles de entre las variables no medidas podrán calcularse y cuáles no; (ii) cuáles de las variables medidas son redundantes, es decir, tales que si se perdiera su medida aún podrían calcularse las mismas variables no medidas e incluso ella misma. Además aprovecha la información obtenida para simplificar la resolución numérica del problema de reconciliación. Veamos los pasos del método propuesto.

1. Descomposición QR de la matriz \mathbf{B}

La descomposición QR de una matriz $\mathbf{B}[M \times J]$ consiste en encontrar una matriz ortogonal $\mathbf{Q}[M \times M]$, una matriz de permutación $\Pi[J \times J]$ y una matriz trapezoidal superior $\mathbf{R}[M \times J]$ con elementos diagonales de magnitud decreciente, tales que

$$\mathbf{B}\Pi = \mathbf{QR} \quad (18)$$

Las matrices \mathbf{Q} y \mathbf{R} quedan particionadas como

$$\mathbf{Q} = [\mathbf{Q}_1 \quad \mathbf{Q}_2] \quad \mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (19)$$

siendo $\mathbf{R}_1[L \times L]$ una matriz triangular superior cuyo rango L obtenido como resultado es también el rango de \mathbf{B} . En las librerías profesionales de cálculo numérico pueden encontrarse procedimientos adecuados para realizar esta operación. Nosotros hemos utilizado las subrutinas LQRRR y LQERR de la librería matemática IMSL.⁸ Una comparación de la descomposición QR con otros métodos puede verse en Quintana y Quintana.⁹

2. Sistema equivalente de restricciones

A partir de la matriz de permutación el conjunto de las variables no medidas $\mathbf{y}[J]$ puede particionarse en dos subconjuntos $\mathbf{y}_1[L]$ e $\mathbf{y}_2[J - L]$ mediante

$$\Pi^T \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \quad (20)$$

Sustituyendo $\mathbf{B} = \mathbf{QR}\Pi^{-1}$ en la ec. (16), premultiplicando por \mathbf{Q}^{-1} y teniendo en cuenta que $\mathbf{Q}^{-1} = \mathbf{Q}^T$ y $\Pi^{-1} = \Pi^T$, resulta

$$\begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{bmatrix} \mathbf{A} \mathbf{x} + \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \end{bmatrix} \mathbf{c} \quad (21)$$

$$\begin{bmatrix} \mathbf{Q}_1^T \mathbf{A} & \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{Q}_2^T \mathbf{A} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_1^T \mathbf{c} \\ \mathbf{Q}_2^T \mathbf{c} \end{bmatrix} \quad (22)$$

Por tanto el sistema lineal original dado por la ec. (16) queda transformado en el conjunto de los dos siguientes

$$\mathbf{Q}_1^T \mathbf{A} \mathbf{x} + \mathbf{R}_1 \mathbf{y}_1 + \mathbf{R}_2 \mathbf{y}_2 = \mathbf{Q}_1^T \mathbf{c} \quad (23)$$

$$\mathbf{Q}_2^T \mathbf{A} \mathbf{x} = \mathbf{Q}_2^T \mathbf{c} \quad (24)$$

Analizando estos sistemas de ecuaciones podemos clasificar las variables.

3. Clasificación de las variables no medidas

Como hemos dicho, la descomposición QR de la matriz \mathbf{B} permite obtener el valor de L que es el rango de \mathbf{B} y \mathbf{R}_1 . Si $L = J$, donde J es el número de variables no medidas, entonces todas las variables no medidas podrán calcularse a partir de las ecuaciones del modelo y de las variables medidas. En caso contrario $L < J$ y al menos las $J - L$ variables no medidas correspondientes al subconjunto \mathbf{y}_2 no podrán calcularse. Premultiplicando la ec. (23) por \mathbf{R}_1^{-1} y despejando \mathbf{y}_1 , resulta

$$\mathbf{y}_1 = -\mathbf{R}_1^{-1} \mathbf{Q}_1^T \mathbf{A} \mathbf{x} - \mathbf{R}_1^{-1} \mathbf{R}_2 \mathbf{y}_2 + \mathbf{Q}_1^T \mathbf{c} \equiv -\mathbf{A}_L \mathbf{x} - \mathbf{R}_{12} \mathbf{y}_2 + \mathbf{c}_L \quad (25)$$

Analizando las filas de la matriz \mathbf{R}_{12} podemos clasificar las variables pertenecientes al subconjunto \mathbf{y}_1 . Si la fila correspondiente a una variable en dicha matriz está formada por ceros, entonces la variable podrá calcularse. Por el contrario, si algunos de los elementos de la fila no son cero, no podrá calcularse, pues al menos necesitaríamos medir el valor de las variables correspondientes en \mathbf{y}_2 para hacerlo. Definiendo como $\mathbf{A}_{L_0}[L_0 \times I]$, $\mathbf{y}_0[L_0]$ y $\mathbf{c}_{L_0}[L_0]$ a la matriz y vectores resultantes de eliminar las filas correspondientes a las variables no calculables en $\mathbf{A}_L[L \times I]$, $\mathbf{y}_1[L]$ y $\mathbf{c}_L[L]$, podemos escribir

$$\mathbf{y}_0 = -\mathbf{A}_{L_0} \mathbf{x} + \mathbf{c}_{L_0} \quad (26)$$

4 Clasificación de las variables medidas

Comenzaremos definiendo $\mathbf{A}_1[H \times I]$ y $\mathbf{c}_1[H]$, siendo $H \equiv M - L$ el denominado grado de redundancia del problema de reconciliación, reformulando la ec. (24)

$$\mathbf{Q}_2^T \mathbf{A} \mathbf{x} = \mathbf{Q}_2^T \mathbf{c} \rightarrow \mathbf{A}_1 \mathbf{x} = \mathbf{c}_1 \quad (27)$$

La variable medida x_j será redundante y participará en la reconciliación de datos, es decir su valor estimado podrá ser distinto al medido, si su correspondiente columna en la matriz \mathbf{A}_1 contiene al menos un elemento no nulo. En caso contrario dicha variable será clasificada como no redundante.

Reconciliación lineal con variables no medidas

El problema de reconciliación original dado por la ec. (17) se transforma a la forma equivalente

$$\begin{aligned} \text{minimizar } & F = (\mathbf{x}_m - \mathbf{x})^T \mathbf{Q}_m^{-1} (\mathbf{x}_m - \mathbf{x}) \\ \text{sujeto a } & \mathbf{A}_1 \mathbf{x} = \mathbf{c}_1 \\ & \mathbf{y}_0 = -\mathbf{A}_{L_0} \mathbf{x} + \mathbf{c}_{L_0} \end{aligned} \quad (28)$$

donde $\mathbf{y}_0[L_0 \times 1]$ contiene las variables no medidas calculables. Este puede resolverse en dos etapas. Primero se realiza la reconciliación de datos

$$\begin{aligned} \text{minimizar } & F = (\mathbf{x}_m - \mathbf{x})^T \mathbf{Q}_m^{-1} (\mathbf{x}_m - \mathbf{x}) \\ \text{sujeto a } & \mathbf{A}_1 \mathbf{x} = \mathbf{c}_1 \end{aligned} \quad (29)$$

por el procedimiento descrito anteriormente, obteniéndose los siguientes resultados

$$\mathbf{r} = \mathbf{A}_1 \mathbf{x}_m - \mathbf{c}_1 \quad [H \times 1] \quad (30)$$

$$\mathbf{Q}_r = \mathbf{A}_1 \mathbf{Q}_m \mathbf{A}_1^T \quad [H \times H] \quad (31)$$

$$\lambda = -\mathbf{Q}_r^{-1} \mathbf{r} \quad [H \times 1] \quad (32)$$

$$\mathbf{v} = -\mathbf{Q}_m \mathbf{A}_1^T \lambda \quad [I \times 1] \quad (33)$$

$$\mathbf{Q}_v = \mathbf{Q}_m \mathbf{A}_1^T \mathbf{Q}_r^{-1} \mathbf{A}_1 \mathbf{Q}_m \quad [I \times I] \quad (34)$$

$$\mathbf{x} = \mathbf{x}_m - \mathbf{v} \quad [I \times 1] \quad (35)$$

$$\mathbf{P} = \mathbf{I} - \mathbf{Q}_m \mathbf{A}_1^T \mathbf{Q}_r^{-1} \mathbf{A}_1 \quad [I \times I] \quad (36)$$

$$\mathbf{Q}_x = \mathbf{P} \mathbf{Q}_m \mathbf{P}^T \quad [I \times I] \quad (37)$$

Posteriormente se determinaran las variables no medidas calculables y su matriz de covarianza

$$\mathbf{y}_0 = -\mathbf{A}_{L_0}\mathbf{x} + \mathbf{c}_{L_0} \quad [L_0 \times 1] \quad (38)$$

$$\mathbf{R} = \mathbf{A}_{L_0}\mathbf{P} \quad [L_0 \times I] \quad (39)$$

$$\mathbf{Q}_{y_0} = \mathbf{R}\mathbf{Q}_m\mathbf{R}^T \quad [L_0 \times L_0] \quad (40)$$

Reconciliación no lineal

Sea $\mathbf{f}(\mathbf{z}) = \mathbf{0}$ el conjunto de M ecuaciones a cumplir por el conjunto de variables $\mathbf{z}[N \times 1]$ que describen la operación del proceso analizado. Sea $\mathbf{x}[I \times 1]$ el subconjunto de variables medidas e $\mathbf{y}[J \times 1]$ su complementario que incluye a las no medidas.

De momento supondremos que el número de variables no medidas J es igual a M (pudiendo incluir en ellas algunas medidas descartadas) y que las $I = N - M$ medidas seleccionadas son independientes. Aunque no sea necesariamente el método a aplicar para resolver el sistema de ecuaciones, calculando el vector \mathbf{y} , describiremos el de Newton por razones teóricas. En este método se aproximan las funciones f_i por su desarrollo en serie de Taylor truncado después de las primeras derivadas para dar el sistema de ecuaciones lineales

$$\mathbf{J}(\mathbf{f}, \mathbf{y}^*)\Delta\mathbf{y} = -\mathbf{f}(\mathbf{y}^*) \quad (41)$$

donde $J_{ij}(f, y) \equiv \partial f_i / \partial y_j$, $\Delta\mathbf{y} = \mathbf{y} - \mathbf{y}^*$, \mathbf{y}^* es la aproximación disponible a la solución e \mathbf{y} la nueva solución calculada que se espera este más próxima a la solución real. Resolviendo para $\Delta\mathbf{y}$ puede aplicarse un proceso iterativo con

$$\mathbf{y}^* \leftarrow \mathbf{y}^* + \Phi\Delta\mathbf{y} \quad (42)$$

hasta satisfacer el criterio de convergencia

$$|\Delta y_j / y_j| < \epsilon \quad \forall j = 1, \dots, M \quad (43)$$

El número de iteraciones necesario dependerá de lo pequeño que sea ϵ y del factor de relajación Φ . Resulta conveniente utilizar $\Phi < 1$ en las primeras iteraciones cuando no se disponga de una buena aproximación inicial para el vector \mathbf{y}^* .

La incertidumbre de los resultados puede calcularse a partir de las incertidumbres de los datos

$$\sigma^2(y_j) = \sum_i \left(\frac{\partial y_j}{\partial x_i} \right)^2 \sigma^2(x_i) \quad (44)$$

suponiendo exactas las ecuaciones empleadas en el modelo. Si parte de los resultados corresponde a variables cuyas medidas han sido descartadas, la comparación medida–resultado nos permitirá covalidar ambos o detectar la posible presencia de errores de medida.

En el caso más general, dado el conjunto de medidas, podrá ocurrir que algunas de ellas sean redundantes y que entre las variables no medidas unas sean calculables y otras no. Para resolver el problema de reconciliación no lineal vamos a aprovechar el método de reconciliación lineal presentado antes. Aplicaremos un procedimiento iterativo. En cada iteración procederemos a linealizar las ecuaciones y a estimar una nueva solución por reconciliación lineal hasta satisfacer el criterio de convergencia impuesto.

Sea $(\mathbf{z}^*)^T = [(\mathbf{y}^*)^T, (\mathbf{x}^*)^T]$ el punto inicial de una iteración cualquiera (en la primera iteración tomaremos $\mathbf{x}^* = \mathbf{x}_m$ y valores adecuados para \mathbf{y}^*). El sistema de ecuaciones lineales que aproxima las ecuaciones no lineales en torno a dicho punto viene dado por

$$\approx \mathbf{f}(\mathbf{y}, \mathbf{x}) = \mathbf{f}(\mathbf{y}^*, \mathbf{x}^*) + \mathbf{J}(\mathbf{f}, \mathbf{y}^*)(\mathbf{y} - \mathbf{y}^*) + \mathbf{J}(\mathbf{f}, \mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \quad (45)$$

Imponiendo la condición de que las ecuaciones deben satisfacerse, es decir, $\approx \mathbf{f}(\mathbf{y}, \mathbf{x}) = \mathbf{0}$, resulta

$$\mathbf{J}(\mathbf{f}, \mathbf{y}^*)\mathbf{y} + \mathbf{J}(\mathbf{f}, \mathbf{x}^*)\mathbf{x} = -\mathbf{f}(\mathbf{y}^*, \mathbf{x}^*) + \mathbf{J}(\mathbf{f}, \mathbf{y}^*)\mathbf{y}^* + \mathbf{J}(\mathbf{f}, \mathbf{x}^*)\mathbf{x}^* \quad (46)$$

Identificando los coeficientes de esta ecuación con los correspondientes al caso lineal [ec. (16)], resulta

$$\mathbf{B} = \mathbf{J}(\mathbf{f}, \mathbf{y}^*) \quad (47)$$

$$\mathbf{A} = \mathbf{J}(\mathbf{f}, \mathbf{x}^*) \quad (48)$$

$$\mathbf{c} = -\mathbf{f}(\mathbf{y}^*, \mathbf{x}^*) + \mathbf{J}(\mathbf{f}, \mathbf{y}^*)\mathbf{y}^* + \mathbf{J}(\mathbf{f}, \mathbf{x}^*)\mathbf{x}^* \quad (49)$$

La resolución del problema de reconciliación lineal nos proporcionará nuevos valores de \mathbf{x} e \mathbf{y} . Sólo cambiarán de valor las variables medidas redundantes y las variables no medidas calculables. Con $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^*$ e $\Delta\mathbf{y} = \mathbf{y} - \mathbf{y}^*$ comprobaremos el criterio de convergencia impuesto y en caso de no cumplirse procederemos a estimar el punto inicial de la siguiente iteración: $\mathbf{x}^* \leftarrow \mathbf{x}^* + \Phi\Delta\mathbf{x}$, $\mathbf{y}^* \leftarrow \mathbf{y}^* + \Phi\Delta\mathbf{y}$.

El análisis de propagación de la incertidumbre y la detección de errores de medida no ofrece ninguna dificultad adicional, debiendo aplicarse los procedimientos explicados para el caso lineal a los resultados de la última iteración.

Veverka y Madron³ presentan otros métodos para resolver el problema de reconciliación no lineal y analizan las condiciones necesarias para que pueda garantizarse la resolución de un problema concreto.

INCERTIDUMBRE DE LOS RESULTADOS Y ERRORES DE MEDIDA

Ajustabilidad

La ajustabilidad de la variable medida j es

$$\omega_j \equiv 1 - \left[\frac{\sigma_x(j)}{\sigma_{x_m}(j)} \right] \quad (50)$$

Dada la relación entre las varianzas de las medidas, ajustes y estimados de las variables medidas

$$\sigma_x^2(j) = \sigma_{x_m}^2(j) - \sigma_v^2(j) \leq \sigma_{x_m}^2(j) \quad (51)$$

la ajustabilidad de las variables medidas no redundantes es nula y la ajustabilidad de las variables redundantes es menor que la unidad.

La ajustabilidad nos informa sobre la posibilidad de calcular correctamente una variable en el caso de que por la razón que fuese se perdiera su medida. También mide la facilidad con que puede detectarse un error grosero de medida. Por tanto, estaremos interesados en que las variables medidas tengan valores elevados de ajustabilidad. Para conseguirlo deberemos aumentar el número de variables medidas.

Propagación de la incertidumbre

La magnitud de la incertidumbre de los resultados es función de la incertidumbre de las medidas. Las matrices de covarianza nos proporcionan información sobre la incertidumbre de los resultados.

Sea $\sigma_{y_0}^2(j)$ la varianza estimada de una variable no medida calculable j . Entonces podemos escribir que el intervalo de confianza que contiene el valor real de $y_0(j)$ con una probabilidad mayor al 68 %, 95 % y 99 % viene dado, respectivamente, por

$$y_0(j) \pm \sigma_{y_0}(j) \quad y_0(j) \pm 2\sigma_{y_0}(j) \quad y_0(j) \pm 3\sigma_{y_0}(j) \quad (52)$$

Lógicamente estamos interesados en reducir al máximo dicho intervalo de confianza. Para ello tenemos que conseguir que disminuya $\sigma_{y_0}(j)$, lo cual pasa por reducir la incertidumbre de las medidas. Si queremos actuar selectivamente reduciendo la incertidumbre de un resultado en particular, convendrá conocer sobre qué medidas actuar prioritariamente. Es posible, al menos de forma aproximada, expresar la varianza $\sigma_{y_0}^2(j)$ como una función de las varianzas de las medidas en la forma

$$\sigma_{y_0}^2(j) = \sum_i R(j, i)^2 \sigma_{x_m}^2(i) \quad (53)$$

donde los $R(j, i)$ son los elementos de la matriz \mathbf{R} . La expresión anterior nos permite determinar qué porción relativa de la varianza se debe a la imprecisión (varianza) de las medidas

$$\theta(j, i) = R(j, i)^2 \left[\frac{\sigma_{x_m}^2(i)}{\sigma_{y_0}^2(j)} \right] \quad (54)$$

Calculados los valores de $\theta(j, i)$, conviene ordenarlos de mayor a menor despreciando aquellos menores que 0,03 (3 %). Normalmente la suma de los restantes superará el 95 %. Para disminuir la incertidumbre del valor estimado $y_0(j)$ deberá actuarse sobre la incertidumbre de las medidas con mayores cuotas $\theta(j, i)$.

Un análisis paralelo para las variables medidas conduce a las ecuaciones siguientes

$$x(j) \pm \sigma_x(j) \quad x(j) \pm 2\sigma_x(j) \quad x(j) \pm 3\sigma_x(j) \quad (55)$$

$$\theta(j, i) = P(j, i)^2 \left[\frac{\sigma_{x_m}^2(i)}{\sigma_x^2(j)} \right] \quad (56)$$

donde los $P(j, i)$ son los elementos de la matriz \mathbf{P} . Lógicamente la cuota dominante corresponderá ahora a la propia variable analizada, pero otras medidas también pueden aportar cuotas importantes. En el caso de variables medidas no redundantes se cumplirá que $\theta(j, j) = 1$.

Detección de errores de medida

Sólo pueden detectarse errores groseros en las medidas redundantes. Cuanto mayor sea el número de variables medidas, mayor número de ellas serán redundantes y mayor la probabilidad de detectar errores de medida cuando existan.

La presencia de errores groseros de medida se establece con las pruebas siguientes:²

1. Test global
si $F > \chi_{1-\alpha}^2(H)$ se detecta error.
2. Test sobre medidas redundantes
si $\varepsilon_{vj} \equiv |v_j/\sigma_{vj}| > \kappa_\alpha(D)$ se detecta error en la medida j (D es el número de valores de ε_{vj} diferentes).

En la Tabla I se anotan los estadísticos necesarios para aplicar dichos test con $\alpha = 0,05$ (5 %), lo que implica una probabilidad mayor que $1 - \alpha = 0,95$ (95 %) de que exista error cuando el test lo indique.

| H, D | $\chi_{1-0,05}^2$ | $\kappa_{0,05}$ | H, D | $\chi_{1-0,05}^2$ | $\kappa_{0,05}$ |
|--------|-------------------|-----------------|--------|-------------------|-----------------|
| 1 | 3,84 | 1,96 | 15 | 25,0 | 2,93 |
| 2 | 5,99 | 2,24 | 20 | 31,4 | 3,02 |
| 3 | 7,82 | 2,39 | 25 | 37,0 | 3,08 |
| 4 | 9,49 | 2,49 | 30 | 43,8 | 3,14 |
| 5 | 11,1 | 2,57 | 35 | 49,8 | 3,18 |
| 6 | 12,6 | 2,63 | 40 | 55,8 | 3,22 |
| 7 | 14,1 | 2,68 | 45 | 61,8 | 3,25 |
| 8 | 15,5 | 2,73 | 50 | 67,5 | 3,28 |
| 9 | 16,9 | 2,77 | 100 | 124 | ~ 3,3 |
| 10 | 18,3 | 2,80 | 200 | 234 | ~ 3,3 |

Tabla I. Criterios para detectar errores de medida

Cuando se detecte error de medida con el test global, debe actuarse en dos direcciones distintas. Por un lado revisando el proceso de obtención de las medidas redundantes sospechosas de error (y posiblemente también las hipótesis implícitas en el modelo matemático: estado estacionario, ausencia de fugas, etc.). Por otra parte comprobando los resultados del problema de reconciliación, cuando se desechan una a una las medidas sospechosas. Se espera que si una medida sospechosa x_j se declara no medida, se obtendrá como resultado de la reconciliación subsiguiente una reducción drástica de F . Puede demostrarse³ que cuando sólo se elimina una medida redundante x_j , el nuevo valor F^* viene dado por

$$F^* = F - \varepsilon_{vj}^2 \quad (57)$$

Así puede comprobarse el cumplimiento del test global sin necesidad de repetir cálculos. Si ocurre

$$F^* < \chi_{1-\alpha}^2(H - 1) \quad (58)$$

estaremos en condiciones de avalar la hipótesis de que la medida sospechosa eliminada es realmente responsable del error global detectado. No debe pensarse sin embargo que esta forma de proceder es una panacea. Puede ocurrir que ninguna o más de una medida cumplan el criterio anterior. También debe tenerse en cuenta que un error grosero en una medida no redundante no es detectable y si está presente, corromperá los valores calculados del resto de las variables.

EJEMPLO

La Figura 1 muestra el diagrama de flujos de una central termoelectrica simple considerada como ejemplo. El objetivo principal del diagnóstico energético es determinar el rendimiento eléctrico de la central (rpl_{total}) y el rendimiento energético de sus tres seccio-

nes: combustión (rplcomb), caldera (rplcald) y ciclo de potencia (rplciclo).

El modelo matemático no lineal que describe la operación de la central termoeléctrica consta de $M = 22$ ecuaciones y $N = 44$ variables.

En la Tabla II se clasifican las variables para tres casos distintos y se presentan resultados detallados para el último. En las columnas 1, 2 y 3 se muestran los resultados de la clasificación de variables para dichos casos que presentan distintos conjuntos de medidas. El primer caso, columna 1, corresponde a la situación preconizada en la normativa sobre pruebas de rendimiento. El conjunto de variables medidas $I = 22$ resulta suficiente para determinar el estado de la planta. No existen medidas redundantes y todas las variables no medidas $J = 22$ pueden calcularse mediante el modelo matemático. Al no existir medidas redundantes, no podrán detectarse errores de medida. En el segundo caso, columna 2, tal vez el más frecuente en la práctica industrial, se dispone de la medida de ciertas variables. Algunas de ellas resultan redundantes. Ciertas variables no medidas pueden calcularse con el modelo matemático y el resto no. El tercer caso, columna 3, que es para el que se presentan resultados numéricos, todas las variables no medidas son calculables y algunas de las variables medidas son redundantes.

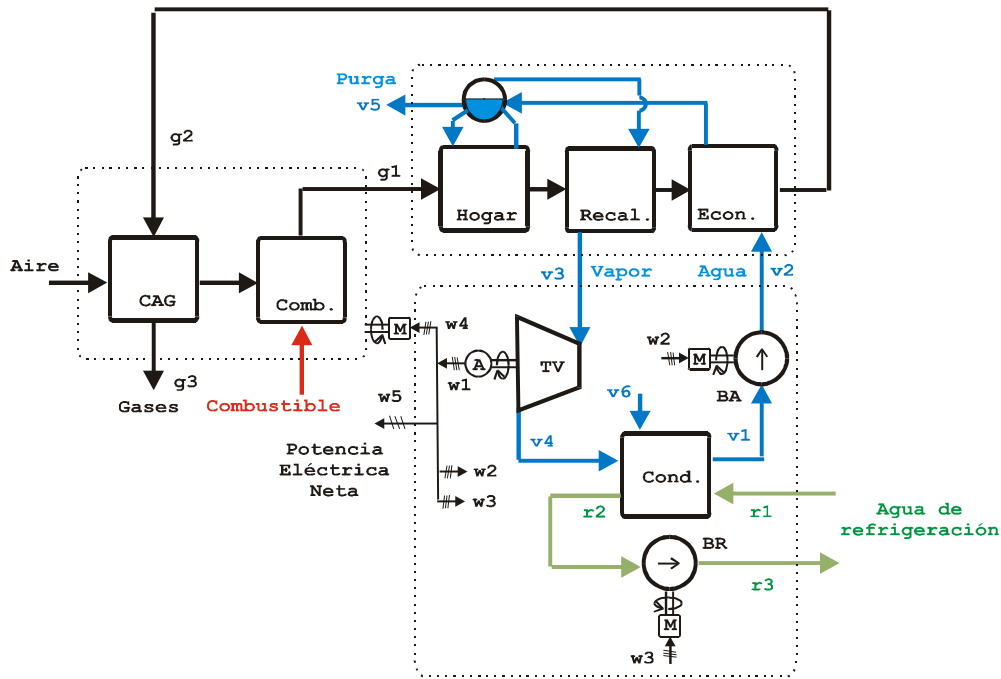


Figura 1. Esquema y variables de la planta analizada

Los datos y resultados detallados correspondientes al último caso se muestran en las cuatro últimas columnas de la Tabla II. En la primera de ellas se incluyen los valores de las medidas x_m y estimaciones iniciales para las variables no medidas y_{apro} . En la segunda se aportan los valores disponibles de las desviaciones típicas de las medidas σ_{x_m} . Con estos datos se lleva a cabo la reconciliación no lineal con el método propuesto obteniéndose, además de la clasificación de variables, los resultados mostrados en las dos últimas columnas. En la primera de ellas se presentan los valores reconciliados de las variables medidas x junto con los valores calculados y de las variables no medidas que satisfacen el sistema de ecuaciones. En la segunda se anotan las desviaciones típicas de los resultados, tanto las correspondientes a las variables medidas σ_x como a las variables no medidas σ_y .

| Variable | 1 | 2 | 3 | $x_m - y_{apro}$ | σ_{x_m} | $x - y$ | $\sigma_x - \sigma_y$ |
|------------------|----|----|----|------------------|----------------|------------|-----------------------|
| rpc (adim.) | MN | MR | MR | 10.000E-3 | 1.000E-3 | 10.196E-3 | 989.396E-6 |
| rca (adim.) | NC | NC | NC | 94.000E-3 | - | 94.059E-3 | 214.048E-6 |
| yo2 (adim.) | MN | MR | MR | 20.000E-3 | 400.000E-6 | 20.001E-3 | 395.205E-6 |
| rnc (kmol/s) | NC | MR | MR | 240.000E-3 | 1.200E-3 | 238.845E-3 | 858.949E-6 |
| rna (kmol/s) | NC | MR | MR | 2.540E+0 | 50.000E-3 | 2.539E+0 | 10.664E-3 |
| rng (kmol/s) | NC | NC | MR | 2.770E+0 | 55.000E-3 | 2.778E+0 | 11.395E-3 |
| ta (K) | MN | MR | MR | 300.000E+0 | 500.000E-3 | 299.981E+0 | 499.803E-3 |
| tg1 (K) | NC | NN | NC | 2.300E+03 | - | 2.332E+03 | 4.264E+0 |
| tg2 (K) | MN | NN | MN | 600.000E+0 | 2.000E+0 | 600.000E+0 | 2.000E+0 |
| tg3 (K) | MN | MR | MR | 400.000E+0 | 1.000E+0 | 400.089E+0 | 997.828E-3 |
| rmv1 (kg/s) | NC | MR | MR | 56.500E+0 | 250.000E-3 | 56.664E+0 | 187.320E-3 |
| rmv3 (kg/s) | NC | MR | MR | 56.000E+0 | 500.000E-3 | 56.165E+0 | 187.836E-3 |
| rmv5 (kg/s) | MN | MR | MR | 500.000E-3 | 25.000E-3 | 499.060E-3 | 24.841E-3 |
| rmv6 (kg/s) | NC | MR | MR | 500.000E-3 | 250.000E-3 | 499.060E-3 | 24.841E-3 |
| rmr (kg/s) | MN | NN | MR | 3.050E+03 | 15.000E+0 | 3.050E+03 | 14.901E+0 |
| tr1 (K) | MN | MN | MR | 300.000E+0 | 300.000E-3 | 299.989E+0 | 215.182E-3 |
| tr2 (K) | MN | MN | MR | 310.000E+0 | 300.000E-3 | 310.011E+0 | 215.182E-3 |
| tcon (K) | MN | MR | MR | 320.000E+0 | 500.000E-3 | 319.724E+0 | 442.900E-3 |
| tcal (K) | NC | MR | MR | 560.000E+0 | 1.000E+0 | 559.491E+0 | 694.342E-3 |
| tv3 (K) | MN | MR | MR | 775.000E+0 | 5.000E+0 | 778.379E+0 | 4.315E+0 |
| tv6 (K) | MN | MN | MR | 300.000E+0 | 300.000E-3 | 300.000E+0 | 300.000E-3 |
| pcon (bar) | NC | MR | MR | 100.000E-3 | 5.000E-3 | 104.194E-3 | 2.286E-3 |
| pcal (bar) | MN | MR | MR | 70.000E+0 | 1.000E+0 | 70.494E+0 | 719.644E-3 |
| pv3 (bar) | MN | MR | MR | 60.000E+0 | 1.000E+0 | 59.933E+0 | 998.743E-3 |
| hv2 (MJ/kg) | NC | NC | NC | 205.000E-3 | - | 204.538E-3 | 1.884E-3 |
| hv4 (MJ/kg) | NC | NN | NC | 2.450E+0 | - | 2.472E+0 | 10.162E-3 |
| hr3 (MJ/kg) | NC | NN | NC | 150.000E-3 | - | 154.549E-3 | 900.007E-6 |
| w1 (MW) | NC | NN | MR | 52.000E+0 | 500.000E-3 | 51.904E+0 | 100.823E-3 |
| w2 (MW) | MN | MR | MR | 600.000E-3 | 10.000E-3 | 599.948E-3 | 9.998E-3 |
| w3 (MW) | MN | NN | MR | 300.000E-3 | 10.000E-3 | 300.040E-3 | 9.998E-3 |
| w4 (MW) | MN | MR | MR | 1.000E+0 | 20.000E-3 | 999.791E-3 | 19.983E-3 |
| w5 (MW) | MN | MN | MR | 50.000E+0 | 100.000E-3 | 50.004E+0 | 98.046E-3 |
| wciclo (MW) | NC | NC | NC | 51.000E+0 | - | 51.004E+0 | 99.907E-3 |
| qcomb (MW) | NC | NC | NC | 182.500E+0 | - | 181.959E+0 | 656.412E-3 |
| qcald (MW) | NC | NC | NC | 180.000E+0 | - | 181.325E+0 | 657.937E-3 |
| qcond (MW) | NC | NN | NC | 127.500E+0 | - | 127.855E+0 | 685.967E-3 |
| rtom2 (adim.) | MN | MR | MR | 900.000E-3 | 30.000E-3 | 899.449E-3 | 29.997E-3 |
| rtom3 (adim.) | MN | MN | MN | 900.000E-3 | 30.000E-3 | 900.000E-3 | 30.000E-3 |
| rtom4 (adim.) | MN | MR | MR | 900.000E-3 | 30.000E-3 | 899.082E-3 | 29.992E-3 |
| rtoal (adim.) | MN | MN | MR | 960.000E-3 | 3.000E-3 | 959.995E-3 | 2.999E-3 |
| rplcomb (adim.) | NC | NC | NC | 950.000E-3 | - | 944.581E-3 | 1.108E-3 |
| rplcald (adim.) | NC | NC | NC | 995.000E-3 | - | 996.518E-3 | 174.360E-6 |
| rplciclo (adim.) | NC | NC | NC | 280.000E-3 | - | 281.283E-3 | 1.157E-3 |
| rpltotal (adim.) | NC | NC | NC | 260.000E-3 | - | 260.933E-3 | 1.068E-3 |

MN → Medida No Redundante, MR → Medida Redundante
 NC → No medida Calculable, NN → No medida No calculable

Tabla II. Datos y resultados del ejemplo

Como ya se ha comentado, disponer de medidas redundantes produce ventajas evidentes a la hora de diagnosticar la operación de una planta de proceso. Las más destacables son las siguientes:

1. *Capacidad de detectar errores de medida*

La calidad de los resultados y por tanto la fiabilidad del diagnóstico depende de la calidad de las medidas. La redundancia de medidas permite detectar errores de medida por comparación entre los valores medidos y reconciliados. En el caso presentado se superan tanto el test global

$$F = 3,55 < \chi_{1-\alpha}^2(H = 9) = 16,9 \quad (59)$$

como el test de medidas redundantes. Obsérvese, sin embargo, que al existir medidas no redundantes no puede ofrecerse una garantía completa.

2. *Certidumbre de los resultados obtenidos*

La presencia de medidas redundantes hace que la desviación típica de los resultados sea más reducida. Considérese el siguiente ejemplo. Si descartamos las nueve medidas presentes en el caso 3 y no presentes en el caso 1 y calculamos los resultados, ahora sin medidas redundantes, se obtendrían los valores indicados en la siguiente tabla para los rendimientos energéticos de las secciones y del sistema global analizado.

| | Combustión | Caldera | Ciclo | CENTRAL |
|--------------------------------|---------------------|---------------------|---------------------|---------------------|
| Convencional ($I = 22$) | $0,9448 \pm 0,0011$ | $0,9965 \pm 0,0002$ | $0,2817 \pm 0,0085$ | $0,2614 \pm 0,0079$ |
| Reconciliación ($I = 31$) | $0,9446 \pm 0,0011$ | $0,9965 \pm 0,0002$ | $0,2813 \pm 0,0012$ | $0,2609 \pm 0,0011$ |

Como puede comprobarse, los rendimientos calculados no varían apreciablemente, pero las desviaciones típicas de los rendimientos del ciclo y de la central son significativamente mayores cuando no hay redundancia de medidas.

3. *Seguridad del método de diagnóstico*

La indisponibilidad de alguna de las medidas redundantes, por razones de mantenimiento o fallo del instrumental utilizado, no afecta a la cantidad de resultados obtenidos, aunque sí a su calidad, según hemos visto en el punto anterior. Esto es muy importante en el caso del diagnóstico en tiempo real de las plantas de proceso sobre todo si sus condiciones de operación en el futuro se programan en función del comportamiento observado.

CONCLUSIONES

En este trabajo se ha mostrado la aplicación de la descomposición QR a la clasificación de variables y reconciliación de datos en ingeniería de procesos. Una vez se dispone de un modelo matemático para la planta analizada y de un conjunto de medidas, puede aplicarse la metodología aquí propuesta sacando el máximo provecho de la información disponible. Como resultado: (i) se detectan y eliminan los errores sistemáticos de medida; (ii) se obtiene un conjunto consistente de medidas ajustadas; (iii) se estiman las variables no medidas calculables por el método de máxima verosimilitud y (iv) se obtienen intervalos de confianza para los resultados. Todas estas tareas, realizadas de modo independiente en los métodos

clásicos de diagnóstico de la operación, quedan ahora relacionadas, lo que aumenta la consistencia de los resultados obtenidos. Otra ventaja importante es que las ecuaciones y los datos quedan totalmente separados. Por tanto, si falta la medida de una variable por fallos de la instrumentación o se dispone de medidas nuevas, no habrá que replantear las ecuaciones del modelo matemático ni la forma de resolverlas. En definitiva, se facilita al ingeniero de procesos obtener la mayor cantidad de información sobre la operación de su planta y conocer su calidad.

El ejemplo mostrado ilustra la aplicación de la metodología propuesta de cuya consolidación da muestra la publicación reciente de libros sobre el tema.³⁻⁵ Aquí nos hemos centrado en sus aspectos matemáticos. Para ver con mayor detalle sus ventajas sobre la calidad del diagnóstico véase su aplicación a calderas de vapor.¹⁰

REFERENCIAS

- 1 C. Crowe, "Data reconciliation—progress and challenges", *J. Proc. Cont.*, Vol. **6**, N° 2/3, pp. 89–98, (1996).
- 2 R.S. Mah, "*Chemical process structures and information flows*", Butterworth, (1990).
- 3 V. Veverka y F. Madron, "*Material and energy balancing in the process industries*", Elsevier, (1997).
- 4 S. Narasimham y C. Jordache, "*Data reconciliation and gross error detection*", Gulf, (2000).
- 5 J. Romagnoli y M. Sánchez, "*Data processing and reconciliation for chemical process operations*", Academic Press, (2000).
- 6 M. A. Lozano y J.A. Remiro, "Performance evaluation of energy systems with data reconciliation", *Información Tecnológica*, Vol. **12**, N° 2, pp. 99–104, (2001).
- 7 M. Sánchez y J. Romagnoli, "Use of orthogonal transformation in data classification–reconciliation", *Computers Chem. Engng.*, Vol. **20**, N° 5, pp. 483–493, (1996).
- 8 Visual Numerics, "*IMSL Fortran subroutines for mathematical applications*, Capítulo 1: "Linear Systems", Math. Library, (1997).
- 9 G. Quintana y E.S. Quintana, "Descomposiciones ortogonales para el cálculo del rango numérico matricial", *Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería*, Vol. **15**, N° 4, pp. 461–470, (1999).
- 10 M. A. Lozano y J.A. Remiro, "Diagnóstico de calderas de vapor con reconciliación de datos", *Ingeniería Química*, N° 372, pp. 117–123, (2000).