UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

USO DE MODELOS DE SIMILITUD PARA DETECCION DE ANOMALIAS Y MODELOS DE PREDICCION EN PROCESOS DE CONCENTRACION DE MINERALES

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELECTRICO

MATÍAS ANDRÉS EUGENÍN CASTILLO

PROFESOR GUÍA

MARCOS ORCHARD CONCHA

MIEMBROS DE LA COMISIÓN

HECTOR AGUSTO ALEGRÍA

JORGE SILVA SANCHÉZ

SANTIAGO DE CHILE

2015

RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL ELECTRICISTA
POR: MATÍAS EUGENÍN C.
FECHA: POR DEFINIR
PROF. GUÍA: Dr. MARCOS ORCHARD CONCHA

"USO DE MODELOS DE SIMILITUD PARA DETECCION DE ANOMALIAS Y MODELOS DE PREDICCION DE VARIABLES EN PROCESOS DE CONCENTRACION DE MINERALES"

En la actualidad, la supervisión de procesos necesita ser cada vez más eficiente. Hoy en día se cuenta con múltiples sensores en cada proceso, los cuales entregan información de algún estado/variable del mismo. Al analizar esta información, es posible encontrar relaciones entre variables y puntos de operación del proceso. Lo anterior posibilita la construcción de modelos de procesos en base solo a información recaudada de los estados/variables. Esta herramienta de modelación es muy útil cuando no se conoce de manera completa, a nivel fenomenológico, el proceso a estudiar.

El presente Trabajo de Titulo está centrado en la realización de modelos para un molino SAG (Semi-Autógeno) en base a estructuras no-paramétricas de similitud. En primer lugar, se desarrolla un algoritmo para la generación de modelos de similitud usando un enfoque basado en los residuos (diferencia entre el valor real y el valor estimado). Posteriormente, se desarrolla una metodología para estimar variables usando modelos de similitud, la predicción de ellas y un posterior análisis de escenario usando un modelo de predicción.

Al momento de generar un modelo para estimar las variables controladas usando datos históricos reales de un molino SAG, se observa que el modelo creado cumple con los criterios de validación. Posteriormente, usando los mismos datos, se procede a crear un modelo de predicción, con el fin de generar un análisis de posibles futuros escenarios. Los resultados muestran que en el nuevo escenario propuesto, se obtiene un mejor desempeño energético del molino SAG estudiado.

Por otro lado, se genera un modelo para la detección de anomalías usando una base de datos con una anomalía identificada. Los resultados de esta modelación muestran que se detecta exitosamente la anomalía en la base de datos. Se propone a futuro, desarrollar una herramienta que sea capaz de realizar un pronóstico, estimando las variables independientes para una mejor predicción de las variables controladas.

AGRADECIMIENTOS

En primer lugar, agradezco a mis padres, Patricio y Julia, que siempre me apoyaron en mis estudios, gracias a ellos conozco el valor del esfuerzo, compromiso y responsabilidad. A mis hermanos, Pato y Dani, que me han acompañado todos estos años en Santiago, gracias por su compañía y apoyo. A la Ali, por apoyarme en todo momento.

A mi Profesor Guía Dr. Marcos Orchard, por siempre darse el tiempo para atender mis dudas, por sus rápidas respuestas y su gran sabiduría, gracias por ayudarme en este trabajo, siempre aprendiendo algo nuevo. Gracias a usted, me entusiasmé con el área de control debido a su gran docencia.

A mis amigos curicanos, que siempre me apoyaron a salir adelante. A mis amigos de la Universidad, Scholz, Nico y Pazos, por los grandes momentos que hemos vivido. En especial a Scholz, por la paciencia que tenía en enseñarme como resolver algunos problemas.

TABLA DE CONTENIDO

ÍNDICE DE FIGURAS

INDICE DE TABLAS

CAPÍTULO 1. INTRODUCCIÓN

1.1 Motivación

Chile es el país con mayor producción de cobre a nivel mundial, el cual, en el año 2014, aportó un 12% al PIB chileno de acuerdo al SONAMI (Sociedad Nacional de Minería). Durante la producción del cobre, se realizan múltiples procesos para extraer su máxima riqueza, donde existen una gran cantidad de equipos que deben realizar esta labor y por lo tanto, debe existir un manera de obtener el mayor rendimiento de estas.

Dada la constante evolución de la tecnología, hoy en día, existen nuevas herramientas para la supervisión de procesos que son más eficientes. Además, se cuenta con múltiples sensores en cada proceso, los cuales entregan información de algún estado/variable. Por otro lado, en la industria, se recolecta toda esta información, almacenándola en servidores y haciendo uso de ella.

En los datos recolectados se puede encontrar tendencias entre variables, ya sea correlación entre variables, como tal vez no. Además, al realizar un análisis de datos, es posible identificar si existió algún cambio temporal entre un par de variables debido a un desperfecto. En la actualidad, existe la posibilidad de construir modelos de procesos en base solo a información recaudada, en donde se encuentran tendencias entre variables mediante un análisis de datos.

1.2 Alcance

El presente Trabajo de Título está centrado en la realización de modelos para un molino SAG usando una modelación basado en similitudes (SBM). Se desarrollará una metodología para estimar variables, la predicción de ellas y un posterior análisis de escenario usando un modelo de predicción. Además, se realizaron pruebas experimentales de los modelos sobre un molino real ubicado en la mina de El Teniente. Se disponen de datos históricos de la operación del segundo molino SAG ubicado en la mina El Teniente, proporcionados por la empresa HONEYWELL Chile.

1.3 Objetivos

El objetivo principal de este trabajo es el desarrollo de modelos de similitud para el desempeño del proceso de molienda.

Para llevar a cabo este objetivo se plantearon una serie de objetivos específicos. Estos objetivos representan los hitos más importantes en el desarrollo del trabajo, estos son:

- Diseño de algoritmo de generación de modelo de similitud usando observaciones representativas del proceso a modelar.
- Estudio del molino SAG: su funcionamiento y variables relevantes.
- Metodología para estimar variables controladas usando un modelo SBM.
- Metodología para detectar anomalías usando modelos SBM.
- Metodología para predecir variables controladas usando modelo de similitud. Con ello, realizar una metodología para analizar posibles escenarios.
- Diseño de herramienta de predicción y análisis de escenario para puesta en línea.

1.4 Indicación sobre confidencialidad

El Trabajo de Título presente se desarrolla dentro de un proyecto de la Empresa HONEYWELL Chile S.A. Debido a la presencia de contratos de confidencialidad de la empresa, existe información relevante referida principalmente a los algoritmos desarrollados y las bases de datos utilizadas que no será entregada de forma íntegra y detallada. Sin embargo, se deja constancia en este documento que el Profesor Guía de la Memoria de Título está en conocimiento de toda la información no presentada.

1.5 Estructura general

El Trabajo de Título está constituido por 5 capítulos. En los primeros dos capítulos se realiza una introducción, se presenta la revisión bibliográfica y el estado del arte sobre las herramientas de detección de anomalías actuales. Se detallan los distintos enfoques posibles para abordar el problema de detección y se presenta un marco teórico sobre la detección de anomalías basada en observadores;

modelación no paramétrica y análisis estadístico multivariable, conceptos utilizados en el desarrollo de este trabajo.

El Capítulo 3 se diseña y describe el algoritmo para la creación de un modelo de similitudes para la estimación y predicción de variables. En primer lugar, se describen las técnicas estadísticas utilizadas en el algoritmo. Posteriormente, se detalla el algoritmo paso a paso, describiendo cada bloque para la creación del modelo SBM. Finalmente se detalla una metodología para la estimación, detección de anomalías, predicción y análisis de escenarios usando modelos de similitud.

El Capítulo 4 presenta la descripción de un molino SAG, el cual es utilizado para probar y validar la herramienta descrita en el capítulo anterior. Además, se entregan los resultados obtenidos en dicho proceso, los cuales son enseñados de forma separada para cada una de las etapas que constituyen la herramienta. Análisis de tales resultados son también presentados en este capítulo.

Finalmente, en el Capítulo 5 se presenta las conclusiones finales del Trabajo de Título y proponiendo trabajo a futuro.

CAPÍTULO 2. REVISIÓN BIBLIOGRÁFICA Y ESTADO DEL ARTE

El presente capítulo tiene por objetivo ubicar al lector en el entorno en el cual se desarrolla este trabajo de título, entregando los antecedentes previos y necesarios para su contextualización.

En primer lugar, en la Sección 2.1, se indican los conceptos generales de un sistema para detección de fallas, conceptos generales de los procesos de concentración de minerales y una breve descripción de palabras recurrentes en Asset Manager. En la Sección 2.2 se describe en términos generales los métodos para evaluar desempeño y detectar anomalías. En la Sección 2.3 se describe el método de utilización de los residuos de modelos para detectar fallas. Finalmente, la Sección 2.4 hace referencia a las técnicas y algoritmos utilizados en los capítulos siguientes para diseñar e implementar la generación de modelos.

2.1 Conceptos generales

Los conceptos descritos a continuación han sido obtenidos desde [11], [16], [23] y [25]. Estos conceptos corresponden a la terminología que históricamente se ha utilizado.

- Modos de operación: son aquellas características que definen la dinámica de un sistema. Un modo de operación normal es cuando las características son las deseadas, y un modo de operación en falla es cuando las características involucran comportamientos inesperados.

- Síntoma (*symptom*): cambios inusuales, en comparación a un comportamiento aceptable o nominal, de las características o parámetros de un sistema. Los efectos observados en una falla son síntomas.

- Anomalía: datos del proceso que escapan del modo de operación normal. Las observaciones discordantes, valores atípicos, *outliers* son ejemplos de anomalías.

- Falla (*fault*): desviaciones no permitidas en al menos una característica o parámetro del sistema en un comportamiento aceptable o normal.

- Perturbación: entrada desconocida y descontrolada que actúa en el sistema, desviándolo de su estado actual.

- Evaluación de desempeño (*performance evaluation*): consiste en determinar el estado de los activos, ya sean equipos o procesos, usando indicadores de condiciones.

- Indicadores de condiciones (*condition indicators*): consisten en reglas (condiciones) que se imponen a las variables de interés para mostrar el estado de los activos.

- Diagnóstico de falla: esquema de monitoreo que consiste en determinar el tipo de falla en un sistema con el mayor detalle posible: detectarla al momento que ocurre (detección de falla), encontrar su origen aislando las componentes del sistema cuando este no está en su operación nominal (aislamiento de fallas) y finalmente estimar el tamaño y tipo o naturaleza de la falla (identificación de fallas).

- Detección de fallas (*fault detection*): tiene por objetivo encontrar patrones en el sistema que indiquen que este no está en su operación normal, tan pronto como sea posible.

- Aislamiento de fallas (*fault isolation*): consiste en encontrar la causa de la falla detectada, aislando componentes del sistema cuando no está en su operación nominal.

- Identificación de fallas (*fault identification*): tiene como objetivo estimar el tamaño y tipo de la falla encontrada.

- Residuos: diferencias entre salidas estimadas obtenidas a través de un modelo del proceso y las salidas con datos reales. Las características de los residuos pueden determinar si el proceso se encuentra en falla o no.

- Observación: conjunto de mediciones tanto de las variables manipuladas como de las variables controladas.

- Proceso de molienda: consiste en utilizar grandes equipos giratorios o molinos en forma cilíndrica con la finalidad de reducir el tamaño del mineral, de dos formas diferentes:

molienda convencional o molienda semi-autógena. En esta etapa, al material mineralizado se le agregan agua en cantidades suficientes para formar un fluido lechoso y los reactivos necesarios para realizar el proceso siguiente que es la flotación.

- Molino SAG: es un equipo que recibe el mineral directamente desde el chancador primario y se mezcla con agua y cal. Este material es reducido de tamaño gracias a la acción del mismo material mineralizado presente en partículas de variados tamaños y por la acción de numerosas bolas de acero. Estas bolas son lanzadas en caída libre cuando el molino gira, logrando un efecto conjunto de chancado y molienda más efectivo y con menor consumo eléctrico.

2.2 Evaluación de desempeño y alerta temprana de anomalías

2.2.1 Métodos de modelos del proceso

Los métodos basados en modelos de proceso para detección de falla encuentran relaciones matemáticas entre señales de entrada $U(k)$ y señales de salida $Y(k)$, con k representando un índice temporal, para extraer información y así encontrar cambios inesperados que podrían ser causados por fallas, tal como muestra el esquema general de la Figura 2.1. Estos métodos extraen características como parámetros θ, variables de estado x o residuos r, las que son comparadas con sus valores nominales para detectar eventuales cambios y fallas.

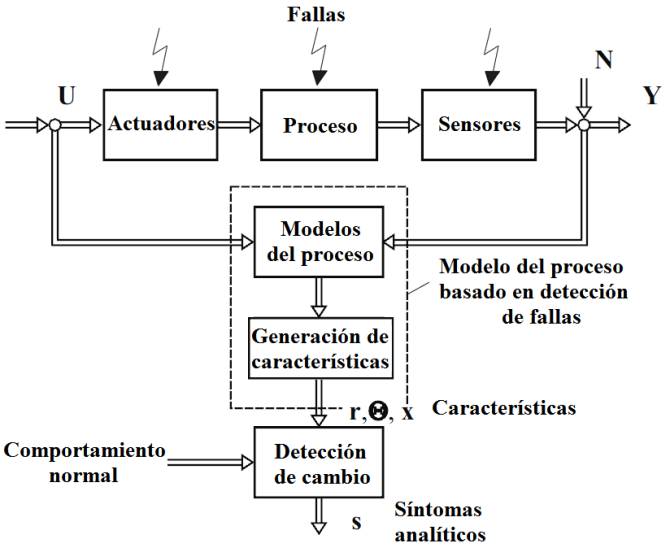


Figura 2.1 : Esquema de los métodos de detección de falla con modelos del proceso. [11]

2.2.2 Método basado en modelo de señales

Los métodos basados en modelos de señales pueden ser aplicados cuando se presentan señales que muestran oscilaciones ya sea de naturaleza armónica, estocástica o ambas, y se producen anomalías en los actuadores que originan cambios en estas mediciones.

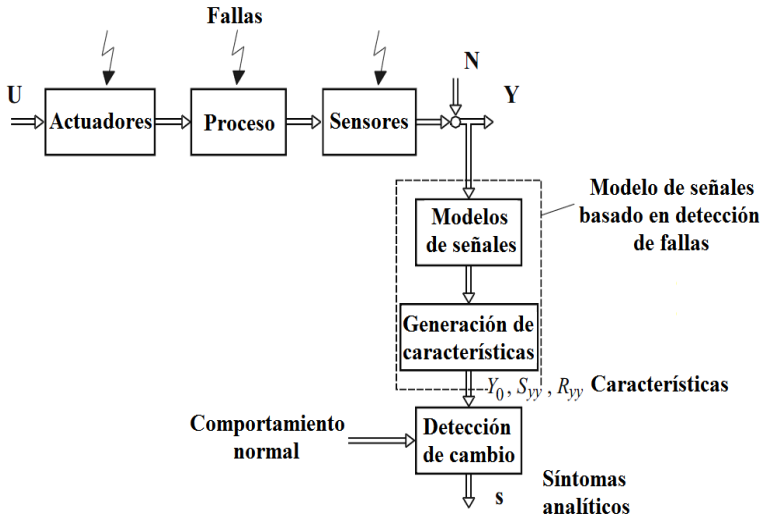


Figura 2.2 : Esquema de los métodos de detección de falla con modelos de señales. [25]

Tal como se observa en la Figura 2.2, los métodos de detección de fallas basados en modelos de señales se clasifican en tres tipos de acuerdo a las características de las señales: periódicas, no estacionarias o estocásticas. En el caso en que las señales sean periódicas se utilizan técnicas como el análisis de Fourier, funciones de correlación, la transformada rápida de Fourier (FFT) o la estimación del espectro de máxima entropía. La transformada *wavelet* es útil cuando las señales no son estacionarias. Finalmente, en el caso en que las señales son estocásticas se utiliza el análisis de correlación a través de la función de autocorrelación, el análisis del espectro de densidad obtenido como la transformada de Fourier de la función de autocovarianza o la estimación de parámetros de señales con modelos de tipo ARMA (*Autoregressive Moving Average Model*).

2.2.3 Método de análisis multivariado

Las técnicas basadas en el análisis de datos lineales tales como PCA (*Principal Component Analysis*) y PLS (*Partial Least Squares*) o análisis en datos no Gaussianos como ICA (*Independant Component Analysis*), que son métodos de trasformación que reducen el número de dimensiones del sistema.

Entre las opciones para aplicar estos métodos se encuentran:

- Detección de cambios utilizando las proyecciones que entrega PCA.
- Detección de cambios en las variables del espacio original, obtenidas a partir de las proyecciones.
- Análisis de residuos entre la variable original y la variable obtenida a partir de las proyecciones.

2.3 Método de los residuos de modelos

Un esquema de detección y aislamiento de fallas, utilizando análisis en los residuos (Figura 2.3), contempla principalmente dos etapas: obtención de residuos y determinación de anomalía. El esquema comienza con una selección de las características del proceso que serán modeladas con alguna de las técnicas de modelación ya vistas. A continuación, las salidas estimadas de estos modelos y_{f_est} se comparan con las salidas reales del proceso. A través de esta comparación se obtienen las variables residuales. En el caso ideal en que estos residuos son cero.

$$\boldsymbol{r = y_f - y_{f_{est}}} \qquad \textbf{(2.1)}$$

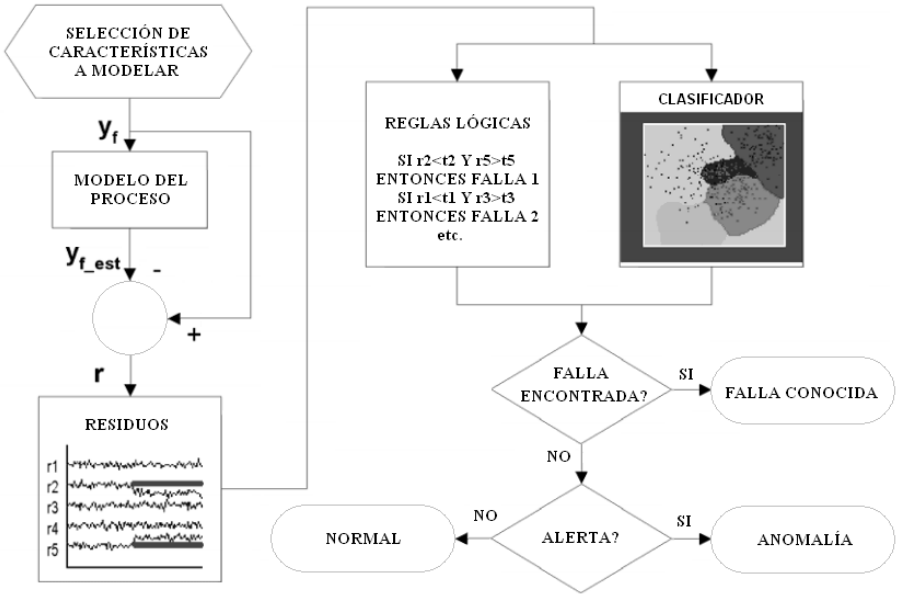


Figura 2.3 : Esquema de un sistema FDI utilizando método de los residuos. [28]

De acuerdo a los resultados que entregan estos métodos (reglas lógicas o clasificador), se podrá concluir si corresponde a una falla conocida, a una anomalía desconocida hasta el momento, o se trata simplemente de una falsa alarma y el proceso sigue estando en modo nominal.

2.4 Herramientas utilizadas

2.4.1 Error cuadrático medio

El error cuadrático medio (ECM) de un estimador de una variable es un concepto ampliamente utilizado para medir la diferencia entre dicha estimación $\hat{X}$ y el valor real X. El ECM es una función que se calcula como el valor esperado del error cuadrático.

$$\boldsymbol{ECM = E\left[\left(X - \hat{X}\right)^2\right]} \quad \textbf{(2.2)}$$

El error cuadrático medio se calcula con un número M de variables reales X_i y estimadas $\hat{X}_i$, para una observación fija k. El promedio sobre todas las observaciones es utilizado como criterio para determinar si un modelo es mejor que otro o no.

$$\boldsymbol{ECM(k) = \frac{1}{M}\sum_{i=1}^{M}\left(X_i(k) - \hat{X}_i(k)\right)^2} \quad \textbf{(2.3)}$$

2.4.2 Error porcentual relativo

El error porcentual relativo (EPR) de un estimador de una variable es un concepto muy utilizado en la industria para medir la diferencia entre la estimación $\hat{X}$ y el valor real X. El error porcentual se calcula de la siguiente manera.

$$\boldsymbol{EPR = \frac{X - \hat{X}}{X} * 100 = 100 * \left(1 - \frac{\hat{X}}{X}\right)} \quad \textbf{(2.4)}$$

Este error nos entrega una noción más intuitiva del error de estimación del proceso, ya que entrega un porcentaje de desviación entre el valor real y el valor estimado con respecto al valor real. Además, también, es utilizado como criterio para evaluar el desempeño de un modelo.

2.4.3 Raíz de error cuadrático medio (RMSE)

La raíz del error cuadrático medio representa la desviación de los residuos. Este indicador es un buen indicador de la precisión del modelo.

$$\boldsymbol{RMSE = \sqrt{MSE} = \sqrt{E\left[\left(X - \widehat{X}\right)^2\right]} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(X_i(k) - \widehat{X}_i(k)\right)^2}} \quad \textbf{(2.5)}$$

2.4.4 Coeficiente de variación (CV)

El coeficiente de variación es un concepto utilizado en ingeniería para medir de manera estandarizada la dispersión de una variable. Esto nos ayuda a entender de manera más intuitiva que tanto varía la variable estudiada.

$$\boldsymbol{CV = \frac{\sigma}{\mu} = \frac{RMSE}{\bar{x}}} \quad \textbf{(2.6)}$$

Esta fórmula nos indica que tan desviado se encuentra la variable con respecto a su promedio. Entonces si el coeficiente de variación es alto, implica que los datos se encuentran muy dispersos, y viceversa.

2.4.5 Agrupación por k-medias (*K-means Clustering*)

Es un método que agrupa un conjunto de datos de n elementos en k grupos y que se basa en clasificar de acuerdo a la menor distancia que tiene un elemento a los k centros de los grupos. El algoritmo es el siguiente:

1. Inicializar k medias. Estos serán inicialmente los centros de los grupos.
2. Asignar cada elemento x_f a un grupo S_i de acuerdo a la media u_i más cercana.
3. Recalcular la media u_i como:

$$\boldsymbol{u_i = \frac{1}{|S_i|}\sum_{x_f \in S_i} x_j} \quad \textbf{(2.7)}$$

4. Continuar hasta que no hayan cambios en las medias.
5. Retomar las medias u_i.
6. Fin del algoritmo.

2.4.6 Mínimos cuadrados parciales (PLS)

Es una técnica de modelación [27] cuyo propósito es explicar una o más variables dependientes (Y) en función de un número de variables explicativas o independientes (X). PLS consiste en un gran número de variables independientes que entienden de alguna forma, los efectos dominantes producidos por cambios en la matriz de salida o variables dependientes.

Entonces, sea el sistema descrito como:

$$\boldsymbol{Y = f(X)} \quad \textbf{(2.8)}$$

Donde X son las variables independientes e Y son las variables dependientes. Además, las variables están centradas en X_o, Y_o. Se asume que A es igual al número de componentes relevantes para la predicción. Luego, se pueden definir los pesos w_c (espacio de máxima covarianza), como:

$$\bar{\boldsymbol{w}}_{\boldsymbol{c}} = \boldsymbol{X}_{\boldsymbol{c-1}}^{\boldsymbol{T}}\boldsymbol{Y}_{\boldsymbol{c-1}}; \boldsymbol{c} = \boldsymbol{1, 2, 3, \dots, A} \quad \textbf{(2.9)}$$

$$\boldsymbol{w_c} = \frac{\bar{\boldsymbol{w}}_{\boldsymbol{c}}}{||\bar{\boldsymbol{w}}_{\boldsymbol{c}}||} \quad \textbf{(2.10)}$$

Donde c es el número de la componente relevante para la predicción. Con ello, se puede definir el vector puntaje S_c (score) de la componente c de la siguiente manera:

$$\boldsymbol{S_c = X_{c-1} w_c} \quad \textbf{(2.11)}$$

Finalmente, el vector de carga de la c componente de las variables independientes p_c (*X-loading*) y para las variables dependientes q_c (Y-loadings) se puede escribir como:

$$\boldsymbol{p_c} = \frac{\boldsymbol{X_{c-1}^T S_c}}{\boldsymbol{S_c^T S_c}} \quad \textbf{(2.12)}$$

$$\boldsymbol{q_c} = \frac{\boldsymbol{Y_{c-1}^T S_c}}{\boldsymbol{S_c^T S_c}} \quad \textbf{(2.13)}$$

Luego la actualización para la siguiente componente, queda descrita como:

$$\boldsymbol{X_c = X_{c-1} - S_c p_c^T} \quad \textbf{(2.14)}$$

$$\boldsymbol{Y_c = Y_{c-1} - S_c q_c} \quad \textbf{(2.15)}$$

2.4.7 Análisis de componentes principales (PCA)

El análisis de componentes principales o PCA es una técnica lineal de reducción de dimensiones que captura la máxima variabilidad de los datos, obteniendo los llamados *vectores de carga* [20]. Dada

una matriz de datos $X \in \Re^{NxM}$, con el número N de observaciones y M el número de variables, PCA resuelve el problema de optimización:

$$\max_{v \neq 0} \frac{v^T X^T X v}{v^T v} \quad \textbf{(2.16)}$$

Donde $v \in \Re^M$.

Para obtener la solución de este problema, se debe obtener de la matriz de covarianza , la matriz V que contiene en sus columnas los vectores propios (o vectores de carga) asociados a la matriz diagonal Λ que contiene los valores propios λ_i ordenados de mayor a menor ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M \geq 0$) siendo todos mayores o iguales a cero. Esta matriz se define como:

$$\boldsymbol{S = \frac{1}{n-1} X^T X = V \Lambda V^T} \quad \textbf{(2.17)}$$

Con la finalidad de reducir el ruido que pueden representar las componentes asociadas a los valores propios menores, se selecciona una cantidad a de los primeros valores y vectores propios. Así, sea $P \in \Re^{Mxa}$, con $a \leq M$, la matriz que contiene las primeras a columnas de V. Entonces, las proyecciones de en un espacio de menor dimensión están contenidas en la matriz:

$$\boldsymbol{T = XP} \quad \textbf{(2.18)}$$

La proyección de T en el espacio original de dimensión M es

$$\boldsymbol{\widehat{X} = TP^T} \quad \textbf{(2.19)}$$

La matriz residual E, que captura las variaciones en las observaciones asociadas a las componentes principales con valores propios que no fueron considerados en la matriz P , se calcula como:

$$\boldsymbol{E = X - \widehat{X}} \quad \textbf{(2.20)}$$

Las columnas t_i de la matriz T en el conjunto de entrenamiento, cumplen con las siguientes propiedades:

- $Var(t_1) \geq \cdots \geq Var(t_a)$
- $media(t_i) = 0;\ \forall i$
- $t_i t_k^T = 0;\ \forall i \neq k$
- No existe otra expansión ortogonal de a componentes que capture mayor variación en los datos.

Cuando se incorpora un nuevo vector fila x_n $(1xM)$ a la base de datos, y se desean obtener las primeras a componentes principales, se debe simplemente utilizar la formula descrita abajo. Es muy importante notar que el vector x_n debe ser normalizado utilizando la media y desviación estándar de los datos originales, sin incorporar la información que agrega este vector.

$$\boldsymbol{t_n = x_n P} \quad \textbf{(2.21)}$$

2.4.8 Test estadístico de Hotelling

Sea $x \in \Re^M$. El estadistico de Hotelling T^2 se calcula como:

$$\boldsymbol{T^2 = z^T z} \quad \textbf{(2.22)}$$

Donde z es:

$$\boldsymbol{z = \Lambda^{-\frac{1}{2}} V^T x} \quad \textbf{(2.23)}$$

Λ y V son obtenidos a partir de la matriz de covarianza de x. Donde Λ es la matriz que contiene los valores propios λ_i ordenados de mayor a menor $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M \geq 0)$ siendo todos mayores o iguales a cero y V es la matriz con los vectores propios.

Entonces,

$$\boldsymbol{T^2 = z^T z = x^T V \Lambda^{-1} V^T x} \quad \textbf{(2.24)}$$

Luego se define el indicador de Hotelling para el conjunto de entrenamiento como:

$$\boldsymbol{T_\alpha^2 = \frac{(n-1)^2\left(\frac{m}{n-m-1}\right) F_\alpha(m, n-m-1)}{n\left(1+\left(\frac{m}{n-m-1}\right) F_\alpha(m, n-m-1)\right)}} \quad \textbf{(2.25)}$$

El indicador de Hotelling es utilizado en detección para verificar si el error de estimación se encuentra dentro de una región limitada por un umbral, y así determinar si es aceptable o no. Donde $F_\alpha(g, h)$ es el punto crítico superior $(100 * \alpha)\%$ de la distribución F de Fisher de g y h grados de libertad. Por otro lado, el indicador de Hotelling para el conjunto de no entrenamiento es:

$$\boldsymbol{T_\alpha^2 = \frac{m(n-1)(n+1)}{n(n-m)} F_\alpha(m, n-m)} \quad \textbf{(2.26)}$$

2.4.9 SBM (*Similarity Based Modelling*)

Técnica de modelación no paramétrica, y por ende, no requiere a priori conocimiento de la estructura del sistema a modelar puesto que su implementación está basada en identificar similitudes y relaciones entre las variables de un conjunto de observaciones dado. Sea el sistema estático siguiente:

$$\boldsymbol{y = f(x)} \quad \textbf{(2.27)}$$

Donde $x \epsilon \Re^m$ e $y \epsilon \Re^p$ son las variables de entrada y de salida del sistema, respectivamente, y $f(\cdot)$ es una función desconocida.

A continuación, se definen las matrices de entrenamiento D_i y D_o a través de observaciones de variables de entrada y salida, respectivamente:

$$\boldsymbol{D_i = [x_1, \dots, x_n] \epsilon \Re^{mxn}} \quad \textbf{(2.28)}$$

$$\boldsymbol{D_o = [y_1, \dots, y_n] \epsilon \Re^{pxn}} \quad \textbf{(2.29)}$$

Donde $y_i = f(x_i)$. Los pares $[x_i, y_i]_{i=1,\dots,n}$ deben ser representativos de los puntos de operación del proceso que se desea modelar.

Entonces dado un vector x^*, SBM encuentra una estimación $\hat{y}^*$ de $y^* = f(x^*)$ por medio de una combinación lineal de las columnas de D_o.

$$\boldsymbol{\hat{y}^* = D_o w} \quad \textbf{(2.30)}$$

Donde *w* está definido como:

$$\boldsymbol{w = \frac{\hat{w}}{1^T \hat{w}}} \quad \textbf{(2.31)}$$

$$\boldsymbol{\hat{w} = \left(D_i^T \Delta D_i\right)^{-1} (D_i \Delta x^*)} \quad \textbf{(2.32)}$$

Donde Δ es el operador de similitud. Para dos elementos $A, B \epsilon \mathbb{R}^n$, $A \Delta B \epsilon \Re^+$ debe ser simétrica, alcanzar su máximo en $A = B$ y ser monótonamente decreciente con $\|A - B\|$. El operador de similitud que mejor captura la variabilidad de los datos es el operador triangular saturado definido como:

$$\boldsymbol{A \Delta B = \begin{cases} d - \|A - B\| & \|A - B\| > d + \varepsilon \\ \varepsilon & \|A - B\| \leq d + \varepsilon \end{cases}} \quad \textbf{(2.33)}$$

Donde $\varepsilon > 0$ es un numero pequeño cercano a cero para asegurar que $A \Delta B > 0$, y $d > 0$ es la distancia de Kernel. A pesar que SBM asume que el sistema es estático, es posible adaptarlo para

sistemas dinámicos discretos si se dispone de una secuencia temporal de observaciones. En este caso, el problema puede ser abordado desde dos perspectivas: abandonar las propiedades del sistema dinámico y abordarlo como si fuera estático, o incorporar regresores y considerarlos como una entrada o una salida.

CAPÍTULO 3. IMPLEMENTACIÓN DE HERRAMIENTAS PARA LA ESTIMACIÓN USANDO SBM.

En sistemas complejos no se suele tener toda la fenomenología del proceso, dado que está sujeto a dinámicas no modeladas, fenomenologías desconocidas o parámetros no estimados, etc. Sin embargo, en la práctica, se cuenta con una gran cantidad de datos, con los cuales se pueden encontrar relaciones o patrones y dar a cabo con una buena representación del sistema a modelar.

Durante este capítulo se entregara una descripción detallada de la implementación de una herramienta para la estimación de variables basada en residuos de procesos usando modelación SBM. En primer lugar, Sección 3.1, se presentaran técnicas estadísticas que ayudan al entendimiento de relaciones entre variables. Posteriormente en la Sección 3.2, el algoritmo utilizado para la modelación SBM. Finalmente en la Sección 3.3, las metodologías para la estimación de variables, detección de anomalías, predicción de variables y análisis de escenarios usando modelos SBM.

3.1 Técnicas estadísticas

En la modelación no paramétrica, es necesario obtener una base de datos rica en información, es decir, que la base de datos contenga los puntos de operación del proceso y, además, que contenga perturbaciones en torno a esos puntos de operación. Esto ayudará a entender de mejor manera el proceso ya que representa un mayor dinamismo.

Las técnicas estadísticas utilizadas en este Trabajo de Título son análisis de componentes principales (PCA), mínimos cuadrados parciales (PLS) y test de Hotelling. En primer lugar, el análisis de componentes principales nos permite observar el proceso desde otra perspectiva y así identificar de manera más sencilla los puntos operacionales del proceso. Esto lo logra, creando un nuevo sistema de coordenadas usando una transformación lineal con los datos originales. En este sistema, la primera componente principal (PCA_1) es aquella que representa la mayor variabilidad del proceso, luego la segunda componente principal (PCA_2) representa la segunda mayor variabilidad y así sucesivamente.

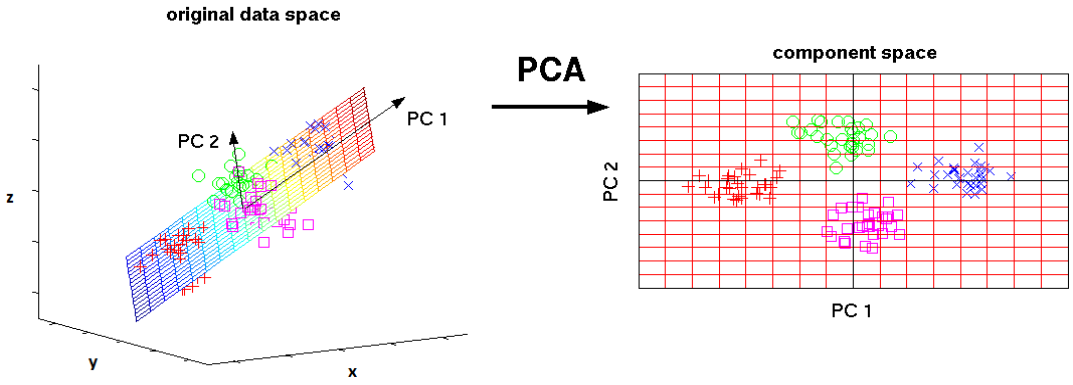


Figura 3.1 : Análisis de componentes principales.

Posteriormente al análisis de componentes principales, existe una reducción o selección de variables realizada por el algoritmo de mínimos cuadrados parciales. Esta técnica permite obtener los vectores de carga para cada variable presente en el proceso, ellos representan que tan significativas son las variables para el modelo a crear.

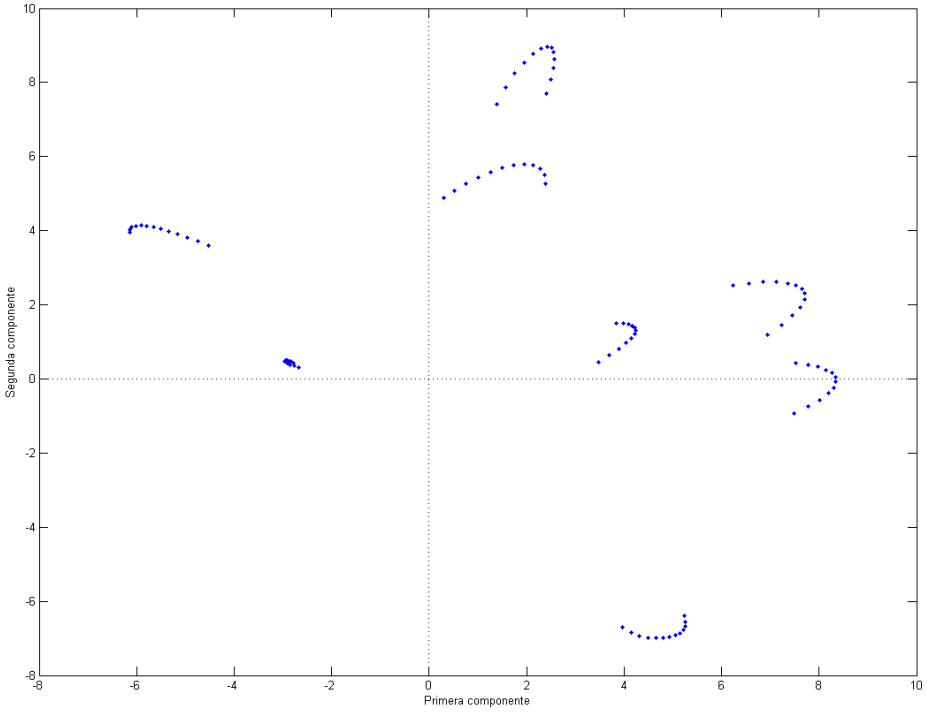


Figura 3.2 : Vectores de carga.

Con ello, se puede crear un umbral para excluir variables poco relevantes (aquellas que se encuentren cercanas al origen). En la Figura 3.3, se puede observar el umbral siendo un círculo de color verde, y los puntos al interior son las variables a eliminar.

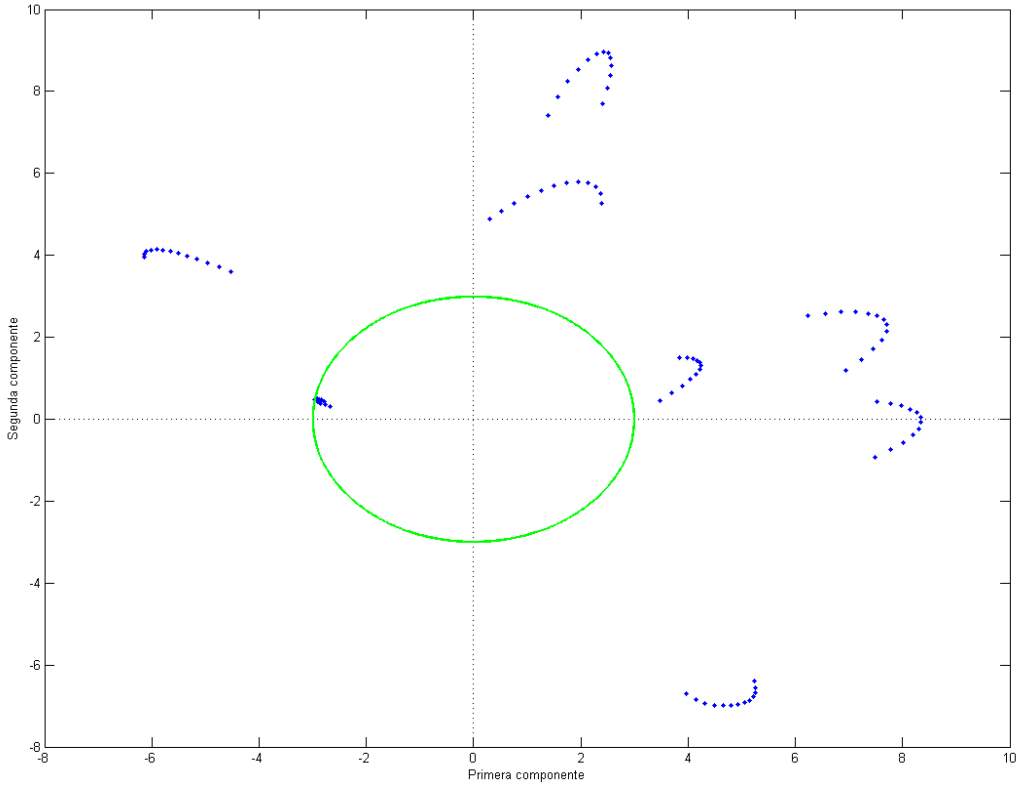


Figura 3.3 : Umbral para exclusión de variables.

Una vez creado el primer modelo SBM, se extrae la matriz de errores de estimación y se integran nuevas observaciones usando el test de Hotelling. Este test transforma los datos usando los valores y vectores propios de la covarianza del error. Luego, compara los nuevos datos con el umbral de Hotelling y se dicta que datos serán agregados a la matriz de entrenamiento (aquellos que superen el umbral, son datos “malos”).

3.2 Algoritmo para modelación SBM

Se presenta el diagrama de bloques de las etapas que son necesarias para crear un modelo de similitudes. Este algoritmo fue desarrollado en el software MATLAB, por lo tanto, el lenguaje de programación utilizado es el que corresponde para aquel software.

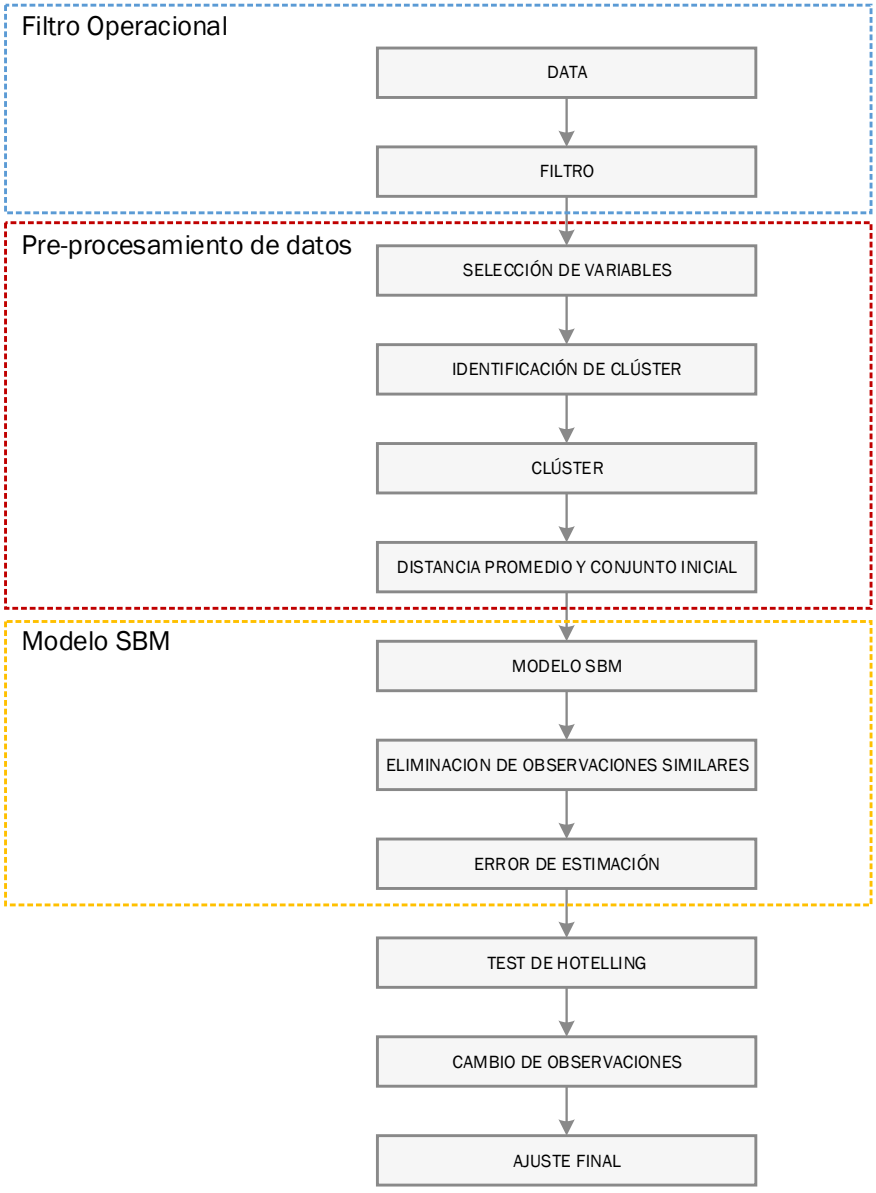


Figura 3.4 : Algoritmo implementado.

En la Figura 3.4, se aprecia el algoritmo diseñado. Los contenedores corresponden a una manera simplificada del algoritmo, por ejemplo: la etapa Filtro Operacional contiene las etapas de Data y Filtro. A continuación se describe cada componente del algoritmo detallado.

3.2.1 Filtro Operacional

3.2.1.1 Data

En primer lugar, se debe disponer de una base de datos con información relevante de la operación del proceso a modelar. Esta base de datos debe tener, si es posible, todos los puntos de operación del proceso, para que de esta manera, el futuro modelo de similitudes estime de manera correcta los valores del proceso.

3.2.1.2 Filtro

En la primera etapa del algoritmo, se eliminan todos los datos tipo NaN (*Not a Number*), esto se puede realizar previamente en el software Microsoft Excel o en MATLAB utilizando la función *isnan* para identificar los valores NaN. Además, se realiza un filtrado operacional de acuerdo a los límites fijados para cada variable por los operarios de la planta.

Posteriormente, se deben eliminar las magnitudes de las variables (x_i). Por lo tanto se normalizaran los datos de la siguiente manera:

$$\overline{\boldsymbol{x}}_{\boldsymbol{i}} = \frac{\boldsymbol{x_i - \min(x_i)}}{\boldsymbol{\max(x_i) - \min(x_i)}} \quad \textbf{(3.1)}$$

Esto garantiza la eliminación de las unidades de ingeniería de las variables, y su valor será entre cero y uno. Se realiza esta normalización para eliminar las futuras preferencia de pesos a variables con magnitudes más grandes.

3.2.2 Pre-procesamiento de datos

3.2.2.1 Selección de variables

En esta etapa, se eliminan variables que son insignificantes para el modelo del proceso. Esto se decide usando el algoritmo de mínimos cuadrados parciales (PLS). De aquí, se estiman los vectores de carga de las primeras componentes y se eliminan las variables que se encuentran cercanas a cero.

3.2.2.2 Identificación de clúster

Los datos suelen tener ciertos patrones y/o grupos a seguir, estos se denominan clústers. Para poder observarlos y así poder distinguir los diferentes puntos operacionales, se utilizara análisis de componentes principales PCA.

3.2.2.3 Clústers (Grupos)

En esta etapa, se estiman los centros de los clústers (o grupos) previamente encontrados con PCA, vale recordar que cada clúster corresponde a un tipo de operación del proceso. Estos centros se pueden calcular utilizando el algoritmo de k-medias.

3.2.2.4 Distancia promedio y conjunto inicial de observaciones

Una vez obtenidos los clústers y sus centros, se estima la distancia promedio entre las observaciones. Esta distancia se calcula entre pares de observaciones, es decir, la distancia de $(2,1), (3,1), \dots, (n,1), (3,2), \dots, (n, n-1)$. Esto puede ser expresado de la siguiente manera.

$$\boldsymbol{d_{promedio} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} [obs(i) - obs(j)]} \quad \textbf{(3.2)}$$

Con la distancia promedio calculada, podemos calcular la distancia de kernel, la cual es utilizada para la operación de similitud. Esta distancia tiene relación con la distancia promedio entre las observaciones, la cual está caracterizada como:

$$\boldsymbol{d_{Kernel} = \frac{d_{promedio}}{\beta};\ \ \beta = 1, 2, 3, \dots, 10} \quad \textbf{(3.3)}$$

El factor β será determinante para la operación de similitud, ya que un valor pequeño ($\beta = 1$) implicara que la d_{Kernel} será igual a $d_{promedio}$, entonces el modelo podría estimar que una observación es símil o parecida a otra cuando en realidad no lo son. Por el otro lado, si β es grande (cercano a 10), implicara que la d_{Kernel} es pequeña en comparación a $d_{promedio}$, entonces el modelo puede no reconocer o no encontrará similitud entre las observaciones, y además, las matrices de entrenamiento serán grandes, consumiendo bastante recurso computacional.

Por otro lado, con la distancia de kernel obtenida, se puede crear el conjunto inicial para nuestro modelo SBM. Este conjunto se estima usando los centros de los clusters previamente calculados y una cierta cantidad de datos ($X\%$) cercanos a cada centro.

$$\boldsymbol{Conjunto} = \{\boldsymbol{C_1}, \boldsymbol{C_2}, \dots, \boldsymbol{C_n}\}$$

$$\boldsymbol{C_i} = \boldsymbol{c_i} * \frac{\boldsymbol{X*N}}{\boldsymbol{100}}; \qquad \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{c_i} \tag{3.4}$$

Donde N corresponde a la cantidad total de datos, c_i corresponde al porcentaje de datos perteneciente al clúster i y C_i corresponde a la cantidad de datos a agregar al conjunto por parte del clúster i.

3.2.3 Modelo SBM

3.2.3.1 Modelo SBM

En esta etapa, se crea el modelo SBM y se evalúa usando la función de similitud con los datos pertenecientes a la matriz de entrenamiento. La función de similitud entrega un valor que corresponde a la similitud entre la observación candidata y las observaciones de entrenamiento. Posteriormente, se obtiene el vector de pesos y se pondera con la matriz de entrenamiento de salida, obteniendo la salida estimada. Finalmente, se escriben los errores de estimación para cada observación.

3.2.3.2 Eliminación de observaciones similares

Al ver el vector de pesos máximos en el modelo, se puede observar que son mayores a 1 puesto que la matriz inversa de la similitud $\left(D_i^T \Delta D_i\right)^{-1}$ tiene valores negativos, haciendo, por consiguiente, que algunos pesos sean negativos y así, al normalizar, otorga un mayor peso a algunas observaciones en particular.

Entonces, se plantea la opción de eliminar las observaciones que se parezcan a otras, de esta manera cada observación es única, es decir, la operación similitud entre x_j y x_k es cercana a cero ($x_j \Delta x_k \approx 0$). Al eliminar dichas observaciones, la matriz inversa de la similitud es una matriz diagonal con valor igual al inverso de la distancia de kernel y de esta manera, se asegura que los pesos máximos no superen el valor 1.

En primer lugar para eliminar observaciones parecidas, se realiza el cálculo de la matriz $D_i^T \Delta D_i$, posteriormente, se determina que si el valor de cada elemento supera al 1% del valor de la distancia de kernel, entonces se elimina una de las dos observaciones involucradas.

3.2.3.3 Errores de estimación

Al estimar las salidas del proceso usando el modelo SBM creado, se tendrán errores entre el valor real del proceso y el valor estimado del modelo, esto se denominaran como errores de estimación. Con ellos, se puede construir una matriz de error de estimación de la siguiente manera:

$$\boldsymbol{E} = \begin{bmatrix} \boldsymbol{E(1,1)} & \cdots & \boldsymbol{E(1,m)} \\ \vdots & \boldsymbol{E(j,k)} & \vdots \\ \boldsymbol{E(n,1)} & \cdots & \boldsymbol{E(n,m)} \end{bmatrix} \quad \textbf{(3.5)}$$

$$\boldsymbol{E(j,k) = y_{real}(j,k) - y_{est}(j,k)} \quad \textbf{(3.6)}$$

Donde E es la matriz de errores de estimación, $E(j,k)$ es el valor del error de estimación de la observación j perteneciente a la variable k.

3.2.4 Test de Hotelling

Con la matriz de errores obtenida, se procede a realizar un test de Hotelling. Este test crea un estadístico T^2, el cual es una transformación lineal de los datos y en conjunto con el umbral de Hotelling, servirá para agregar nuevas observaciones a la matriz de entrenamiento. Aquí, se agregaran aquellas observaciones que se encuentren por debajo del umbral de Hotelling, y a la vez, más cerca del umbral. Además, solo en casos específicos, serán consideradas algunas observaciones que superen el umbral, estas serán aquellas observaciones consecutivas que superen el umbral.

3.2.5 Cambio de observaciones iniciales

Para la selección de observaciones del primer modelo de similitudes no se utilizó una herramienta de estadística. Por lo tanto, se realizará un cambio de observaciones de esas matrices de entrenamiento iniciales. En primer lugar, se realiza un test de Hotelling de la matriz de entrenamiento de entrada, esto determinara el estadístico T^2 de los datos perteneciente a la matriz de entrenamiento de entrada. Posteriormente, se realiza un test de Hotelling a la matriz de errores de estimación del último modelo ejecutado, de ahí se seleccionan una cierta cantidad de observaciones (por ejemplo, un quinto de la cantidad de la matriz de entrenamiento inicial) usando el criterio descrito en la Sección 3.2.3.3. Finalmente, se realiza el intercambio, donde se quitan observaciones, mal catalogadas por el test de Hotelling hecho a la matriz de entrenamiento inicial, y se agregan las nuevas observaciones, obtenidas del test de Hotelling hecho a la matriz de errores de estimación, a nueva matriz de entrenamiento.

Con ello, se ejecuta el nuevo modelo SBM y se calculan los errores cuadráticos medios y errores porcentuales relativos; si estos son menores a los errores del modelo anterior, entonces se conservan los cambios y viceversa.

3.2.6 Ajustes finales

Posterior al cambio de observaciones, se inicia la etapa de ajustes finales, donde se pueden agregar observaciones particulares a la matriz de entrenamiento. Recordar que al agregar observaciones a la matriz de entrenamiento puede ayudar a cubrir zonas que no estaban consideradas en el modelo. Además, tener en cuenta que no sean observaciones similares a la matriz de entrenamiento, sino el algoritmo las descartará como indica en la Sección 3.2.3.2.

Finalmente, el algoritmo principal entrega múltiples elementos, de los cuales los más importantes son los siguientes.

- $D_i \in \Re^{n\,x\,m}$: Matriz con m variables de entrada y n observaciones.
- $D_o \in \Re^{n\,x\,p}$: Matriz con p variables de salida y n observaciones.
- $E \in \Re^{n\,x\,p}$: Matriz con p variables de salida y n observaciones.
- $Mse \in \Re^{7\,x\,p}$: Matriz con p variables de salida y 7 modelos.
- $Ep \in \Re^{7\,x\,p}$: Matriz con p variables de salida y 7 modelos.
- $d \in \Re$: Distancia de Kernel.

Los elementos mencionados son utilizados en la herramienta de creación de modelo SBM. Donde D_i, D_o y d son los elementos que conforman un modelo SBM y pueden estimar las variables de salida en el instante de tiempo actual dado las variables de entradas. Por otro lado, Mse y Ep son los elementos que validan el modelo SBM creado.

La matriz Mse contiene los valores del error cuadrático medio de cada uno de los modelos SBM creados hasta obtener el modelo de similitud final. A su vez, la matriz Ep contiene los valores de los errores porcentuales relativos promedios de cada uno de los modelo SBM generados hasta obtener el modelo final.

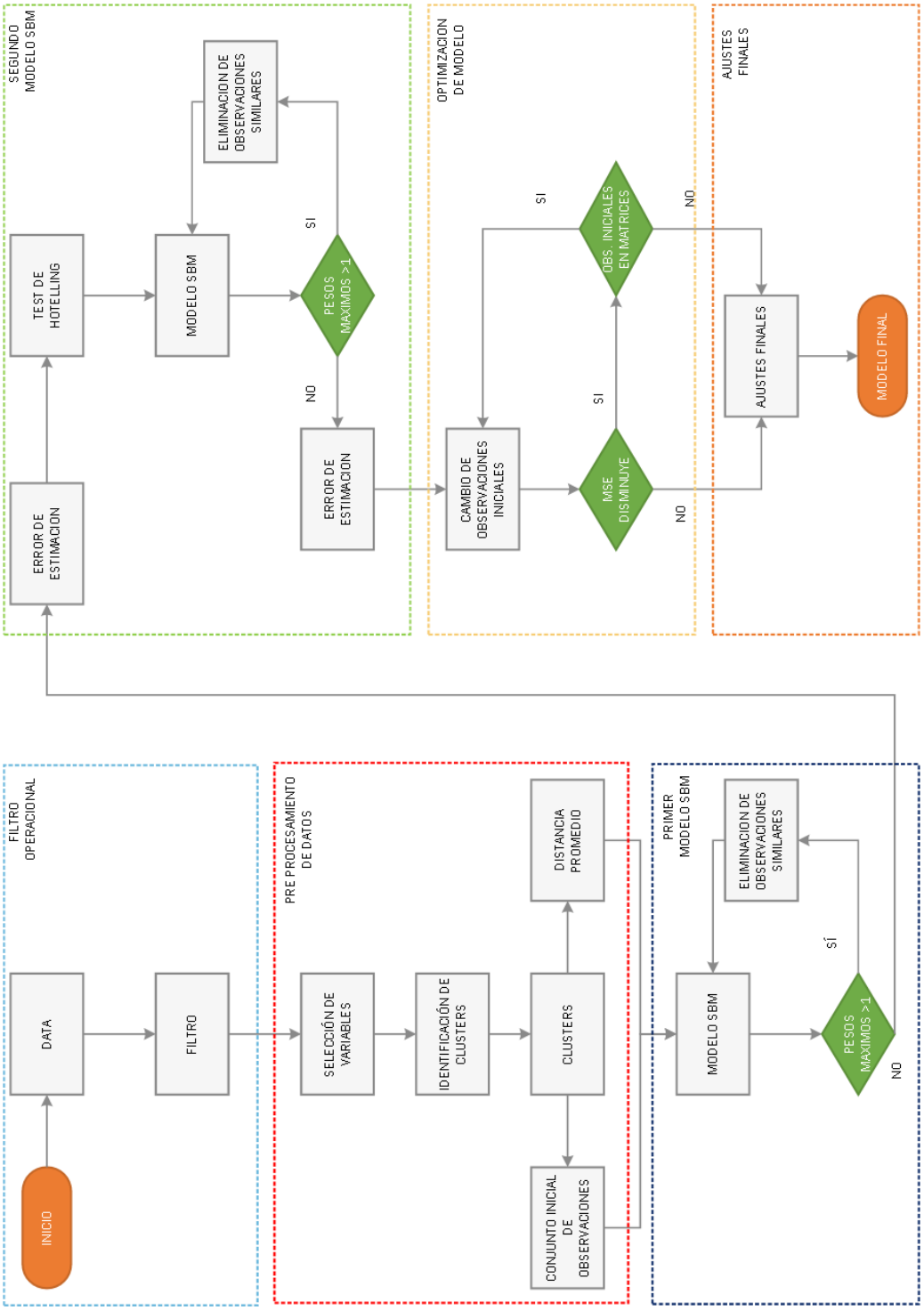


Figura 3.5 : Algoritmo de creación de modelos

3.3 Metodologías para estimación, predicción y análisis de escenarios

A continuación se presentan las metodologías realizadas para cada propósito, partiendo por la estimación de variables controladas, seguido por la detección de anomalías, luego por la predicción de variables y con ello, el análisis de escenarios.

3.3.1 Estimación de variables

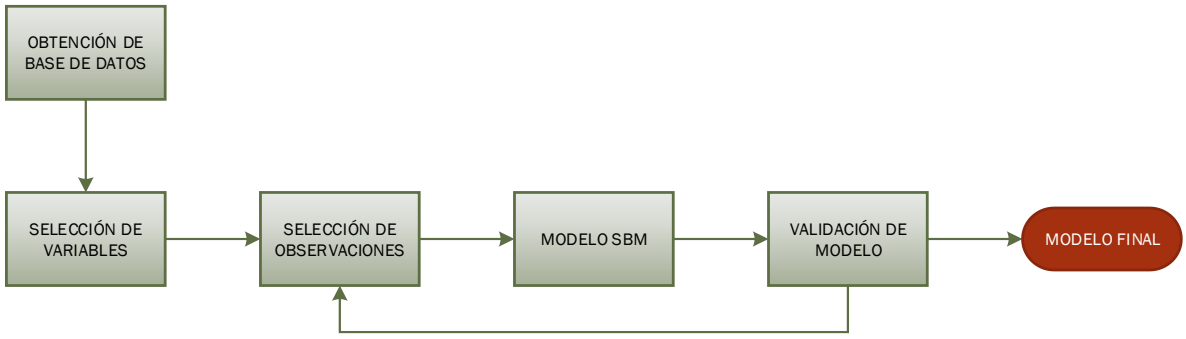


Figura 3.6 : Metodología de estimación.

3.3.1.1 Obtención de base de datos

Se tiene que poseer una base de datos rica en información donde se cubran todos los puntos de operación del proceso, y luego, filtrar los valores no numéricos. Posteriormente, se filtran los datos en base a la operación de la planta.

3.3.1.2 Selección de variables y observaciones

Elegir las variables relevantes para la modelación del proceso mediante PLS, es decir, aquellas variables que tengan peso para el modelo. Luego, las observaciones para la matriz de entrenamiento de entradas se tienen las variables independientes (manipuladas) hasta el instante actual y las variables dependientes (controladas) hasta el instante $(t-1)$.

$$\boldsymbol{D_{i_{est}}(obs_t) = \left[x_1(t), x_2(t), \dots, x_n(t), y_1(t-1), y_2(t-1), \dots, y_p(t-1)\right]} \qquad (3.7)$$

La selección de observaciones, posterior al primer modelo SBM creado, se realiza con un test de Hotelling utilizando la matriz de errores de estimación. Serán seleccionadas aquellas observaciones que se encuentren por debajo el umbral del test con 95% de confianza.

3.3.1.3 Modelo SBM

Este modelo estima las salidas (variables controladas) en el instante t usando las matrices de entrenamiento $D_{i_{est}}$ y $D_{o_{est}}$.

$$x^*(t) \rightarrow \begin{bmatrix} \hat{w} = \left(D_{i_{est}}^T \Delta D_{i_{est}}\right)^{-1}\left(D_{i_{est}}^T \Delta x^*\right) \\ w = \frac{\hat{w}}{1^T * \hat{w}} \\ y_{est} = D_{o_{est}} \cdot w \end{bmatrix} \rightarrow y_{est}(t) \qquad (3.8)$$

3.3.1.4 Validación del modelo

Se estiman los errores de estimación para cada observación, y con ello, se estiman los errores cuadráticos medios y porcentuales para validar el modelo. Si se aprueba el criterio de error cuadrático medio o el criterio de error porcentual relativo entonces el modelo está listo, de lo contrario, se regresa a la selección de observaciones.

3.3.2 Detección de anomalías

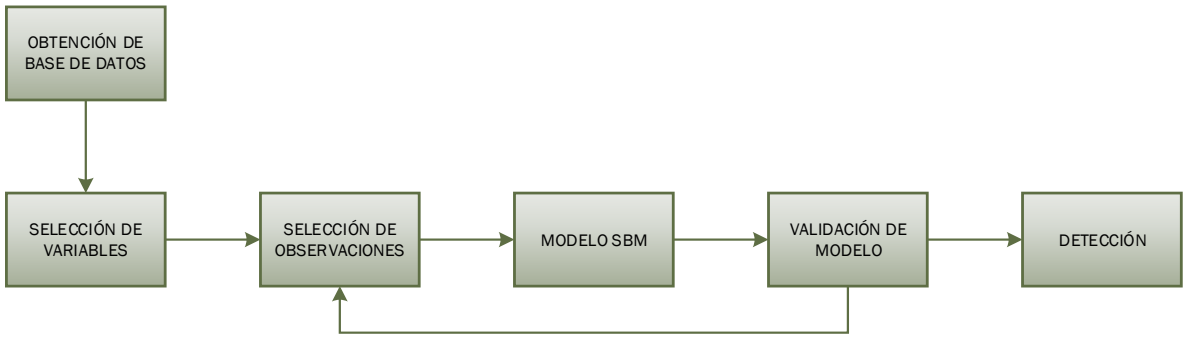


Figura 3.7 : Metodología de detección de anomalías.

3.3.2.1 Modelo SBM

Se crea un modelo SBM usando datos de operación normal de la planta, utilizando la misma metodología de estimación de variables. Este modelo puede que contenga unas matrices de

entrenamiento diferentes al modelo de estimación de variables, ya que este último puede contener datos de operación (normal/anormal) de la planta.

3.3.2.2 Validación del modelo

Se estiman los errores de estimación de datos con operación normal, y con ello, se estiman los errores cuadráticos medios y porcentuales para validar el modelo. Si se aprueba el criterio de error cuadrático medio y/o el criterio de error porcentual relativo entonces el modelo está listo, de lo contrario, se regresa a la selección de observaciones.

3.3.2.3 Detección

Una vez aprobada la validación del modelo, se realizan pruebas con datos de operación anormal de la planta. En estos momentos, se revisan los errores de estimación del modelo, donde deberían ser mayores en comparación a los errores de estimación de operación normal, debido a que no identifica el punto de operación actual. Posteriormente, se realiza un test de Hotelling y si entrega que los errores de estimación con datos anormales superan el umbral de Hotelling, entonces se trata de una anomalía.

3.3.3 Predicción de variables

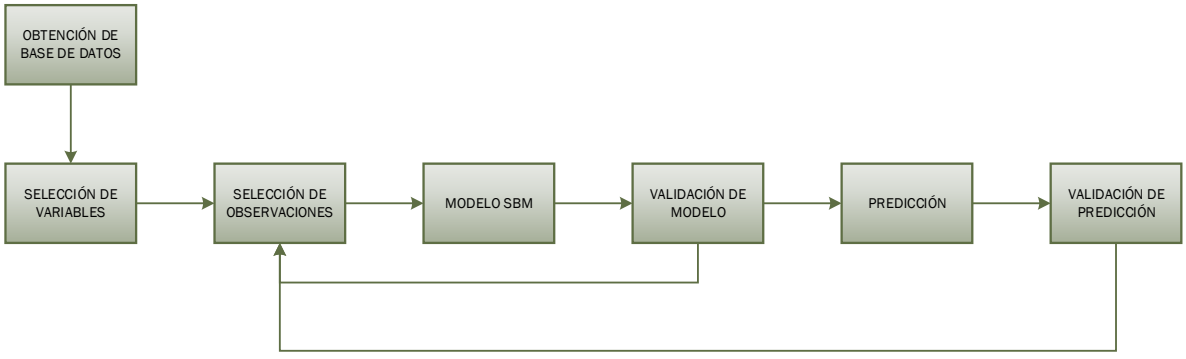


Figura 3.8 : Metodología de predicción.

3.3.3.1 Obtención de base de datos

Se tiene que poseer una base de datos rica en información donde se cubran todos los puntos de operación del proceso, y luego, filtrar los valores no numéricos. Posteriormente, se filtran los datos en base a la operación de la planta.

3.3.3.2 Selección de variables y observaciones

Elegir las variables relevantes para la modelación del proceso mediante PLS. Luego para la observación para la matriz de entrenamiento de entradas se tienen las variables independientes (manipuladas) hasta el instante t y las variables dependientes (controladas) hasta el instante t.

$$\boldsymbol{D_{i_{pred}}(obs_t) = \left[x_1(t), x_2(t), \dots, x_n(t), y_1(t), y_2(t), \dots, y_p(t)\right]} \quad \textbf{(3.9)}$$

La selección de observaciones, posterior al primer modelo SBM creado, se realiza con un test de Hotelling utilizando la matriz de errores de estimación. Serán seleccionadas aquellas observaciones que se encuentren por debajo el umbral del test con 95% de confianza.

3.3.3.3 Modelo SBM

Este modelo estimara la salida (variables controladas) en el instante $t + 1$ usando las matrices de entrenamiento $D_{i_{pred}}$ y $D_{o_{pred}}$.

$$\boldsymbol{x^*(t) \rightarrow \begin{bmatrix} \hat{w} = \left(D_{i_{pred}}^T \Delta\, D_{i_{pred}}\right)^{-1} \left(D_{i_{pred}}^T \Delta\, x^*\right) \\ w = \frac{\hat{w}}{1^T * \hat{w}} \\ y_{est} = D_{o_{pred}} \cdot w \end{bmatrix} \rightarrow y_{est}(t+1)} \quad \textbf{(3.10)}$$

3.3.3.4 Validación del modelo

Se estiman los errores de estimación para cada observación, y con ello, se estiman los errores cuadráticos medios y porcentuales para validar el modelo. Si se aprueba el criterio de error cuadrático medio o el criterio de error porcentual relativo entonces el modelo está listo, de lo contrario, se regresa a la selección de observaciones.

3.3.3.5 Predicción

Se utiliza el modelo creado para la predicción de las variables controladas manteniendo las variables manipuladas congeladas.

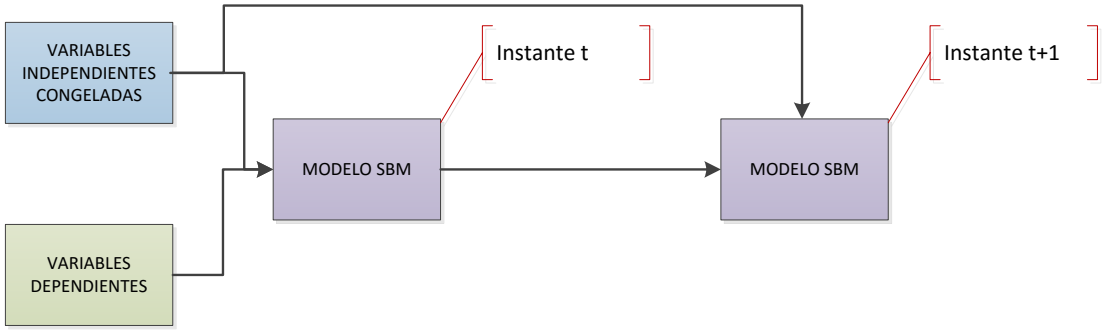


Figura 3.9 : Predicción.

En la Figura 3.9Figura 3.9, se observan que las variables independientes congeladas hasta el instante t en conjunto con las variables dependientes en el instante t serán utilizadas para estimar las variables controladas en el instante $(t+1)$ usando el modelo SBM. Posteriormente, se utilizaran las mismas entradas independientes en conjunto con las salidas estimadas recientemente para estimar las variables de salida en el instante $(t+2)$.

3.3.3.6 Validación de la predicción

Se verifica que la predicción hecha por el modelo es adecuada. Si no aprueba el criterio para la predicción, el cual consiste en un error porcentual relativo menor al 3%, se regresa a la selección de observaciones y se disminuyen los pasos a predecir.

3.3.4 Análisis de escenarios

Utilizando el mismo modelo creado para la predicción de variables.

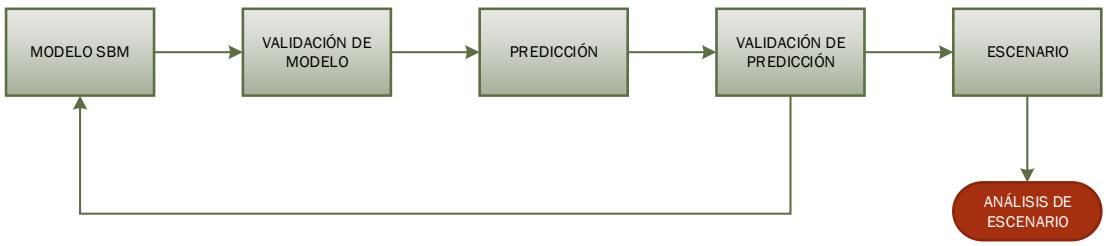


Figura 3.10 : Metodología de análisis de escenarios.

3.3.4.1 Escenario

Se mantienen congeladas las variables manipuladas, a excepción de una de ellas. Esta variable puede aumentar o disminuir su valor pero siempre manteniéndola dentro de los límites de operación. Posteriormente de su modificación, se mantiene congelada en ese valor. Desde ese entonces, se realiza predicción con el nuevo escenario.

3.3.4.2 Análisis

Se analiza el escenario obtenido, observando cómo afecta a las variables controladas. Si el escenario es favorable, es decir, entrega salidas con mejor desempeño, se guarda la configuración para futuros usos.

CAPÍTULO 4. PRUEBAS Y RESULTADOS

En este capítulo se entregaran los resultados de la aplicación de las herramientas descritas en el capítulo anterior. Para ello, se necesita tener datos de un proceso para la realización de un modelo usando modelación SBM. Entonces, en la Sección 4.1 se detallara la descripción de un proceso minero y su base de datos históricos que serán utilizadas para modelar el proceso. En la Sección 4.2, se entregara los resultados obtenidos del modelo de estimación generado a partir de las herramientas descritas en la Sección 3.3.1. En la Sección 4.3, se entregaran los resultados obtenidos del modelo de detección de anomalías generado de acuerdo a las herramientas descritas en la Sección 3.3.2. Finalmente, en la Sección 4.4, se entregaran los resultados obtenidos del modelo de predicción de variables controladas generado a partir de las herramientas definidas en la Sección 3.3.3.

4.1 Descripción del proceso minero estudiado y de los datos utilizados

La concentración de minerales tiene por objetivo enriquecer las especies mineralógicas útiles de un mineral, mediante la eliminación de componentes estériles. La concentración se divide en cuatro procesos: Chancado, Molienda, Flotación y Espesadores. Dado los alcances de este trabajo, se estudió el proceso de molienda. En este proceso, existen 2 tipos de molienda: convencional y SAG. La primera, consiste en utilizar molinos de bolas de acero o barras de acero.

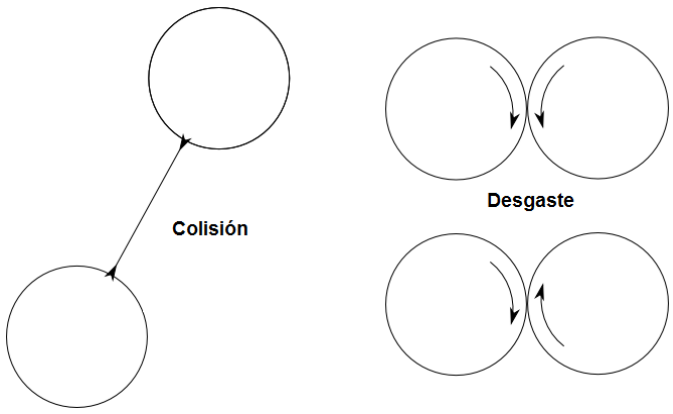


Figura 4.1 : Fuerzas de colisión y desgaste en molino para el proceso de molienda.

En el caso del molino de bolas, se ingresan bolas de acero al molino y con el movimiento rotacional, chocan con el material, provocando que el mineral se muela. En el molino de barras, el material ingresa al molino, donde existen grandes barras de acero que se encuentra girando y muelen el mineral.

Por otro lado, la molienda SAG o semiautógena (Semi: a medias. Autógena: se origina a sí mismo) consiste en la reducción del tamaño del material usando el material mismo y algunas bolas de acero (12% de bolas, normalmente en los molinos de bolas se utiliza un 35%). Esta molienda permite que se pueda pasar del chancado primario a flotación sin emplear etapas intermedias de chancado secundario y terciario para reducir el tamaño del mineral.

4.1.1 Molino SAG

El molino SAG centrifuga el material y lo eleva por las paredes internas del molino gracias a elementos levantadores en rotación (*lifters*), hasta el punto que la gravedad lo despega y lo impulsa en una caída parabólica, produciendo así una ola continua que impacta con el mineral en la zona inferior del molino. El golpe continuo de las bolas de acero y el material mismo, disminuyen continuamente el tamaño de las rocas, hasta el momento de su expulsión a través de las parrillas adosadas a la tapa de descarga. En la salida del molino, se obtienen granulometrías hasta 180 micrones, que permiten la liberación de la mayor parte de los minerales de cobre en forma de partículas.

Figura 4.2 : Representación del proceso del molino SAG.

En caso de que, para ciertos tamaños (sobre-tamaños de harnero o “trommel”), el proceso de molienda no sea efectivo, es necesario enviar el mineral a un chancador de pebbles. Con esto, el material reducirá su tamaño y será reingresado al molino.

4.1.2 Bases de datos utilizadas

Las bases de datos utilizadas en la elaboración de los modelos de estimación y predicción fueron proporcionadas por Honeywell Chile S.A. Estas bases contemplan datos históricos sobre el funcionamiento de un molino SAG. A continuación se detalla la base de datos.

Característica	Base de Datos BD1	Base de Datos BD2
Número de variables	26	26
Número de observaciones	44640 (41140 en Operación)	44640 (36840 en Operación)
Tiempo de muestreo	1[min]	1[min]
Anomalía	No	Si

Tabla 4.1 : Bases de datos utilizadas.

Las bases de datos contienen información de operación del molino SAG 2 de la mina El Teniente durante el mes de Diciembre del 2014, donde existe una anomalía, y el mes de Mayo del 2015, donde se encuentra en operación normal. En ella, se encuentran datos de operación y no operación, donde posteriormente serán filtrados.

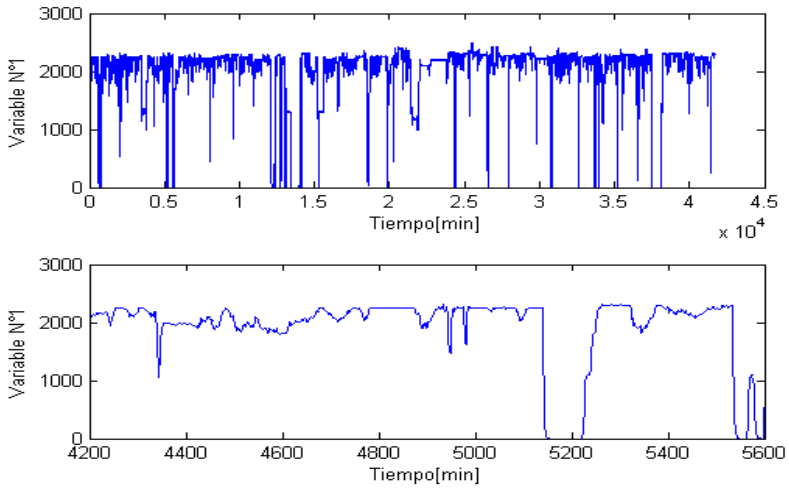


Figura 4.3 : Gráfico de operación del molino.

En la Figura 4.3, se observa el molino en dos instancias: en operación y en apagado. Los datos cuando el molino no se encuentre encendido, serán filtrados. Además, se conoce que el proceso del molino SAG tiene un periodo de asentamiento de 12 minutos, es decir, si se realiza un cambio en las variables manipuladas, este se verá reflejado en las variables controladas 12 minutos más adelante.

4.2 Resultados obtenidos en prueba de algoritmo de estimación

Posteriormente a la descripción del proceso y la base de datos, se procede a presentar los resultados del algoritmo de estimación de variables controladas, etapa por etapa, con el objetivo de observar y evaluar la evolución de la modelación usando SBM.

En primer lugar, para la modelación del proceso, se deben identificar en la base de datos cuales serán nuestras variables de entrada y de salida del proceso. Además, se consideraran las observaciones/mediciones pasadas de cada variable hasta el instante $(t-12)$, y con ello, se considerara cada medición anterior como una nueva variable. Las cuales se detallan a continuación.

Variables de entrada N°	1	6	11	16	21	26	31	36	41	46	51	56	61
	2	7	12	17	22	27	32	37	42	47	52	57	62
	3	8	13	18	23	28	33	38	43	48	53	58	63
	4	9	14	19	24	29	34	39	44	49	54	59	64
	5	10	15	20	25	30	35	40	45	50	55	60	65

Variables de salida N°	66	69	72	75	78	81	84	87	90	93	96	99
	67	70	73	76	79	82	85	88	91	94	97	100
	68	71	74	77	80	83	86	89	92	95	98	101

Tabla 4.2 : Variables de entrada y salida del proceso.

Con ello, se presentan los resultados iniciales, correspondientes la etapa de pre procesamiento de los datos. En primer lugar, se realiza una selección de variables como indica la Sección 3.2.2.1, entregando los pesos de carga de las variables de entrada al modelo, tal como se aprecia en el siguiente gráfico.

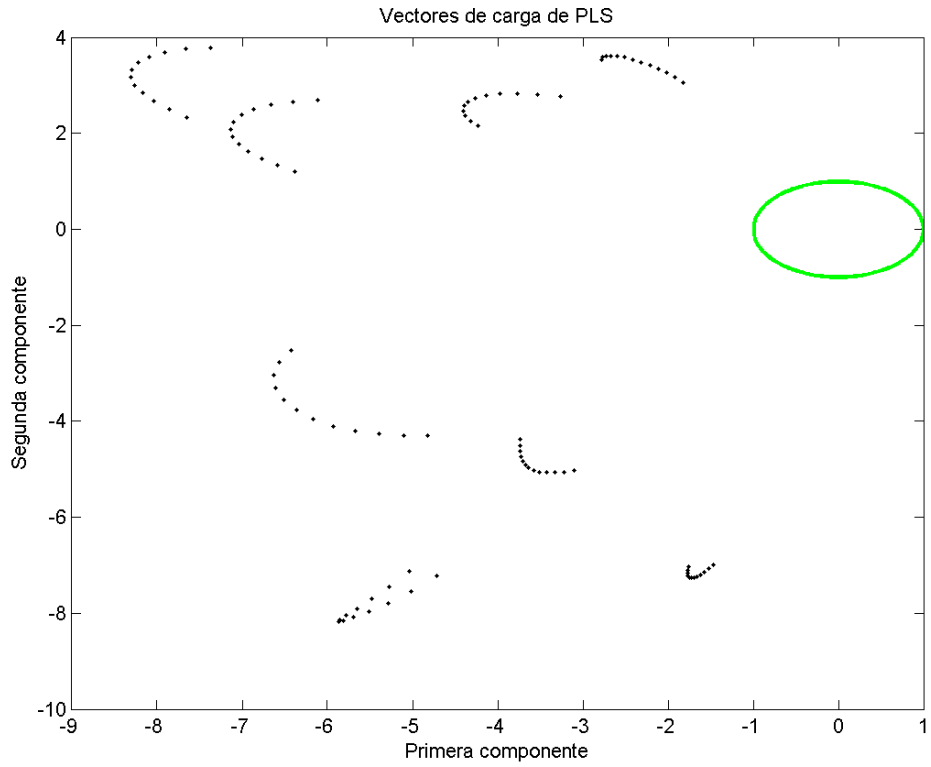


Figura 4.4 : Vectores de carga (Primera componente vs Segunda componente).

Como se puede apreciar de la Figura 4.4, las variables se ordenan en grupos y en forma parabólica. Esto se debe a que cada grupo corresponde a una variable en particular y sus regresores, es decir, un grupo puede ser $\{X_n(t), X_n(t-1), \dots, X_n(t-12)\}$. Además, se puede observar que no existe un conjunto de variables que se encuentran adentro del círculo en color verde, esto indica que las variables graficadas son relevantes para la modelación.

Posteriormente, se realiza un análisis de componentes principales (PCA) para la detección de grupos (clusters) de datos, los cuales representarían diferentes puntos de operación del molino.

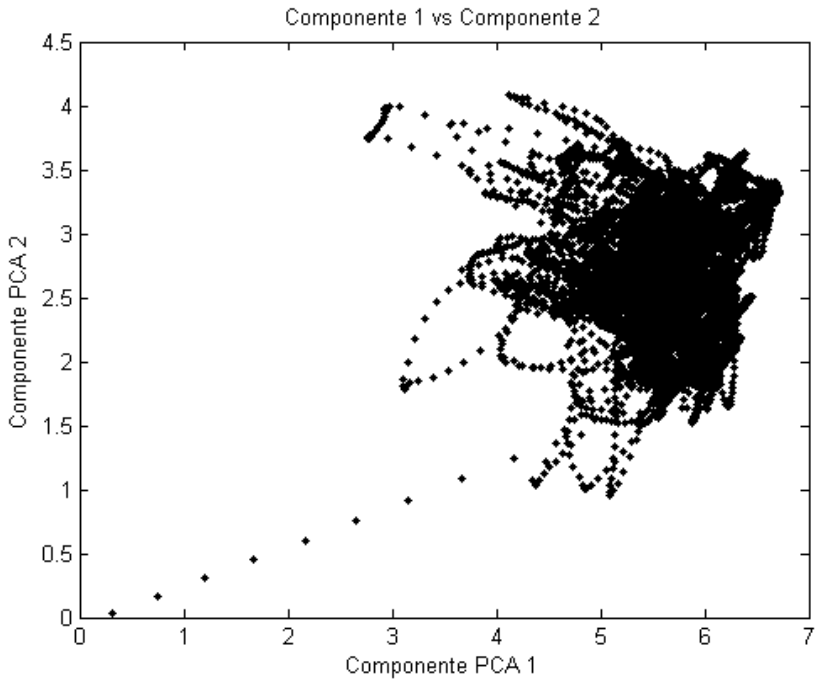


Figura 4.5 : Gráfico de primera y segunda componente del análisis PCA.

En la Figura 4.5, se puede observar las dos primeras componentes del análisis PCA que corresponden a un 60% de la representatividad de los datos, en ellas se puede apreciar que existe un gran cluster. Además, se observaron las otras componentes para verificar que existiese un solo cluster. Por otro lado, la línea punteada que se dirige al cluster representa el transiente del proceso.

De aquí en adelante, sólo falta calcular la distancia de kernel que será utilizada para comparar observaciones, la cual esta descrita en la Sección 3.2.2.4 y la creación de las primeras matrices de entrenamiento D_i y D_o. Se elegirá un conjunto inicial para aquellas observaciones que satisfagan la siguiente restricción:

$$\boldsymbol{d_{Kernel} \leq \left|\left|obs - centro_{cluster}\right|\right| \leq 2 \cdot d_{Kernel}} \quad \textbf{(4.1)}$$

Además, los datos a seleccionar serán aquellos que se encuentren más cerca del centro del cluster. El centro del cluster se puede calcular usando el promedio de cada variable o la función *kmeans* de MATLAB® para $k = 1$, ya que comprobamos que existe un punto de operación. Se seleccionaron un 4% de los datos (1646 observaciones). Posteriormente a la etapa de pre procesamiento, se inicia la etapa de creación del modelo SBM. A continuación se presentan los resultados del modelo usando un 4% de la base de datos.

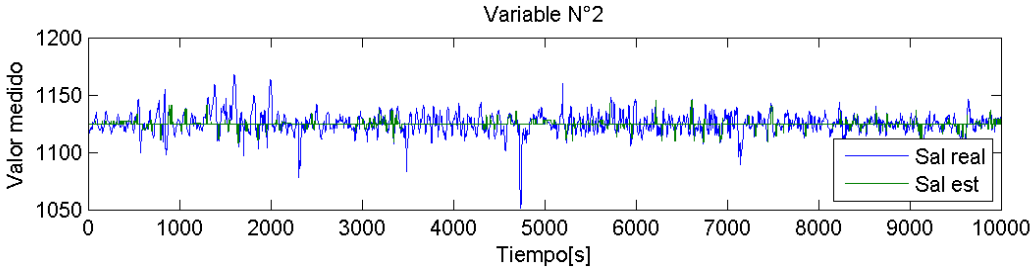


Figura 4.6 : Salida real y salida estimada del modelo con 4% de los datos.

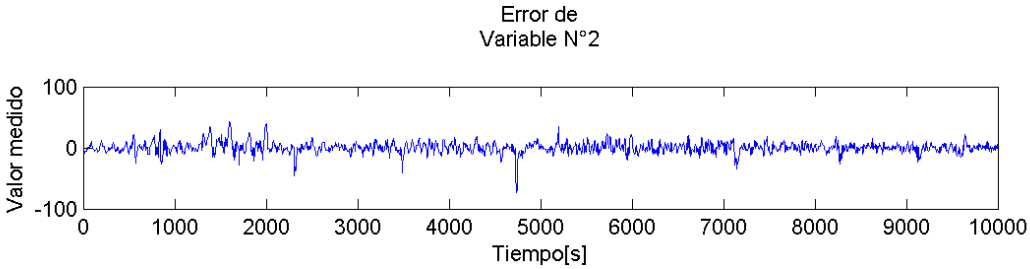


Figura 4.7 : Gráfico de los errores de estimación del modelo con 4% de los datos.

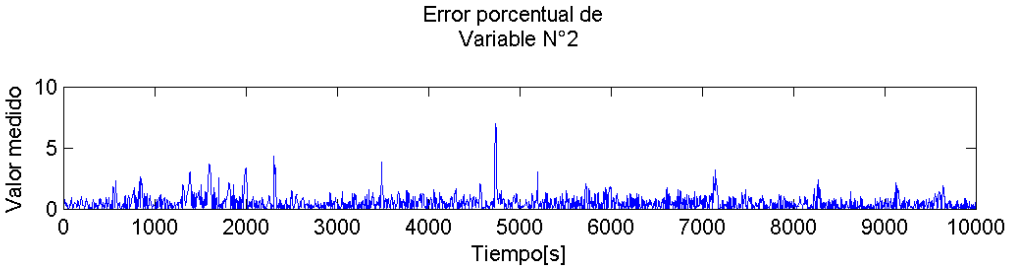


Figura 4.8 : Gráfico de los errores porcentuales del modelo con 4% de los datos.

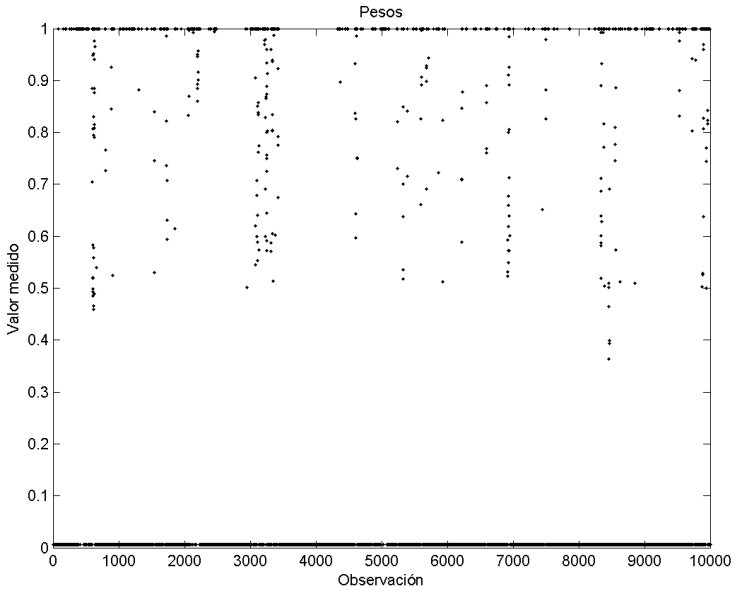


Figura 4.9 : Gráfico de los pesos máximos del modelo con 4% de los datos.

Se puede observar que el modelo estima mal las salidas del proceso. Esto se debe a que sus matrices de entrenamiento todavía no contienen mucha información del proceso. Esto se puede ver reflejado en el gráfico de pesos máximos (Figura 4.9), donde la mayoría de los puntos se encuentran cercanos a cero. Por otro lado, existe una cantidad de pesos máximos con valores igual a uno, debido a que existen observaciones que se encuentran adentro del área de similitud de una única observación perteneciente a la matriz de entrenamiento de entradas.

Luego, se procede a la etapa de integración de nuevas observaciones. Para esto, se realizara un test de Hotelling usando la matriz de residuos o errores de estimación del modelo SBM. A continuación se grafica el estadístico T^2 relacionado con las observaciones.

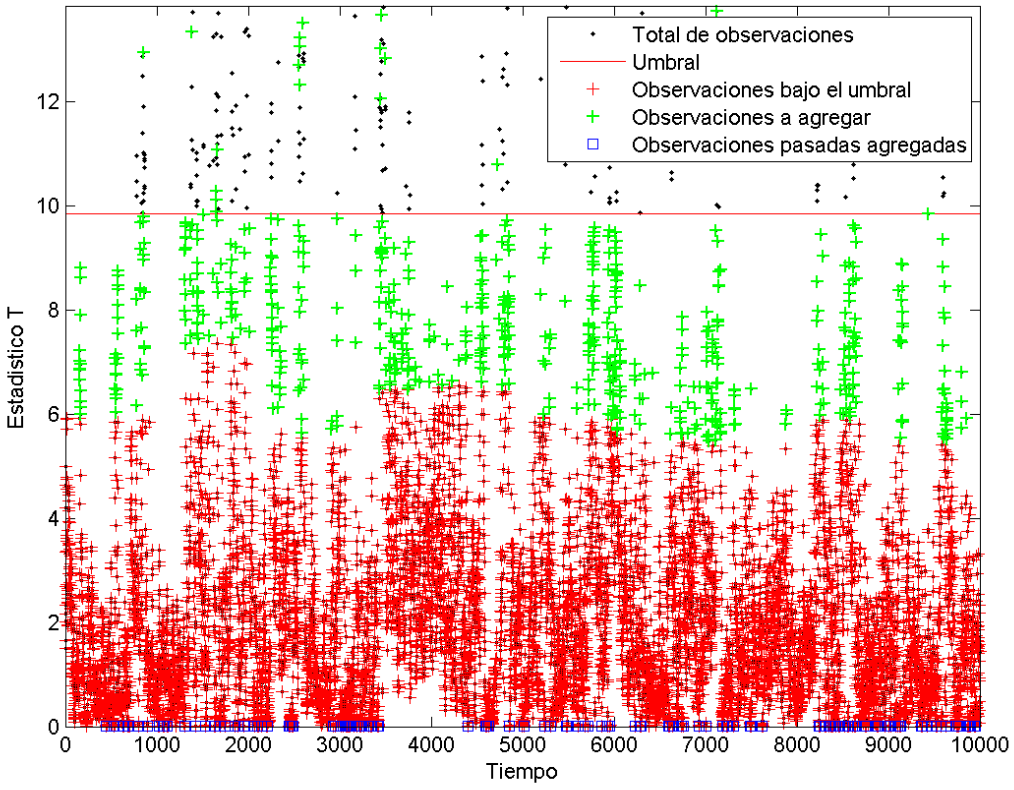


Figura 4.10 : Test de Hotelling del modelo con 4% de los datos.

En la Figura 4.10, se observa que los puntos en rojo son las observaciones bajo el umbral (línea roja), en verde son posibles observaciones a agregar, en azul las observaciones ya pertenecientes a la matriz de entrenamiento y en negro las observaciones sobre el umbral. Se determina que las observaciones a agregarse serán aquellas que se encuentran por debajo del umbral de Hotelling y, a la vez, que se encuentren cerca al umbral, ya que presentaran una mayor variabilidad al modelo, bajo este criterio

se agregaran 6% de nuevas observaciones. Además, se agregaran aquellos conjuntos de observaciones consecutivas que se encuentren sobre el umbral (ya que esto sería un error permanente de la estimación, dado que significa que el modelo no reconoce esa condición de operación. Si bien el hecho que las observaciones superan el umbral significa que los datos son anómalos con respecto al modelo, durante el periodo de entrenamiento o evolución del modelo, se consideran los datos como condición de operación normal).

Con ello, se realiza un segundo modelo SBM con 10% de los datos (4114), es decir, el 4% de los datos del modelo original más el 6% de los datos agregados por el test de Hotelling, y posteriormente, se iniciara la etapa de iteración, donde se busca cambiar las matrices de entrenamiento inicial por unas matrices de entrenamiento basadas en el criterio del test de Hotelling.

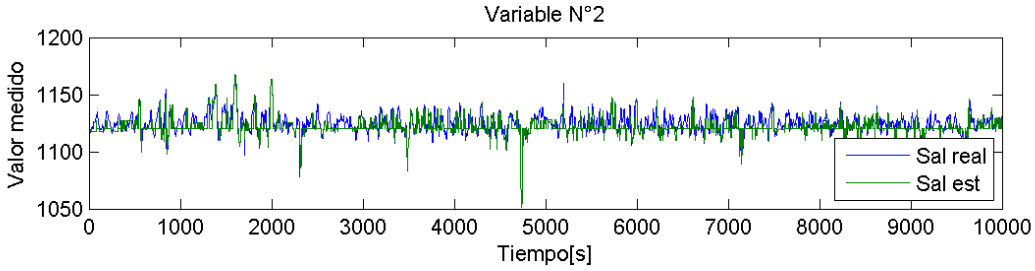


Figura 4.11 : Salida real y salida estimada del modelo con 10% de los datos.

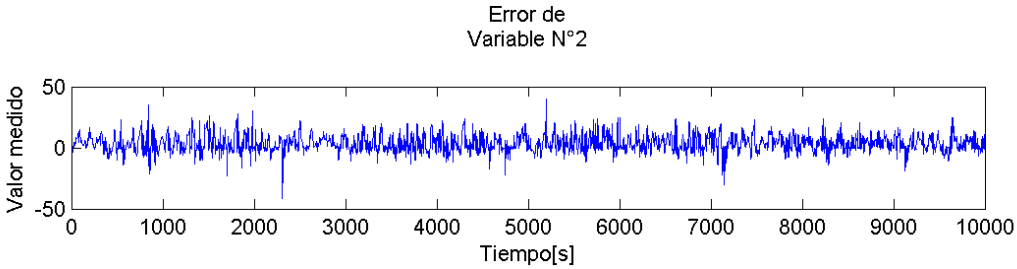


Figura 4.12 : Errores de estimación del modelo con 10% de los datos.

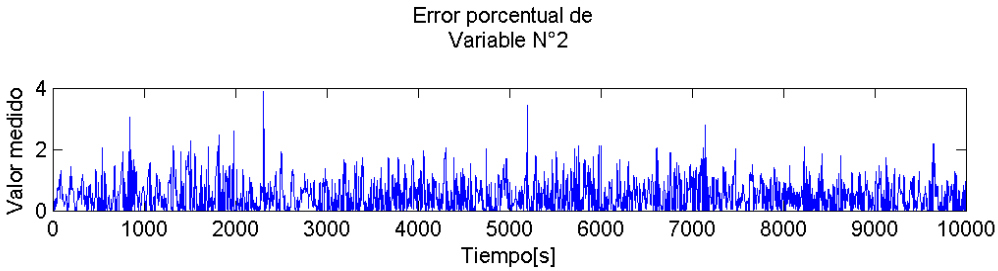


Figura 4.13 : Errores porcentuales del modelo con 10% de los datos.

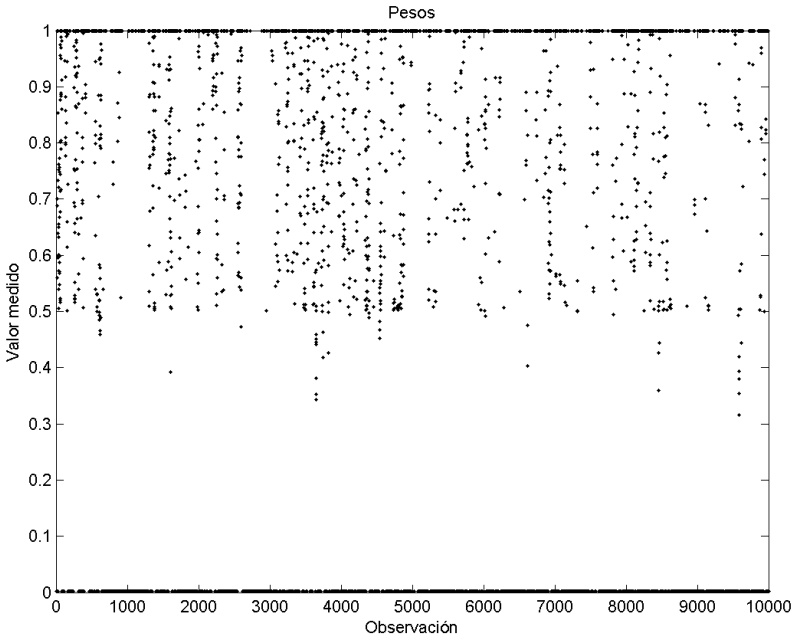
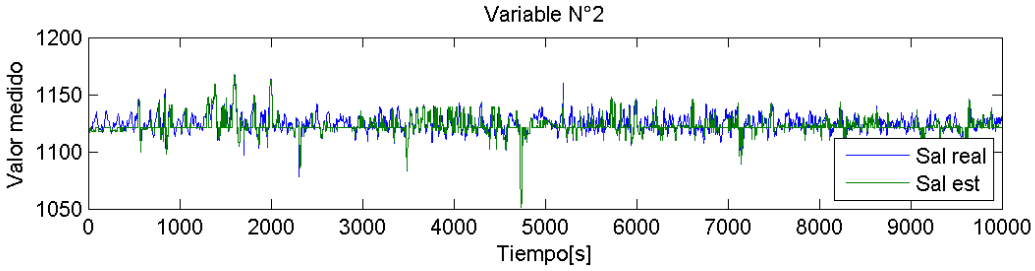


Figura 4.14 : Gráfico de los pesos máximos del modelo con 10% de los datos.

Como se puede observar de los resultados del segundo modelo SBM creado, la estimación mejoró pero aun así no sigue siendo buena. Esto se debe a que todavía las matrices de entrenamiento no contienen toda la información del proceso. Entonces, se procede a cambiar el conjunto inicial, es decir, las matrices de entrenamiento de entradas y de salidas originales del modelo (representación con 4% de los datos). Para realizar cambios de observaciones (cambios en las matrices de entrenamiento), se usó el criterio del mínimo error cuadrático medio para validar el cambio sugerido, es decir, se compara el error cuadrático del modelo con $H\%$ de datos iniciales y $J\%$ de datos usando Hotelling con el modelo de $(H-1)\%$ de datos iniciales y $(J+1)\%$ de datos usando Hotelling. Si el último modelo tiene el menor error cuadrático medio, entonces se mantiene ese modelo, en caso contrario, se revierten los cambios.



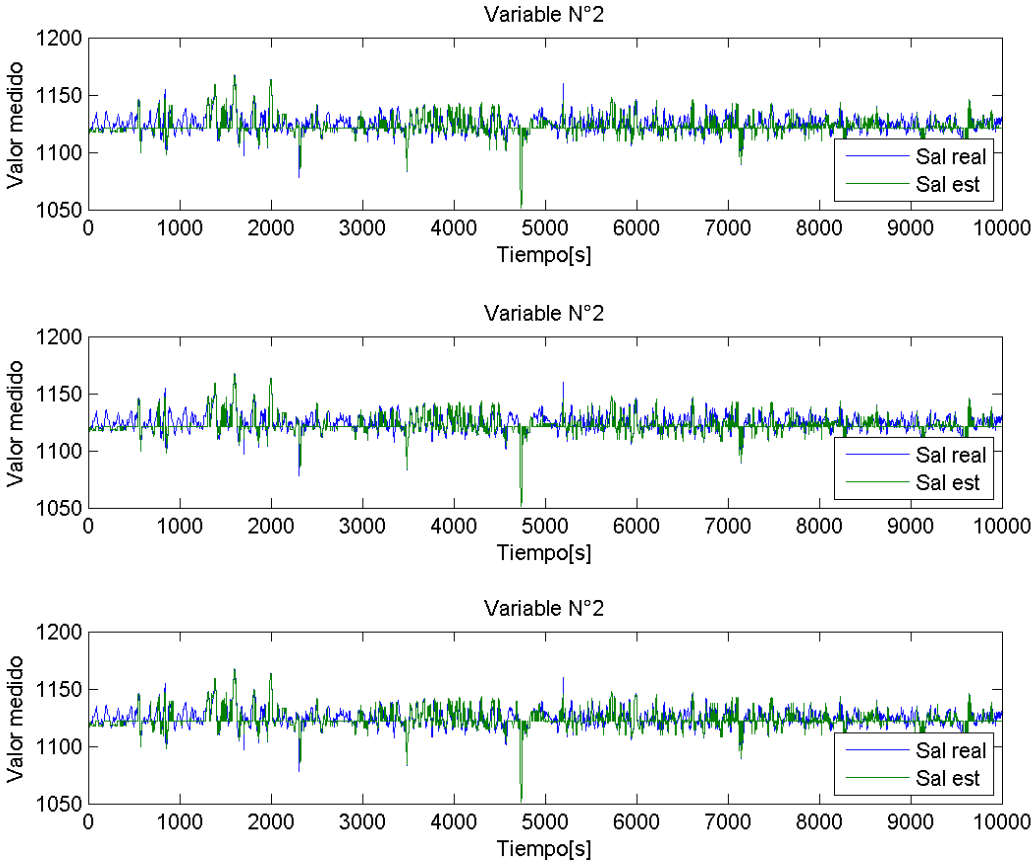


Figura 4.15 : Gráfico de las salidas del modelo con el conjunto inicial cambiado.

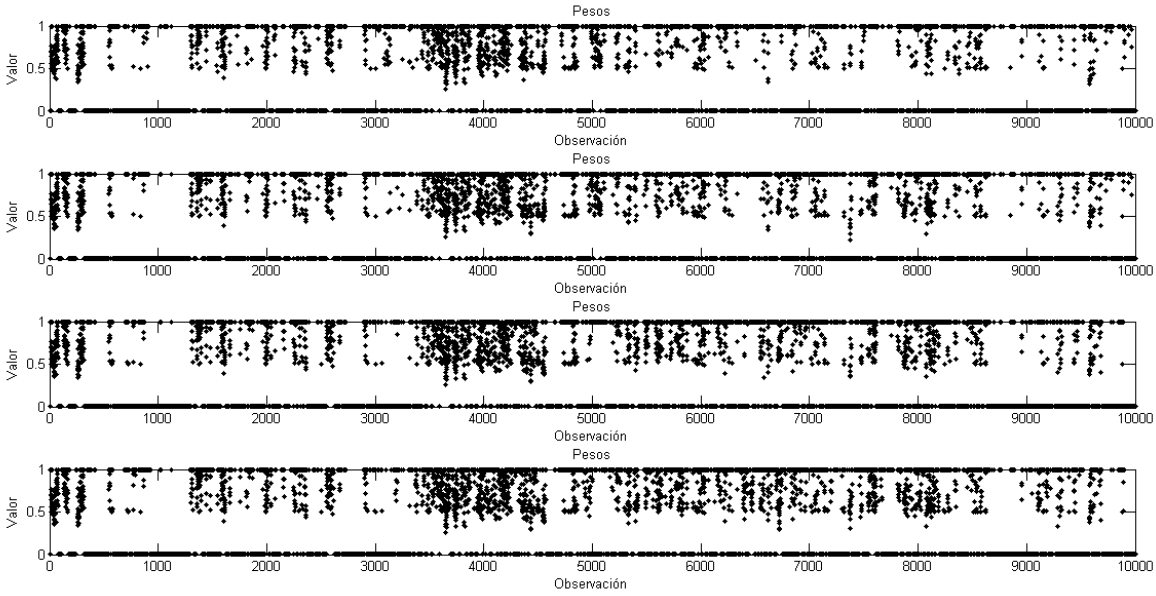


Figura 4.16 : Pesos máximos de las iteraciones del modelo.

Se puede observar en la Figura 4.15 alguna pequeña mejoría en la estimación después de cada iteración, pero aun así no reconoce bien el proceso. Por otro lado, en la Figura 4.16, se observan los pesos máximos de la etapa de iteraciones de cambio de matrices de entrenamiento del modelo, además se pueden observar las variaciones entre cada iteración, como por ejemplo, entre los intervalos de 5000 a 9000, se observa considerablemente el aumento de pesos máximos, lo que significa que el modelo reconoce mejor esa zona.

Finalmente, se conoce a priori que no existen anomalías en esta base de datos, entonces se agregaron las observaciones con peso menor a $0.1w_{max}$ para obtener el modelo final del molino SAG.

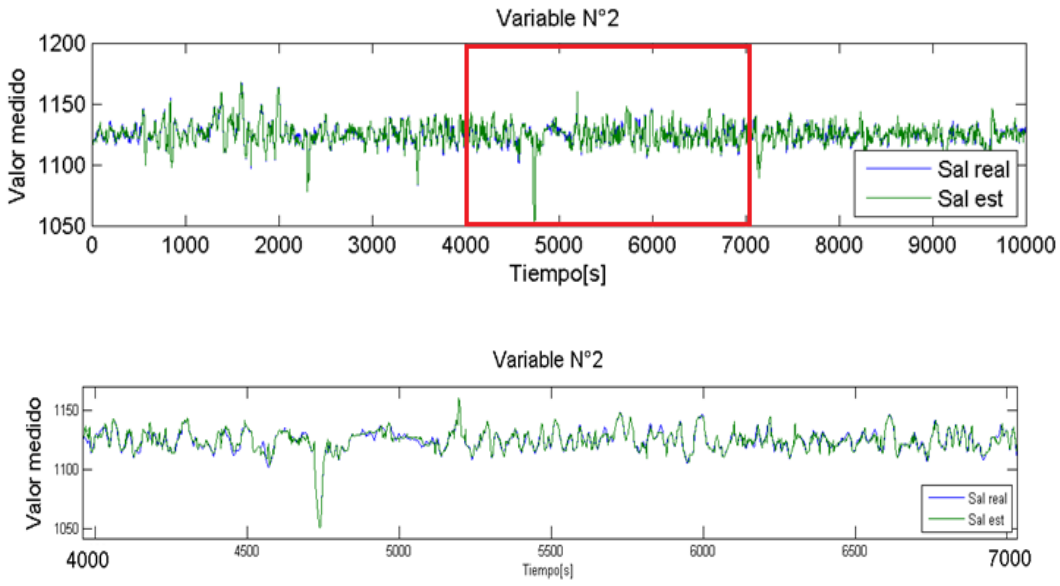


Figura 4.17 : Salida real y salida estimada del modelo final con 18% de los datos.

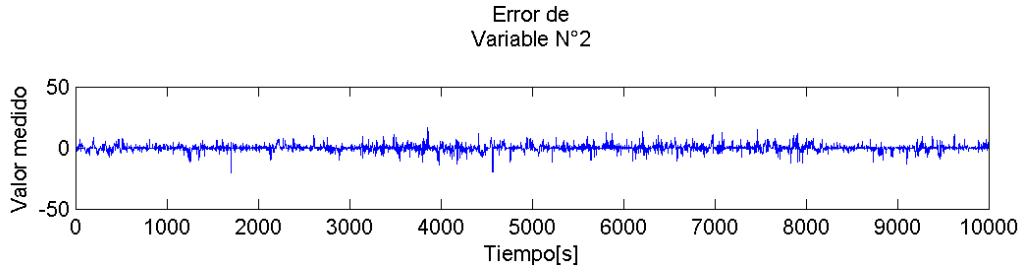


Figura 4.18 : Gráfico de los errores de estimación del modelo con 18% de los datos.

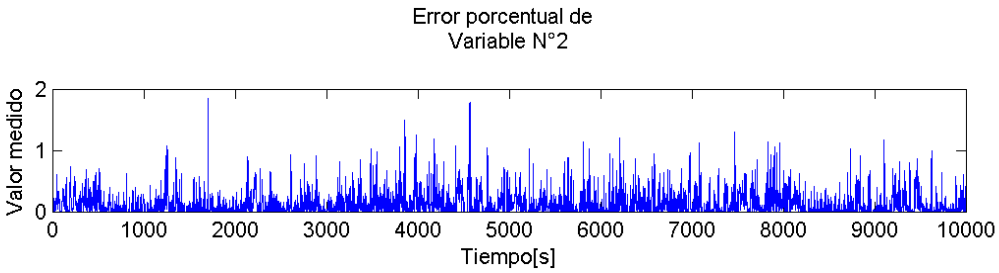


Figura 4.19 : Gráfico de los errores porcentuales del modelo con 18% de los datos.

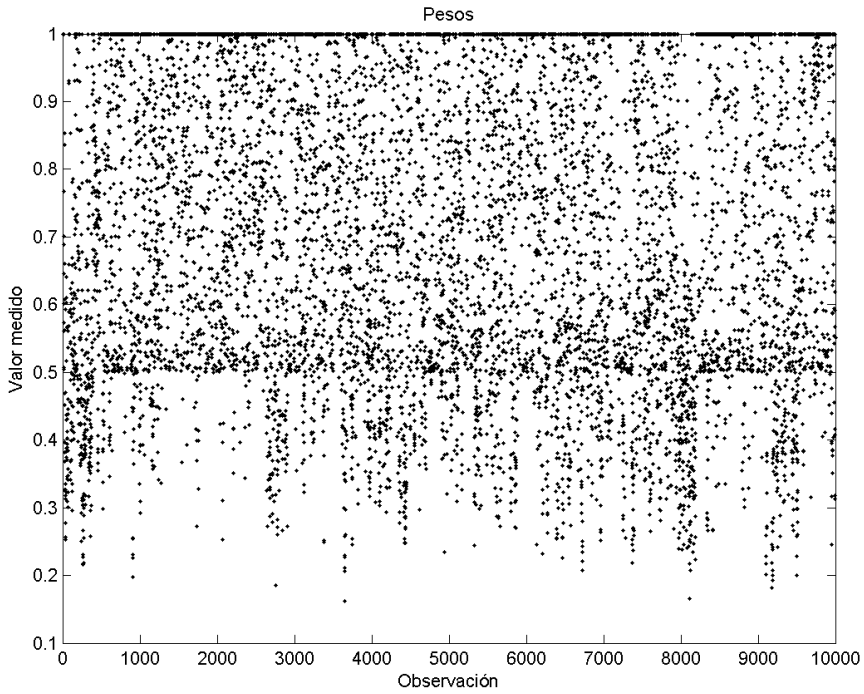


Figura 4.20 : Gráfico de los pesos máximos del modelo con 18% de los datos.

Como se puede observar, finalmente, el modelo estima correctamente el proceso tal como muestra la Figura 4.17. A continuación, se presentan las tablas de la evolución de los errores de estimación del modelo.

	Primer modelo	**Segundo modelo**	**Primera iteración**	**Segunda iteración**	**Tercera iteración**	**Cuarta iteración**	**Ajuste final**
MSE	84.474	59.413	51.753	44.38	40.262	35.336	8.6036
Error porcentual (%)	0.56344	0.49212	0.45928	0.41908	0.39713	0.36753	0.18412

Tabla 4.3 : Resumen del modelo de estimación.

4.3 Resultados obtenidos en prueba de detección de anomalía

Se utilizaron los datos de la Base de datos BD2, donde existe una anomalía. En primer lugar, se identifica alguna variable que presente un cambio fuera de lo normal.

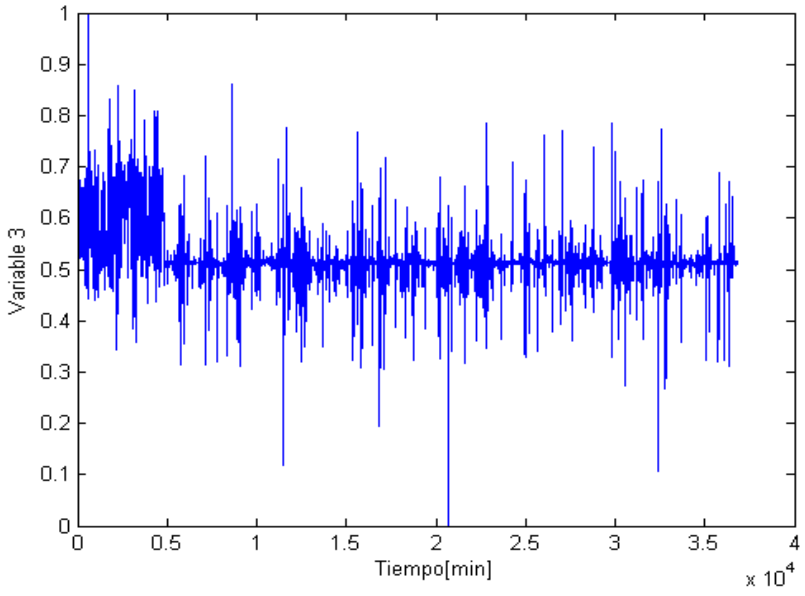


Figura 4.21 : Grafico de variable N°3.

Como se puede observar en la Figura 4.21, la variable N°3 tiene 2 partes relevantes: intervalo entre [1,5000] y intervalo entre [5001,36840]. La primera parte tiene una forma sinodal con mucho ruido y la segunda está centrada en el valor 0.5 (normalizado) con ruido agregado. Por lo tanto, existe un cambio en la dinámica de la variable presente, el cual se caracteriza como una anomalía, ya que no debería existir tal cambio.

Con esta información, podemos definir un intervalo de entrenamiento donde se encuentre en operación normal, el cual es entre [1,4900]. Con ello, crear un modelo SBM y luego validar la detección de la anomalía en los instantes de tiempo posteriores. Como se observó en el modelo de estimación, todas las variables involucradas resultaban relevantes para la modelación, seguiremos el mismo supuesto ya que se quiere buscar una anomalía. Entonces para la creación del modelo, no se realizará la selección de variables pero se seguirán el resto de los pasos del algoritmo de generación de modelos. Los resultados se muestran a continuación.

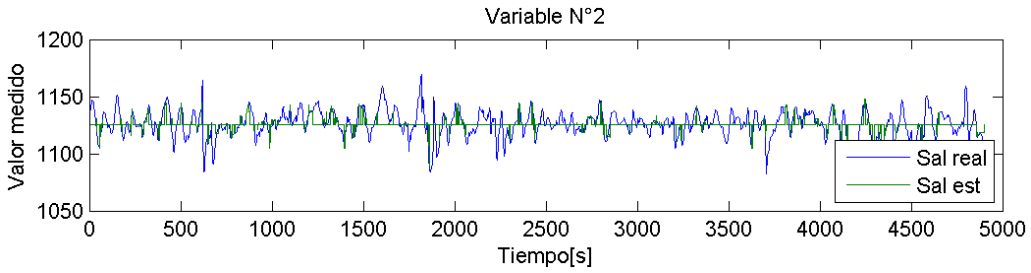


Figura 4.22 : Salida real y estimada del modelo con 4% de los datos.

Como se observa en la Figura 4.22, el modelo solo estima algunas zonas, principalmente las de entrenamiento. Esto se debe a que el conjunto inicial no contiene la información necesaria para representar el proceso.

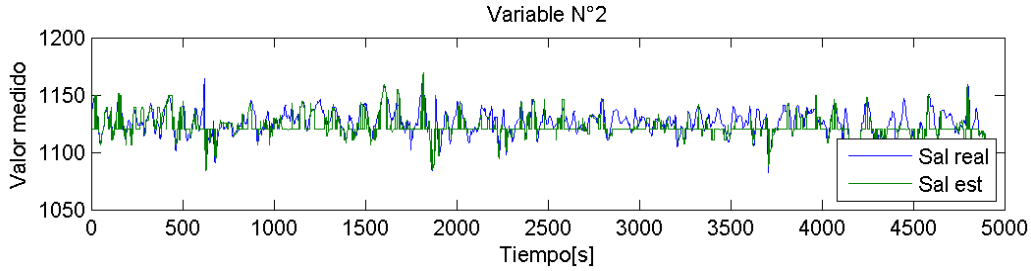


Figura 4.23 : Salida real y estimada del modelo con 10% de los datos.

Se aprecia en la Figura 4.23, que el modelo actual es mejor que el anterior, dado que estima una mayor cantidad de observaciones, a grandes rasgos, se podría decir que estima alrededor de un 60% del proceso.

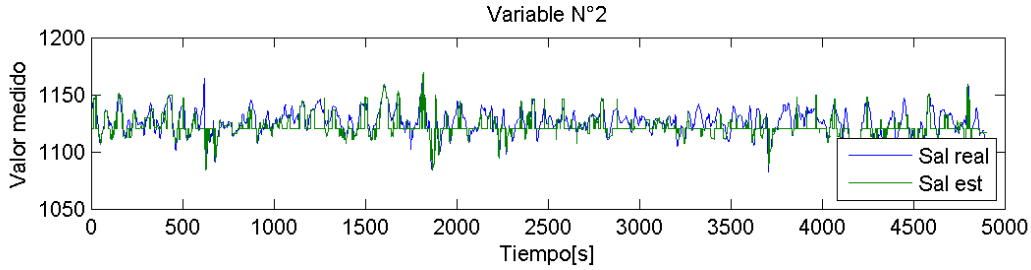


Figura 4.24 : Salida real y estimada del modelo con 10% de los datos (matrices cambiada).

En la Figura 4.24, se observa el modelo con datos de entrenamiento seleccionados solamente con test de Hotelling. A diferencia con el modelo anterior, se observa que el último modelo estima correctamente zonas que antes estimaba de manera errónea como en el intervalo [100,500].

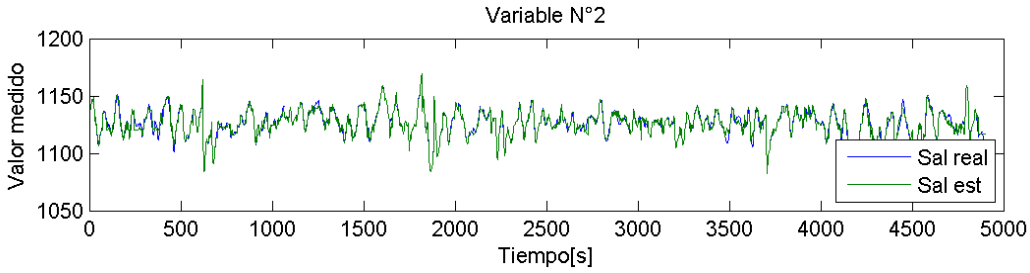


Figura 4.25 : Salida real y estimada del modelo final de detección.

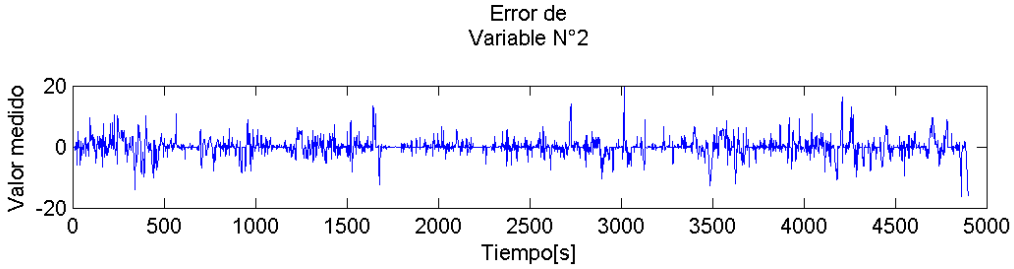


Figura 4.26 : Error de estimación del modelo final de detección.

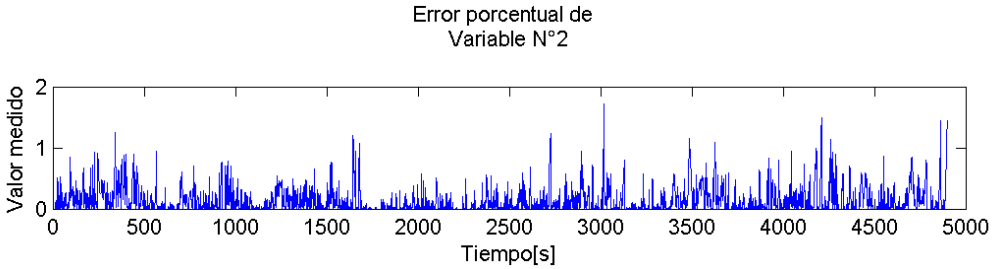


Figura 4.27 : Error porcentual del modelo final de detección.

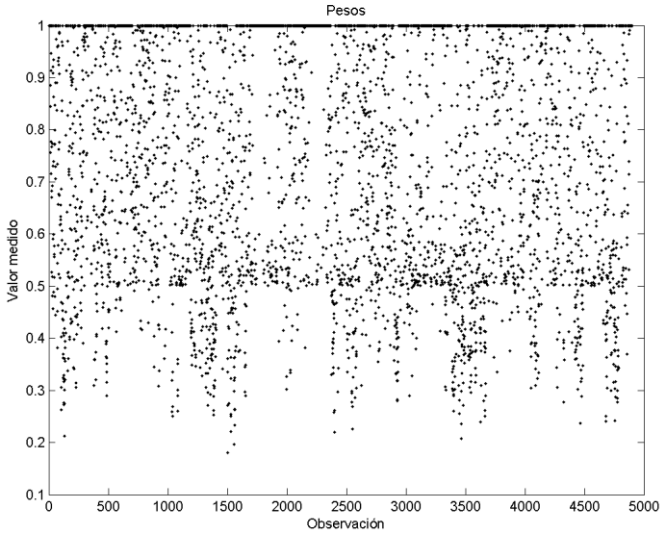


Figura 4.28 : Pesos máximos del modelo final de detección.

En la Figura 4.25, se aprecia que el modelo final estima correctamente las salidas controladas del proceso en operación normal. Además, en las Figura 4.26 y la Figura 4.27, se pueden observar los errores de estimación y los errores porcentuales relativos, respectivamente. En cuanto al error de estimación, no supera los 20 psi, lo que corresponde a un error porcentual menor al 2%. Por otro lado en la Figura 4.28, se observan los pesos máximos para cada observación según el modelo final del proceso, reconoce todos los puntos, ya que para hacer este modelo se agregaron las observaciones que tenían un peso máximo menor al $0.1w_{max}$.

	Primer modelo	**Segundo modelo**	**Primera iteración**	**Segunda iteración**	**Tercera iteración**	**Cuarta iteración**	**Ajuste final**
MSE	123.87	84.46	70.62	67.25	55.10	53.64	10.265
Error porcentual (%)	3.71615	1.21816	0.94978	0.83847	0.75761	0.74489	0.19069

Tabla 4.4 : Resumen del modelo de deteccion de anomalias.

Se procederá a mostrar los resultados de la detección de anomalía. A continuación se presenta el grafico desde el instante 4901 hasta el instante 8000. De tal manera, existe una zona con operación normal y una zona de operación anormal.

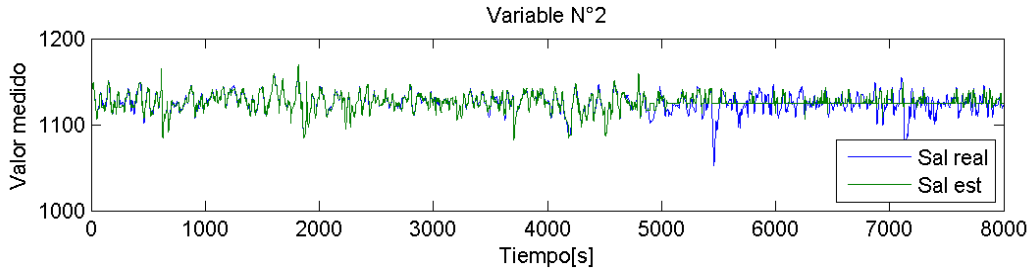


Figura 4.29 : Salida real y estimada de la prueba de detección de anomalías.

Como se puede observar en la Figura 4.29, a partir del instante 5000, el modelo estima de manera incorrecta la variable dependiente. Esto se debe a que no es una condición normal, por lo tanto no se ha entrenado para reconocer ese modo de operación, de otra forma, el modelo debería estimar correctamente la variable.

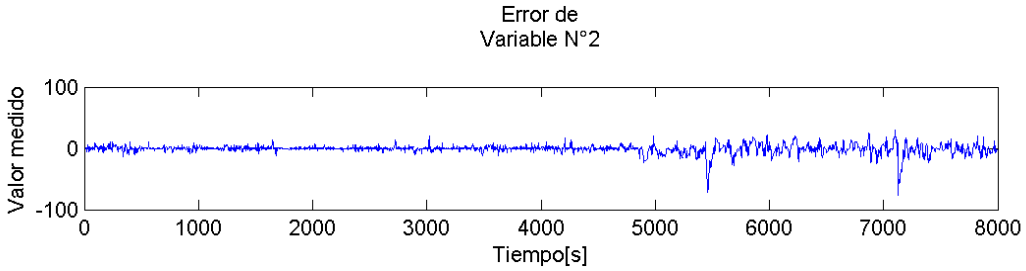


Figura 4.30 : Error de estimación de la prueba de detección de anomalías.

En la Figura 4.30, se pueden observar los errores de estimación del modelo. A partir del instante 5000, los errores de estimación son mayores, ya que el modelo no reconoce las observaciones candidatas en la matriz de entrenamiento.

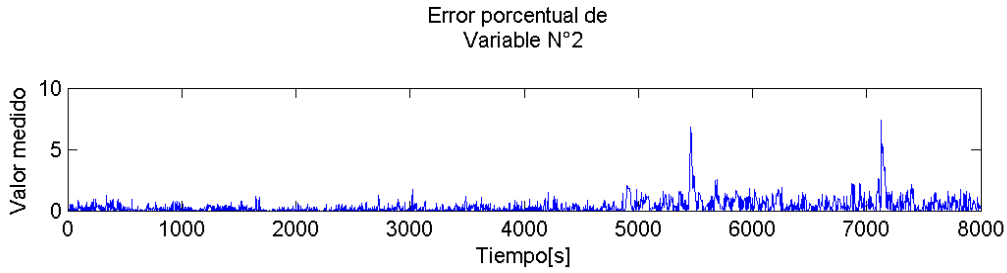


Figura 4.31 : Error porcentual de la prueba de detección de anomalías.

Por otro lado, en la Figura 4.31, se aprecian los errores porcentuales y es posible apreciar que desde el instante 5000 aumentan considerablemente, debido a que es una zona de operación anormal.

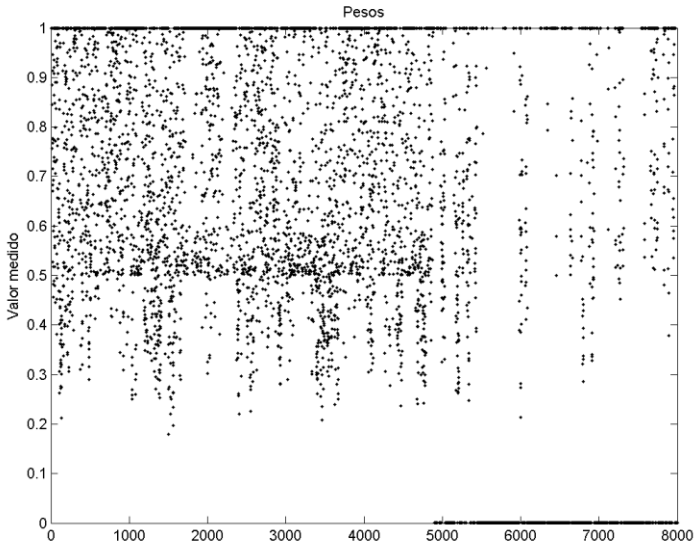


Figura 4.32 : Pesos máximos de la prueba de detección de anomalías.

En la Figura 4.32, se puede observar los pesos máximos de la prueba de detección de anomalías. En ella, se observa claramente como el modelo estima correctamente hasta el instante 4900, el periodo de entrenamiento, y posteriormente no reconoce la operación, dado que la mayoría de los pesos máximos se encuentran cercanos a cero.

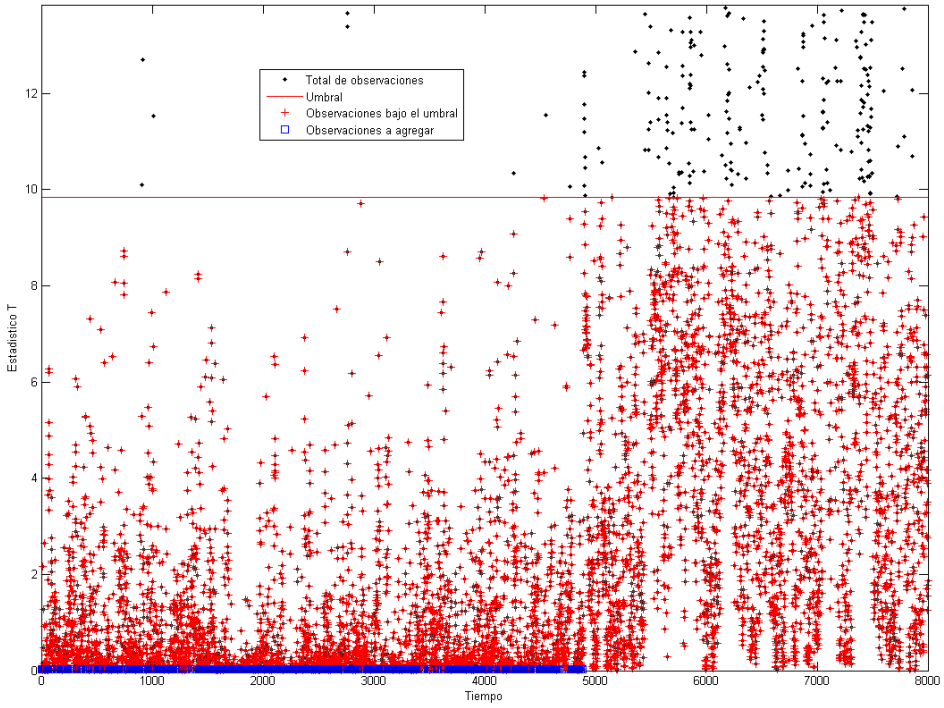


Figura 4.33 : Test de Hotelling para detección de anomalías.

En la Figura 4.33, se observa el test de Hotelling realizado con los errores de estimación de la prueba de detección de anomalías. En ella, se aprecia que una gran cantidad de observaciones se encuentran sobre el umbral pasado el instante 5000, lo que confirma que se detecta una anomalía en el proceso.

4.4 Resultados obtenidos en prueba de algoritmo de predicción

Utilizando la base de datos BD1, se procedió a realizar un modelo SBM que fuese capaz de predecir las variables controladas. Para esto se realizó un ajuste a las observaciones de las matrices de entrenamiento, donde se integraron las salidas del instante t a la matriz de entrenamiento de entradas y se agregaron las salidas en el instante $(t+1)$ en la matriz de entrenamiento de salida.

En la predicción, se mantienen congeladas las variables independientes. Es decir, sólo se actualizan las entradas del modelo que son variables dependientes. En la siguiente figura se puede observar la actualización:

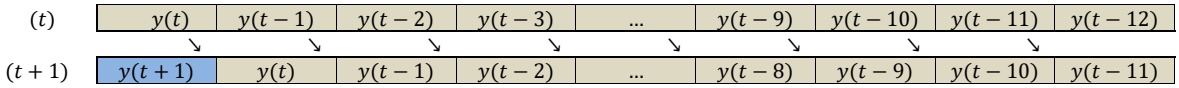


Figura 4.34 : Actualización del vector de entrada candidato para predicción.

En la Figura 4.34, se observan dos instantes: las entradas de variables dependientes en el instante t y en el instante $(t+1)$. En el instante t se predice usando el modelo SBM, obteniendo las salidas en el instante $(t+1)$. En este último instante, se actualiza el vector de entradas de variables dependientes integrando la salida del modelo SBM ($y(t+1)$, la que se encuentra en color azul), y con esta nueva entrada, nuevamente se predicen las salidas. Al cabo de 12 pasos se tendrá un vector de entradas dependientes que serán completamente estimaciones (predicciones) de las salidas del proceso. A continuación se presentan los resultados finales de la creación del modelo de predicción.

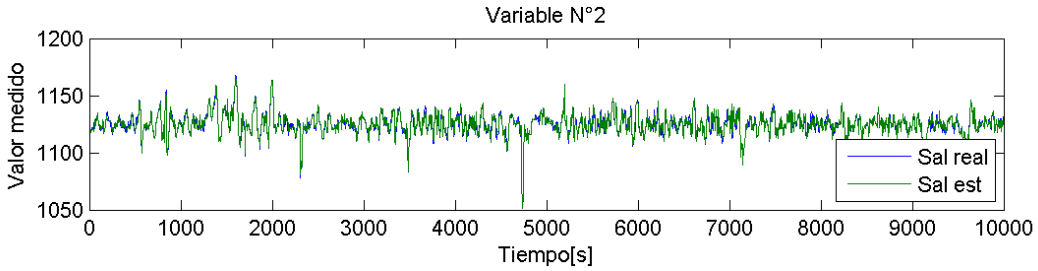


Figura 4.35 : Salida real y estimada del modelo final de predicción.

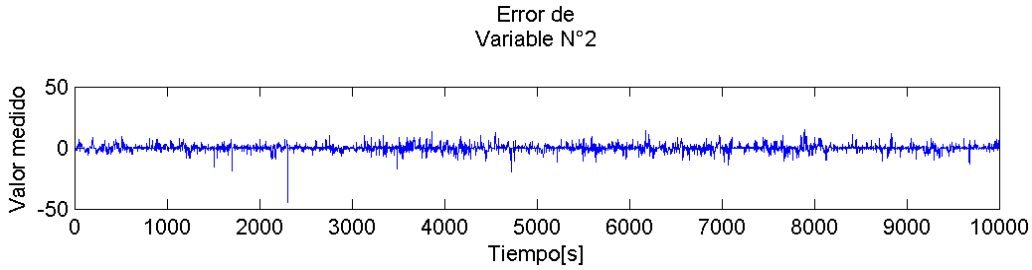


Figura 4.36 : Error de estimación de modelo final de predicción.

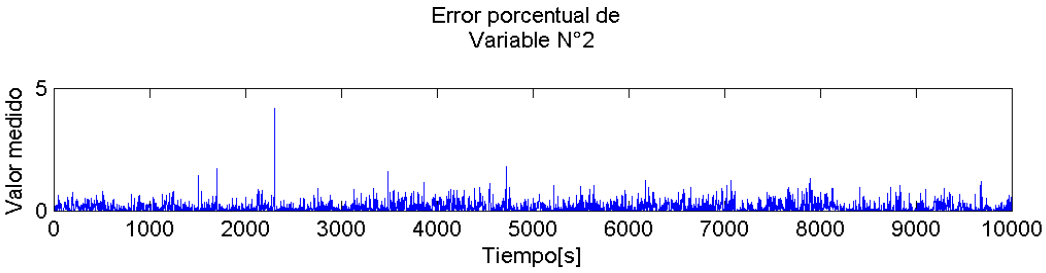


Figura 4.37 : Error porcentual de modelo final de predicción.

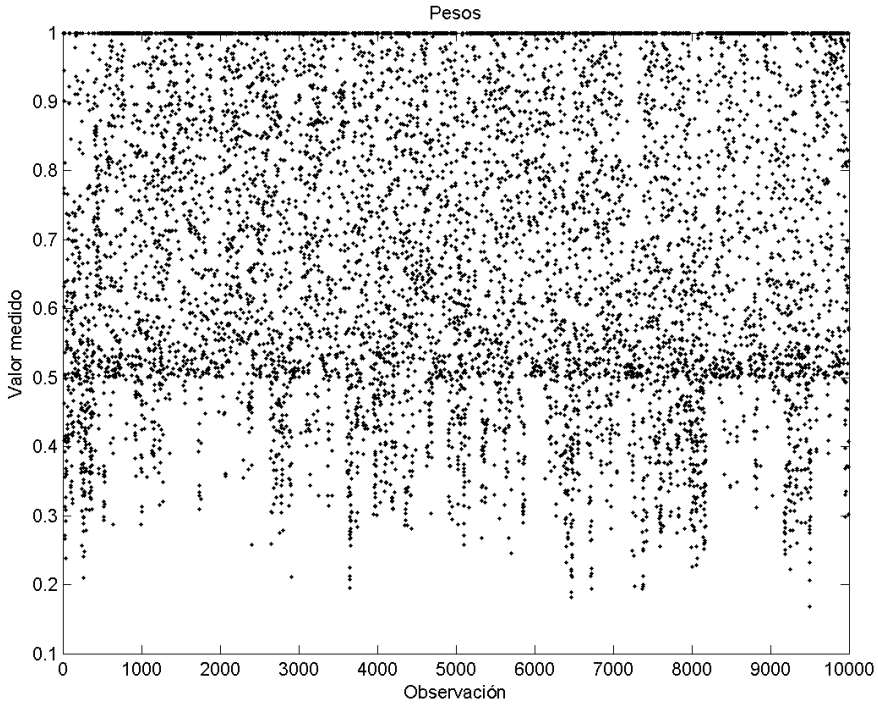


Figura 4.38 : Pesos máximos de modelo final de predicción.

	Primer modelo	Segundo modelo	Primera iteración	Segunda iteración	Tercera iteración	Cuarta iteración	Ajuste final
MSE	77.612	62.146	54.25	49.305	44.571	40.693	8.4617
Error porcentual (%)	0.54328	0.53226	0.48589	0.46709	0.44291	0.42117	0.17291

Tabla 4.5 : Resumen del modelo de predicción.

El modelo final resultante realiza correctamente su función de estimar a un paso, además, aprueba el criterio del residuos, donde se tiene un error cuadrático medio de 8.4617 y un error porcentual relativo promedio de 0.1729%.

Posteriormente, se validara la capacidad de predicción del modelo, tomando algún tramo aleatorio para realizar predicción. En nuestro caso, elegimos múltiples intervalos donde se realizaran predicciones hasta 10 pasos en cada instante. Posteriormente, se calcularán los errores de estimación asociados a cada observación, y con ello, estimar el error cuadrático medio y error porcentual para determinar si el modelo cumple con su objetivo.

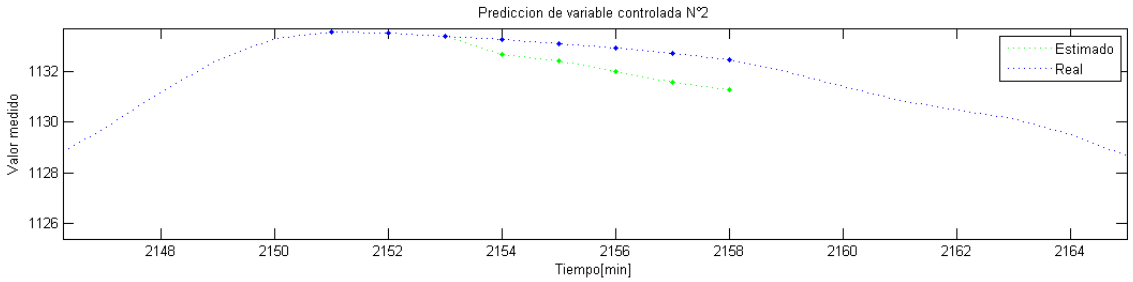


Figura 4.39 : Predicción de la segunda variable controlada.

En la Figura 4.39, se aprecia que la predicción son los últimos 5 valores estimados, donde se tiene un error de estimación en la primera predicción, si bien es menor al 1%, al estimar la segunda predicción se obtendrá una salida con un mayor error, dado que la entrada al modelo ya posee errores. Al transcurrir más pasos, estos errores se van acumulando, provocando una mayor dispersión de los errores.

	Paso 1	**Paso 2**	**Paso 3**	**Paso 4**	**Paso 5**
MSE	0.36213	0.40593	0.53935	0.71144	0.84453
EPP	0.05310	0.05918	0.07925	0.09782	0.1032

Tabla 4.6 : Resumen de los errores de predicción.

Como se puede observar en la Tabla 4.6, los errores cuadráticos medios (MSE) y los errores porcentuales promedio (EPP) van aumentando a medida que aumentan los pasos. Entonces, de acuerdo al criterio como predicción válida del proceso, se contemplaran 5 pasos. Los valores correspondiente a la Tabla 4.6, corresponde a valores no normalizados de las variables, al ser normalizados se tiene un error porcentual promedio menor al 5%.

Una vez realizado el modelo de predicción y su validación, se procede a iniciar la etapa de analizar posibles escenarios. Esto se realiza manteniendo las variables de entradas, ya sea independiente o dependiente, congeladas en un valor actual, exceptuando una de las variables independientes. Esto es debido a que a las variables independientes se les pueden fijar valores, siempre y cuando se encuentre

en los límites de operación. Con esto se espera encontrar un mejor desempeño del molino SAG. A continuación se presentan los resultados del análisis de escenarios.

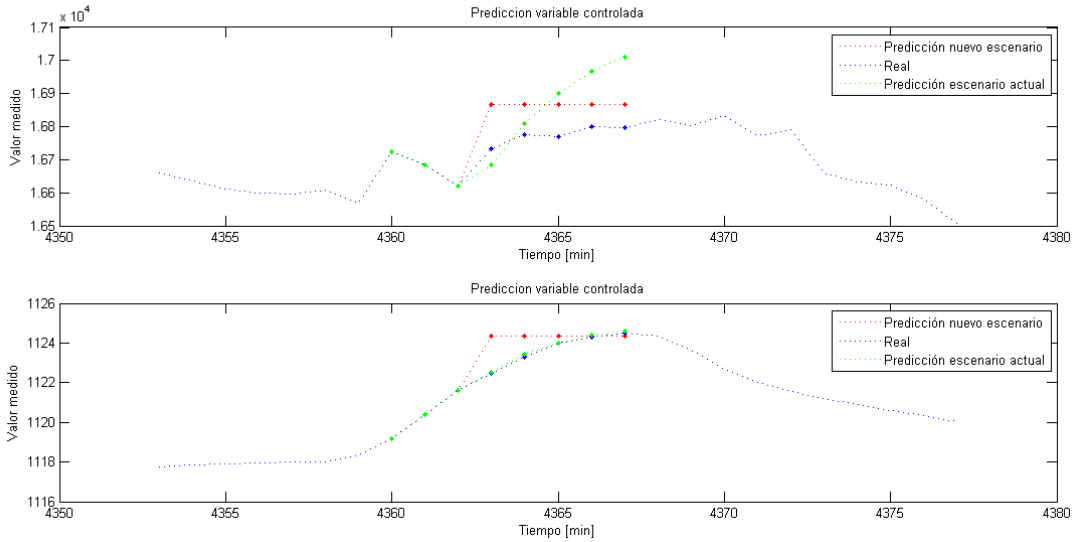


Figura 4.40 : Predicción con nuevo escenario.

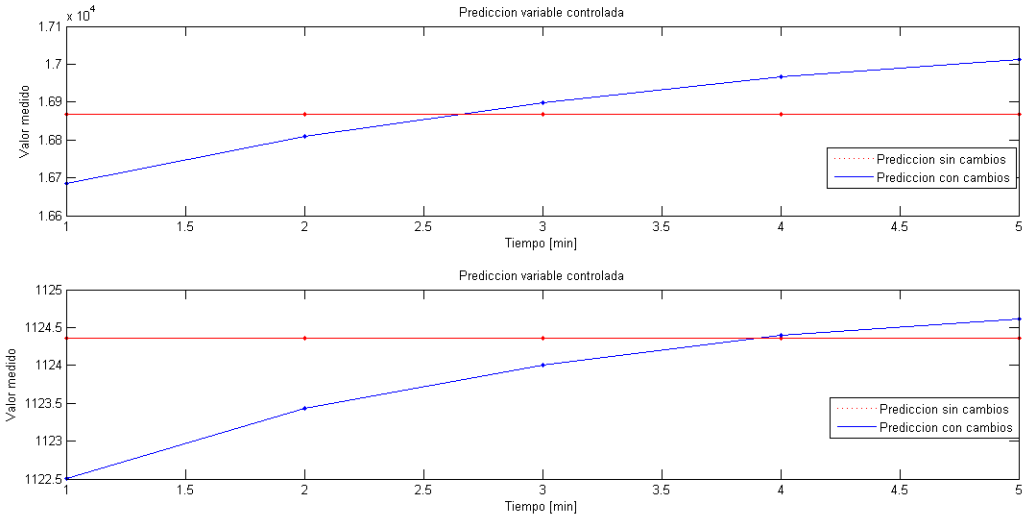


Figura 4.41 : Comparación entre escenario actual y nuevo escenario.

Como nuevo escenario se plantea aumentar el tonelaje por hora en un 5%, con esto se genera las predicciones para ese escenario. En la Figura 4.40, se observa la salida controlada N° 1, la que corresponde a la potencia consumida y la salida controlada N°2, la cual corresponde a la presión del molino. En la imagen, se aprecia que al aumentar el tonelaje, trae consigo un aumento en la potencia

consumida y en la presión de descanso del molino. Por otro lado en la Figura 4.41 , se aprecia la comparación entre la predicción del escenario actual y la predicción del nuevo escenario (aumento del 5% en el tonelaje). Para determinar si es un mejor desempeño que el escenario anterior, se observa la variable de consumo de energía específico, el cual es la potencia consumida con respecto al tonelaje en el molino.

$$CEE = \frac{Potencia}{Tonelaje} \quad (4.2)$$

	Escenario Actual	**Nuevo escenario**
Consumo energético especifico (CEE)	7.2186	7.1252

Tabla 4.7 : Consumo de energía específico.

En la Tabla 4.7, se observa el consumo de energía específico para el escenario actual, es decir, el modo de operación que actual del molino, y el CEE del nuevo escenario con un tonelaje aumentado en un 5%. Se aprecia que con el nuevo escenario, el CEE disminuye en un 1.31%, siendo que aumentó el tonelaje, por lo tanto con ese nuevo escenario se obtiene un mejor desempeño.

CAPÍTULO 5. CONCLUSIONES

El monitoreo y análisis de procesos es un tema relevante en la industria, sobre todo en Chile, país con mayor producción de cobre a nivel mundial. En la minería, como toda industria, se busca minimizar costos operacionales y humanos. Por ello, se desarrollaron herramientas para la estimación, detección de anomalías y predicción de variables usando modelos de similitud (SBM). Dado que se disponían de una gran base de datos, se decidió utilizar una herramienta de modelación no paramétricas para modelar los procesos, ya que por lo general, operan bajo ciertos límites de operación y gracias a los datos se pueden encontrar ciertos patrones entre variables, como a la vez, patrones de operación del proceso.

En primer lugar, se realizó un algoritmo para generar un modelo basado en similitud, el cual pudiera estimar las salidas (variables controladas) del proceso usando la primera base de datos, la cual contiene mediciones de operación normal del sistema. Los resultados muestran cómo evolucionaba el modelo SBM hasta obtener un modelo que estima de manera correcta (bajo un criterio de error cuadrático medio y error porcentual relativo) las salidas del proceso. Esto nos demuestra la potencialidad que tiene esta modelación en particular, ya que gracias a la adaptabilidad que tiene, nos permite cambiar las matrices de entrenamiento para generar diferentes propósitos.

En segundo lugar, se realizó un modelo de detección de anomalías usando modelación SBM con datos de operación normal del proceso, pertenecientes a la segunda base de datos (BD2). Con este modelo, se demostró su capacidad de detectar anomalías en un proceso de planta al observar los resultados del test de Hotelling, ya que se había definido un umbral de Hotelling que determina si un dato es normal o anormal, donde si el estadístico T^2 para esa estimación supera el umbral de Hotellling, entonces será una anomalía en el proceso.

En tercer lugar, usando la misma base de datos que el modelo de estimación de variables (BD1) y aprovechando la versatilidad de la modelación SBM, se realizó un algoritmo para crear un modelo capaz de predecir las salidas del proceso utilizando un nuevo modelo basado en similitud. Con solo cambiar las variables pertenecientes a cada matriz de entrenamiento, es posible generar un nuevo modelo que predice a un paso hasta 5 pasos con un error porcentual relativo asociado menor al 1% (en valores no normalizados). Con ello, se pueden realizar diferentes escenarios para buscar un mejor rendimiento o desempeño del molino SAG.

En cuarto lugar, utilizando el mismo modelo para predecir variables, se realizaron algunos posibles nuevos escenarios y lograr obtener un mejor desempeño del molino SAG. Para crear un nuevo escenario se mantuvieron congeladas las entradas al modelo, pero se modificaron una sola variable independiente, agregando un $\pm 5\%$ de su valor actual. Luego, el modelo generó salidas hasta cinco instantes de tiempo más adelante. Para validar su mejor desempeño, se observó la variable de consumo de energía específico, la cual nos permite observar la cantidad de potencia consumida con respecto al tonelaje procesado en el molino. Entonces si la potencia se mantiene constante pero aumenta el tonelaje, el consumo de energía especifico disminuirá y se considerara un mejor desempeño, pero también se puede dar el caso en que el tonelaje se mantenga constante y la potencia aumente, entonces el consumo de energía también aumenta y se considerara un peor desempeño.

Finalmente, se propone como trabajo a futuro, implementar una herramienta que pueda realizar pronósticos en vez de predicción, con el fin de crear un modelo usando similitudes que sea capaz de detectar anomalías de manera anticipada. Esto se puede lograr integrando algún algoritmo que estime los valores futuros de las variables independientes de tal forma que la predicción a realizar estime las salidas del proceso usando los valores futuros de las variables manipuladas. Además, se propone a futuro, realizar una comparación entre el futuro pronóstico y las predicciones hechas con modelos de similitud.

REFERENCIAS

1. Alejandro, León; “Detección de anomalías en procesos industriales usando modelos basados en similitud” Diciembre 2011.

2. Fuentealba, Sebastián; “Diseño e implementación de un sistema supervisor con modelos basados en similitud para la detección y aislamiento de fallas en turbina a gas natural”, Julio 2012.

3. Tomás Carricajo , Felipe Kripper , Marcos E. Orchard , Luis Yacher, Rodrigo Paredes; “Anomaly Detection in Gas Turbine Compressor of a Power Generation Plant using Similarity-based Modeling and Multivariate Analysis” Annual Conference of the Prognostics and Health Management Society 2013.
4. Gonzalez G. D., M. Orchard, J.L. Cerda, A. Casali & G. Vallebuona (2003), “Local models for soft-sensors in a rougher flotation bank," Minerals Engineering, vol. 16, no.5, pp. 441-453.

5. Jackson J. E. (1991), “A users guide to principal components,” Wiley.

6. S. Wegerich and X. Xu, “A performance comparison of similarity based and kernel modeling techniques,” in Proc. of MARCON 2003, TN, May 2003.

7. Liuling Gong, Dan Schonfeld; “Space Kernel Analysis”, University of Illinois at Chicago, Dept. of Electrical and Computer Engineering 851 S Morgan St, Chicago, IL 60607

8. Jozsef Bokor; Zoltan Szabo; “Fault detection and isolation in nonlinear systems”, Octubre 2009.

9. Pivoso M. J. & Kosanovich K. A. (1994), “Applications of multivariate statistical methods to process monitoring and controller design,” Int. J. of Control, vol. 59, pp. 743-765

10. Shengwei Wang; Fu Xiao; “Detection and diagnosis of AHU - sensor faults using PCA method”, Enero 2004.

11. Rolf Isermann; “Model- based fault-detection and diagnosis - status and applications”, Diciembre 2004.

12. Beale G. O. & Kim J. H. (2002), "Fisher discriminant analysis and the T2 statistic for process fault detection and classification," Industrial Electronics Society, IEEE 2002 28th Annual Conference, vol. 3, pp. 1995- 2000, 5-8.

13. Chiang L. H., Russell E. L. & Braatz R. D. (2001), “Fault Detection and Diagnosis in Industrial Systems,” Springer Verlag London Limited.

14. Fuente M. J., Garcia-Alvarez D., Sainz-Palmero G. I. & Villegas T. (2009), "Fault detection and identification method based on multivariate statistical techniques," Emerging Technologies & Factory Automation 2009, IEEE Conference, pp.1-6, 22-25.

15. Rolf Isermann; “Supervision, fault-detection and fault-diagnosis methods”, Marzo 1997.

16. V. Chandola, A. Banerjee, V. Kumar, "Anomaly Detection: A Survey", Department of Computer Science and Engineering University of Minnesota, Agosto 2009.

17. John MacGregor; Ali Cinar; “Monitoring, fault diagnosis, fault - tolerant control and optimization”, Junio 2012.

18. Philip Nelson; Paul Taylor; John McGregor; “Missing data methods in PCA and PLS Score calculations with incomplete information”, Enero 1996.

19. R. Patton, P. Frank, R. Clarke,” Fault diagnosis in dynamic systems: theory and application”, Prentice-Hall, 2000.

20. Theordora Kourti; John McGregor; “Process analysis, monitoring and diagnosis, using multivariate projection methods”, Noviembre 1994.

21. Theordora Kourti; John McGregor; “Statistical process control of multivariate processes”, Enero 1995.

22. Theordora Kourti; John McGregor; Paul Nomikos; “Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS”, 1995.

23. Daniel Hodouin; “Methods for automatic control, observation and optimization in mineral processing plants”, Noviembre 2010.

24. Wenyi Wang; David Forrester; Peter Frith; “A generalized machine fault detection method using unified change detection”, 2014.

25. Isermann R., Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance. Heidelberg: Springer, 2006, pp. 5-19.

26. Isermann R.; Ballé P.,"Trends in the Application of Model-Based Fault Detection and Diagnosis of Technical Processes". Control Engineering Practice, vol. 5, no. 5, pp. 709-719, May 1997.

27. Isermann, R., "Process fault detection based on modeling and estimation methods – A survey," Automatica, vol.20, no.4, pp.387-404, July 1984.

28. S. Wegerich, "Similarity based modeling of time synchronous averaged vibration signals for machinery health monitoring", IEEE Aerospace Conference, vol. 6, pp. 3654- 3662, Marzo 2004.

29. Oppenheim, A. V.; Willsky, A. S., Signal and Systems, New Jersey: Prentice Hall, 1996, pp. 38-56.

30. Tahir Mehmood⋆, Kristian Hovde Liland, Lars Snipen, Solve Sæbø ," A review of variable selection methods in Partial Least Squares Regression", April 2012.