

NUEVA ESTRATEGIA PARA IDENTIFICACION DE ERRORES SISTEMATICOS EN LAS MEDICIONES

J. J. Elwart, R. Galdeano, C.R. Alvarez, A. Brandolin y M.C. Sánchez^a

^a*Planta Piloto de Ingeniería Química, PLAPIQUI (UNS-CONICET)
Universidad Nacional del Sur
Camino La Carrindanga Km 7, Bahía Blanca, Buenos Aires
msanchez@plapiqui.edu.ar*

1 - Resumen

En este trabajo se presenta una estrategia destinada a detectar e identificar errores sistemáticos de magnitud constante (biases) en las mediciones de un proceso que opera en estado estacionario. La técnica utiliza el estadístico de Hotelling para detectar la presencia de biases. El método del vecino más próximo en condiciones de control estadístico se emplea para identificar los sensores defectuosos.

El procedimiento se basa exclusivamente en el uso de datos históricos, por lo tanto no se requiere el modelado del proceso, ni tampoco conocer la función de densidad de probabilidad de los errores de las mediciones.

En contraposición con los métodos basados en principios fundamentales, es posible identificar errores sistemáticos en todas las mediciones, sin la necesidad de distinguir entre las redundantes y no redundantes.

El desempeño de la estrategia propuesta se evalúa utilizando una red de intercambio de calor y los resultados indican que la técnica resulta exitosa en la identificación de biases simples y múltiples.

2 - Introducción

Dado que las mediciones presentan errores aleatorios y sistemáticos, se han realizado contribuciones importantes durante las últimas cuatro décadas a fin de reducir su efecto perjudicial sobre las estimaciones de las variables de un proceso. En la actualidad es una práctica común en las industrias de proceso reconciliar las mediciones, de manera tal que éstas resulten consistentes con los balances de masa y energía. A fin de obtener estimaciones más exactas, es necesario tomar acciones conducentes a reducir el efecto de los errores sistemáticos, tales como los biases.

Se han propuesto diferentes estrategias basadas en modelos para detección e identificación de biases (Narasimhan y Jordache (2000), Romagnoli y Sánchez (2000)). Estas metodologías comparan la operación actual de la planta con la predicha por el modelo matemático que la representa mediante la aplicación de tests de hipótesis estadísticos. Entre éstos se destacan el Test Global, Test de las Mediciones, Test de los Nodos, Test de Razón de Verosimilitudes Generalizado, Test de Bonferroni y Test de las Componentes Principales, entre otros. Algunos de estos test han sido incorporados en procedimientos desarrollados para identificar la presencia de múltiples errores sistemáticos, tales como las estrategias de Eliminación en Serie, Compensación en Serie y Compensación Simultánea.

Además se ha establecido la Teoría de Equivalencia de Errores Sistemáticos (Bagajewicz y Jiang, 1998). Esta postula que dos conjuntos de errores sistemáticos son equivalentes cuando producen el mismo efecto sobre el procedimiento de reconciliación de datos. Por lo

tanto, cuando se identifica un conjunto de biases, existe la misma posibilidad de que la localización verdadera de éstos se encuentre en uno de sus conjuntos equivalentes, lo que introduce incertidumbre en la tarea de identificación.

El estadístico de Hotelling (T^2) es muy utilizado en Control Estadístico de Procesos para detectar si el proceso se encuentra fuera de control, pero no ofrece asistencia como herramienta de identificación. Para tal fin se han propuesto diferentes técnicas que calculan la contribución de cada variable al valor del estadístico cuando éste supera el correspondiente valor crítico. Algunas estrategias operan en el espacio definido por las variables originales y otras emplean variables latentes.

En relación con los métodos que calculan las contribuciones al estadístico T^2 empleando las variables reales, Mason y col. (1995,1997) presentaron una estrategia de tipo combinatorial, que proporciona un número importante de descomposiciones diferentes lo cual incrementa la complejidad del procedimiento de identificación. Por el contrario, Alvarez y col. (2007) descompusieron el estadístico en una suma única de las contribuciones de cada variable al valor de T^2 , denominada OSS (Original-Space Strategy). Esta provee una clara comprensión de las contribuciones positivas y negativas, y estima un límite para las negativas. En base a la descomposición previamente mencionada, Sánchez y col. (2008) propusieron una estrategia de identificación de biases en los sensores para procesos en estado estacionario que utiliza solamente la información disponible en la base de datos históricos del proceso. Recientemente Alvarez y col. (2008) presentaron una estrategia de identificación de fallas, que utiliza la ubicación óptima del vecino más cercano a un vector de mediciones fuera de control para determinar las posibles causas de la falla.

3 - Objetivos

En este trabajo se presenta una estrategia para detección e identificación de biases en los sensores de un proceso que opera en condiciones de estado estacionario. El procedimiento utiliza la técnica de identificación de fallas propuesta por Alvarez y col. (2008). Se analiza el desempeño del método propuesto para la identificación de errores sistemáticos en una red de intercambio de calor.

4 - Metodología

La estrategia consta de dos etapas. La primera se ejecuta fuera de línea y proporciona los parámetros de la población de referencia que define la operación normal del proceso y el valor crítico del estadístico.

Durante la segunda etapa, se calcula el estadístico T^2 en línea en cada instante de muestreo. Si éste excede el valor crítico, se detecta la presencia de uno o varios biases. Luego, se calcula la ubicación del vecino más próximo a la medición que se encuentra en control (NICN: nearest in control neighbour). Se resuelve un problema de optimización para encontrar el NICN y se evalúan las diferencias en las posiciones del vector de observaciones y el NICN para cada variable. Esta información indica cuan alejada del estado normal se encuentra la observación y cuál es la mínima distancia y dirección que podría ser explicada por la ocurrencia de la falla. Aquellas variables que se encuentran más alejadas de la condición normal se identifican como sospechosas.

Cada etapa se compone de los siguientes pasos:

a) Etapa fuera de línea

1. Se construye la población de referencia formada por I vectores de observaciones. Cada muestra se obtiene como un promedio de las mediciones correspondientes a una ventana móvil de longitud H .

2. La población de referencia se estandariza y se calcula la matriz de correlación \mathbf{R} .

3. Se evalúa el estadístico T^2 para cada miembro de la población de referencia. Luego se estima la función de densidad de probabilidad de T^2 usando el método kernel de estimación. Dado el nivel de significado del test, α , se selecciona el cuartilo $(1-\alpha)$ de la distribución como el valor crítico de T^2 (T^2_{crit}).

b) Etapa en línea

1. Detección

En cada etapa de muestreo se obtiene una nueva observación \mathbf{z}_{new} como el promedio de las mediciones incluidas en una ventana móvil de longitud H . El estadístico T^2 se calcula, luego de estandarizar \mathbf{z}_{new} , usando la Ec. 1. Si éste resulta mayor que T^2_{crit} , se ha detectado la presencia de biases.

$$T^2 = \mathbf{z}_{new}^T \mathbf{R}^{-1} \mathbf{z}_{new} \quad (1)$$

2. Identificación

Se calculan las coordenadas en el espacio de las variables originales del NICN. Este es el vecino más próximo a \mathbf{z}_{new} que se encuentra en control. Para encontrar el NICN se minimiza la distancia de Mahalanobis entre \mathbf{z}_{new} y NICN sujeto a que el estadístico T^2 para el NICN sea menor o igual a T^2_{crit} . Se resuelve el siguiente problema de optimización:

$$\begin{aligned} \text{Min} \quad & (\mathbf{z}_{new} - \mathbf{x}_{NICN})^T \mathbf{R}^{-1} (\mathbf{z}_{new} - \mathbf{x}_{NICN}) \\ \text{s.t.} \quad & \mathbf{x}_{NICN}^T \mathbf{R}^{-1} \mathbf{x}_{NICN} \leq T^2_{crit} \end{aligned} \quad (2)$$

Luego se calculan las diferencias absolutas entre las componentes de los vectores \mathbf{z}_{new} y NICN, y se evalúa la media de las diferencias. Se consideran como variables sospechosas aquellas para las cuales la diferencia absoluta resulta mayor al límite τ , que se define a continuación

$$\tau = K * des \quad (3)$$

En esta ecuación, des es la media de las diferencias absolutas entre \mathbf{z}_{new} y NICN. Dado que el Número Promedio de Errores Tipo I (AVTI) se fija en 0.05 cuando no hay presencia de biases en las mediciones, el parámetro K se evalúa por simulación a fin de satisfacer esta condición. En consecuencia, el número de falsas alarmas está acotado para cada valor del parámetro I .

5 – Resultados

Mediante un procedimiento de simulación se evaluó el desempeño de la estrategia propuesta para identificar biases en las mediciones correspondientes a la red de

intercambio de calor (Swartz, 1989) representa en la Fig. 1. Esta se utiliza para calentar la corriente A mediante las corrientes de proceso B, C y D.

El sistema presenta 16 variables medidas y 14 no medidas, las cuales están relacionadas por 17 restricciones de igualdad (balances de masa y energía alrededor de los intercambiadores de calor, mezcladores y divisores). El desvío estándar de los caudales másicos y temperaturas se asumen iguales a 2% del valor medido y 0.75°C, respectivamente. Una descripción completa del ejemplo está disponible en la literatura (Romagnoli y Sánchez, 2000).

La aplicación de procedimientos de clasificación de variables al conjunto de mediciones de la red de intercambio de calor revela la existencia de variables medidas redundantes y no redundantes. Debe notarse que la identificación de errores sistemáticos no puede realizarse para las observaciones no-redundantes utilizando estrategias basadas en modelos, ya que éstas no participan en el conjunto de ecuaciones de reconciliación.

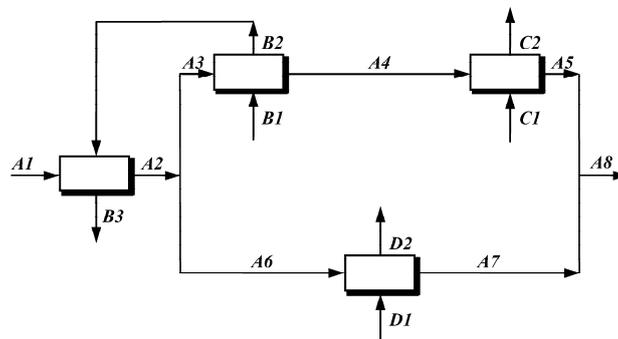


FIGURA 1: Red de intercambio de calor

La evaluación del desempeño del método consistió en examinar 100 configuraciones de biases. Para cada configuración particular, se analizaron 100 ventanas móviles de datos formadas agregando ruido y biases a los valores verdaderos de las variables. Se generaron horizontes de datos de longitud fija $H=5$, los cuales se ordenaron en una matriz. En cada instante de muestreo, el primer vector de observaciones se elimina de la matriz y se incorpora el vector de la última medición en la última fila.

En relación con los errores aleatorios, se utilizaron para generarlos tres distribuciones de probabilidad distintas (Normal, Uniforme y Laplace). El método Kernel se empleó para aproximar la función de densidad de probabilidad del estadístico T^2 y el valor T^2_{crit} se calculó asumiendo un nivel de significado $\alpha = 0.05$.

La ubicación y la magnitud de los biases de cada configuración se seleccionó en forma aleatoria. Se simuló la presencia de 1 y 2 errores sistemáticos. La magnitud de los biases simulados se encuentra en el rango $[4\sigma - 7\sigma]$, siendo σ el desvío estándar de los errores de las mediciones. En otros trabajos presentes en la literatura se empleó el rango $[5\sigma - 25\sigma]$ que resulta demasiado amplio.

Se utilizaron tres medidas de desempeño ampliamente difundidas en la literatura. Estas son: el Número Promedio de Errores Tipo I (AVTI), la Potencia Global (OP) y la Fracción Esperada de Identificación Perfecta (OPF), que se definen a continuación:

$$AVTI = \frac{\text{No. of gross errors incorrectly identified}}{\text{No. of simulation trials}} \quad (4)$$

$$OP = \frac{\text{No. of gross errors correctly identified}}{\text{No. of gross errors simulated}} \quad (5)$$

$$OPF = \frac{\text{No. of trials with perfect identification}}{\text{No. of simulation trials}} \quad (6)$$

La Tabla 1 presenta las medidas de desempeño de la estrategia propuesta para la identificación de biases, tanto en mediciones redundantes como no-redundantes, de la red de intercambio de calor.

TABLA 1: Medidas de desempeño

Distribución	#B	I	K	AVTI	OP	OPF
Normal	1	90	3.64	0	1	1
	1	120	3.53	0	1	1
	1	180	3.28	0.0001	1	0.9999
	1	300	3.11	0	1	1
	2	90	3.64	0	0.9843	0.9687
	2	120	3.53	0	0.9891	0.9783
	2	180	3.28	0	0.9989	0.9978
	2	300	3.11	0	0.9996	0.9991
Uniforme	1	90	3.41	0	1	1
	1	120	3.29	0	1	1
	1	180	3.07	0	1	1
	1	300	2.88	0.0001	1	0.9999
	2	90	3.41	0	0.9950	0.9899
	2	120	3.29	0	0.9960	0.9921
	2	180	3.07	0	0.9997	0.9994
	2	300	2.88	0	1	0.9999
Laplace	1	90	4.03	0.0004	0.9890	0.9886
	1	120	3.82	0.0009	0.9968	0.9959
	1	180	3.60	0.0006	0.9975	0.9969
	1	300	3.48	0.0023	0.9984	0.9961
	2	90	4.03	0.0001	0.8120	0.6421
	2	120	3.82	0	0.8709	0.7453
	2	180	3.60	0	0.9378	0.8772
	2	300	3.48	0.0002	0.9535	0.9073

La segunda columna de la tabla muestra el número de biases simulados. Las siguientes dos columnas indican el número de muestras de la población de referencia (I) y el parámetro K utilizado para calcular el valor límite τ , respectivamente. Posteriormente se presentan las medidas de desempeño de la identificación AVTI, OP y OPF.

Si las técnicas de identificación de biases se ejecutan habitualmente en línea, cuando se detecta la presencia de errores sistemáticos, en general, existe un solo error. La Tabla 1 muestra valores de OP mayores a 0.99 cuando se simula la presencia de un solo bias para todas las distribuciones de probabilidad de errores aleatorios. Esto evidencia que la estrategia identifica exitosamente el bias en el instrumento defectuoso, lo cual es de gran importancia en la práctica. Si se aumenta el número de errores sistemáticos simulados, la OP disminuye para el mismo valor de I . Esto es particularmente notorio para la distribución de Laplace.

El parámetro I afecta la medida OPF, ésta es la capacidad de la estrategia de identificar perfectamente el conjunto de sensores defectuosos; si I aumenta OPF también aumenta. Por lo tanto, la actualización de la población de referencia en línea, mediante la incorporación de observaciones de estados en control, mejora el desempeño de la identificación, y en consecuencia reduce el número de instrumentos sospechosos que deben ser inspeccionados por el sector de mantenimiento.

En relación con el valor del parámetro K , éste se ajusta a fin de mantener un AVTI igual a 0.05 cuando se satisface la hipótesis nula. Su valor disminuye para tamaños de muestras mayores, dado que se dispone de mejores estimaciones de los parámetros poblacionales en estos casos.

6 - Conclusiones

En este trabajo se presenta una estrategia para identificación de biases en las mediciones de un proceso que opera en estado estacionario. Se propone utilizar el estadístico T^2 para detectar la presencia de biases en el conjunto de mediciones, e identificar las observaciones defectuosas examinando la menor distancia existente entre la medición y su vecino más cercano que se encuentra en control.

La estrategia se basa en el uso de datos históricos, evita la tarea de modelado del proceso no lineal y los errores de linealización ocasionados cuando estrategias desarrolladas para sistemas lineales se emplean en identificación de errores sistemáticos en sistemas no lineales.

Dado que la estrategia no requiere conocer el modelo del proceso, la técnica permite identificar biases en observaciones redundantes y no-redundantes, y evita la incertidumbre de localización explicada por la Teoría de Equivalencia de Errores Sistemáticos. Además este método es aplicable cualquiera sea la distribución de probabilidad de los errores aleatorios de las mediciones.

La metodología propuesta muestra muy buen desempeño en la identificación de biases para los procesos analizados.

Los futuros trabajos se centrarán en el desarrollo de fórmulas alternativas para calcular los límites de identificación de variables sospechosas.

7 – Bibliografía

- Alvarez, R.; Brandolin, A.; Sánchez, M. (2007) On the Variable Contributions to the D-statistic. Chemometr. and Intell. Lab. System., 88, 89-196.
- Alvarez, R.; Brandolin, A.; Sánchez, M.; Puigjaner, L. (2008) A Nearest In Control Neighbour Based Method to Estimate Variable Contributions to the Hotelling's Statistic. Proceedings of 2008 AIChE Annual Meeting, Philadelphia, USA, November 16–21.

- Bagajewicz, M.; Jiang, Q. (1998) Gross error modeling and detection in plant linear dynamic reconciliation. *Comp. & Chem. Engng.*, 22, 1789-1810.
- Mason, R. L.; Tracy, N. D.; Young, J. C. (1995) Decomposition of T^2 for multivariate control charts interpretation. *J. Qual. Technol.*, 27, 99-108.
- Mason, R. L.; Tracy, N. D.; Young, J. C. (1997) A practical approach for interpreting multivariate T^2 control chart signals. *J. Qual. Technol.*, 29, 396-406.
- Narasimhan, S.; Jordache, C. (2000) Data reconciliation and gross error detection. Houston: Gulf Publishing Company.
- Romagnoli, J.; Sánchez, M. (2000) Data processing and reconciliation for chemical process operations. San Diego: Academic Press.
- Sánchez, M.; Alvarez, R.; Brandolin, A. (2008) A MSPC procedure for bias identification in steady state processes. *AIChE Journal*, 54, 8, 2082-2088.
- Swartz, C. L. E. (1989) Data reconciliation for generalized flowsheet applications. Presented at the National Meeting of the American Chemical Society. Dallas.