

**IDENTIFICACIÓN DE LA CAUSA DE ERRORES GRUESOS: POR FALLA EN  
EL SISTEMA DE MEDICIÓN O POR FUGA, EN UNA RED DE TRANSPORTE  
DE FLUIDOS**

**ANA MARIA GONZÁLEZ MANTILLA  
ROYER ARLEY MATEUS GALINDO**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISCOQUÍMICAS  
ESCUELA DE INGENIERÍA QUÍMICA  
BUCARAMANGA  
2015**

**IDENTIFICACIÓN DE LA CAUSA DE ERRORES GRUESOS: POR FALLA EN  
EL SISTEMA DE MEDICIÓN O POR FUGA, EN UNA RED DE TRANSPORTE  
DE FLUIDOS**

**ANA MARIA GONZÁLEZ MANTILLA  
ROYER ARLEY MATEUS GALINDO**

**Trabajo de grado para optar el título de  
Ingeniero químico**

**Director  
CARLOS EDUARDO GARCÍA SÁNCHEZ  
Ingeniero Químico, Ph.D**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISCOQUÍMICAS  
ESCUELA DE INGENIERÍA QUÍMICA  
BUCARAMANGA  
2015**

*A Dios por sus infinitas bendiciones, por permitirme seguir avanzando cuando me faltaban las fuerzas, porque me mostró los caminos que debía tomar a lo largo de esta carrera que ya culmina.*

*A mí amado padre Luis Felipe, porque sus enseñanzas quedaron grabadas en mi corazón. Aunque ya no se encuentre entre nosotros, todavía puedo sentir su gran amor y puedo escuchar en mi mente sus palabras de aliento y cariño, Porque me enseñó a ser una persona correcta y avanzar sin importar las circunstancias, a no permitir la derrota en mi vida. Fuiste el mejor ejemplo que pude tener a mi lado.  
A ti Phi...*

*A mis hermanos y mi mamá por su apoyo incondicional.*

*A mis compañeros de carrera y amigos, porque fueron una parte fundamental en este sueño, porque su apoyo y comprensión marcaron la diferencia.*

*Ana María González Mantilla*

*A Dios primeramente, Por haberme permitido llegar hasta este punto y darme la oportunidad de alcanzar mis objetivos y culminar esta gran etapa de mi vida.*

*A Mi padre Alirio Mateus, Por los ejemplos de perseverancia y constancia que lo caracterizan y que me ha infundado siempre, por el valor mostrado para salir adelante Por guiarme por el buen camino con su ejemplo junto a sus consejos y sacrificios.*

*A Mi madre Olibe Galindo, Quien siempre me motivo a seguir adelante con sus palabras y esfuerzo, y a pesar de las dificultades que se presentaron siempre estuvo ahí apoyándome.*

*A Mi abuela Ana María Noguera, quien me acogió en su hogar durante la mayor parte de esta travesía, convirtiéndose en una segunda madre para mí.*

*A mis hermanos que con su ejemplo y compañía hicieron del trayecto para culminar esta meta algo más ameno.*

*A mis profesores, amigos y demás familiares que hicieron parte de esta gran aventura convirtiéndose en ejemplo de aprendizaje en este camino que deja alegres recuerdos y anécdotas plasmados en mi memoria.*

***Royer Arley Mateus Galindo***

## **Agradecimientos**

A la UNIVERSIDAD INDUSTRIAL DE SANTANDER, por habernos dado la oportunidad de ser parte de esta gran familia y a los diferentes profesores que hicieron parte de esta meta que ya culmina, porque más que conocimientos nos enseñaron a perseverar y a mirar las cosas desde un punto de vista diferente, a ver soluciones donde los demás ven problemas.

A la Corporación CDT de Gas por brindarnos la oportunidad de realizar este proyecto en sus instalaciones, con la asesoría y acompañamiento de sus diferentes ingenieros.

Al Doctor en Ingeniería Química CARLOS EDUARDO GARCIA SANCHEZ por sus incontables recomendaciones, sugerencias, correcciones y asesorías a lo largo de este trabajo.

Al Ingeniero Mecánico JOSE AUGUSTO FUENTES OSORIO por direccionar este proyecto.

Al semillero del Grupo de Investigación en Fluidos y Energía por sus aportes y contribuciones en cada avance presentado.

A nuestros padres, familiares y amigos por su gran apoyo y compañía en este proceso.

## CONTENIDO

	Pág.
INTRODUCCION .....	15
1. DESCRIPCION METODOLOGICA .....	19
1.1. DESCRIPCION GLOBAL DE LA METODOLOGIA.....	19
1.2. METODOLOGIA PARA ESCOGER LA FUNCION OBJETIVO CON MAYOR RENDIMIENTO CON LA DEG MEDIANTE LA RD.....	20
1.2.1. Reconciliación de Datos (RD) .....	20
1.2.2. Detección de Errores Gruesos (DEG).....	21
1.3. CRITERIOS DE EVALUACION PARA FUNCIONES OBJETIVO .....	24
1.4. IDENTIFICACION DE LA POSIBLE CAUSA DEL ERROR GRUESO.....	26
2. ANALISIS DE RESULTADOS .....	26
2.1. VALIDACIÓN DE LA METODOLOGÍA PARA LA ELECCIÓN DE FUNCIÓN OBJETIVO.....	26

2.1.1. Ejemplo 1 .....	26
2.1.2. Ejemplo 2 .....	28
2.1.3. Ejemplo 3 .....	29
2.1.4. Ejemplo 4 .....	31
2.2. IDENTIFICACIÓN DE LA CAUSA DEL ERROR GRUESO, MEDIANTE LA METODOLOGÍA GLOBAL PLANTEADA.....	33
3. CONCLUSIONES .....	39
4. RECOMENDACIONES.....	40
BIBLIOGRAFIA .....	44
ANEXOS.....	47

## LISTA DE TABLAS

Pág.

<b>Tabla 1.</b> Funciones objetivo para la RD.....	21
<b>Tabla 2.</b> Criterios de evaluación. ....	24
<b>Tabla 3.</b> Resultados del algoritmo desarrollado con el Ejemplo 1. ....	28
<b>Tabla 4.</b> Resultados del algoritmo desarrollado con el Ejemplo 2. ....	29
<b>Tabla 5.</b> Resultados del algoritmo desarrollado con el Ejemplo 3. ....	31
<b>Tabla 6.</b> Resultados del algoritmo desarrollado con el Ejemplo 4. ....	31
<b>Tabla 7.</b> Valores de Caudal y Presión. ....	34
<b>Tabla 8.</b> Distancia de tramos de tubería en la red. ....	34
<b>Tabla 9.</b> Resultados de la DEG en la red de agua. ....	35
<b>Tabla 10.</b> Resultados del algoritmo para determinar la causa del error grueso. ...	37
<b>Tabla 11.</b> Valores verdaderos de las corrientes y los balances del Ejemplo 1. ....	52
<b>Tabla 12.</b> Valores verdaderos del proceso en el Ejemplo 2. ....	53
<b>Tabla 13.</b> Valores verdaderos del proceso en el Ejemplo 3. ....	54
<b>Tabla 14.</b> Valores verdaderos del proceso en el Ejemplo 4. ....	56

## LISTA DE FIGURAS

	<b>Pág.</b>
<b>Figura 1.</b> Descripción global de la metodología. ....	19
<b>Figura 2.</b> Procedimiento para hallar la posible causa del error grueso. ....	25
<b>Figura 3.</b> Diagrama de la red de medición del Ejemplo 1.....	27
<b>Figura 4.</b> Diagrama de la red de medición del Ejemplo 2.....	29
<b>Figura 5.</b> Diagrama de la red de medición del Ejemplo 3.....	30
<b>Figura 6.</b> Diagrama de la red de medición del Ejemplo 4. ....	31
<b>Figura 7.</b> Diagrama red de medición de agua. ....	33
<b>Figura 8.</b> Histograma de frecuencia de los rangos del error agregado, de los 72 casos de errores gruesos simulados. ....	37

## LISTA DE ANEXOS

<b>Anexo A.</b> Funciones objetivo en la RD.....	47
<b>Anexo B.</b> Pruebas de hipótesis empleadas para DEG.....	48
<b>Anexo C.</b> Información necesaria para realizar la RD, de los sistemas propuestos. .....	52

## RESUMEN

**TITULO:** IDENTIFICACIÓN DE LA CAUSA DE ERRORES GRUESOS: POR FALLA EN EL SISTEMA DE MEDICIÓN O POR FUGA, EN UNA RED DE TRANSPORTE DE FLUIDOS\*

**AUTORES:** ANA MARIA GONZÁLEZ MANTILLA, ROYER ARLEY MATEUS GALINDO\*\*

**PALABRAS CLAVES:** Detección de errores gruesos, reconciliación de datos, funciones objetivo, causa de los errores gruesos.

### DESCRIPCION

Es imperativo identificar la existencia de errores en cualquier tipo de proceso, incluyendo el transporte y distribución de fluidos. En este trabajo, se presentan dos métodos para la Detección de Errores Gruesos (DEG) en operaciones en estado estacionario, que emplean los valores que resultan de la Reconciliación de Datos (RD); una planteada por los autores y otra encontrada en la literatura, donde se evalúan diferentes funciones objetivo (Mínimos Cuadrados, Cauchy, Fair, Lorentzian y Normal Contaminada). Adicionalmente, se propone una metodología para determinar la posible causa del error grueso, clasificándolo como fuga o como falla en el sistema de medición, que son las causas más comunes de los errores gruesos en un sistema o red de transporte de fluidos, a través de análisis de rangos de caudal y presión.

Los resultados finales muestran que la metodología con la red presentada tiene una selectividad de DEG del 100 % con la función objetivo Fair con el método A (propuesto por los autores) de los errores gruesos simulados, y un porcentaje de identificación global de la causa del error grueso de 58% y una selectividad de detección del 82% que tiene en cuenta los errores gruesos correctamente identificados y su causa, lo que lleva a concluir que es posible implementar esta metodología en un sistema de medición en tiempo real, debido a los beneficios económicos que trae el saber que medidores o secciones del proceso tienen inconsistencias.

---

\* Trabajo de grado

\*\* Facultad de Ingenierías Físicoquímicas. Escuela de Ingeniería Química. Director: Carlos Eduardo García Sánchez, Ingeniero Químico, Doctor en Ingeniería Química.

## ABSTRACT

**TITLE:** IDENTIFICATION OF THE CAUSE OF GROSS ERRORS: BY FAIL IN MEASUREMENT SYSTEM OR LEAK, IN A FLUIDS TRANSPORT NETWORK\*

**AUTHORS:** ANA MARIA GONZÁLEZ MANTILLA, ROYER ARLEY MATEUS GALINDO\*\*

**KEYWORDS:** Gross error detection, Data reconciliation, Objective functions, Cause of gross errors.

### DESCRIPTION

It is imperative to identify the existence of errors in any type of process, including the transportation and distribution of fluids. In this project, two methods are presented for the Detection of Gross Errors (DGE) in steady state operations, which use the resulting values from the Reconciliation Data (RD); one of them is proposed by the authors and the other one is found in the literature; these methods evaluate different objective functions (Least Squares, Cauchy, Fair, Lorentzian, and Normal Contaminated). Additionally, a methodology is presented to determine the possible causes of the gross errors, classifying it as leak or failure in the measurement system, which are the most common causes of gross error in a system or fluids transport network, through analysis of flow and pressure ranges.

The final results showed that the methodology with the presented network has a DGE selectivity of 100% with the objective function Fair with the A method (proposed by the authors), identifying gross errors simulated and a global identification percentage of the gross error's cause of 58% and a detection selectivity of 82% which takes into account gross errors correctly identified and their cause, leads to conclude that this methodology can be implemented in a measurement system in real time, because of the economic benefits it brings to know that gauges or process sections are inconsistencies.

---

\* Degree Work

\*\* Faculty of Engineering physic-chemical. School of Chemical Engineering. Director: Carlos Eduardo Garcia Sanchez. Chemical Engineer, Doctor in Chemical Engineering.

## INTRODUCCION

En los procesos productivos es de suma importancia realizar diferentes balances: de masa, de energía, de composición, entre otros, para garantizar seguridad, calidad, control y buen funcionamiento de las operaciones. Para que los datos reales sean coherentes con los modelos matemáticos de los balances, es imperativo contar con mediciones confiables. Sin embargo, los sistemas de medición no siempre cuentan con la misma precisión, las medidas presentan errores aleatorios (incertidumbre), e incluso pueden presentarse errores sistemáticos en las mediciones; estos últimos se conocen como errores gruesos, generando que los balances no cierren.

La incertidumbre es el nivel de duda sobre el valor de una medición, y es ocasionado por los errores aleatorios y la falta de control que se presenta sobre diferentes magnitudes de influencia [1]. La magnitud o señal de la incertidumbre no se puede predecir: si la medida se repite con el mismo instrumento bajo condiciones idénticas, los valores obtenidos serán diferentes. Los errores aleatorios se presentan con alta frecuencia y su magnitud usualmente es pequeña; se pueden presentar por diferentes causas como fluctuaciones en el suministro de energía, ruido en la transmisión de la red y conversión de señal, cambios en las condiciones ambientales, entre otras causas. Por otra parte, los errores gruesos son causados por eventos no aleatorios, como mal funcionamiento del instrumento de medida, mala calibración, desgaste o corrosión del sensor, depósitos sólidos, sensores sucios, fugas y demás. Si la medida es repetida con el mismo instrumento bajo condiciones idénticas, la contribución sistemática del error grueso será la misma. Los errores gruesos se presentan con menos frecuencia pero su magnitud típicamente es mayor que los errores aleatorios. En presencia de estos errores en la Reconciliación de Datos (RD) puede conducir a grandes ajustes que

se realizan en los valores medidos y los resultados estimados son inexactos e incluso incoherentes [2].

Actualmente existen programas informáticos que le permiten a las industrias mejorar la confiabilidad de los valores medidos mediante los balances del proceso. El uso de programas contribuye a disminuir las inversiones necesarias para mejorar los balances, debido a que la adquisición de equipos de última tecnología que proporcionen incertidumbres muy bajas representa inversiones altas. Otra ventaja importante de los programas es que hacen posible identificar qué secciones del proceso presentan medidas que no son confiables, y es que es fundamental determinar si las diferencias en las ecuaciones de balances son debidas a errores aleatorios o errores gruesos, para lo cual se emplean técnicas como la RD y la Detección de Errores Gruesos (DEG) simultáneamente.

La RD es una técnica de pre-procesamiento de datos con el objetivo de reducir el efecto de errores sistemáticos, mediante la explotación de las relaciones que existen entre las variables de un proceso, como los balances de masa y energía [3]. La RD es sensible a la redundancia de un sistema; una variable es redundante si además de ser medida puede ser estimada, al menos de una forma, por medio de las otras variables medidas [4]. Los mejores resultados aplicando la RD se obtienen cuando el sistema tiene alta redundancia; por otra parte, el rendimiento de la reconciliación baja cuando hay presencia de errores gruesos.

La DEG tiene el objetivo de identificar los valores atípicos o valores que se desvían en gran manera de su media. El objetivo de la detección de errores no es evitar que ocurran errores, sino evitar que haya errores sin detectar [5]. La presencia de errores gruesos en la RD dispersa el error afectando a las variables libres de este tipo de error, alejándolas de su valor convencionalmente verdadero (valores con los cuales los balances cierran); a lo anterior se le conoce como

'*smearing effect*'. El grado en que afecta este fenómeno depende de la redundancia del sistema, la ubicación del error, las desviaciones estándar de las medidas y del tamaño del error [2].

Diferentes autores han trabajado las técnicas de RD y DEG simultáneamente, [6-11] se observa la necesidad de emplear las dos metodologías, pero han sido pocos los que han desarrollado su metodología en la identificación de la causa del error grueso. Al detectar un error grueso en una red de mediciones es necesario conocer la causa de este, para tomar los correctivos pertinentes. Las dos causas generales por las que se presenta un error grueso son: por fuga o por falla en el sistema de medición.

Bagajewicz y Jiang [4] presentan un método que se basa en la identificación en serie con compensación colectiva en cada etapa de identificación, permitiendo conocer la causa de los errores gruesos que se detectan; Sánchez *et al.* [12] se basaron en el Test Global para determinar si el sistema tiene o no errores gruesos, y mediante la reconciliación de datos en estado estacionario es posible tener un estimado del tamaño de la fuga y las variables del proceso afectadas; Narasimhan y Mah [13] por medio del método de razón de verosimilitud generalizada (*the Generalized Likelihood Test* de hipótesis *Ratio*, GLR) a través de la prueba estadística de razón de verosimilitud proporcionan un marco general para identificar cualquier tipo de error grueso (por sesgos en los instrumentos y controladores, fugas y la insuficiencia de los controladores de la medición); Miao *et al.* [14] mediante una metodología que se fundamenta en el apoyo del vector regresión (*Support Vector Regression*, SVR) realizan la reconciliación simultánea para la estimación de desviaciones o fugas constantes. Los artículos que se mencionan anteriormente fundamentan sus metodologías en o pruebas

estadísticas, y no emplean en la identificación de la causa del error grueso medidas de presión.

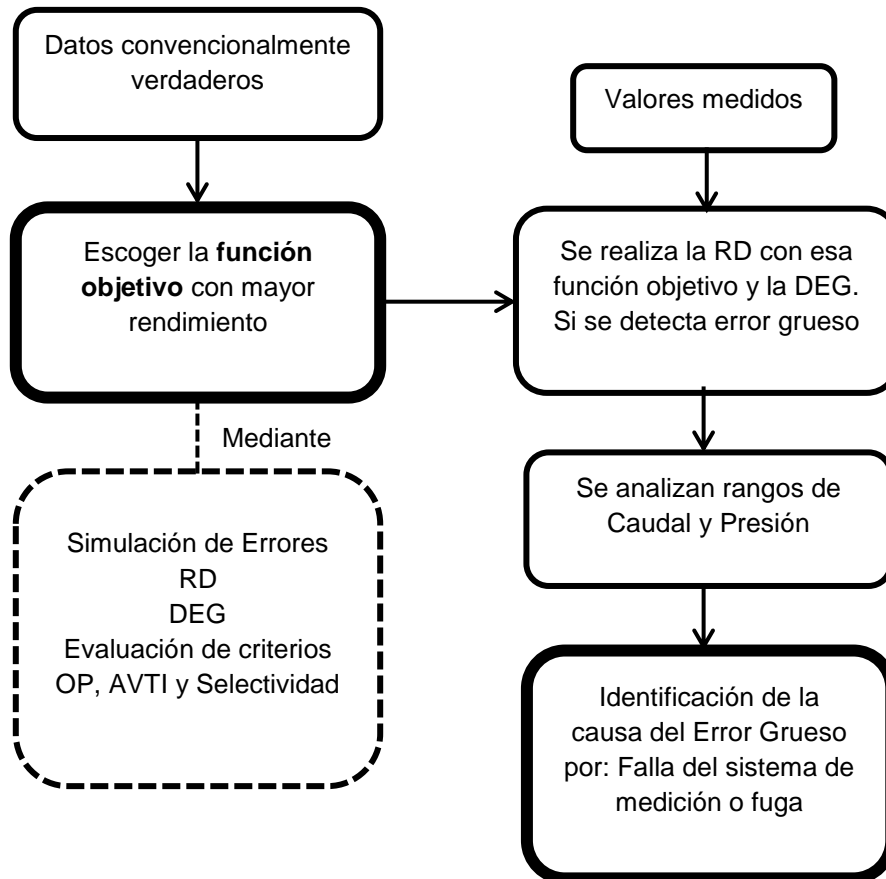
Con este trabajo se pretende identificar la causa del error grueso presentado en una red, ya sea por falla en el sistema de medición o por fuga. Para cumplir con lo anterior se requiere plantear una metodología para la DEG por medio de la RD, que evalúe diferentes funciones objetivo, con la cual sea posible identificar que función objetivo es la más adecuada al sistema analizado mediante los criterios *Overall Power (OP)*, *Average number of Type I error (AVTI)* y Selectividad. Esto permitirá determinar la causa del error grueso por medio de la definición de rangos de caudal y presión donde debería encontrarse el valor de estas variables, para el sistema analizado.

## 1. DESCRIPCION METODOLOGICA

### 1.1. DESCRIPCION GLOBAL DE LA METODOLOGIA

Este trabajo requiere la ejecución de dos etapas importantes: la escogencia de la función objetivo que más se adecúe a la red de medición a analizar, y la identificación de la causa del error grueso. Como se aprecia en la **Figura 1**, estas etapas requieren de una secuencia determinada.

**Figura 1.** Descripción global de la metodología



## 1.2. METODOLOGIA PARA ESCOGER LA FUNCION OBJETIVO CON MAYOR RENDIMIENTO CON LA DEG MEDIANTE LA RD

Se requiere contar con las restricciones (balances de masa) de la red de medición a analizar y los valores convencionalmente verdaderos de caudal y presión. Para evaluar la aptitud de las funciones objetivo, se recurre al método Monte Carlo, determinando el número de simulaciones a realizar, se escoge aleatoriamente una variable agregándole un error que oscila entre el -10 y 10% de la variable, por iteración. Se asume que se simuló un error grueso si el valor absoluto del error excede el 1% del valor real, y posteriormente se realiza la RD.

**1.2.1. Reconciliación de Datos (RD).** La RD ajusta las medidas redundantes, de modo que se obedezcan las leyes de conservación y cualquier otra restricción que incorpore el modelo matemático de una red de medición [15]. Los métodos más usados para la RD son: método de minimización de funciones, método de incertidumbre y matrices. El método de minimización de funciones se emplea en este trabajo.

Este método ajusta las mediciones de forma que los errores se reduzcan y que las leyes de conservación de la naturaleza sean obedecidas [7]. Consiste en la minimización de una combinación lineal de funciones dependientes de los errores en cada medidor  $\sum W_i \rho(\varepsilon_i)$ , donde  $W_i$  es el factor de ponderación,  $\rho$  es una función objetivo que depende de  $\varepsilon_i$  que es el error estándar, que se calcula mediante  $\varepsilon_i = (y_i - x_i)/\sigma_i$ , siendo  $y_i$  el valor medido,  $x_i$  el valor reconciliado y  $\sigma_i$  la desviación estándar del medidor  $i$ . En el Anexo A se presenta más información acerca de las funciones objetivo.

Las funciones objetivo más usadas para RD en redes de medición se encuentran en la **Tabla 1**.

**Tabla 1.** Funciones objetivo para la RD.

FUNCION DE MINIMIZACIÓN	ECUACIÓN
Mínimos Cuadrados (MC)	$\rho = \varepsilon_i^2$
Cauchy	$\rho = \ln \left( 1 + \frac{\varepsilon_i^2}{2,3849^2} \right)$
Fair	$\rho = \left  \frac{\varepsilon_i^2}{1,3998} - \ln \left( 1 + \frac{\varepsilon_i^2}{1,3998} \right) \right $
Normal Contaminada (NC)	$\rho = -\ln \left[ 0,765e^{\left(-\frac{\varepsilon_i^2}{2}\right)} + 0,0235e^{\left(-\frac{\varepsilon_i^2}{200}\right)} \right]$
Lorentzian (Lor)	$\rho = -\frac{1}{1 + (\varepsilon_i^2/13,52)}$

La herramienta computacional utilizada para realizar todo el presente trabajo, fue el lenguaje de programación PYTHON bajo la interfaz de usuario Winpython, que ha sido desarrollada bajo la licencia de código abierto [16]. Para la RD se emplea un factor de ponderación igual para cada medida y se utiliza el método SLSQ (*Sequential Least Squares Programming*) de optimización, originalmente planteado por Kraft [17].

**1.2.2. Detección de Errores Gruesos (DEG).** Para la DEG se emplearon dos métodos: el método A consiste en una propuesta de los autores, y el método B es un método encontrado en la literatura, que emplea pruebas de hipótesis [18].

**Método A.** El método A para la DEG emplea dos criterios que de ser cumplidos, se declara error grueso detectado. Para lo anterior es necesario calcular dos variables, la diferencia absoluta entre variable medida y reconciliada, y el promedio de estas diferencias, de la siguiente forma:

$$Dif_{p,i} = |Medida_{p,i} - Reconciliado_{p,i}|$$

$$\overline{Dif_{p,i}} = \frac{\sum_1^n |Medida_{p,i} - Reconciliado_{p,i}|}{n} = \frac{\sum_1^n Dif_{p,i}}{n}$$

Donde  $p$  es la posición de la variable,  $i$  la iteración y  $n$  el número de las variables medidas. El criterio uno para declarar DEG es  $Dif_{p,i} > 1\% Medida_{p,i}$ , parte del Reglamento Único de Transporte de gas natural (RUT), donde se establece que la incertidumbre de las mediciones de un sistema de distribución de gas natural debe ser no mayor al 1%.

**Método B.** En el método B para la DEG, implementa Test de Hipótesis con ayuda del planteamiento: el error estándar sigue una distribución normal estándar, con lo cual los errores normalizados están dentro de un intervalo de confianza  $(1 - \alpha)$  con un nivel de significancia elegido  $\alpha$ . Cualquier valor del error normalizado que cae fuera de esa región de confianza, se declara un valor atípico o error grueso [2]. Para profundizar sobre los Test de hipótesis, ver Anexo B.

Definiendo la hipótesis nula  $H_0$ , como la ausencia de errores gruesos y la hipótesis alternativa  $H_1$ , como la contraria a la hipótesis nula. Bajo este criterio se desglosan varios test para la DEG entre los que se encuentran: Test Global, Test Nodal, Test de Medida, entre otros [19].

En este tipo de Test se pueden presentar dos tipos de errores: el Error Tipo I, que es rechazar la hipótesis nula cuando ésta es cierta, y el Error Tipo II, que consiste en aceptar la hipótesis nula siendo ésta falsa. Con el fin de disminuir estos errores se utiliza un nivel de significancia modificado  $\beta$  (siendo  $\beta = 1 - (1 - \alpha)^n/2$ ), donde  $n$  es el número de nodos o variables medidas dependiendo el Test al que se aplique [19], con el cual disminuye tanto la probabilidad de existencia del error tipo I como del tipo II.

El método B emplea dos test de hipótesis en conjunto, el Test Nodal y el Test de Medida [18]. Buscando una detección más selectiva, se evalúan todos los nodos o balances para determinar en cuales se sospecha que contienen error grueso. Para ello, con ayuda del vector de medidas  $Y$ , de la matriz de balances  $M$ , su respectiva transpuesta  $M^T$  y la matriz varianza-covarianza del error  $S_r$ , se calcula el criterio estadístico del Test Nodal  $Z_{N_j}$  y se compara con el valor crítico del Test  $Z_{N_c}$ , si el criterio estadístico supera el valor crítico se demarca el nodo con sospecha de error grueso.

Posteriormente, se procede a realizar el Test de Medida para determinar la ubicación de la variable donde se presenta el error grueso. Para el cálculo del criterio estadístico  $Z_{M_i}$  se usan las matrices  $M, M^T, S_r$ , su inversa  $S_r^{-1}$  como el vector  $Y$ , junto al vector de variables reconciliadas  $X$ . Si el valor del criterio estadístico  $Z_{M_i}$  supera su valor crítico  $Z_{M_c}$ , se confirma la existencia de un error grueso en la medida  $i$ .

Después de realizar la DEG por los dos métodos, se comparan posición e iteración de los errores gruesos simulados y detectados.

### 1.3. CRITERIOS DE EVALUACION PARA FUNCIONES OBJETIVO

Para evaluar los métodos de DEG con las diferentes funciones objetivo se evalúan los criterios *Overall Power* (OP), *Average number of Type I Error* (AVTI) y Selectividad, que se aprecian en la **Tabla 2**.

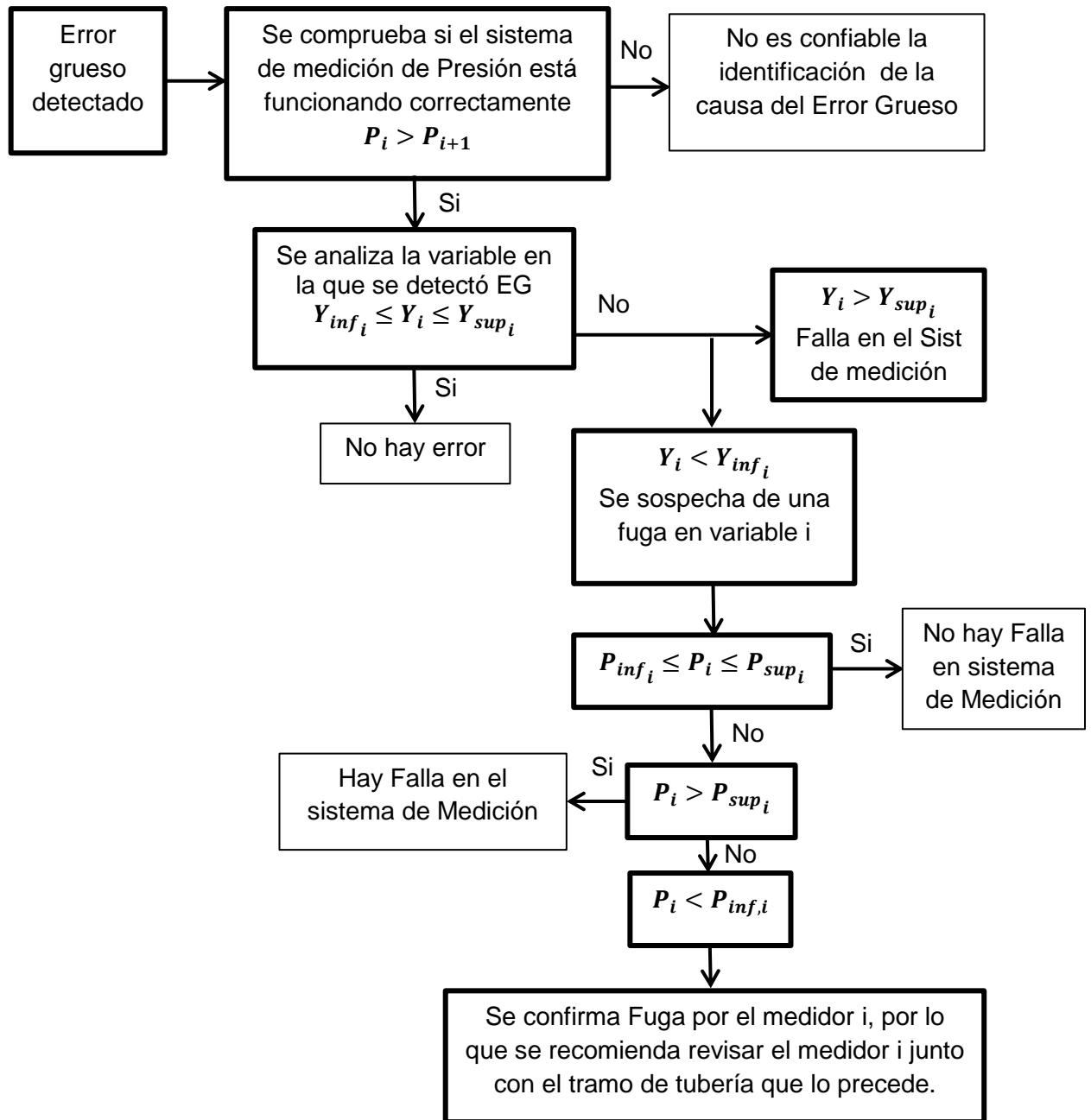
**Tabla 2.** Criterios de evaluación.

CRITERIOS DE EVALUACION	
OP	$\frac{\text{Número de errores gruesos correctamente detectados}}{\text{Número de errores gruesos simulados}}$
AVTI	$\frac{\text{Número de errores gruesos erróneamente identificados}}{\text{Número de simulaciones realizadas}}$
Selectividad	$\frac{\text{Número de errores gruesos correctamente identificados}}{\text{Número de errores gruesos detectados}}$

Castillo, D. & Gamboa, J. Aplicación de BRAM como modelo para el análisis de una Red de Gas Natural mediante Reconciliación de Datos. *En*: MET&FLU, 2010. Vol. 3, P. 42-53

Se espera que el criterio OP sea igual o cercano a 1, AVTI igual o cercano a 0 y Selectividad igual o cercano a 1; siendo los valores óptimos para escoger la función objetivo con mayor rendimiento, independiente del método de DEG.

**Figura 2.** Procedimiento para hallar la posible causa del error grueso.



#### **1.4. IDENTIFICACION DE LA POSIBLE CAUSA DEL ERROR GRUESO**

Después de detectar un error grueso con la función objetivo de mayor rendimiento, se procede a analizar rangos establecidos tanto de caudal como de presión, el procedimiento para este análisis se puede apreciar en la **Figura 2**. En los casos específicos donde se presenta fuga, en los sistemas de medición se registrará tanto caída de caudal como de presión, por lo cual se confirma fuga cuando el caudal y la presión están por debajo del límite inferior preestablecido, en cierto punto de medición. Cuando hay falla negativa en el sistema de medición (no por fuga), se espera que la presión se mantenga en el rango establecido.

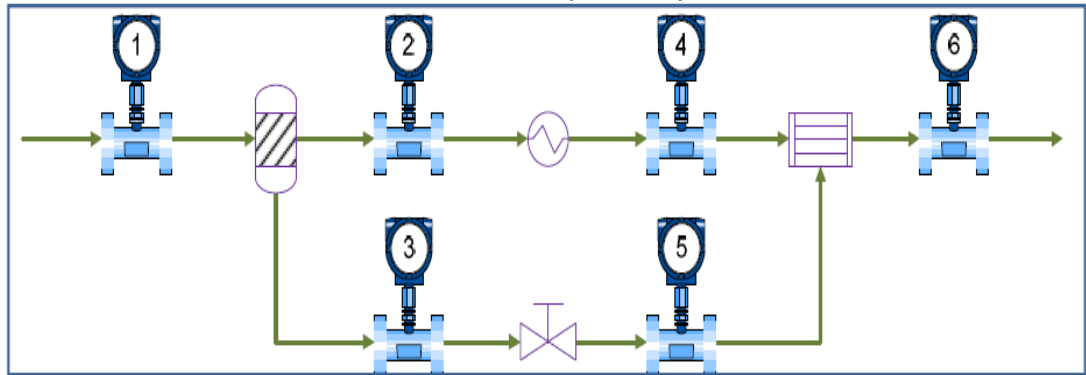
## **2. ANALISIS DE RESULTADOS**

### **2.1. VALIDACIÓN DE LA METODOLOGÍA PARA LA ELECCIÓN DE FUNCIÓN OBJETIVO**

Para realizar la validación de la metodología, esta se aplicó a 4 diferentes redes de medición, en el Anexo C se presenta la información de las redes, que se necesita ingresar en el algoritmo, como valores convencionalmente verdaderos de caudal o flujo másico, balances o restricciones del sistema y los valores críticos empleados para el método B de DEG.

**2.1.1. Ejemplo 1.** El Ejemplo 1 es un sistema de intercambiador de calor con *bypass* presentado por [2], y se aprecia en la **Figura 3**.

**Figura 3.** Diagrama de la red de medición del Ejemplo 1.



Castillo, D. Planteamiento de una metodología para la evaluación y control de los balances en las redes de gas natural. Tesis de pregrado. Ingeniera química. Bucaramanga: Universidad Industrial de Santander. Facultad de ingenierías fisicoquímicas. 2010.

Como se puede observar en la **Tabla 3**, las funciones objetivo con mejor rendimiento son Cauchy y Lorentzian, con el método A de DEG. Los criterios de evaluación son muy satisfactorios, *Overall Power* (OP) igual 1 indica que el número de errores gruesos correctamente identificados es igual al número de los errores simulados, AVTI de 0,117 indica que el programa detectó errores incorrectamente, aunque fueron pocos en consideración al número de simulaciones realizadas, y Selectividad igual a 0,882 indica que el programa detectó correctamente más de la mitad de los errores detectados.

Comparando los criterios de evaluación previamente expuestos con los presentados por [20], se evidencia una mejoría en la DEG, con distinta función objetivo.

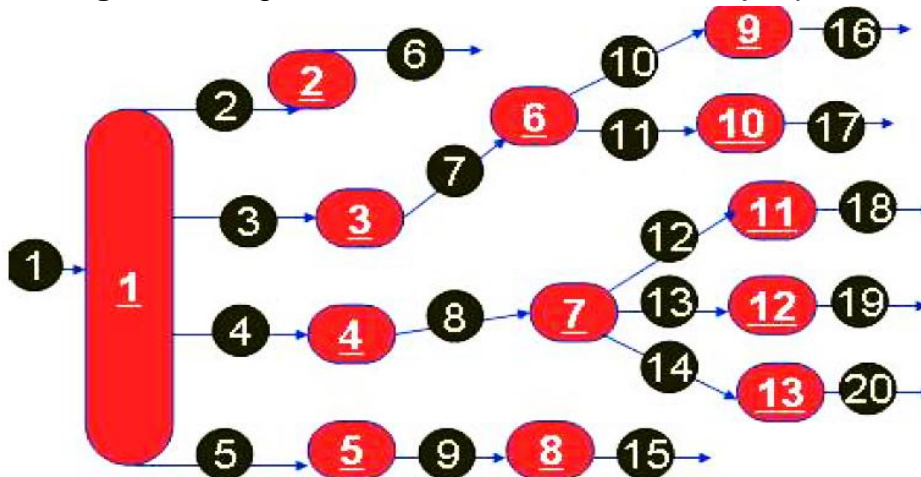
**Tabla 3.** Resultados del algoritmo desarrollado con el Ejemplo 1.

Número de Simulaciones	Método A					Método B				
	MC	Cau	Fair	NC	Lor	MC	Cau	Fair	NC	Lor
<b>E. Gruesos Simulados</b>	879					874				
<b>Función</b>	MC	Cau	Fair	NC	Lor	MC	Cau	Fair	NC	Lor
<b>Errores Detectados</b>	1876	996	1172	1123	996	734	535	535	546	535
<b>Correctamente Detectados</b>	817	879	879	737	879	535	535	535	503	535
<b>Detectados Incorrectamente</b>	1059	117	293	386	117	199	0	0	43	0
<b>OP</b>	0,929	1,000	1,000	0,838	1,000	0,612	0,612	0,612	0,575	0,612
<b>AVTI</b>	1,059	0,117	0,293	0,386	0,117	0,199	0,000	0,000	0,043	0,000
<b>Selectividad</b>	0,435	0,882	0,750	0,656	0,882	0,728	1,000	1,000	0,921	1,000

**2.1.2. Ejemplo 2.** Este ejemplo es presentado en [6], donde los autores (Tiria y Vega) plantean modelos de redes de medición con diferente complejidad. En la **Figura 4** se puede apreciar que los puntos negros representan los puntos donde se realizan las mediciones y que los rojos simbolizan equipos de un proceso.

Con los resultados presentados en la **Tabla 4** se puede determinar que la función objetivo con mejor rendimiento es Lorentzian, con el método A de DEG. Los criterios de evaluación con esta función reflejan que la metodología empleada es apropiada para este tipo de diagrama de medidas. OP igual a 1 indica que se detectaron correctamente todos los errores grueso simulados, AVTI cercano a 0 señala que el programa detecta incorrectamente algunos errores, que asume que son gruesos, cuando no lo son, por el efecto del “*smearing effect*”, la Selectividad de 0.939 señala que el programa es favorable, indicando que de los 911 errores gruesos detectados, 855 fueron correctamente identificados y 56 incorrectamente, de las 1000 simulaciones realizadas.

Figura 4. Diagrama de la red de medición del Ejemplo 2.



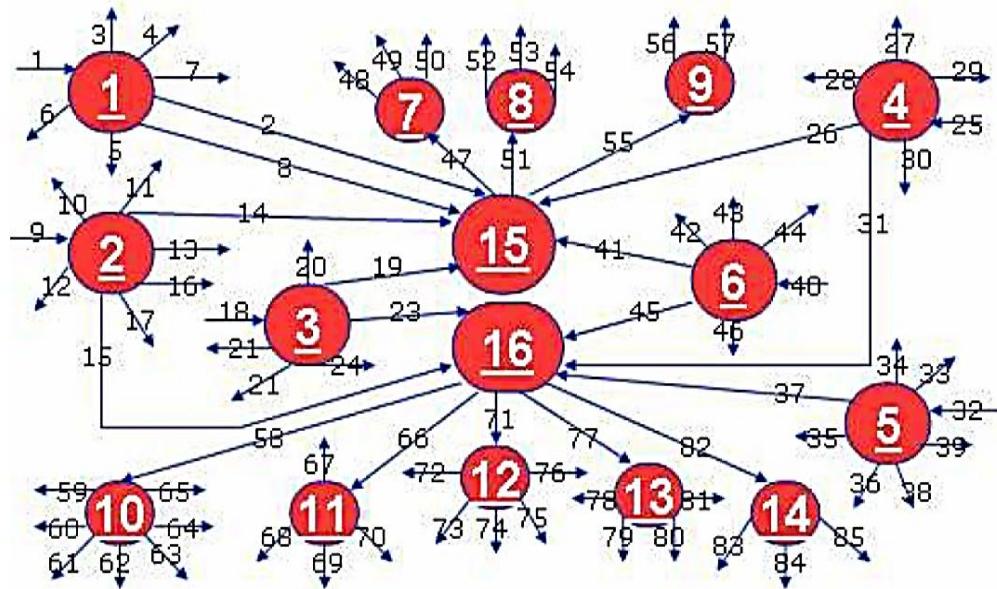
Tiria, S. & Vega, S. Desarrollo de un modelo de identificación de error grueso en sistemas complejos de balance de masa. Trabajo de pregrado. Ingeniero químico. Bucaramanga: Universidad Industrial de Santander. Facultad de ingenierías fisicoquímicas. 2007.

Tabla 4. Resultados del algoritmo desarrollado con el Ejemplo 2.

Número de Simulaciones	Método A					Método B				
	MC	Cau	Fair	NC	Lor	MC	Cau	Fair	NC	Lor
E. Gruesos Simulados	855					859				
Errores Detectados	2756	910	1077	1178	911	1245	621	905	641	616
Correctamente Detectados	843	854	853	847	855	559	588	591	559	564
Detectados Incorrectamente	1913	56	224	331	56	686	33	314	82	52
OP	0,986	0,999	0,998	0,991	1,000	0,650	0,684	0,688	0,650	0,656
AVTI	1,913	0,056	0,224	0,331	0,056	0,686	0,033	0,314	0,082	0,052
Selectividad	0,306	0,938	0,792	0,719	0,939	0,449	0,946	0,653	0,872	0,915

2.1.3. Ejemplo 3. Este ejemplo también es propuesto por [6], y se puede observar en la Figura 5.

**Figura 5.**Diagrama de la red de medición del Ejemplo 3.



Tiria, S. & Vega, S. Desarrollo de un modelo de identificación de error grueso en sistemas complejos de balance de masa. Trabajo de pregrado. Ingeniero químico. Bucaramanga: Universidad Industrial de Santander. Facultad de ingenierías fisicoquímicas. 2007.

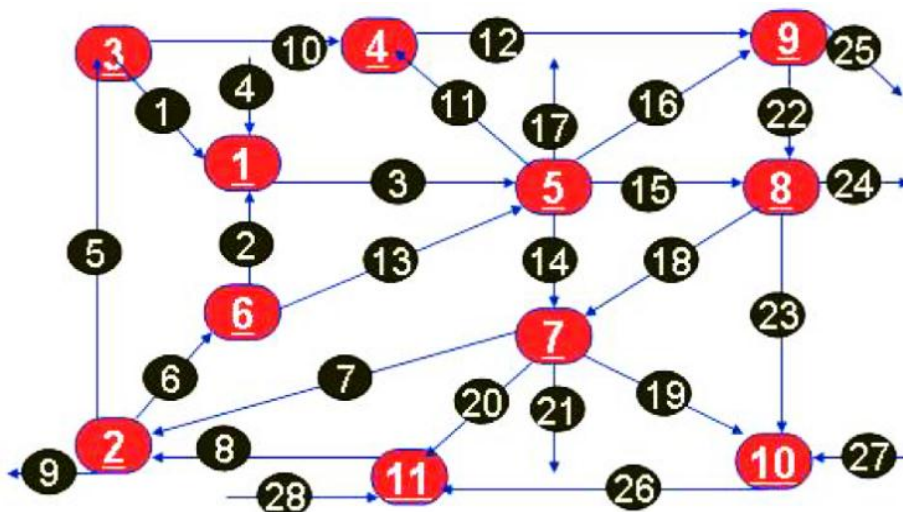
Los resultados que arroja el algoritmo para el Ejemplo 3 se encuentran en la **Tabla 5**. Los criterios de evaluación que se presentan no son los más satisfactorios, por lo que no es posible escoger una función objetivo con un rendimiento aceptable. El OP arroja valores muy bajos, 0,326 en promedio, cuando estos debería ser cercanos a 1; el criterio AVTI presenta valores de 1,83 en promedio, siendo 0 el valor esperado con cada función objetivo; y la Selectividad reporta valores 0,184 en promedio, siendo los valores cercanos a 1 los que generan confianza en la metodología. Lo anterior señala que la metodología empleada en el programa no es apropiada para este tipo de red de medición.

**Tabla 5.** Resultados del algoritmo desarrollado con el Ejemplo 3.

Número de Simulaciones	Método A					Método B				
	1000					1000				
E. Gruesos Simulados	903					908				
Función	MC	Cau	Fair	NC	Lor	MC	Cau	Fair	NC	Lor
Errores Detectados	770	2059	888	791	5347	1659	3960	1732	1667	2430
Correctamente Detectados	253	286	269	189	243	316	492	326	286	297
Detectados Incorrectamente	517	1773	619	602	5104	1343	3468	1406	1381	2133
OP	0,280	0,316	0,297	0,209	0,269	0,348	0,541	0,359	0,315	0,327
AVTI	0,517	1,773	0,619	0,602	5,104	1,343	3,468	1,406	1,381	2,133
Selectividad	0,328	0,138	0,302	0,238	0,045	0,190	0,124	0,188	0,171	0,122

**2.1.4. Ejemplo 4.** Este ejemplo se puede apreciar en la **Figura 6**, es originalmente presentado por [21].

**Figura 6.** Diagrama de la red de medición del Ejemplo 4.



Tiria, S. & Vega, S. Desarrollo de un modelo de identificación de error grueso en sistemas complejos de balance de masa. Trabajo de pregrado. Ingeniero químico. Bucaramanga: Universidad Industrial de Santander. Facultad de ingenierías fisicoquímicas. 2007

**Tabla 6.** Resultados del algoritmo desarrollado con el Ejemplo 4.

Número de Simulaciones	Método A					Método B				
	1000					1000				
E. Gruesos Simulados	918					902				
Función	MC	Cau	Fair	NC	Lor	MC	Cau	Fair	NC	Lor
Errores Detectados	1297	566	661	572	759	427	329	324	357	419
Correctamente Detectados	484	484	490	481	395	319	329	318	311	197
Detectados Incorrectamente	813	82	171	91	364	108	0	6	46	222
OP	0,527	0,527	0,533	0,523	0,430	0,353	0,364	0,352	0,344	0,218
AVTI	0,813	0,082	0,171	0,091	0,364	0,108	0,000	0,006	0,046	0,222
Selectividad	0,373	0,855	0,741	0,840	0,520	0,747	1,000	0,981	0,871	0,470

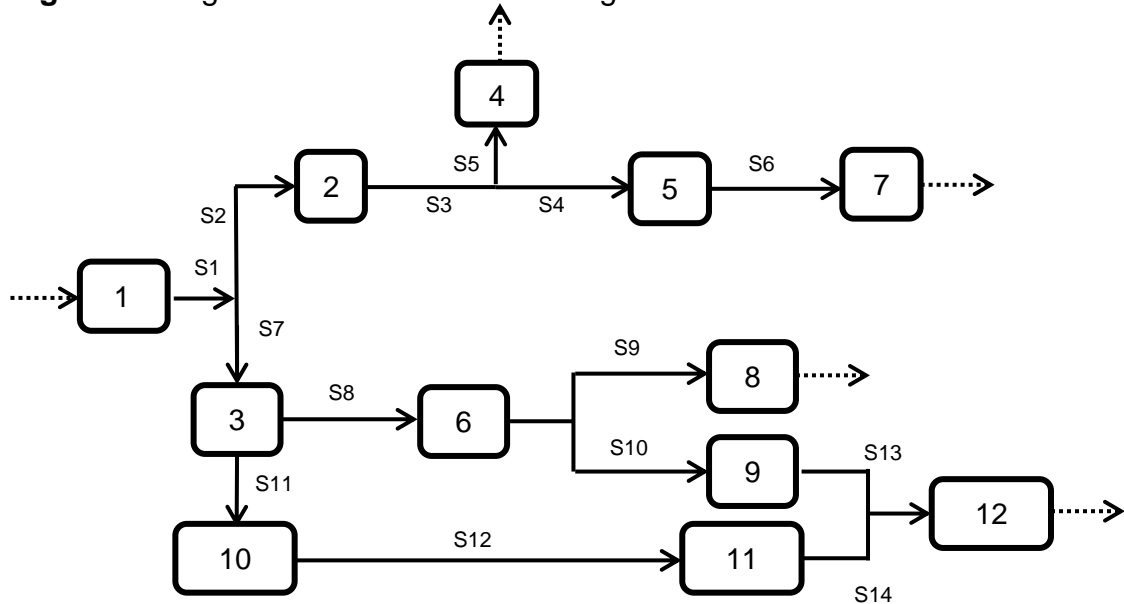
En la **Tabla 6** se encuentran los resultados del Ejemplo 4 que arroja el algoritmo con la metodología planteada. El criterio OP muestra un promedio de valores igual a 0,417, conociendo que para ser aceptables, sus valores deben ser cercanos a 1, AVTI presenta valores de 0,19 en promedio, siendo 0 el valor esperado con cada función objetivo; y la Selectividad reporta valores de 0,74 en promedio, siendo los valores cercanos a 1 los esperados.

Para este ejemplo o red de medición, las funciones objetivo con resultados apropiados son: Cauchy y Normal Contaminada, con el método A de DEG. Sin embargo el criterio OP arroja valores de 0,52 (para las dos funciones), lo que indica que se detecta correctamente un poco más de la mitad de los errores simulados, cuando se espera que detecten todos o la mayoría de estos.

## 2.2. IDENTIFICACIÓN DE LA CAUSA DEL ERROR GRUESO, MEDIANTE LA METODOLOGÍA GLOBAL PLANTEADA

La metodología global se prueba en una red de transporte de agua simulada en Aspen Hysys. En la **Figura 7** se muestra el diagrama de la red de medición.

**Figura 7.** Diagrama red de medición de agua.



Las consideraciones tomadas al diseñar la red fueron: agua pura que entra a 296,15K y a 1241 kPa (180 psi), con diámetro de 0,102 m (4 in, *Schedule 40*), sin cambio de altura. En la **Tabla 7** se encuentran los valores convencionalmente verdaderos de caudal y los valores estimados de presión por el simulador. En la **Tabla 8** las distancias de los segmentos de tubería.

**Tabla 7.** Valores de Caudal y Presión.

Variable	Caudal [m <sup>3</sup> /h]	Presión [kPa]	Variable	Caudal [m <sup>3</sup> /h]	Presión [kPa]
1	7200	1241	7	2448	1053
2	3600	1067	8	1022,4	1079
3	3600	1082	9	849,6	1079
4	1152	1061	10	1728	1079
5	2448	1061	11	1728	1075
6	1872	1080	12	2577,6	1073

**Tabla 8.** Distancia de tramos de tubería en la red.

Segmento	Longitud [M]	Segmento	Longitud [M]
S1	900	S8	150
S2	300	S9	72
S3	120	S10	108
S4	12	S11	240
S5	60	S12	258
S6	300	S13	120
S7	6	S14	120

Las siete restricciones de la red de transporte de agua se encuentran a continuación, donde  $F_i$  es el caudal en m<sup>3</sup>/h, en los diferentes puntos de medición  $i$ , pues se asume densidad constante.

$$F_1 - F_2 - F_3 = 0$$

$$F_2 - F_4 - F_5 = 0$$

$$F_5 - F_7 = 0$$

$$F_3 - F_6 - F_{10} = 0$$

$$F_7 - F_8 - F_9 = 0$$

$$F_{10} - F_{11} = 0$$

$$F_9 - F_{11} - F_{12} = 0$$

Los resultados de la DEG empleando la RD mediante los 2 métodos mencionados se presentan en la **Tabla 9**, para el método se emplean los criterios de prueba de  $Z_{NC} = 2,68$  y  $Z_{MC} = 2,85$ . Para el método B se emplea un nivel de significancia  $\alpha$  de 5%, aplicado en [10]. Se evidencia notoriamente que la función Fair en conjunto con el método A de DEG son los más adecuados para implementar en el algoritmo (para analizar casos donde se presentan tanto fugas como fallas en el sistema de medición en la red de agua), puesto que presenta los resultados más apropiados tanto para OP como en Selectividad, valores que priman en la selección de la función objetivo para la RD.

**Tabla 9.** Resultados de la DEG en la red de agua.

	MÉTODO A					MÉTODO B				
	MC	Cau	Fair	NC	Lor	MC	Cauc	Fair	NC	Lor
<b>Número de Simulaciones</b>	1000					1000				
<b>Número de Errores Simulados</b>	884					906				
<b>Función</b>	MC	Cau	Fair	NC	Lor	MC	Cauc	Fair	NC	Lor
<b>Errores Detectados</b>	2349	890	879	2425	1861	2669	1551	1849	1285	1389
<b>Correctamente Detectados</b>	817	858	879	849	830	861	858	849	854	856
<b>Detectados Incorrectamente</b>	1532	32	0	1576	1031	1808	693	1000	431	533
<b>OP</b>	0,924	0,970	0,994	0,960	0,938	0,950	0,947	0,937	0,942	0,944
<b>AVTI</b>	1,532	0,032	0,000	1,576	1,031	1,808	0,693	1,000	0,431	0,533
<b>Selectividad</b>	0,347	0,964	1,000	0,350	0,446	0,326	0,553	0,459	0,664	0,616

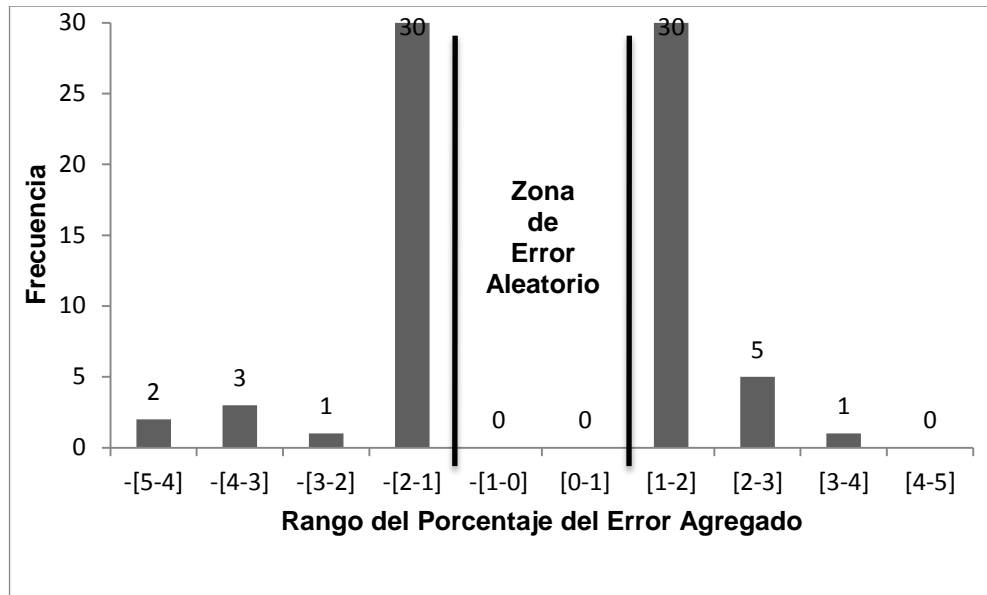
El rango establecido de caudal es 1% tanto inferior como superior del valor convencionalmente verdadero, por la razón previamente expuesta, tomando como base el RUT. El rango establecido de presión es 0,5% inferior y superior de los valores de presión estimados por el simulador Aspen Hysys, debido a que los medidores de presión convencionales con una buena precisión tienen desviaciones estándar no mayores al 0,5%.

Mediante 72 casos, 36 casos de fugas y 36 casos de fallas de sistema de medición, se probó el algoritmo para determinar la causa del error grueso. Para lo anterior fue necesario simular 72 errores gruesos. En la **Figura 8** se presenta, el número de casos con su respectivo rango de porcentaje agregado, al valor convencionalmente verdadero. En el medio se encuentra la zona del error aleatorio, ya que en este trabajo se considera error grueso cuando el valor absoluto del porcentaje del error supera el 1%.

Para poder determinar la causa del error grueso, primero se comprueba si el sistema de medición de presión está funcionando correctamente, de manera que  $P_i > P_{i+1}$ , siendo la presión aguas arriba mayor que aguas abajo. Cuando se comprueba que está funcionando correctamente, se pasa a determinar la causa del error grueso.

Como ya se mencionó, las fugas generan caída de caudal y presión, mientras que los errores gruesos causados por desviaciones tanto positivas como negativas respecto al valor real (en cuanto a caudal), no presentan variación de presión, por lo cual representan fallas en el sistema de medición. En la Tabla 6 se presentan los resultados del algoritmo para determinar la causa del error grueso.

**Figura 8.** Histograma de frecuencia de los rangos del error agregado, de los 72 casos de errores gruesos simulados.



**Tabla 10.** Resultados del algoritmo para determinar la causa del error grueso.

	Fugas	Falla Sist. de Medición	Total
<b>Casos</b>	36	36	72
<b>Errores gruesos detectados</b>	24	27	51
<b>Correctamente Detectadas</b>	18	24	42

Para determinar la eficiencia del algoritmo para detectar la causa del error grueso se calcularon dos selectividades, global y de detección, de la siguiente manera:

$$S_G = \frac{\text{Numero de fugas y fallas correctamente detectadas}}{\text{Numero de fugas y fallas o errores gruesos simuladas}} = \frac{18 + 24}{72} = 0.58$$

$$S_D = \frac{\text{Numero de fugas y fallas correctamente detectados}}{\text{Numero total de errores gruesos detectados}} = \frac{18 + 24}{24 + 27} = 0.82$$

La selectividad global muestra que es posible detectar más de la mitad de las fugas y fallas simuladas en la red de agua, con un 58%. Sin embargo, la selectividad de detección es mayor que la global, debido a que solo tiene en cuenta los errores gruesos detectados, por lo que el algoritmo tiene un alto potencial para determinar la causa del error grueso, cuando se detectan en la red de agua, con un 82%.

### 3. CONCLUSIONES

La combinación de los Test estadísticos, nodal y de medida, son una buena forma para la DEG. Sin embargo, se comprueba que la metodología planteada por los autores (método A) genera mejores resultados en las diferentes funciones objetivo, en los 4 ejemplos presentados junto con la red de agua, presentando un método más sencillo con el que se disminuye el tiempo de corrida, además posee una manera menos compleja de ingresar los datos y las restricciones del sistema.

Con los resultados arrojados para los ejemplos es posible identificar la función objetivo más adecuada a las redes de medición, mediante los criterios de evaluación OP, AVTI y Selectividad. Se aprecia que el rendimiento de estas funciones varía de acuerdo al tipo de red, por razones como: la redundancia en el sistema, la ubicación y tamaño de los errores.

La metodología planteada a lo largo de este trabajo, muestra que es posible detectar y determinar la causa del error grueso, ya sea fuga o falla en el sistema de medición, en la mayoría de los casos, como se evidencia en la red de agua, por medio del análisis de rangos de caudal y presión.

Lo anterior refleja que es factible implementar esta metodología en un sistema de medición en tiempo real, lo que ayudaría a detectar los errores gruesos en el momento en que se presenten, sumando los beneficios económicos que traería el saber que medidores necesitan revisión o calibración inmediata y que tramos de tubería de una red necesitan inspección y/o reparación, por pérdidas de material.

#### 4. RECOMENDACIONES

- Probar la metodología planteada en diferentes redes reales de medición de transporte de fluidos.
- Emplear el factor de compresibilidad  $Z$  para probar la metodología en una red de Gas Natural.
- A la hora de programar (escribir el código en determinado lenguaje de programación) optimizar el número de líneas para disminuir el tiempo de ejecución.
- En los próximos trabajos se recomienda hacer partícipe a estudiantes de otras ingenierías, como ingeniería de sistemas, para montar la metodología en un programa solido (software).

## CITAS BIBLIOGRAFICAS

- [1] BIPM, IEC, et al. Evaluation of measurement data-guide to the expression of uncertainty in measurement. En: Joint Committee for Guides in Metrology (JCGM 100: 2008, GUM 1995 with minor corrections). 2008.
- [2] Narasimhan S, Jordache C. Data Reconciliation & Gross Error Detection. An intelligent use of process data. Gulf Publishing Company. 2000.
- [3] Martini, A., Sorace, A., Traverso, A. & Massardo, A. Data Reconciliation for power systems monitoring: Application to a microturbine-based test rig. En: Applied Energy, 2013. Vol. 111, p.1152-1161.
- [4] Bagajewicz, M. & Jiang, Q. Gross error modeling and detection in plant linear dynamic reconciliation. En: Computers & Chemical Engineering, 1998. Vol. 22, p. 1789-1809.
- [5] Tomasi, W. Sistemas de comunicaciones electrónicas. Edición 4. Prentice-Hall. 2003.
- [6] Tiria, S. & Vega, S. Desarrollo de un modelo de identificación de error grueso en sistemas complejos de balance de masa. Trabajo de pregrado. Ingeniero químico. Bucaramanga: Universidad Industrial de Santander. Facultad de ingenierías fisicoquímicas. 2007.
- [7] Castillo, D. & Gamboa, J. Aplicación de BRAM como modelo para el análisis de una Red de Gas Natural mediante Reconciliación de Datos. En: MET&FLU, 2010. Vol. 3, p. 42-53
- [8] Sun, S., Huang, D. & Gong, Y. Gross Error Detection and Data Reconciliation using Historical Data. En: Procedia Engineering, 2011. Vol. 15, p. 55-59.
- [9] Badillo, J. Desarrollo de una herramienta para la evaluación y control de los balances en las redes de distribución de gas natural, basada en la

reconciliación de datos y la detección de errores gruesos. Trabajo de pregrado. Ingeniero químico. Bucaramanga: Universidad Industrial de Santander. Facultad de ingenierías fisicoquímicas. 2012.

- [10] Badillo, J., Chaves, A., & Fuentes, J. Computational tool for material balances control in natural gas distribution network. En: Ciencia, Tecnología y Futuro CT&F, 2013. Vol. 5, p. 31-46.
- [11] Jiang, X., Liu, P. & Li, Z. Data reconciliation and gross error detection for operational data in power plants. En: Energy, 2014. Vol. 75, p. 14-23.
- [12] Sánchez. M., Romagnoli, J., Jiang, Q. & Bagajewicz, M. Simultaneous estimation of biases and leaks in process plants. En: Computers & Chemical Engineering, 1999. Vol. 23, p. 841-857.
- [13] Narasimhan, S. & Mah RHS. Generalized likelihood ratio method for gross error identification. En: AIChE Journal, 1987. Vol. 33, p. 1514-1521.
- [14] Miao, Y., Su, H., Wang, W. & Chu, J. Simultaneous data reconciliation and joint bias and leak estimation based on support vector regression. En: Computers & Chemical Engineering, 2011. Vol. 35, p. 2141-2151.
- [15] Lozano, M. & Remiro, J. Diagnóstico de calderas de vapor. Aplicación de la técnica de reconciliación de datos. En: Ingeniería Química, 2000. Vol. 372, p. 117-123.
- [16] PYTHON software foundation. Python: 3.4.3. [Online] Python, 2001-2015. [Beaverton- Oregon, E.U.] Python Software Foundation, Septiembre, 2015. [citado 14 Octubre 2015] Available from Internet: <https://www.python.org/downloads/release/python-343/>.
- [17] Kraft, D. A software package for sequential quadratic programming. En: DLR German Aerospace Center, Institute for flight mechanics, 1988.

- [18] Mei, C., Su, H. & Chu, J. An MT-NT Combined Method for Gross Error Detection and Data Reconciliation. En: Chinese Journal Chemical Engineering, 2006. Vol. 14, p. 592-596.
- [19] Bagajewicz, M. Smart Process Plants. Software and Hardware Solutions for Accurate Data and Profitable Operations. McGraw-Hill Companies 2010.
- [20] Castillo, D. Planteamiento de una metodología para la evaluación y control de los balances en las redes de gas natural. Tesis de pregrado. Ingeniera química. Bucaramanga: Universidad Industrial de Santander. Facultad de ingenieras fisicoquímicas. 2010.
- [21] Serth, R. & Heenan, W. Gross error detection and data reconciliation in steam-metering Systems. En: AIChE Journal, 1986. Vol. 32, p. 733-742.
- [22] Huber, P.J. Robust Statistics. John Wiley & Sons, New York 1981.
- [23] Mah, R.S.H. & A.C. Tamhane. Detection of gross errors in process data. En: AIChE Journal, 1982. Vol. 28, p. 828-830.

## BIBLIOGRAFIA

- BADILLO HERREA, Jesús David; CHAVES, Arlex y FUENTES OSORIO, José Augusto. Computational tool for material balances control in natural gas distribution network. *Ciencia Tecnología y Futuro*. [Online] June, 2013. Vol. 5 N° 2 [citado 6 marzo 2015] p. 31-46. Available from Internet. SCIELO – Scientific Electronic Library Online: <http://www.scielo.org.co/scielo.php?lng=en>
- BAGAJEWICZ, Miguel & JIANG Qiyu. Gross error modeling and detection in plant linear dynamic reconciliation. *Computers & Chemical Engineering*. [Online] November, 1998. Vol. 22 [Citado Agosto 2015] p. 1789-1809.
- BAGAJEWICZ, Miguel J. Smart Process Plants. Software and Hardware Solutions for Accurate Data and Profitable Operations. McGraw-Hill Companies. 2010.
- BIPM, IEC, et al. Evaluation of measurement data-guide to the expression of uncertainty in measurement. JCGM 100:2008, GUM 1995 with minor corrections. [Online] September, 2008. First edition [Citado Septiembre 2015].
- CASTILLO BLANCO, Diana Marcela. Planteamiento de una metodología para la evaluación y control de los balances en las redes de gas natural. Tesis de Grado Ingeniera Química. Bucaramanga: Universidad Industrial de Santander, Facultad de Ingenierías Físico químicas, 2010.
- CASTILLO BLANCO, Diana Marcela y GAMBOA JAIMES, René. Aplicación de BRAM como modelo para el análisis de una Red de Gas Natural mediante Reconciliación de Datos. *MET&FLU*. [En línea] 2010. Vol. 3 [Citado Septiembre 2015] p. 42-53.

- HUBER, P.J. Robust Statistics. John Wiley & Sons, New York, 1981.
- XIAOLONG, Jiang; PEI, Liu & ZHENG, Li. Data reconciliation and gross error detection for operational data in power plants. *Energy*. [Online] October, 2014. Vol.75 [citado Agosto de 2015] p. 14-23.
- KRAFT, D. A software package for sequential quadratic programming. German Aerospace Center. Institute for Flight Mechanics, Koln, Germany, 1988.
- LOZANO SERRANO, Miguel Angel & REMIRO HERNÁNDEZ, Jesús Angel. Diagnóstico de calderas de vapor. Aplicación de la técnica de reconciliación de datos. *Ingeniería Química*. [En línea] Octubre, 2000. Vol.372 [Citado Septiembre 2015] p. 117-123.
- MAH, R.S.H. & A.C. Tamhane. Detection of gross errors in process data. *AIChE Journal*. [Online] September, 1982. Vol. 28 [Citado Julio 2015], p.828-830.
- MARTINI, A.; SORCE, A.; TRAVERSO, A. & MASSARDO A. Data Reconciliation for power systems monitoring: Application to a microturbine-based test rig. *Applied Energy*. [Online] November, 2013. Vol. 111, [Citado Junio 2015] p. 1152-1161.
- MEI, Congli; SU, Hongye & CHU, Jian. An MT-NT Combined Method for Gross Error Detection and Data Reconciliation. *Chinese Journal Chemical Engineering*. [Online] 2006. Vol 14, [Citado Mayo 2015] p. 592-596.
- MIAO, Yu; SU, Hongye; WANG, Wei & CHU, Jian. Simultaneous data reconciliation and joint bias and leak estimation based on support vector regression. *Computers & Chemical Engineering*. [Online] October, 2011. Vol 35, [Citado Julio 2015] p. 2141-2151.
- NARASIMHAN, S. & MAH, RHS. Generalized likelihood ratio method for gross error identification. *AIChE Journal*. [Online] September, 1987. Vol. 33 [citado Julio 2015] p. 1514-1521.

- NARASIMHAN, S. & JORDACHE, C. Data Reconciliation & Gross Error Detection. An intelligent use of process data. Gulf Publishing Company. 2000.
- PYTHON SOFTWARE FOUNDATION. Python: 3.4.3 [Online] Python, 2001-2015. [Beaverton- Oregon, E.U.] Python Software Foundation, September, 2015. [Citado 14 Octubre 2015] Available from Internet: <https://www.python.org/downloads/release/python-343/>
- SÁNCHEZ, M., ROMAGNOLI, J., JIANG, Q., & BAGAJEWICZ, M. Simultaneous estimation of biases and leaks in process plants. *Computers & Chemical Engineering*. [Online] July, 1999. Vol. 23 [Citado Agosto 2015] p. 841-857.
- SERTH, R. & HEENAN, W. Gross error detection and data reconciliation in steam-metering Systems. *AIChE Journal*. [Online] May, 1986. Vol. 32 [Citado Agosto 2015] p. 733-742.
- SUN, S., HUANG, D. & GONG, Y. Gross Error Detection and Data Reconciliation using Historical Data. *Procedia Engineering*. [Online] 2011. Vol. 15 [Citado Julio 2015] p. 55-59.
- TIRIA, Sergio y VEGA, Sergio. Desarrollo de un modelo de identificación de error grueso en sistemas complejos de balance de masa. Tesis de Grado Ingeniera Química. Bucaramanga: Universidad Industrial de Santander, Facultad de Ingenierías Físico químicas, 2007.
- TOMASI, Wayne. Sistemas de comunicaciones electrónicas (Edición 4) Prentice-Hall, 2003.

## ANEXOS

### Anexo A. Funciones objetivo en la RD

En el enfoque tradicional de la RD, se considera que el error de medición sigue una distribución normal con media en cero y varianza conocida, todas las deducciones estadísticas se basan en éste concepto, empleando una función objetivo. Sin embargo, si se producen desviaciones de esta distribución ideal por presencia de cualquier error grueso, aunque sea pequeño, este invalida la base estadística del estimador y genera estimaciones contaminadas (*smearing effect*).

Hay varias familias de funciones objetivas que pueden utilizarse para resolver el problema de RD en presencia de errores gruesos. Funciones objetivo con puntos de ruptura apropiados (por ejemplo, distribuciones normales con media diferente de cero, o combinación de dos distribuciones normales teniendo en cuenta el efecto que generan los errores gruesos) se pueden utilizar de manera que el error grueso no se extienda por todo el conjunto de datos.

Una de estas familias de funciones se desglosa a partir de la función objetivo general de máxima verosimilitud propuesta por Huber [22].

$$\min \left\{ L = \sum_{i=1}^n \rho(\varepsilon_i) \right\}$$

Donde  $\varepsilon$  es el error estándar, dado por

$$\varepsilon = \left( \frac{y_i - x_i}{\sigma_i} \right)$$

Donde  $y_i$  es la variable medida,  $x_i$  la variable reconciliada y  $\sigma_i$  su desviación estándar.

Por lo tanto, una función monótona  $\rho$  puede ser utilizado para formular la RD de modo que los errores gruesos tengan poco efecto sobre la estimación de las mediciones.

## ANEXO B. Pruebas de hipótesis empleadas para DEG

Las técnicas estadísticas más comúnmente utilizados para la detección de errores gruesos se basan en pruebas de hipótesis. En dicha detección, la hipótesis nula  $H_0$  declara que no hay errores gruesos presentes, y la hipótesis alternativa  $H_1$  declara que uno o más errores gruesos están presentes en el sistema. Todas las técnicas estadísticas para decidir entre las dos hipótesis hacen uso de un estadístico de prueba, que es una función de las mediciones y cuya distribución de probabilidad depende de la población original y del parámetro sobre el cual se desea realizar la conclusión. La prueba de hipótesis compara el valor del estadístico de prueba con un valor umbral predefinido por el nivel de significancia ( $\alpha$ ) escogido, y la hipótesis nula se rechaza o se falla en rechazar, respectivamente, dependiendo de si el estadístico excede el umbral o no. El valor umbral también se conoce como el criterio de prueba o el valor crítico de la prueba.

En estas pruebas se pueden presentar dos clases de errores:

- Error tipo I: Rechazar la hipótesis nula, cuando la hipótesis nula es cierta.
- Error tipo II: Fallar en rechazar la hipótesis nula, siendo ésta falsa.

Una serie de pruebas estadísticas se derivan de este principio estadístico básico y son capaces de detectar errores gruesos. Pero no todas las pruebas estadísticas son capaces de identificar los diferentes tipos y la localización de los errores gruesos. Algunas pruebas estadísticas básicas son capaces de detectar sólo los errores de medición (sesgos). Otras pruebas estadísticas sólo pueden detectar errores de modelo de proceso o fugas [2].

### GLOBAL TEST

Este método no puede identificar cuál variable posee el error grueso, y se basa en el análisis del vector residual  $r$ .

$$r = My$$

Donde  $y$  es el vector de las variables medidas y  $M$  es la matriz de los coeficientes lineales de las restricciones o balances.

La matriz covarianza de este vector está dada por:

$$J = MS_r M^T$$

Donde  $S_r$  es la matriz covarianza del error y  $M^T$  es la matriz transpuesta de  $M$ .

Las covarianzas para las restricciones lineales se pueden despreciar debido a la poca relación entre medidores [19], quedando solo las varianzas de cada medición.

En ausencia de errores gruesos, la siguiente variable

$$\chi^2 = rJ^{-1}r$$

Bajo  $H_0$ , la estadística anterior sigue una  $\chi^2$ -distribución con  $\vartheta$  grados de libertad, donde  $\vartheta$  es el rango de la matriz  $M$ , si el criterio del test es escogido como  $\chi_{1-\alpha, \vartheta}^2$ , donde  $\chi_{1-\alpha, \vartheta}^2$  es el valor crítico de la distribución  $\chi^2$  con el nivel de significancia elegido  $\alpha$ , entonces se rechaza  $H_0$  y un error grueso es detectado, si  $\chi^2 > \chi_{1-\alpha, \vartheta}^2$ . Esta elección del criterio de prueba garantiza que la probabilidad de error tipo I para esta prueba es inferior o igual a  $\alpha$ . La prueba global combina todos los residuos de las restricciones en la obtención de la estadística de prueba, y por lo tanto da lugar a una prueba multivariada o colectiva [2].

## TEST NODAL

La prueba Nodal investiga los residuos de las ecuaciones de los balances con un criterio estadístico  $Z_N$ . En ausencia de error grueso el vector de residuos sigue una distribución normal estándar bajo  $H_0$ .

$$r_N = MY$$

Donde la matriz covarianza de  $r_N$  esta dada por

$$J = MS_r M^T$$

Se espera entonces que la siguiente variable siga una distribución normal estándar

$$Z_{N_j} = \frac{r_{N_j}}{\sqrt{J_{jj}}}$$

Por lo tanto, si  $Z_{N_j}$  es más grande que el valor crítico ( $Z_{1-\alpha/2}$ ) basado en el nivel de confianza  $\alpha$ , se concluye que hay al menos un error grueso en el conjunto de mediciones que participan en el balance del nodo correspondiente  $j$ .

Puesto que múltiples pruebas se realizaron con el mismo valor crítico, aumenta la probabilidad de que una de las pruebas puede ser rechazada, incluso si no hay errores graves presentes. En otras palabras, la probabilidad de error tipo I será más que el valor especificado de  $\alpha$ . Si queremos controlar la probabilidad de error de tipo I, el siguiente nivel de significación modificado  $\beta$  puede usarse y el criterio del test para todas las restricciones puede ser escogido como  $Z_{1-\beta/2}$ .

$$\beta = 1 - (1 - \alpha)^{1/m}$$

## TEST DE MEDIDA

El test está basado en la suposición de que los errores aleatorios en las mediciones son independientes y distribuidos normalmente con media en cero. Este método está basado en la medida del ajuste del vector  $a$  el cual bajo la hipótesis nula  $H_0$ , se espera que su valor sea cero [19].

$$a = Y - X$$

Donde  $Y$  es el vector de las variables medidas y  $X$  es el vector de las variables reconciliadas.

La matriz covarianza de este vector está dada por:

$$W = S_r \cdot M^T (M \cdot S_r \cdot M^T)^{-1} M \cdot S_r$$

$Z_i$  es la prueba estática y sigue una distribución normal estándar a falta de errores gruesos.

$$Z_i = \frac{|a|}{W_{i,i}}$$

En esta prueba el criterio es  $Z_{B/2}$  y se define como el valor crítico de la distribución normal estándar para un nivel de confianza modificado  $\beta$

$$\beta = 1 - (1 - \alpha)^{1/n}$$

Donde  $n$  es el número de columnas de la matriz de restricciones. Si  $Z_i$  es mayor que  $Z_{1-\beta/2}$  la medición  $i$  es sospechosa de contener un error grueso.

### **TEST DE MEDIDA DE MÁXIMA POTENCIA**

Mah y Tamhane [23] en 1982, plantearon una transformación lineal para la cual el test de medida tiene máxima potencia, que reduce el número de errores Tipo II. Esta transformación es:

$$d = S_r^{-1} a$$

$$W = M^T (M \cdot S_r \cdot M^T)^{-1} M$$

Donde  $S_r^{-1}$  es la matriz inversa de  $S_r$ . La estadística asociada está dada por:

$$|Z_i| = \frac{|d_i|}{\sqrt{W_{i,i}}}$$

Que tiene una distribución normal estándar. Este test se aplica de la misma manera que el test de medida normal.

**ANEXO C. Información necesaria para realizar la RD, de los sistemas propuestos.**

Los balances y los valores convencionalmente verdaderos que se emplearon para aplicar la RD se presentan a continuación, de acuerdo al ejemplo o red analizada.

**Ejemplo 1**

**Tabla 11.** Valores verdaderos de las corrientes y los balances del Ejemplo 1.

Variable	Valor Verdadero de Flujo	Balances
1	100	$F_1 + F_2 - F_3 = 0$ $F_2 - F_4 = 0$ $F_3 - F_5 = 0$ $F_4 + F_5 - F_6 = 0$
2	64	
3	36	
4	64	
5	36	
6	100	

Para el método B en el ejemplo 1, empleando las pruebas de hipótesis con un nivel de significancia  $\alpha = 0,05$ . Se estiman niveles de significancia modificados  $\beta$  tanto para el test nodal como el test de medida, garantizando así mejores resultados. Con estos valores se encuentran los siguientes criterios estadísticos  $Z_{N_c} = 2,49$  y un  $Z_{M_c} = 2,63$  para cada test respectivamente:

## Ejemplo 2

**Tabla 12.** Valores verdaderos del proceso en el Ejemplo 2.

Variable	Valor Verdadero [Kg/h]	Variable	Valor Verdadero [Kg/h]
1	800	11	40
2	100	12	100
3	100	13	100
4	400	14	200
5	200	15	200
6	100	16	60
7	100	17	40
8	400	18	100
9	200	19	100
10	60	20	200

Los balances de flujo másico empleados en el Ejemplo 2 son:

$$F_1 - F_2 - F_3 - F_4 - F_5 = 0$$

$$F_9 - F_{15} = 0$$

$$F_2 - F_6 = 0$$

$$F_{10} - F_{16} = 0$$

$$F_3 - F_7 = 0$$

$$F_{11} - F_{17} = 0$$

$$F_4 - F_8 = 0$$

$$F_{12} - F_{18} = 0$$

$$F_5 - F_9 = 0$$

$$F_{13} - F_{19} = 0$$

$$F_7 - F_{10} - F_{11} = 0$$

$$F_{14} - F_{20} = 0$$

$$F_8 - F_{12} - F_{13} - F_{14} = 0$$

Para el método B de DEG en el Ejemplo 2, empleando un nivel de significancia  $\alpha = 0,05$ , se calcula un  $\beta$  nodal y un  $\beta$  de medida. En este sistema se presenta un número de nodos o balances, igual a 13, con un número total de 20 variables medidas. Con estos betas se lee en la tabla de distribución normal estándar un  $Z_{N_c} = 2,88$  y un  $Z_{M_c} = 3$ .

### Ejemplo 3

**Tabla 13.** Valores verdaderos del proceso en el Ejemplo 3.

Variable	Valor Verdadero [kg/h]	Variable	Valor Verdadero [kg/h]	Variable	Valor Verdadero [kg/h]	Variable	Valor Verdadero [kg/h]
1	4365,3	23	3745,17	45	1410,95	67	29,84
2	491,4	24	0,02	46	3,13	68	443,47
3	545,25	25	5312,32	47	1902,22	69	1023,91
4	988,22	26	515	48	269,89	70	1958,43
5	1372,87	27	623,73	49	1631,39	71	1389,99
6	2,81	28	1202,8	50	0,94	72	718,59
7	51,35	29	136,54	51	1686,75	73	1,79
8	913,4	30	4,98	52	452,84	74	83,74
9	6515,11	31	2829,26	53	1228,82	75	394,33
10	611,97	32	5634,73	54	5,1	76	191,54
11	438,33	33	215,82	55	638,97	77	3210,72
12	514,88	34	623,28	56	293,39	78	48,76
13	1522,89	35	1165,96	57	345,58	79	583,18
14	1122,54	36	144,62	58	1497,55	80	802,48
15	2219,98	37	3294,18	59	7,48	81	1776,29
16	76,07	38	57,21	60	308,69	82	3945,63
17	8,45	39	133,67	61	285,28	83	111,81
18	6441,22	40	2924,86	62	273,88	84	2022,57
19	749,4	41	436,2	63	1,28	85	1811,25
20	628,72	42	315,57	64	588,14		
21	1220,52	43	753,93	65	32,81		
22	97,38	44	5,08	66	3455,65		

Los balances de flujo másico empleados en el Ejemplo 3 son:

$$\begin{aligned}
 F_1 - F_2 - F_3 - F_4 - F_5 - F_6 - F_7 - F_8 &= 0 \\
 F_9 - F_{10} - F_{11} - F_{12} - F_{13} - F_{14} - F_{15} - F_{16} - F_{17} &= 0 \\
 F_{18} - F_{19} - F_{20} - F_{21} - F_{22} - F_{23} - F_{24} &= 0 \\
 F_{25} - F_{26} - F_{27} - F_{28} - F_{29} - F_{30} - F_{31} &= 0 \\
 F_{32} - F_{33} - F_{34} - F_{35} - F_{36} - F_{37} - F_{38} - F_{39} &= 0 \\
 F_{40} - F_{41} - F_{42} - F_{43} - F_{44} - F_{45} - F_{46} &= 0 \\
 F_{47} - F_{48} - F_{49} - F_{50} &= 0 \\
 F_{51} - F_{52} - F_{53} - F_{54} &= 0 \\
 F_{55} - F_{56} - F_{57} &= 0 \\
 F_{58} - F_{59} - F_{60} - F_{61} - F_{62} - F_{63} - F_{64} - F_{65} &= 0 \\
 F_{66} - F_{67} - F_{68} - F_{69} - F_{70} &= 0 \\
 F_{71} - F_{72} - F_{73} - F_{74} - F_{75} - F_{76} &= 0 \\
 F_{77} - F_{78} - F_{79} - F_{80} - F_{81} &= 0 \\
 F_{82} - F_{83} - F_{84} - F_{85} &= 0 \\
 F_2 + F_8 + F_{14} + F_{19} + F_{26} + F_{41} - F_{47} - F_{51} - F_{55} &= 0 \\
 F_{15} + F_{23} + F_{31} + F_{37} + F_{45} - F_{58} - F_{66} - F_{71} - F_{77} - F_{82} &= 0
 \end{aligned}$$

Para el método B de DEG en el Ejemplo 3, se emplea el mismo nivel de significancia del ejemplo 2, y se estiman los  $\beta$  (nodal y de medida). Aumentado el número de nodos y variables medidas: A 16 y 85 respectivamente, debido al tamaño de esta nueva red. Con los valores  $\beta$  se determina por tablas que el  $Z_{N_c} = 2,93$  y el  $Z_{M_c} = 3,33$  en este sistema.

## Ejemplo 4

**Tabla 14.** Valores verdaderos del proceso en el Ejemplo 4.

Variable	Valor Verdadero [kg/h]	Variable	Valor Verdadero [kg/h]	Variable	Valor Verdadero [kg/h]	Variable	Valor Verdadero [kg/h]
1	0,86	8	164,05	15	60	22	5,45
2	1	9	0,83	16	23,64	23	2,59
3	111,82	10	52,41	17	32,73	24	46,63
4	109,96	11	14,86	18	16,23	25	85,46
5	53,27	12	67,27	19	7,95	26	81,32
6	112,27	13	111,27	20	10,5	27	70,78
7	2,32	14	91,86	21	87,32	28	72,23

El sistema o red del ejemplo 4 presenta un número de 11 nodos y 28 variables medidas, con lo cual, usando el método B de DEG se estima unos nuevos valores para los  $\beta$  (nodal y de medida) con el mismo valor de significancia mencionado anteriormente, obteniendo con ellos un  $Z_{N_c} = 2,82$  y el  $Z_{M_c} = 3,09$  leídos en tablas.

Los balances de flujo másico empleados en el Ejemplo 4 son:

$$F_1 + F_2 + F_4 - F_3 = 0$$

$$F_{14} + F_{18} - F_7 - F_{20} - F_{21} - F_{19} = 0$$

$$F_7 + F_8 - F_5 - F_6 - F_9 = 0$$

$$F_{15} + F_{22} - F_{24} - F_{18} - F_{23} = 0$$

$$F_5 - F_1 - F_{10} = 0$$

$$F_{12} + F_{16} - F_{22} - F_{25} = 0$$

$$F_{10} + F_{11} - F_{12} = 0$$

$$F_{19} + F_{23} + F_{27} - F_{26} = 0$$

$$F_3 + F_{13} - F_{11} - F_{17} - F_{16} - F_{15} - F_{14} = 0$$

$$F_{26} + F_{20} + F_{28} - F_8 = 0$$

$$F_6 - F_2 - F_{13} = 0$$