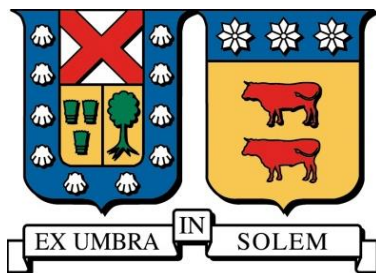


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE INGENIERÍA QUÍMICA Y AMBIENTAL

VALPARAÍSO - CHILE



**DESARROLLO DE UN SENSOR VIRTUAL DINÁMICO DE LEY DE
CONCENTRADO EN UNA COLUMNA DE FLOTACIÓN USANDO
METODOLOGÍA PLS**

**MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL QUÍMICO**

ROBINSON RAÚL ROJAS SANTIBÁÑEZ

**PROFESOR GUÍA: DR. LUIS BERGH O.
PROFESOR CO-REFERENTE: DR. CLAUDIO ACUÑA P.**

Valparaíso, diciembre del 2023

RESUMEN

Para que los sistemas de control puedan monitorear y controlar los procesos de forma exitosa, se deben alimentar oportunamente con información confiable respecto las de variables críticas. En específico para controlar una columna de flotación se debe medir en línea la ley de concentrado, lo que requiere equipos de gran inversión y mantención. La opción tradicional de medir la concentración mediante métodos de laboratorio, queda limitada también por la frecuencia y retraso de la información. Una alternativa es el uso de un sensor virtual, que es un modelo capaz de predecir variables utilizando otras con mayor disponibilidad, en este sentido el objetivo de este trabajo es desarrollar y evaluar un sensor virtual dinámico para la ley de concentrado en una columna de flotación empleando la metodología de PLS Dinámico (Projection to Lattice Structure), dicha metodología incluye valores pasados de las variables en el bloque de entrenamiento para capturar la dinámica del proceso, superando las limitaciones de otras regresiones convencionales.

La aplicación de esta investigación está basada en simulaciones del proceso, para lo cual se dispone de un simulador que utiliza modelos fenomenológicos que permiten predecir en estado estacionario la ley de concentrado, adicionalmente este es acoplado con un modelo de primer orden que permite calcular la trayectoria transiente de la ley de concentrado. El simulador cuenta con 8 variables disponibles: 5 de la alimentación y 3 operacionales. Para la colección de datos se caracteriza una columna de flotación cilíndrica con 14 m de altura y 5 m de diámetro, en la cual se realizan perturbaciones del tipo escalón con todas las variables disponibles, esto para incluir efectos tanto de la alimentación como operacionales.

La implementación de los modelos se realizó utilizando el *software* SOLO. El primer caso con modelos tipo FIR, los cuales no incluyen *lags* de la ley de concentrado, este fue abordado con enfoques tanto global como locales. La determinación de las zonas locales se realizó mediante la clasificación de los residuos del modelo global usando arboles de decisión, logrando una disminución del error desde 0,6807 a 0,4138. Adicionalmente se ha determinado un modelo del tipo ARX, el cual, si incluye *lags* de la ley de concentrado, este modelo presenta un error de 0,2737. Sin embargo, muchas veces no se cuenta con información anterior de la variable objetivo como input para el modelo, luego, la alternativa es calcular de forma recursiva la predicción utilizando el valor de la predicción anterior, obteniéndose en este caso un error de 0,8314. La propuesta de

una alternativa recursiva, con la precaución de reiniciar el sensor en intervalos regulares para mitigar la propagación del error, busca abordar esta limitación.

Respecto de la coherencia de los vectores de regresión, las variables con mayor relevancia tienen que ver con la ley de entrada, la velocidad superficial de gas y la altura de espuma, lo que es consistente con la fenomenología esperada, ya que son variables que gobiernan el proceso. Para todos los casos la varianza acumulada del bloque X, ha sido superior al 80%, y en relación al bloque Y supera el 95%. Respecto de los errores absolutos máximos en los cuatro casos no sobrepasan el umbral de 2 (%Cu) a excepción del modelo ARX calculado de forma recursiva.

Se realizó un análisis de propagación en relación a la incertidumbre de la medición, para lo cual se establecen niveles para el error de medición desde 1% a 40%. Para el sensor tipo FIR la variable con mayor influencia corresponde a la ley de entrada, generando una variación en la ley de concentrado desde 0,14% a 5,74%. Para el modelo ARX, la variable más sensible corresponde a la ley de concentrado del periodo anterior, generando una variación en la respuesta desde 0,21% a 8,41%. Lo anterior indica que en una aplicación real es importante contar con programas y estrategias que aseguren la calidad de la medición para disminuir los errores de predicción.

En caso de mantener controlada la incertidumbre de mediciones y controlar la propagación del error el modelo más representativo es ARX-Recursivo ya que el modelo que lo describe presenta un menor error, y adicionalmente presenta una menor complejidad al incluir solo 1 *lag*. Si no se puede controlar la propagación del error lo recomendado es la utilización del modelo Local FIR.

Respecto de una aplicación en un entorno real se esperan diferencias en cuanto a la calidad y cantidad de la información disponible, por tanto, se deben verificar cuidadosamente datos anormales mediante los estadígrafos Q-Residuals y Hottelling's. Además, en un proceso real existen variables no medidas que tienen influencia en el proceso, y al no ser incluidas en el modelo deterioran la calidad de la predicción.

Si se logra disminuir la incertidumbre en las variables que alimentan el sensor, mejorar la calidad de la información disponible y caracterizar lo mejor posible variables, la aplicación de un sensor virtual podría aportar con una inmediatez en la información disponible y podrían ser instalados en sistemas de control avanzados en cualquier tipo de industria.

LISTA DE CONTENIDO

RESUMEN.....	II
LISTA DE CONTENIDO.....	IV
LISTA DE TABLAS	VII
LISTA DE FIGURAS	VIII
NOMECLATURA.....	X
1. INTRODUCCIÓN	1
1.1. CONTEXTO.....	1
1.1.1. <i>Beneficio de minerales</i>	<i>1</i>
1.1.2. <i>Situación minera actual en Chile y sus indicadores claves</i>	<i>2</i>
1.1.3. <i>Sistemas de control en flotación.....</i>	<i>3</i>
1.2. OBJETIVOS.....	6
1.2.1. <i>Objetivo General</i>	<i>6</i>
1.2.2. <i>Objetivos específicos</i>	<i>6</i>
1.3. ESTRUCTURA DEL DOCUMENTO	6
2. ANTECEDENTES GENERALES	8
2.1. PROCESO DE FLOTACIÓN EN COLUMNAS	8
2.1.1. <i>Fenomenología.....</i>	<i>8</i>
2.1.2. <i>Equipos.....</i>	<i>9</i>
2.2. OBJETIVOS DEL PROCESO METALÚRGICO	11
2.3. VARIABLES DEL PROCESO	12
2.4. SISTEMAS DE CONTROL EN FLOTACIÓN COLUMNAR	15
2.4.1. <i>Medición de la ley de concentrado</i>	<i>16</i>
2.5. MÉTODOS DE PROYECCIÓN	18
2.5.1. <i>Análisis de componentes principales (PCA)</i>	<i>19</i>
2.5.2. <i>Partial Least Square (PLS)</i>	<i>22</i>
2.5.3. <i>PLS Dinámico</i>	<i>24</i>

2.6.	SIMULACIÓN EN PROCESOS DE FLOTACIÓN	28
2.7.	ÁRBOLES DE DECISIÓN.....	29
3.	MATERIALES Y METODOLOGIAS	33
3.1.	SIMULADOR.....	33
3.2.	DEFINICIÓN DE VARIABLES DE PROCESO DISPONIBLES EN SIMULADOR	34
3.2.1.	<i>Variables de entrada</i>	34
3.2.2.	<i>Variables de salida</i>	36
3.3.	RESPUESTA TRANSIENTE.....	37
3.3.1.	<i>Sistema de primer orden</i>	38
3.3.2.	<i>Constante de tiempo</i>	39
3.3.3.	<i>Retardo por transporte</i>	40
3.4.	DISEÑO DE EXPERIMENTOS	43
3.4.1.	<i>Selección de variables y sus niveles</i>	43
3.4.2.	<i>Matriz de experimentos</i>	45
3.4.3.	<i>Generación del set de datos</i>	46
3.5.	IMPLEMENTACIÓN MODELO PLS DINÁMICO	48
3.5.1.	<i>Variables del modelo PLS y frecuencia de muestreo</i>	48
3.5.2.	<i>Software SOLO</i>	49
3.6.	MODELO FIR.....	51
3.6.1.	<i>Bloque X e Y</i>	51
3.6.2.	<i>Componentes principales</i>	52
3.6.3.	<i>Límites estadígrafos T^2 y Q</i>	54
3.6.4.	<i>Análisis de residuos</i>	55
3.7.	MODELO CORREGIDO.....	56
3.7.1.	<i>Caracterización de los residuos</i>	57
3.7.2.	<i>Clasificación de los residuos</i>	58
3.7.3.	<i>Construcción de modelos locales</i>	60
3.7.4.	<i>Límites estadígrafos T^2 y Q</i>	60
3.7.5.	<i>Análisis de residuos</i>	63
3.8.	MODELO ARX.....	63
3.8.1.	<i>Bloque X e Y</i>	63

3.8.2.	<i>Componentes principales</i>	64
3.8.3.	<i>Límites estadígrafos T^2 y Q</i>	65
3.8.4.	<i>Análisis de residuos</i>	66
3.8.5.	<i>Predicción recursiva</i>	67
3.8.6.	<i>Convergencia del modelo</i>	68
4.	DISCUSIÓN DE RESULTADOS	70
4.1.	RESUMEN DE MODELOS	70
4.2.	DISTRIBUCIÓN DE RESIDUOS	71
4.3.	ANÁLISIS DE PROPAGACIÓN DEL ERROR	73
4.4.	CONSOLIDADO	75
5.	CONCLUSIONES	78
6.	RECOMENDACIONES	80
7.	REFERENCIAS	83
8.	ANEXOS	87
	ANEXO 1 : VECTORES DE REGRESIÓN	87
	ANEXO 2 : CONSTRUCCIÓN MODELOS LOCALES	90

LISTA DE TABLAS

3-1 Variables de alimentación disponibles y sus rangos.....	34
3-2 Variables de operación disponibles y sus rangos.....	35
3-3 Parámetros geométricos de la columna simulada	36
3-4 Matriz de variables y sus niveles	44
3-5 Dimensiones de la columna de flotación simulada	44
3-6 Resultados metalúrgicos obtenidos	47
3-7 Detalle de las variables consideradas en el modelo PLS.	49
3-8 Resumen estadístico de las componentes principales en modelo FIR	53
3-9 Límites banda de aceptación y N ° de observaciones	57
3-10 Modelos locales: Subconjuntos de aplicación y observaciones.....	59
3-11 Resumen de CP, Varianza acumulada de X e Y, RMSECV en cada modelo local.....	60
3-12 Resumen estadístico de las componentes principales en modelo ARX.....	64
4-1 Indicadores de los modelos obtenidos	70
8-1 Vector de regresión Modelo FIR	87
8-2 Vector de regresión Modelo Local 1	87
8-3 Vector de regresión Modelo Local 2	88
8-4 Vector de regresión Modelo Local 3	88
8-5 Vector de regresión Modelo ARX	89
8-6 Resumen estadístico de las componentes principales en modelo Local 1	90
8-7 Resumen estadístico de las componentes principales en modelo Local 2	90
8-8 Resumen estadístico de las componentes principales en modelo Local 3	91

LISTA DE FIGURAS

1-1 Producción de cobre en Chile por producto año 2005-2020 (COCHILCO, 2021)	3
2-1 Diagrama conceptual flotación.	9
2-2 Esquema de celdas mecánicas (Yianatos & Vinnet, 2015).....	10
2-3 Esquema de una columna de flotación.....	11
2-4 Diagrama de control: Estrategia estabilizadora (Bergh & Yianatos, 2003).....	16
2-5 Analizador Courier® 5X/6X SL (Metso Outotec, 2023).....	17
2-6 Esquema funcionamiento simulador metalúrgico.....	28
2-7 Esquema de un árbol de decisión.....	30
3-1 Interfaz gráfica simulador SiMetCol	33
3-2 Esquema simulador estático utilizado.....	37
3-3 Esquema entrada y salida sistema.....	38
3-4 Desplazamiento del agregado en la zona de espuma	41
3-5 Esquema integración del simulador estático y dinámico.....	43
3-6 Resultados simulación: Ley de concentrado y niveles operacionales	46
3-7 Modulo análisis regresión PLS	50
3-8 Modelo FIR vs Simulador, Q Residuals	54
3-9 Modelo FIR vs Simulador, Hotelling's Reducido	55
3-10 Residuos Modelo FIR	56
3-11 Modelo árbol de decisión: Errores modelo FIR.....	58
3-12 Modelos Locales vs Simulador, Q Residuals	61
3-13 Modelos Locales vs Simulador, Hotelling's Reducido.....	62
3-14 Residuos Modelos Locales	63
3-15 Modelo ARX vs Simulador, Q Residuals	65
3-16 Modelo ARX vs Simulador, Hotelling's Reducido	66
3-17 Residuos Modelo ARX	67
3-18 Residuos Modelo ARX-Recursivo	68
3-19 Efecto del valor inicial en la convergencia del modelo ARX.....	68
4-1 Distribución de errores para los modelos generados	72
4-2 Grafios P-P : Verificación normalidad de residuos	73
4-4 Análisis propagación del error de medición en Ley de concentrado modelo FIR.....	74

4-5	Análisis propagación del error de medición en Ley de concentrado modelo ARX.....	75
4-6	Comparación datos experimentales vs predicción de modelos.....	77

NOMECLATURA

- C : Flujo másico de concentrado
- F : Flujo másico de alimentación
- m : Número de observaciones
- n : Número de variables regresoras
- p_i : *Loadings* del bloque X
- R : Recuperación de cobre
- X : Matriz de datos con m observaciones y n columnas
- x_c : Ley del mineral valioso en el concentrado
- x_f : Ley del mineral valioso en la alimentación
- λ_j : Eigenvalue o valor propio asociado a cada *loadings*
- t_i : *Scores* del bloque X
- Q : Estadígrafo Q - *Residuals*
- T^2 : Estadígrafo Hotelling's – T^2
- Y : Matriz de datos con m observaciones y q columnas
- k : Número de variables de salida u objetivo
- w_i : *Weights* del modelo PLS
- u_i : *Scores* del bloque Y
- q_i : *Loadings* del bloque Y
- h : Número de *lags* o valores pasados PLS dinámico

1. INTRODUCCIÓN

1.1. Contexto

1.1.1. Beneficio de minerales

Los minerales por definición son sustancias inorgánicas naturales que poseen una composición y estructura químicamente definida, los cuales se han formado bajo diversos procesos de condensación, solidificación y disolución en un rango de tiempo milenario. Luego definimos la mena como un material natural del que se pueden extraer uno o más minerales con un beneficio económico. La mena está constituida por minerales cuya extracción representa y produce un beneficio económico, y de otros los cuales carecen valor o su aprovechamiento sería demasiado costoso. Los primeros se les denomina minerales de mena y a los segundos minerales de ganga (Wills & Finch, 2016).

Entendiendo lo anterior, se le llama beneficio de minerales al conjunto de operaciones que se deben llevar a cabo para convertir el producto geológico que se extrae desde la mina en mineral útil para los procesos propios de la metalúrgica extractiva. Estos procesos usualmente se pueden clasificarse como sigue:

- Explotación del yacimiento
- Procesos de liberación y conminución (Chancado, molienda, clasificación, acondicionamiento)
- Procesos de separación o selectivos (Flotación, lixiviación, separación magnética, entre otros)

Los procesos de liberación y conminución tienen como objetivo homogeneizar el material extraído desde la mina, reducir el tamaño de partícula y por consiguiente exponer o liberar el material de interés desde la ganga preparándolo así para la etapa siguiente. Por consiguiente, el proceso de separación tiene como objetivo separar las partes de valor para obtener una porción enriquecida con una mayor cantidad de material con valor, obteniéndose así un concentrado y un descarte, también llamado relave o cola, reduciendo la cantidad de material que debe ser tratada en posteriores etapas (Hodouin, et al., 2000).

Estos procesos de concentración dependen mucho del tipo de mena, para el caso de Chile, la concentración de minerales en especial del cobre se da principalmente en dos menas, las del tipo oxidado y las sulfuradas, la primera usualmente se encuentra en minas de rajo, mientras que las sulfuradas se dispone en menas subterráneas, de esta forma las rutas de concentración más utilizadas en Chile corresponden a lixiviación o flotación respectivamente.

1.1.2. Situación minera actual en Chile y sus indicadores claves

La minería se posiciona como uno de los principales sectores económicos que sustentan la economía chilena, dentro de los últimos años 20 años la participación de este sector se ha mantenido fluctuando entre un 10-11%, en el cual la minería del cobre aporta con 9% (COCHILCO, 2021), sobrepasando por sí solo a otras actividades económicas como por ejemplo transporte y construcción.

Las exportaciones de cobre se han mantenido de forma constante en más o menos 5500 kilotoneladas métricas finas, que consisten principalmente en productos de concentrado de cobre, cátodos vía SX-EW y productos de fundición. Sin embargo, como se puede apreciar en la Figura 1-1, la producción de cátodos vía SW-EX, obtenidos a partir de menas con alto contenido de minerales oxidados se han vuelto más escasos, por consiguiente, la producción de concentrados de cobre obtenidos vía flotación ha aumentado, en efecto desde el año 2010 hasta el 2022, el aumento fue de un 22%, cifra no menor considerando que se trata de una exportación de poco valor agregado.

El panorama futuro no es muy distinto, se espera que siga en aumento, ya que los sulfuros de cobre son cada vez más abundantes, mientras que los óxidos de cobre son cada vez más escasos, según proyecciones de Cochilco, estos últimos pasarán de ser un 30,8% en el 2015 a un 12% en el año 2027 (COCHILCO, 2017).

Sumando a lo anterior, si bien es posible procesar vía lixiviación minerales de cobre sulfurados, se tienen principalmente dos dificultades, su baja cinética y además que este tipo tiende a formar una capa que impide que el agente lixivante reaccione con el metal. Existen algunas técnicas para lixiviar sulfuros, mediante la utilización de bacterias como catalizadores, llamada biolixiviación, consiguiendo un bajo costo, ausencia de emisiones tóxicas, un menor consumo de agua, a costa de

condiciones de proceso muy especiales de temperatura, presión y pH, en conjunto de un prolongado de tiempo de proceso hacen de su uso en el horizonte cercano improbable.

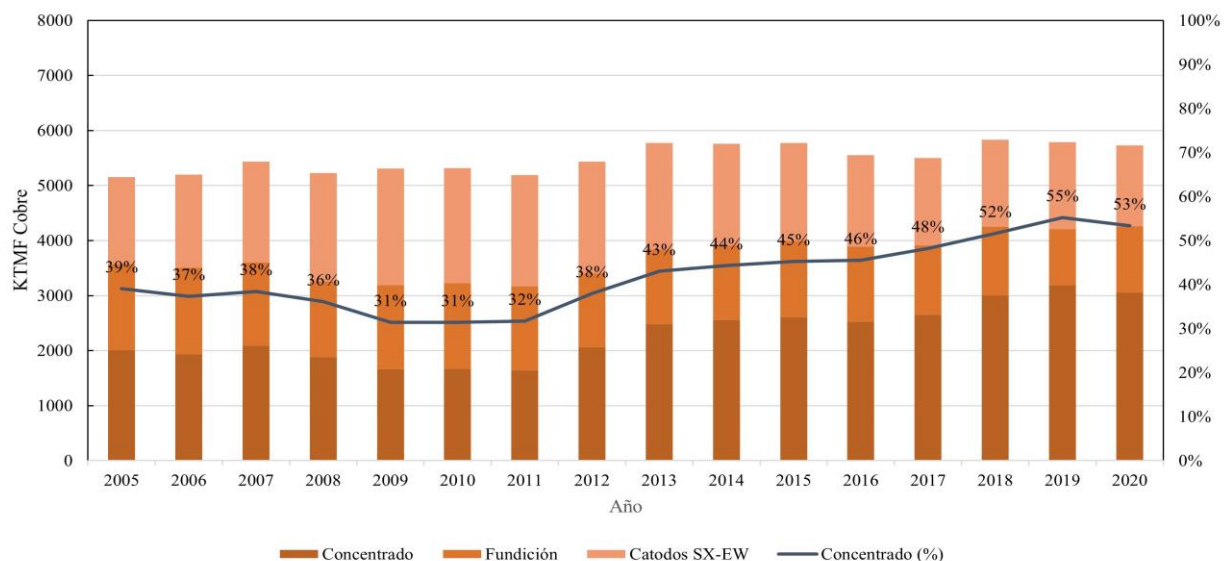


Figura 1-1 Producción de cobre en Chile por producto año 2005-2020 (COCHILCO, 2021)

El uso de agua es un punto importante, ya que el uso de dicho recurso en el proceso de concentración, alcanzó 0,357 metros cúbicos por tonelada de mineral procesado (COCHILCO, 2021), luego considerando este y los antecedentes mencionados, que la flotación como proceso de concentración es clave, por lo que mejorar la eficiencia es parte del desafío que tiene el negocio minero del cobre.

1.1.3. Sistemas de control en flotación

Los sistemas de control se han utilizado frecuentemente en los procesos mineros, principalmente estos sistemas tienen como objetivo reducir la variabilidad en los objetivos metalúrgicos: recuperación y ley de concentrado, a su vez una reducción en variabilidad reporta una serie de beneficios, como por ejemplo mejoras en la productividad, indicadores operacionales más consistentes que permiten una mejor toma de decisiones y reducción de costos (Yianatos & Vinnet, 2015).

Para establecer estos sistemas de control y supervisión, típicamente se utiliza la modelación de procesos, dentro lo de los cuales están:

- Modelos empíricos basados en experiencias anteriores

- Modelos estacionarios o transiente
- Modelos lineales o no-lineales
- Modelos de lógica difusa

Con lo anterior, resulta un factor importante la calidad y disponibilidad de la información con que se alimentan tales modelos, en este caso el proceso de flotación, impone algunas condiciones de trabajo adversas para los instrumentos de medición, por ejemplo, la naturaleza corrosiva de los minerales a procesar que puede afectar negativamente las mediciones y la vida útil de los instrumentos, sumado a los acotados tiempos de mantención y en algunos casos la poca accesibilidad al instrumento para realizar una rutina de mantenimiento, condiciones que pueden afectar en la calidad de la información disponible.

Por otra parte, la variabilidad intrínseca del proceso, asociada principalmente a los cambios en las características del mineral en la alimentación, como, por ejemplo, granulometría, composición mineralógica del mineral (Moolman, et al., 1996), así también las perturbaciones propias del sistema, implican que generalmente el control sea impreciso en la mayoría de las situaciones.

En relación a los controles que se han utilizado para los procesos de flotación, con mayor frecuencia han tenido un enfoque en (Shean & Cilliers, 2011):

- Flujo de aire en equipos
- Nivel de pulpa
- Dosificación de reactivos
- pH
- Flujo del agua de lavado en equipos de limpieza
- Velocidad de descarga de espuma

Al respecto se han logrado controles efectivos en columnas de flotación, usando una combinación de controles en otras variables, por ejemplo, controlar el nivel de pulpa, flujo de gas y agua de lavado con el fin de mantener la ley de concentrado de cobre en control (Bergh & Yianatos, 2003). Es decir, el control puede realizarse mediante la regulación de variables intermedias. Sin embargo,

la tendencia actual en el control es medir directamente la variable metalúrgica de interés por ejemplo la ley del concentrado, y de esta forma, según la respuesta obtenida, retroalimentar y actuar sobre otras variables de proceso, ya sea flujo de gas, flujo agua de lavado, dosificación y nivel de espuma, con el propósito de controlar la variable objetivo.

Expuesto lo anterior, entonces resulta interesante en la industria, de contar con sistemas de control y especialmente de mediciones con una alta confiabilidad y disponibilidad para poder alimentar los lazos de control de forma consistente, y en especial de las variables metalúrgicas, en este caso la ley de concentrado de cobre.

En las últimas décadas se han desarrollado diversas técnicas para la medición de esta variable, de forma tradicional y utilizadas a la fecha, se cuentan con técnicas de análisis en laboratorio por medio de la toma sistemática de muestras que son analizadas directamente en laboratorios. Si bien es un proceso estandarizado, está sujeta a errores experimentales, además de un costo asociado a cada muestra, y su principal desventaja la latencia en que se tienen los resultados lo cual dificulta la toma de decisiones en tiempo real.

En las técnicas más actuales, se tienen los analizadores en línea, diseñados y fabricados por compañías asociadas al mundo de la minería y flotación como por ejemplo Outotec, FLSmidth, entre otros, estos equipos operan mediante la difracción de rayos X (XRD), para lo cual se analiza la pulpa de concentrado en tiempo real. La desventaja de estos equipos es su alto costo de inversión y mantención, rutinas de mantenimiento exigentes y medios adversos de trabajo, por otra parte, usualmente se utiliza el mismo equipo para medir una o más corrientes, por lo que es probable que las mediciones se vean afectadas por contaminación de muestras anteriores resultando en mediciones con un grado de incertidumbre que resultará perjudicial para el control.

Una de las alternativas es el uso de técnicas de estadística en donde a partir de la información de variables de proceso más accesibles, generar un sensor basado en datos o *soft-sensor*. En ese sentido, se han aplicado con éxito algunas técnicas de estadística multivariable, que, utilizando métodos de proyección, han logrado, por ejemplo, la detección de fallas de instrumentación y operacionales en línea de forma efectiva (Bergh & Acosta, 2009).

En este trabajo, se buscará predecir la concentración de concentrado en una columna de flotación, mediante la construcción de un sensor virtual dinámico sustentado en los métodos de proyección.

Para lo cual se dispone de un simulador fenomenológico del proceso, que permitirá la recolección de datos para calibrar y luego evaluar el comportamiento del sensor frente a distintos escenarios operacionales.

1.2. Objetivos

1.2.1. Objetivo General

Desarrollar un sensor virtual dinámico, que pueda predecir la concentración de cobre en línea, en una columna de flotación, utilizando PLS.

1.2.2. Objetivos específicos

- Diseñar experimentos para diferentes condiciones de la alimentación al proceso y de las variables operacionales disponibles en el simulador.
- Generar datos, usando el simulador metalúrgico, que relacionen las variables independientes del proceso con el objetivo metalúrgico de ley de concentrado.
- Construir el sensor virtual en base a modelos en variables latentes, de acuerdo con el método estadístico de componentes principales.
- Evaluar el comportamiento del sensor virtual frente a distintos escenarios operacionales.

1.3. Estructura del documento

A continuación, se presenta una breve descripción de los capítulos que conforman este trabajo:

Capítulo 1 - Introducción: En este capítulo se expone el contexto en donde el trabajo genera su aporte y se presentan los distintos objetivos del trabajo.

Capítulo 2 - Antecedentes generales: Este apartado incluye todos los aspectos consultados respecto de la literatura y estado del arte relacionado con la implementación del sensor virtual, se presenta y discute la fenomenología del proceso de flotación, incluyendo variables y objetivos del proceso, la situación actual del control de proceso en relación con las estrategias y los analizadores en línea. Además, se incluye la teoría asociada a los métodos de proyección: PCA, PLS y su variante dinámica utilizada para esta implementación, finalizando con una introducción a la clasificación de datos mediante árboles de decisión.

Capítulo 3 - Materiales y metodología: Se detalla el flujo de trabajo realizado en la investigación para la construcción de los modelos predictivos, incluyendo la descripción de las herramientas computacionales utilizadas. El capítulo incluye desde la descripción de las variables disponibles en la simulación, la presentación del diseño experimental, siguiendo con la obtención de los datos, hasta la implementación de los modelos dinámicos FIR con su variante de modelos locales, y la obtención del modelo ARX.

Capítulo 4 - Discusión de resultados: Se discuten los resultados obtenidos por los modelos, presentando la comparación del desempeño de los modelos en base a distintos indicadores de ajuste como la distribución de errores, varianza capturada y errores de predicción.

Capítulo 5 – Conclusiones: En las conclusiones se presentan los principales resultados y aprendizajes obtenidos en la implantación del sensor virtual.

Capítulo 6 – Recomendaciones: Respecto del capítulo de recomendaciones se entregan algunas observaciones respecto de la aplicación de la metodología que fueron levantadas durante la implementación y principalmente tienen relación con el uso de esta metodología en un sitio de prueba real.

Anexos: Al final del documento se agregan dos anexos, en el primero se presentan los vectores de regresión obtenidos en los distintos modelos, y el segundo anexo presenta la información detallada de cada modelo respecto de indicadores de ajuste como la varianza capturada y errores de calibración.

2. ANTECEDENTES GENERALES

2.1. Proceso de flotación en columnas

2.1.1. Fenomenología

La flotación es un proceso físico químico de separación y concentración de minerales que ha sido ampliamente utilizando en la minería para la obtención de minerales como el cobre y el molibdeno.

La flotación se sustenta en la interacción que existe entre burbujas y partículas finas suspendidas en un medio acuoso (pulpa) (Yianatos & Vinnert, 2015). La pulpa ingresa al sistema, y producto de la disipación de energía, las partículas de mineral colisionan con las burbujas, de esta forma las partículas de interés se adhieren a la superficie de la burbuja generando un agregado partícula-burbuja.

Esta interacción se relaciona directamente con las características de hidrofobicidad de las partículas por lo que se trata de un proceso selectivo, en este caso mientras más hidrófoba sea la partícula más flotable será y por consiguiente habrá una mayor adhesión de estas partículas a las burbujas. Por el contrario, una partícula que posea mayor afinidad por el medio líquido será menos flotable.

El uso de reactivos juega un rol importante también en este proceso ya que, mediante la adición de químicos o componentes es posible modificar las características superficiales de las partículas, aumentando la hidrofobicidad de la clase mineralógica de interés y de esta forma se puede aumentar la selectividad del proceso a conveniencia y también obtener un mejor desarrollo del proceso de colección (Yianatos & Vinnert, 2015).

Luego de la formación del agregado partícula-burbuja este asciende o flota a consecuencia de tener una menor densidad respecto del medio en el que está contenido, llegando hasta la zona de espuma, en donde las partículas de interés son descargadas al concentrado por medio de un rebalse. Por otro lado, las partículas más hidrofílicas en su mayor parte ganga, forman parte de la corriente de relave o colas.

Desde el punto de vista conceptual, el proceso puede visualizarse en dos etapas: reacción y separación, según lo apreciado en la Figura 2-1, está la zona de colección en donde se alimenta la pulpa, junto con el aire y los reactivos químicos, en la cual el mineral entra en contacto con las

burbujas formándose así el agregado partícula-burbuja, las partículas que no forman el agregado son descartados por las colas.

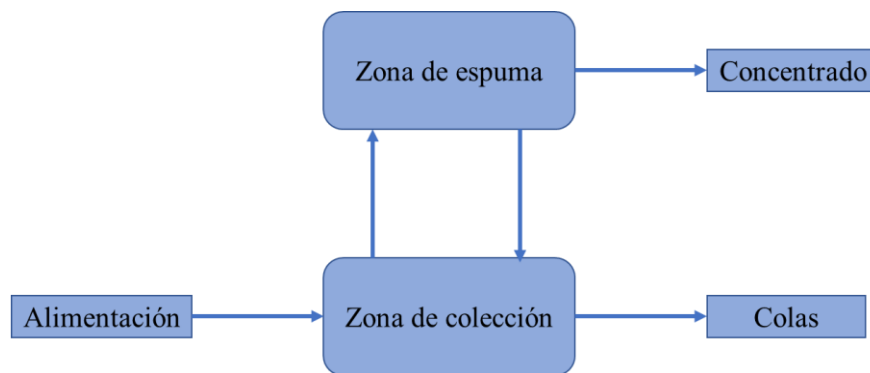


Figura 2-1 Diagrama conceptual flotación.

La zona de espuma o lavado es formada por los agregados que ascienden hasta la superficie del medio líquido, en donde se produce la descarga del concentrado por sobre nivel, en esta zona también la ganga que logró ser arrastrada y en menor cantidad partículas de interés son rechazadas hacia la zona de colección, es decir se desprenden del agregado partícula-burbuja debido al drenaje del líquido.

2.1.2. Equipos

Los equipos de flotación deben cumplir el objetivo de separar materiales de interés en sistemas multicomponente con participación de varias fases. Para ello deben ser capaces de reproducir los procesos explicados en el punto anterior.

En ese sentido, durante su diseño se deben considerar los siguientes requisitos, generar una adecuada suspensión de las partículas, generar una dispersión del gas en pequeñas burbujas y uniformes, luego de la formación del agregado partícula-burbuja, deben proveer una zona de cierta calma, para que se forme una interfase pulpa-espuma estable, y finalmente deben tener la capacidad de transportar la espuma hacia el rebalse con el fin de generar una corriente de salida con una alta concentración del mineral de interés.

En la industria se han utilizado comúnmente dos tipos, celdas mecánicas y celdas neumáticas, las cuales se describen de forma breve a continuación:

Celdas mecánicas: La principal característica de estos equipos es que tienen un sistema de agitación mecánico dedicado para mantener las partículas en suspensión, generar la distribución de burbujas y favorecer la formación de los agregados partícula-burbuja. Actualmente la fabricación es usualmente en estanques cilíndricos con volúmenes de hasta 300 metros cúbicos e inclusive más según los últimos desarrollos (Yianatos & Vinnet, 2015).

Como se puede apreciar en la Figura 2-2, se pueden subclasificar de acuerdo con la forma en que el aire ingresa al equipo: están las del tipo auto-aspirante (2-2a) en donde poseen un rotor, el cual crea una zona de baja presión generando que el aire atmosférico ingrese al sistema. Por otra parte, están las del tipo forzado, en donde el aire ingresa por medio de un sistema de inyección externo como un compresor o similar (2-2b).

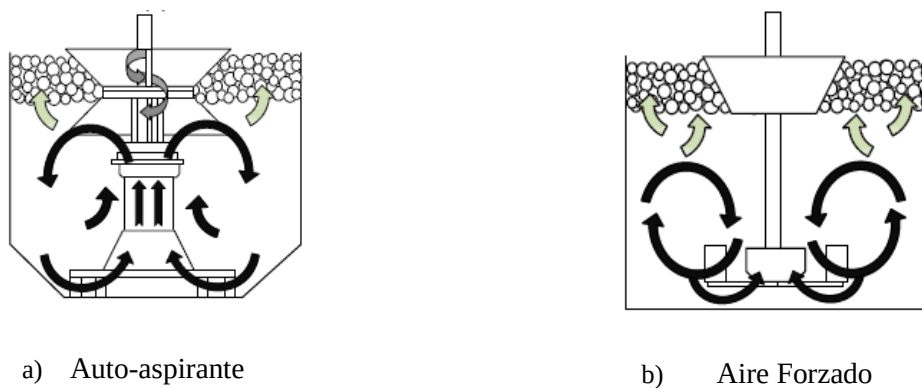


Figura 2-2 Esquema de celdas mecánicas (Yianatos & Vinnet, 2015).

Celdas neumáticas: La principal diferencia con las celdas mecánicas es precisamente la ausencia de un agitador mecánico, el aire debe ser suministrado mediante inyectores que usualmente se encuentra en la parte baja del equipo, por lo general requieren de agua de lavado ya que en la industria se utilizan principalmente para el aumento de la ley del concentrado final, es decir se puede obtener una mayor ley con respecto a su contraparte mecánica.

Otra diferencia significativa es respecto a su geometría como se puede apreciar en la Figura 2-3, su relación altura/diámetro efectivo es mayor que en las celdas mecánicas (Yianatos & Vinnet, 2015), por lo que refieren también a estas como columnas.

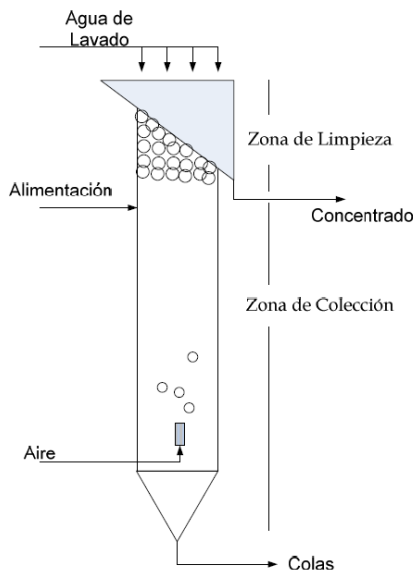


Figura 2-3 Esquema de una columna de flotación.

2.2. Objetivos del proceso metalúrgico

El resultado metalúrgico en un proceso de concentración se puede evaluar utilizando dos índices, uno que tiene que ver con la productividad y el otro con la calidad del producto final, a continuación, se definen ambos, recuperación y ley del mineral en el concentrado.

- **Recuperación**

En el caso de la recuperación se define como la fracción del mineral de interés que se recuperó en el concentrado, esto con relación al mineral de interés total que fue alimentado al equipo, este indicador puede ser calculado en términos de un balance de masa como se ve puede ver en la Ec. 2-1.

$$R = \frac{C \cdot x_c}{F \cdot x_f} \quad (\text{Ec. 2-1})$$

- Ley de concentrado

En cuanto a la ley de mineral corresponde a la fracción de mineral de interés dentro de una corriente, es decir la ley del concentrado es la razón entre la masa de mineral de interés en el concentrado y la masa de mineral total que se encuentra en el concentrado.

La ley del concentrado y la recuperación se encuentran relacionadas de forma inversa en una condición normal de operación, en sentido de que, al aumentar la recuperación, aumenta la cantidad de ganga que es arrastrada al concentrado disminuyendo así la ley del concentrado final. De forma inversa se tiene que al aumentar la pureza del concentrado se perderá material valioso arrastrado en las colas.

2.3. Variables del proceso

A continuación, se describen las principales variables que permiten simular y controlar el funcionamiento de una columna de flotación:

- Tiempo de residencia

El tiempo de residencia es uno de los factores que afectan la ley de concentrado y de esta forma también la recuperación, el tiempo medio de residencia puede estimarse mediante la relación entre el volumen en la zona de colección y el flujo volumétrico de la pulpa. En ese sentido, el tiempo de residencia se puede controlar modificando por ejemplo el flujo de alimentación, el porcentaje de sólidos en la alimentación o la altura de la zona de colección.

El tiempo de residencia para las partículas sólidas, en el caso de las celdas mecánicas es similar al del líquido, sin embargo, en una columna de flotación no es así, en este caso el tiempo de residencia depende fuertemente de la velocidad de sedimentación, de esta forma el tiempo de residencia aumenta al disminuir el tamaño de partícula, aproximándose a tiempo de residencia del líquido para partículas muy finas (Yianatos & Diaz, 2011).

- Características de la alimentación (Ley, granulometría, liberación)

En este caso, la distribución de tamaño de las partículas a procesar es uno de los factores que afectan la recuperación, ya que determinaran la interacción con las burbujas para formar los agregados burbuja-partícula, en ese sentido las partículas finas ($<20\text{ }\mu\text{m}$) (Yianatos & Vinnet,

2015) afectan disminuyendo la recuperación, ya que, si bien tienen un mayor grado de liberación y un mayor tiempo de residencia, al ser tan pequeñas tienen baja inercia por lo que tienden a moverse con el líquido y no con las burbujas.

En el otro extremo partículas muy gruesas también tienen afectan negativamente, ya que tienen un menor tiempo de residencia, menor grado de liberación y una menor eficiencia en la colección ya que debido a su mayor masa, es más probable que por la turbulencia en los equipos se desprendan de las burbujas y estas retornen a la zona de colección.

La liberación hace referencia a que tan expuesta se encuentra la superficie del mineral de interés dentro de los sólidos que constituyen la pulpa, este es un factor determinante en el proceso de flotación, ya una mayor superficie expuesta facilita que el proceso sea más selectivo aprovechando las características hidrofóbicas de las partículas.

La ley de la alimentación corresponde al porcentaje en peso de la especie mineralógica de interés, proveniente en la pulpa lo cual es muy importante ya que determinará a la cantidad disponible de material para poder ser recuperado.

- Reactivos

La adición de reactivos en el proceso de flotación es necesaria para poder tener un control más efectivo en la selectividad y en la estabilidad de la espuma, luego los reactivos más utilizados en el proceso son: colectores para la captura selectiva de minerales, espumantes para reducir el tamaño de burbuja sumado a generar una espuma mucho más estable y por último reguladores de pH, se debe cuidar por supuesto que la adición sea controlada, la cual puede ser realizarla de forma manual o automática a lo largo del proceso.

- Dispersión del aire

La dispersión del aire en el equipo es uno de los factores más influyentes en la eficiencia y calidad del proceso, en este aspecto se pueden desprender algunos parámetros, se tiene el flujo superficial de aire, este parámetro se define como la tasa volumétrica de aire por unidad de área, este parámetro indica la cantidad de área superficial disponible en la superficie de las burbujas, por lo que un aumento de este valor, tiene como consecuencia un aumento en la recuperación. Sin embargo, este aumento está limitado, por algunas condiciones de borde en el proceso, un valor

excesivo puede generar zonas de turbulencia elevadas propiciando la coalescencia de burbujas, un aumento en el tamaño de las burbujas y la pérdida de interfase.

El segundo parámetro que tiene incidencia en la dispersión del aire es el gas holdup, el cual corresponde a la fracción volumétrica de aire contendida en alguna zona del equipo, usualmente este hace referencia a la zona de colección, este parámetro tiene relación directa con la velocidad superficial de aire, el tipo de espumante utilizado, la carga de sólidos y el tamaño de las burbujas, de esta forma la distribución de tamaño de burbuja es otro parámetro de interés ya que tienen un efecto en la eficiencia de colección y transporte de partículas, el uso de burbujas pequeñas y homogéneas aumentan la cinética de colección y transporte de partículas, eso sí, se debe procurar que no sean extremadamente pequeñas ya que las burbujas muy pequeñas presentan una menor velocidad de ascenso, pudiendo ser incluso menor al descenso de la pulpa, disminuyendo la recuperación. Por el contrario, un tamaño muy grande propicia el arrastre de ganga hacia el concentrado disminuyendo la eficiencia del proceso.

- Profundidad de espuma

La profundidad del lecho de espuma toma relevancia en cuanto a la selectividad del proceso, ya que modifica la cantidad de mineral valioso que es descargado al concentrado, en este sentido una mayor profundidad implica un mayor tiempo de residencia en la zona de limpieza, así favoreciendo el drenaje de agua contenida en la espuma, reduciendo la recuperación de agua y por consiguiente disminuyendo el arrastre de ganga, aumentando así la ley del concentrado final.

A su vez el incremento en el tiempo de residencia en la zona de limpieza aumenta la probabilidad de que las partículas se desprendan de las burbujas, provocando un efecto de disminución en la recuperación. Al contrario, una disminución de la profundidad aumentará la recuperación y provocará una disminución en la ley del concentrado, luego con la profundidad de espuma se puede controlar el objetivo metalúrgico establecido para cada etapa de flotación.

- Agua de lavado

El agua de lavado es un recurso utilizado en la flotación columnar que se agrega generalmente sobre el tope o interior de la espuma, el agua de lavado busca evitar el arrastre de ganga fina hacia la corriente de concentrado, disminuyendo la concentración de insolubles. La adición de agua fresca reemplaza el agua de alimentación que es recuperada en el concentrado, también aumenta la

estabilidad al disminuir la coalescencia de las burbujas. Para que la adición sea efectiva, la adición del agua de lavado debe realizarse de forma homogénea, abarcando toda el área de espuma, y la cantidad que se adiciona debe permitir que el lecho de espuma sea estable.

- Bias

Respecto de lo anterior, se desprende el Bias, que representa el flujo neto de agua que fluye a través del colchón de espuma, se considera un Bias positivo cuando el flujo de agua de lavado es mayor que el flujo de agua recuperada en el concentrado, generando un flujo neto descendente de agua, este es uno de los principales responsables del aumento en la ley del concentrado, ya que evita el arrastre hidráulico de ganga hacia la parte superior de la columna.

2.4. Sistemas de control en flotación columnar

En la industria de procesamiento de minerales es común el uso de equipos de flotación en parte a su eficiencia, en donde el resultado final dependerá directamente de las condiciones de operación y las características del mineral a procesar. En general, el control en los equipos de flotación busca maximizar tanto la ley del concentrado como la recuperación, sin embargo, como fue explicado anteriormente, de acuerdo con las curvas de separabilidad los objetivos metalúrgicos se mueven de forma inversa, ya que al aumentar la recuperación se disminuye la ley de concentrado, y si se trata de aumentar la ley del concentrado, disminuirá la recuperación, por lo que se debe buscar la mejor estrategia de acuerdo con las condiciones de cada operación.

Al menos para el caso de la concentración de cobre las características del producto final serán las que determinarán el precio, por lo que se hace importante contar con un proceso controlado para disminuir la variabilidad del producto y de esta forma se obtendrán mayores beneficios económicos. Sin embargo, la medición de estos índices en tiempo real requiere de equipos con un alto costo de inversión, que, sumado la complejidad de funcionamiento y mantenimiento sean opciones poco usadas como información de entrada para una estrategia de control.

Debido a lo anterior, es común encontrar que las principales estrategias en equipos de flotación se enfoquen en el control de otras variables u objetivos secundarios como flujos de aire y agua de lavado, profundidad de espuma, además de un constante análisis de concentraciones y pH, esto con el propósito de mantener una operación estable y obtener un concentrado con una ley en un rango aceptable. Un ejemplo simplificado con la instrumentación tipo se presenta en la Figura 2-4.

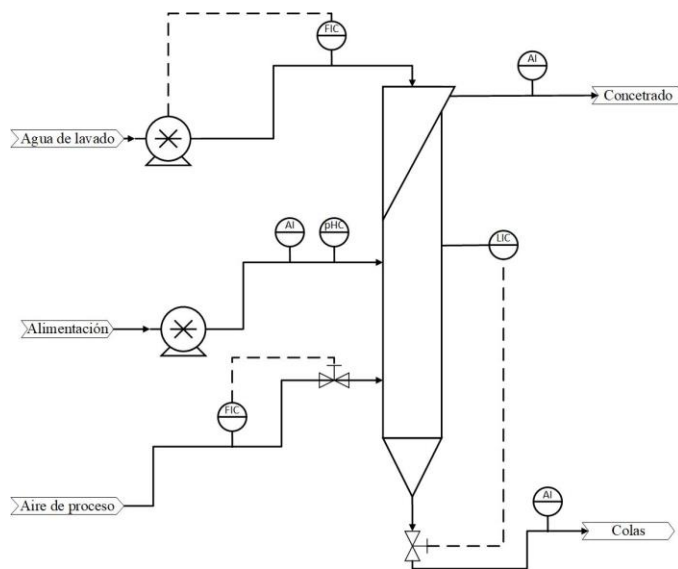


Figura 2-4 Diagrama de control: Estrategia estabilizadora (Bergh & Yianatos, 2003).

Aquellas estrategias son llamadas estabilizadoras (Bergh & Yianatos, 2003), sin embargo, este tipo de estrategias de control usualmente resulta en una alta variabilidad del producto final debido a la falta de mediciones directas, sumado a otros factores que no son considerados como una alta interacción entre variables y perturbaciones intrínsecas del proceso ejemplo de ello son los frecuentes cambios en la alimentación.

2.4.1. Medición de la ley de concentrado

Dada la tendencia a controlar mediante un lazo de control cerrado, se hace importante que la variable objetivo este lo suficientemente disponible y en la frecuencia adecuada como para ser usada en el control de procesos.

En específico, para la medición de la ley de cobre y minerales, se ha utilizado con frecuencia y se siguen utilizando, técnicas de laboratorio y análisis instrumental, por medio de la toma sistemática de muestras en cada una de las líneas de interés. Para ello se siguen procedimientos según normas internacionales que en la mayoría de los casos requieren la digestión de las muestras en un medio ácido y su posterior análisis. Si bien son procedimientos estandarizados, requieren de personal dedicado a el análisis de las muestras, con un costo asociado a cada muestra, entre los \$800 a \$1000 (Cayulef, 2017), y un tiempo de espera propio del análisis experimental, siendo esta su principal desventaja y limitante con relación al uso en el control directo del proceso.

Respecto del interés por contar mediciones en línea es que las principales empresas que se dedican al diseño y fabricación de equipos, como Outotec, FLSmidth han desarrollado y actualmente comercializan analizadores en línea, que utilizan la difracción de rayos X para determinar la ley de cobre en tiempo real. Estos equipos son instalados en la operación y se conectan directamente a las corrientes del proceso de interés, es común que un equipo tenga capacidad de analizar hasta 20 líneas de proceso como se ve en el ejemplo de la Figura 2-5, desde la cual envían una pequeña muestra hacia el equipo y este es capaz de analizarla entregando la concentración de cobre en tiempo real. La principal desventaja de estos equipos es que son muy costosos, requiriendo una inversión de capital mayor que fácilmente puede alcanzar el millón de dólares e incluso más (Metso Outotec, 2023). Además, por su método de operación requieren de rutinas de mantenimiento que son realizadas por personal altamente especializado lo que eleva sus costos de mantención, por otra parte, al recibir muestras desde distintas corrientes de la operación, es posible que sus mediciones puedan tener algún sesgo debido a la contaminación que puede existir entre una medición y otra.



Figura 2-5 Analizador Courier® 5X/6X SL (Metso Outotec, 2023).

Por otra parte, durante los últimos años con el aumento de la información medida en las plantas industriales, es que los investigadores e ingenieros han empezado a hacer uso de esta información para la generación de modelos predictivos, en el contexto industrial estos modelos son llamados *soft sensors* o sensores virtuales. El rango de aplicación de estos sensores es amplio, sin embargo, su principal aplicación es la predicción de variables de proceso que pueden ser determinadas a una frecuencia menor a la necesaria para el control o que solo están disponibles fuera de línea, lo importante es que esta predicción puede ser realizada a partir de otras variables con una mayor facilidad de muestreo. La obtención de estos modelos puede ser realizada mediante algoritmos de

aprendizaje supervisado o métodos estadísticos, en donde uno de los más populares son los métodos de proyección (Jolliffe, 2002).

Por tanto, debido a la inviabilidad de tener mediciones de laboratorio con una frecuencia de muestreo adecuada, y además de los altos costos de un equipo analizador en línea, es que la aplicación de un sensor virtual mediante métodos de proyección se presenta como una alternativa adecuada para poder contar información sobre la ley de concentrado en tiempo real para una columna de flotación.

2.5. Métodos de proyección

La búsqueda por alcanzar seguridad, calidad de producto y consistencia en los procesos industriales ha llevado a establecer un sinnúmero de controles en los procesos como un aspecto clave para alcanzar tales objetivos. Tradicionalmente, se han utilizado métodos, como el control automático, que tenían como objetivo solo controlar una o más variables de salida a partir de perturbaciones realizadas en variables de entrada, sin embargo, no consideraban tendencias u otros aspectos relacionados a correlaciones entre aquellas variables (MacGregor, 1994).

A su vez con procesos cada vez más complejos se hace importante tener un control estadístico de procesos (SPC) que logre entregar una visión en tiempo real del proceso en la que se distingan tendencias y como se relacionan cada una de las variables en los distintos estados operacionales posibles. Al aumentar la complejidad del proceso, lo hace también la cantidad de variables que se deben monitorear, por lo que es fácil encontrar desde cientos a miles de variables medidas, en parte explicado al incremento de instrumentación instalada, por tanto, los datos disponibles se incrementan también, lo que no necesariamente entregará información útil sobre el estado del proceso, generando así un desafío en la obtención de información relevante (Dudzic, et al., 2000).

Debido a lo anterior, es que los datos disponibles se deben filtrar y sintetizar de algún modo, en la forma de que muestren información de calidad y no redundante, luego se han podido desarrollar metodologías basadas en métodos de proyección que permiten reducir la complejidad del problema, disminuyendo la dimensión del set de datos, como lo son el análisis de componentes principales (PCA) (MacGregor, et al., 2007) y la proyección a estructuras latentes (PLS) (Wold, et al., 1984).

2.5.1. Análisis de componentes principales (PCA)

El análisis de componentes principales es una herramienta que tiene como objetivo reducir la dimensionalidad de un conjunto de datos, constituido por una gran cantidad de variables que usualmente están correlacionadas, el método procura retener la mayor cantidad de información posible con la cual puede explicar la variación entre aquellas variables.

Para cumplir aquel objetivo se aplican transformaciones lineales al grupo de variables originales, de esta forma se obtiene un nuevo conjunto de variables llamadas componentes principales (PC), las cuales no están correlacionadas entre sí. Por tanto, el concepto que hay detrás de la herramienta es tomar un conjunto de datos y realizar una proyección de este en un subespacio con una dimensionalidad más baja que el número total de variables originales que lo describen, esto con la mínima pérdida de información (Jolliffe, 2002).

Para la correcta aplicación del método de componentes principales se han de considerar los siguientes supuestos respecto de los datos (Acosta, 2007):

- La relación que existe entre las variables que se utilizan para la construcción del modelo es lineal en el rango de aplicación de este.
- Las observaciones son independientes unas de otras, es decir el proceso se encuentra en un estado estacionario por tanto la dinámica del proceso no es tomada en cuenta.
- Los errores asociados a la medición de cada variable se distribuyen de forma normal, es decir con una media centrada en cero y varianza constante.

En términos matemáticos la herramienta de componentes principales se sustenta en la descomposición de vectores propios o *eigenvectors* de la matriz de covarianza o correlación en los datos que describen el proceso. Sea entonces $\mathbf{X}$ una matriz con m filas y n columnas, donde cada columna es una variable y cada fila es una muestra.

Luego, si cada una de las n variables contenidas en la totalidad de las muestras han sido auto escaladas, es decir que cada una de las variables ha sido centrada en cero y con varianza unitaria al dividir por su desviación estándar, la Ec. 2-2 será equivalente a la matriz de correlación.

$$Cov(\mathbf{X}) = \frac{\mathbf{X}^T \mathbf{X}}{m - 1} \quad (\text{Ec. 2-2})$$

Trabajar con variables auto escaladas es altamente recomendable, especialmente en sistemas con variables con distintas unidades de medida, ya que de esta forma se pueden analizar todas las variables desde una línea base, y así la contribución de cada una de estas variables a las componentes principales será representativa.

En la descomposición llevada a cabo mediante el método PCA, los vectores p_i son llamados vectores propios de la matriz de covarianza o *loadings* del modelo PCA este parámetro entrega información de cómo se relacionan las variables entre sí, luego se puede escribir la matriz de covarianza como sigue:

$$Cov(\mathbf{X}) \cdot \mathbf{p}_i = \lambda_i \cdot \mathbf{p}_i \quad (\text{Ec. 2-3})$$

Donde λ_i corresponde al valor propio o *eigenvalue* asociado a cada p_i . Además, se define el vector $\mathbf{t}_i$ llamado *scores* que tiene información respecto a cómo las distintas observaciones se relacionan entre ellas. Estos vectores poseen la propiedad de formar un conjunto ortogonal es decir que $\mathbf{t}_i^T \mathbf{t}_i = 1$ para $i = j$ y $\mathbf{t}_i^T \mathbf{t}_j = 0$ para $i \neq j$. Mientras que p_j son ortonormales ($\mathbf{t}_i^T \mathbf{t}_i = 1$ para $i = j$, $\mathbf{t}_i^T \mathbf{t}_j = 0$ para $i \neq j$). Luego para cada par de p_i y t_i se cumple que:

$$\mathbf{X} \mathbf{p}_i = \mathbf{t}_i \quad (\text{Ec. 2-4})$$

Es decir que los *scores* corresponden a una combinación lineal o proyección de las variables originales que están escaladas en los vectores p_i . En ese sentido, cada valor propio λ_i es una medida de la cantidad de varianza capturada y explicada por el par t_i y p_i . Luego se pueden ordenar estos valores propios en un orden descendente, en donde los primeros pares en la descomposición serán capaces de capturar la mayor parte de la información.

En esta interpretación, se puede reescribir entonces la matriz de datos originales en función de los parámetros que se describieron con anterioridad,

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{t}_2 \mathbf{p}_2^T + \dots + \mathbf{t}_k \mathbf{p}_k^T + \mathbf{E}; \quad k \leq \min \{m, n\} \quad (\text{Ec. 2-5})$$

Generalmente se puede notar y es esencialmente lo que se busca con la reducción de variables es que los datos puedan ser descritos adecuadamente utilizando una fracción menor a las variables originales del problema, es decir que k sea menor o igual al número de variables originales.

La dirección de estos vectores o scores corresponde a los vectores propios de la matriz de covarianza de los datos. El valor propio asociado a cada vector propio corresponde a la varianza de cada componente principal, por lo que se puede computar la variabilidad explicada por cada componente principal i como el cociente entre su valor propio y la suma de la totalidad de los valores propios:

$$\%Var(CP_i) = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j} \quad (\text{Ec. 2-6})$$

Como se puede ver el objetivo principal es la reducción de dimensionalidad del problema original, conservando la mayor cantidad de información posible, luego es crucial entonces saber que tan pequeño puede hacerse el subconjunto de componentes principales. Existen algunas reglas comunes que se puede adoptar cuando se está construyendo un modelo PCA, sin embargo, hay dos que son transversalmente más utilizados (Jolliffe, 2002).

El primero tiene que ver con la varianza explicada por cada uno de los componentes, en ese sentido, la regla indica que se deben tomar en cuenta las componentes principales que expliquen la mayor varianza acumulada posible, usualmente se utiliza el 80% o 90%. El segundo criterio, tiene que ver con el análisis del valor obtenido en cada uno de los valores propios, dado que los modelos se suelen escalar a varianza unitaria, cuando un valor propio es menor que la unidad, se considera que contiene menor información que las variables originales y por tanto su información es redundante y no es necesario conservarlos.

La herramienta PCA permite encontrar la relación entre las variables de operación de manera de generar un grupo de componentes principales que ayuden a describir distintos estados operacionales con mayor facilidad, inclusive operando en base a solo las componentes principales de interés, predicción de variables o en la aplicación para el diagnóstico de operaciones.

Una vez hecho el cálculo de los modelos, dentro de la metodología se tienen dos importantes estadígrafos que ayudan en la toma de decisiones, estos indicadores permiten explicar la bondad de ajuste de un modelo cuando trata de describir el proceso de interés, los cuales se presentan a continuación:

- Estadígrafo Q - *Residuals*

El estadígrafo Q cuantifica la falta de ajuste estadístico por el modelo PCA, es decir es una medida de la variabilidad que no es capturada por el modelo, como podemos ver en la Ec. 2-7, este indicador se puede obtener como la suma de los cuadrados de la matriz de residuos determinados por el valor medido y el entregado por el modelo,

$$Q = \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (\text{Ec. 2-7})$$

- Estadígrafo Hotelling's - T^2

El estadígrafo T^2 cuantifica la variabilidad de cada muestra dentro del modelo PCA, se obtiene a partir de la suma de los cuadrados de los *scores* normalizados que resultaron ser significativos en el modelo, por tanto, se calcula como sigue,

$$T^2 = \sum_{j=1}^k \frac{t_j^2}{\lambda_j} \quad (\text{Ec. 2-8})$$

2.5.2. Partial Least Square (PLS)

La metodología PLS de forma similar a PCA busca disminuir la dimensionalidad del problema determinando una cierta cantidad de variables latentes, eso sí a diferencia de PCA, no solamente busca que estas variables latentes expliquen la varianza en la matriz de entrada, sino que también expliquen y sean capaces de predecir una o más variables de interés.

Luego para aplicar la metodología PLS, se necesitan dos matrices de datos, la primera es una matriz predictora $\mathbf{X}$ (m observaciones y n variables), que incluye las variables de entrada y la segunda es una matriz $\mathbf{Y}$ (m observaciones y k variables), la cual agrupa las variables a predecir (MacGregor & Kourti, 1995).

Para su aplicación se deben tomar los mismos supuestos que en la metodología PCA, es decir, que la relación que existe entre las variables para construir el modelo es el lineal en el rango de

aplicación, observaciones en estado estacionario, y una distribución de los errores normal. Además, es aconsejable realizar un pretratamiento de los datos, eliminando observaciones anormales y aplicando un auto escalamiento de las variables por las razones ya explicadas anteriormente.

Siguiendo con la implementación, a partir de la matriz en entrada X , se calcula la matriz t , llamada *scores* la cual representa la relación que existe entre cada una de las muestras, además se calcula una matriz p denominada *loadings*, que representa la relación entre las variables y se determina un vector de peso w que relaciona las matrices t y p de forma ortogonal al igual que en la metodología PCA. De forma análoga se realiza el mismo procedimiento para la matriz de salida Y , en donde la matriz de *scores* y *loadings* en este caso se denomina, u y q respectivamente (Wold, et al., 1984).

De esta forma se pueden formular los modelos que explican X e Y en términos de variables latentes,

$$X = tp^T + E_x \quad (\text{Ec. 2-9})$$

$$Y = uq^T + E_y \quad (\text{Ec. 2-10})$$

Relacionando las ecuaciones anteriores, se tiene el modelo predictivo en forma matricial de la forma,

$$Y = XB, \quad B = WQ^T \quad (\text{Ec. 2-11})$$

En donde B es el vector de regresión, y su determinación se realiza mediante un proceso iterativo el cual se denomina SIMPLS, que es una de las variaciones de la metodología PLS. En este caso se utiliza la matriz de covarianza $S = X^T Y$ para la extracción y eliminación de la información de acuerdo con la cantidad variables latentes elegidas para el modelo, el detalle del algoritmo puede ser consultado en la referencia (De Jong, 1993).

Una vez determinado el modelo PLS para poder evaluar su bondad de ajuste se pueden utilizar los mismos indicadores que para la metodología PCA, es decir $Q \text{ residuals}$ y $Hotelling's T^2$, en donde el primero es un estadístico respecto de la falta de ajuste, y el segundo es una medida de la varianza en cada observación respecto del modelo generado, y por supuesto los análisis aplicables a cualquier regresión como el análisis de residuos.

El uso de la metodología PLS para la calibración de un sensor virtual se presenta como una opción conveniente ya que, por ejemplo:

- Al aplicar la metodología PLS, se obtiene una regresión lineal lo cual resulta útil en términos de la interpretación del modelo, permitiendo así obtener información en cómo se relacionan las variables de entrada respecto de la respuesta de interés y además que la linealidad facilita la metodología de cálculo.
- A diferencia de las regresiones lineales multivariantes comunes que tienden a verse afectadas negativamente por una matriz de datos mal acondicionadas, con altos grados de correlación entre las variables, la regresión PLS resuelve este problema proyectando las variables originales del problema a un espacio de variables latentes ortogonales. Lo anterior es de gran relevancia, ya que en una aplicación a un proceso real es muy común que la información disponible para calibrar el modelo esté altamente correlacionada ya que puede ser difícil ejecutar un diseño experimental, por ejemplo, por costos de los experimentos y disponibilidad de planta para realizarlos.
- Permite construir el modelo en base a una gran cantidad de variables explicativas, lo que resulta de ayuda hoy en día, ya que al aumentar la complejidad de los procesos es común que la cantidad de variables que se monitorean aumenten también, incrementando así la dimensión de la matriz de datos.
- El algoritmo ya está implementado en varios paquetes de código abierto para *Python* y *R*, además en *softwares* comerciales como *MATLAB* con el complemento *PLS Toolbox*, el cual mediante una interfaz permite realizar todos los pasos para obtener un modelo adecuado como el pretratamiento de los datos, aplicar el algoritmo, metodologías de validación y entregar el vector de regresión final sin la necesidad de contar con conocimientos de algún lenguaje de programación.

2.5.3. PLS Dinámico

Desde la formulación original y la primera aplicación de PLS (Wold, et al., 1984) se ha utilizado ampliamente en diversos campos de aplicación y que tienen relación con el control automático, por ejemplo (Lin, et al., 2007) determinaron un sensor virtual para distintas variables que caracterizan la calidad del aire, como NO_x , SO_2 y CO_2 . (Kaspar & Harmon, 1992) ejemplifican la aplicación de

un sensor para un proceso de destilación y otro proceso térmico. (MacGregor & Kourti, 1995) utilizan la metodología también para la predicción y el monitoreo en un proceso de producción de polietileno.

La regresión PLS que se discutió en el punto anterior y de las investigaciones mencionadas, tiene como supuesto que cada una de las observaciones es independiente de las observaciones pasadas, es decir que el proceso se encuentra en estado estacionario.

Sin embargo, en general esta suposición no es válida en procesos químicos o mineros, ya que el resultado del proceso en un instante, no solamente depende del estado actual, sino que también es consecuencia de efectos cumulativos de estados del proceso anteriores, luego si se aplicará la regresión el ajuste no sería capaz de determinar las relaciones dinámicas del proceso dejando gran parte de la varianza sin modelar, por tanto, la predicción tendrá un bajo desempeño cuando sea utilizada en la predicción de sistemas dinámicos.

En este sentido, para la aplicación de un sensor dinámico se hace necesario un método que tome en consideración las relaciones lineales y dinámicas del proceso, algunas modificaciones al algoritmo original se han realizado durante los años, por ejemplo, en investigaciones se ha propuesto la utilización de filtros dinámicos en los datos del bloque X antes de aplicar la metodología PLS, de esta forma poder incluir las relaciones dinámicas intrínsecas del proceso como es en el caso de (Kaspar & Harmon, 1993).

Por otra parte, una de las modificaciones más reconocidas y utilizadas con mayor frecuencia en la literatura, se trata de un método mucho más directo, en el que se incluyen un determinado número de *lags* o valores pasados correspondiente a las variables de entrada y/o salida en la matriz X .

Es el caso de (Qin & McAvoy, 1996) quienes incluyen valores pasados de las variables en el bloque X para la formulación de un modelo dinámico del tipo autorregresivo, (Li, et al., 2014) determinaron un modelo PLS dinámico y lo evaluaron mediante el caso de estudio *Tennessee Eastman Process* que es utilizado comúnmente como *benchmark* de metodologías de detección y predicción, obteniendo resultados prometedores.

La metodología también ha sido utilizada para la predicción y monitoreo de procesos de producción *batches* que inherentemente son del tipo dinámicos (Garcia-Muñoz, et al., 2004). En otras investigaciones, por ejemplo, (Wang, et al., 2010) generaron un sensor virtual mediante la

utilización de PLS dinámico para determinar una variable de proceso en línea la cual resulta difícil de medir por métodos tradicionales en un proceso de refinería de petróleo, situación similar a la que se tiene respecto de la medición del concentrado de cobre en una columna de flotación.

Dependiendo de que, si se incluyen o no valores pasados de la variable de salida en el bloque X, determinará el tipo de modelo obtenido, en el caso de que si se incluyan, resultara en un modelo autorregresivo (ARX), por el contrario, si no se incluyen resultara en un modelo de respuesta finita (FIR) (Qin & McAvoy, 1996). A continuación, se muestra la generalización para cada uno de los modelos:

- Modelo Autorregresivo (ARX)

$$y(t) = f_{ARX}(y(t-1), \dots, y(t-h), u(t), u(t-1), \dots, u(t-h)) \quad (\text{Ec. 2-12})$$

- Modelo de respuesta a impulso (FIR)

$$y(t) = f_{FIR}(u(t), u(t-1), \dots, u(t-h)) \quad (\text{Ec. 2-13})$$

Para la implementación del algoritmo se tienen u_t e y_t , los cuales corresponden a los vectores que incluyen a las variables de entrada y salida del sistema respectivamente, estos en función del tiempo, con $t = 0, 1, \dots, m+h$ por tanto se consideran $m+h+1$ observaciones.

Se definen las matrices de observaciones como,

$$U_t = [u_o \ u_1 \ \dots \ u_{h+m}]^T \quad (\text{Ec. 2-14})$$

$$Y_t = [y_o \ y_1 \ \dots \ y_{h+m}]^T \quad (\text{Ec. 2-15})$$

Luego se forma una matriz ampliada con h valores pasados, para lo cual definimos lo siguiente,

$$U_i = [u_i \ u_{i+1} \ \dots \ u_{i+m}]^T \quad (\text{Ec. 2-16})$$

Por tanto, el bloque X queda como se ve a continuación, en donde se construye una matriz ampliada que considera las matrices de observación para cada uno de los valores pasados a considerar (Dong & Qin, 2015),

$$X_h = [U_h \ U_{h-1} \ \dots \ U_o] \quad (\text{Ec. 2-17})$$

Si se desarrolla la Ec. 2-17, según la matriz de observaciones definida en la Ec. 2-16 finalmente el bloque X queda como sigue,

$$X_h = \begin{bmatrix} u_h & u_{h-1} & \cdots & u_0 \\ u_{h+1} & u_h & \cdots & u_1 \\ \vdots & \vdots & \ddots & \vdots \\ u_{h+m} & u_{h-1+m} & \cdots & u_m \end{bmatrix} \quad (\text{Ec. 2-18})$$

Por otra parte, el bloque Y queda definido como,

$$Y_h = \begin{bmatrix} y_h \\ y_{h+1} \\ \vdots \\ y_{h+m} \end{bmatrix} \quad (\text{Ec. 2-19})$$

Notar en las Ecuaciones 2-18 y 2-19, que, si h es igual a cero, se obtienen los bloques X e Y de la metodología PLS tradicional, por tanto, la generalización es consistente con el algoritmo.

La cantidad de valores pasados que se deben agregar a la matriz dependerá de la dinámica del proceso y también de la frecuencia de muestreo disponible en la variable objetivo, esto a modo de representar correctamente los estados transientes del proceso.

En este sentido, existe una relación entre el número de *lags* y la resolución que tendrá la predicción para cada respuesta transiente del proceso, en teoría se podrían agregar infinitos *lags*, sin embargo, al aumentar la cantidad de valores pasados aumentan también la cantidad de variables regresoras lo que no necesariamente resultará en un mejor modelo. El caso contrario, es decir, sin considerar valores pasados corresponde a la metodología PLS tradicional y como ya se explicó tiene un bajo desempeño cuando existen relaciones dinámicas involucradas en los datos.

Con lo anterior se debe buscar un punto de equilibrio, entre el número de *lags* y la resolución del modelo de acuerdo con la aplicación que se le dará al sensor, en general para el caso de los modelos FIR usualmente la cantidad de información que se debe considerar es el tiempo que toma en estabilizarse la respuesta transiente. Para el caso de los modelos autorregresivos la cantidad de valores pasados es menor, usualmente considerar una cantidad de valores pasados equivalentes al orden del sistema dinámico es suficiente (Ku, et al., 1995).

2.6. Simulación en procesos de flotación

Como se pudo apreciar anteriormente el proceso de flotación depende de una gran cantidad de factores asociados a la naturaleza del proceso como la alimentación, distribución de tamaño de partícula y burbujas, entre otros. Pese a lo anterior, existen distintos modelos para describir el proceso de flotación entre los cuales destacan modelos empíricos, modelos probabilísticos y modelos cinéticos (Yianatos & Vinnert, 2015).

Con los modelos aparecen los simuladores fenomenológicos, que con frecuencia operan según el esquema presentado en la Figura 2-6, en donde se caracteriza la alimentación en distintos parámetros como ley de cobre, ley de molibdeno, porcentaje de sólidos, granulometría y caudal de alimentación.

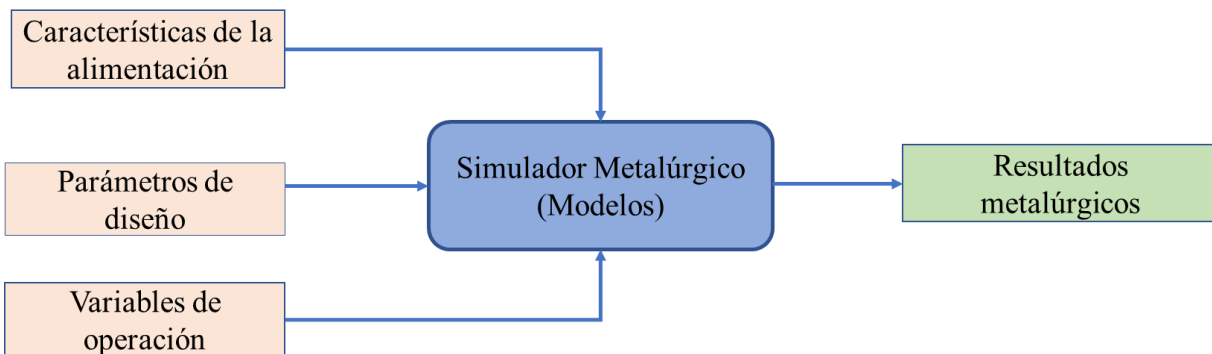


Figura 2-6 Esquema funcionamiento simulador metalúrgico

Por otra parte, se definen los parámetros de diseño, los cuales corresponden a las dimensiones de los equipos y además se cuenta con el set-point de las variables operacionales típicas como profundidad de espuma, flujo de aire, agua de lavado. De esta forma, utilizando la información anterior son capaces de computar los balances de masa, modelos cinéticos para poder calcular y entregar la información de interés, como por ejemplo el desempeño metalúrgico del equipo y caracterizar las corrientes de salida del equipo.

Se debe destacar que en las simulaciones se pueden obtener desviaciones respecto del proceso real, las cuales están explicadas por varios factores, el primero tiene que ver con los supuestos que cada modelo realice, otro factor corresponde a la descripción incompleta de magnitudes, como por ejemplo el tamaño de partícula, el cual muchas veces se estima como un valor, sin embargo no es suficiente para describir la complejidad del proceso ya que mas bien se trata de una distribución de

tamaño y finalmente la existencia de variables no medidas en una planta real que no son incluidas en los modelos. Sin embargo, aun con esas desviaciones son una opción útil cuando no se tiene acceso a datos experimentales para el ajuste de un sensor virtual.

En la actualidad existen diversos tipos de simuladores con distintos niveles de complejidad y campos de aplicación para cada uno, por ejemplo, *JK SimFloat* (Yianatos, et al., 2012), el cual es muy utilizado en la industria de la minería. En este trabajo para obtener datos y calibrar el modelo, se utilizará el simulador fenomenológico *SiMetCol* desarrollado por académicos de la Universidad Técnica Federico Santa María, las capacidades, funcionamiento y uso se desarrollarán más adelante en el apartado de metodología.

2.7. Árboles de decisión

El método de árboles de decisión es un algoritmo de aprendizaje supervisado que es utilizado para generar modelos de clasificación o regresión. Para lo cual se define un set de datos S , con m observaciones, conformado por n variables de entrada y una variable de salida dependiente que puede ser numérica o categórica. El objetivo del algoritmo es determinar un modelo, que para cada una de las m observaciones pueda clasificar o predecir la variable objetivo según el aprendizaje de reglas de decisión simples inferidas a partir de las n variables de entrada (Kotsiantis, 2013).

En la Figura 2-7 se presenta un esquema de un árbol de decisión, el cual consiste en nodos, ramas y hojas. Las hojas o nodos terminales definen la predicción o clase de Y , además determinan un conjunto de reglas que se pueden capturar descendiendo desde el nodo principal hasta la hoja o nodo terminal.

El número de hojas definirá la cantidad de dichos conjuntos de reglas que se pueden extraer desde un árbol de decisión. De esta forma mediante divisiones del set de datos, los árboles de decisión son capaces de entregar reglas lógicas *if-then* con una estructura jerárquica desde el nodo principal hasta los nodos terminales.

En la medida que el árbol se hace más profundo, mayor es la complejidad de estas reglas de decisión, aumentando así también el ajuste del modelo, sin embargo, se debe tener en consideración de no realizar un sobreajuste de los datos, para lo cual existen distintos métodos de podado para obtener un modelo óptimo (Somvanshi, et al., 2016).

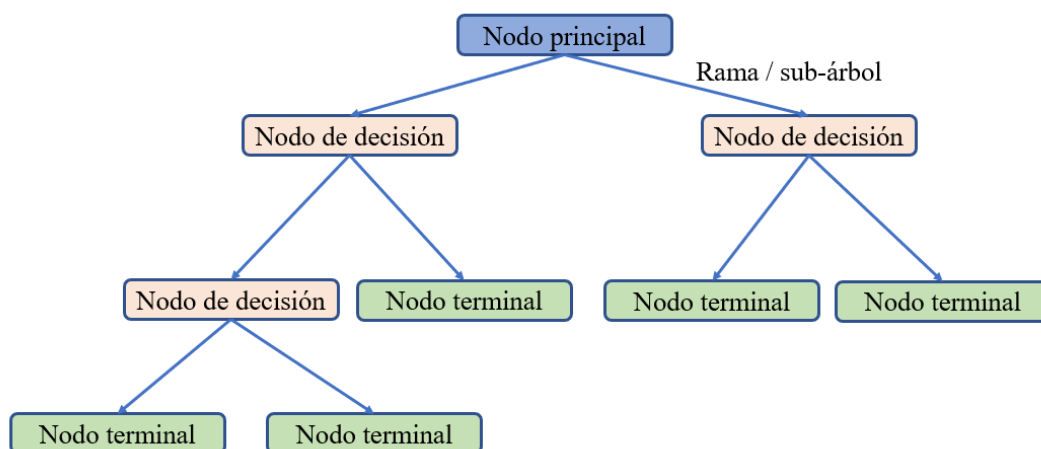


Figura 2-7 Esquema de un árbol de decisión

En el marco de esta investigación los árboles de decisión son utilizados para clasificar los residuos del modelo PLS obtenido, de esta forma se pueden determinar en qué zonas de operación, según las reglas determinadas por el árbol de decisión, corresponden a zonas en donde el modelo global tiene un bajo desempeño y así formular modelos locales en donde sea necesario, esto el con el fin de mejorar la predicción original, dicha aplicación fue realizada en un reactor utilizando modelos locales PLS (Narayanan, et al., 2019).

Siguiendo, para la construcción del modelo de clasificación se deben seguir los siguientes pasos:

i. Construcción del árbol máximo

El método más común para la construcción de un árbol de decisión es el denominado CART, o “*Classification and Regression Trees*” (Breiman, 1984). Este algoritmo puede generar un modelo de clasificación o regresión, solamente con decisiones binarias.

La construcción del árbol se produce generando divisiones binarias en cada uno de los nodos de acuerdo con algún criterio establecido, dicho criterio corresponde una medida que tiene que ver con la impureza del nodo. En este sentido el objetivo es que cada vez que se ejecute la partición de un nodo, los dos nodos resultantes sean menos impuros, de esta forma se va mejorando la predicción o clasificación a medida que se desciende en el árbol. Estas divisiones se realizan hasta que todos los nodos son terminales o hasta que se alcance alguno de los criterios de podado que se

prestan más adelante. El algoritmo evalúa el criterio en cada una de las iteraciones por tanto se trata de un algoritmo del tipo *greedy*.

Para determinar la medida de impureza se utilizan comúnmente alguno de los siguientes índices, el de Gini o el de Entropía, en donde el más ampliamente utilizada es el índice de Gini (Suknovic, et al., 2012), el cual puede ser calculado como sigue:

$$i(t) = \sum_{i,j} p(i|S) \cdot p(j|S), \quad i \neq j \quad (\text{Ec. 2-20})$$

$$\Delta i(t) = i(t) - p[t_a] \cdot i(t_a) - p[t_b] \cdot i(t_b) \quad (\text{Ec. 2-21})$$

Donde:

- t : Nodo
- $k, 1 \dots K$: Clases
- p : Proporción
- $p(k|t)$: Proporción de la clase k en el nodo t .
- t_a, t_b : Nodo a y b generado por la partición

ii. Podado del árbol

Una vez construido el árbol, lo que se tiene es un árbol saturado, el cual tiene muchas ramas o divisiones, ya que el modelo por defecto no se detiene hasta que todos los nodos terminales son puros. Por lo que es necesario reducir su tamaño en un proceso que también es conocido como poda. El objetivo es obtener un árbol con una alta calidad y que sea una adecuada generalización de los datos. Existen algunos criterios que tienen relación con el nivel de podado en el árbol, entre ellos se encuentran (Mingers, 1989):

- Nodos puros (Parámetro por defecto)
- Profundidad máxima del árbol
- Cantidad mínima de observaciones en el nodo padre
- Cantidad mínima de observaciones en el nodo hijo

- Costo-complejidad

3. MATERIALES Y METODOLOGIAS

3.1. Simulador

El simulador utilizado para este trabajo es capaz de simular y determinar los resultados metalúrgicos de una columna de flotación en donde se pueden modificar tanto las condiciones de alimentación y operación para controlar el equipo.

El programa cuenta con diversas funcionalidades que lo hacen útil para la generación del set de datos necesario para el ajuste del modelo. Las cuales pueden ser vistas en la Figura 3-1 en donde se presenta la pestaña principal de la interfaz gráfica, como se ve, el programa permite la caracterización de la corriente de alimentación, además de las variables operacionales y los parámetros de diseño del equipo.

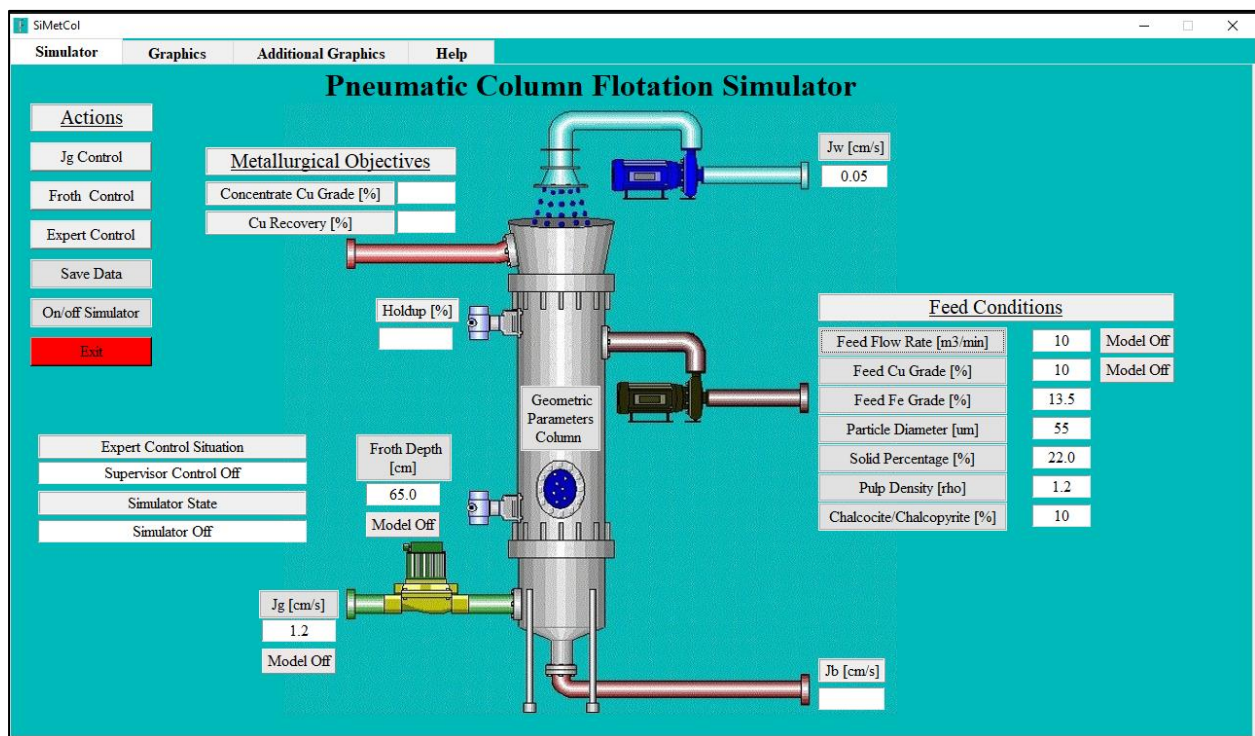


Figura 3-1 Interfaz gráfica simulador SiMetCol

Además, cuenta con la posibilidad de establecer un sistema de control PID para la velocidad superficial del gas y la altura de espuma, como también un sistema experto, detector de estado estacionario, la posibilidad de agregar ruido a las mediciones y exportar los datos generados.

La principal limitante que tiene el simulador respecto con la aplicación de este trabajo, es que los resultados metalúrgicos que entrega son los correspondientes a estados estacionarios del proceso, por tanto, para determinar la dinámica de dichas variables de interés, se acoplará el simulador a un modelo de primer orden para poder obtener la trayectoria entre un estado a otro, a partir de las perturbaciones realizadas en el sistema.

3.2. Definición de variables de proceso disponibles en simulador

3.2.1. Variables de entrada

A continuación, se presentan las variables de entrada disponibles en el simulador separadas de acuerdo con el esquema de un simulador presentando en la Figura 2-6, en general las variables de entrada se refieren a las variables que caracterizan la alimentación y las condiciones de operación, y por tanto pueden ser consideradas como variables predictoras en el modelo final.

- Variables de alimentación

Las variables de alimentación se encargan de caracterizar la corriente de alimentación, considera parámetros de composición en cuanto al cobre y fierro contenido en la corriente, además de propiedades físicas de la pulpa como tamaño de partícula, porcentaje de sólidos y densidad de pulpa. El detalle de las variables y los distintos rangos de operación para cada una se presentan en la Tabla 3-1.

Tabla 3-1 Variables de alimentación disponibles y sus rangos

Variable	Valor mínimo	Valor máximo	Unidad
Flujo de alimentación	5	15	m ³ /min
Ley de cobre en alimentación	5	15	%
Ley de fierro en alimentación	-	-	%
Diámetro de partícula,	20	70	µm
Porcentaje de sólidos	-	-	%
Densidad de pulpa	1,00	1,30	g/mL
Razón calcosina/calcopirita	0	50	%

En el caso de la ley de fierro en la alimentación y el porcentaje de sólidos, si bien se muestran como variables de entrada, no presentan valores máximos o mínimos ya que estas no pueden ser modificadas por el usuario debido a que son función de otras variables y se calculan a partir de la razón calcosina/calcopirita y la densidad de la pulpa respectivamente.

Además, para el caso del flujo de alimentación y la ley de cobre, se pueden considerar modelos estocásticos que a partir de dos parámetros θ y ϕ , pueden aumentar o disminuir la variabilidad de dichas variables, en el sentido de que un aumento del parámetro ϕ modifica la curva estocástica, disminuyendo la influencia del ruido blanco. En el caso del parámetro θ , un aumento de este genera una mayor amplitud del ruido sin modificar la curva estocástica. De esta forma se puede en cierta forma representar la variabilidad intrínseca que tiene la alimentación para el caso de una planta real.

- Variables de operación

Este tipo de variables permiten al usuario definir las condiciones de operación de la columna. En este caso las variables que están disponibles para la simulación corresponden a la velocidad superficial del aire, altura de espuma y la velocidad superficial de agua de lavado. El detalle de los rangos de operación disponibles se presenta en la Tabla 3-2.

Tabla 3-2 Variables de operación disponibles y sus rangos

Variable	Valor mínimo	Valor máximo	Unidad
Velocidad superficial del aire	0,8	3	cm/s
Altura de espuma	30	100	cm
Velocidad superficial de agua de lavado	0	0,1	cm/s

Adicional se cuenta con lazos de control PID para el caso de las variables correspondientes a las velocidades superficiales de aire y agua de lavado, por tanto, se puede utilizar un set point para simular los cambios de estas variables de acuerdo con un lazo de control, es posible ajustar la sintonía de control de estos lazos de acuerdo con los parámetros de controlador K_p y K_i . Por otra parte, el simulador hace uso de estos dos lazos de control para implementar un sistema de control

experto que puede ser programado para optimizar la operación con dos objetivos excluyentes entre sí: mejorar la recuperación o de aumentar la ley de concentrado.

- Variables geométricas

Finalmente se definen las variables geométricas, éstas se relacionan con la forma de la columna que puede ser cilíndrica o rectangular. En la Tabla 3-3 se presentan los rangos mínimos y máximos, para ambos casos, según la sección transversal elegida, para el caso cilíndrico se cuenta con el diámetro y para el caso rectangular el ancho y largo. Notar que existe la posibilidad agregar manualmente el perímetro real de la columna, esto en consideración de columnas de tienen canales de descarga internos con el fin de aumentar el contorno de descarga concentrado.

Tabla 3-3 Parámetros geométricos de la columna simulada

Variable	Valor mínimo	Valor máximo	Unidad
Altura de la columna	1200	1600	cm
Diámetro de la columna	300	700	cm
Ancho de columna	300	700	cm
Largo de columna	240	640	cm
Perímetro real	-	-	cm

3.2.2. Variables de salida

El simulador a partir de las distintas variables de la alimentación y operacionales, computa los distintos indicadores metalúrgicos del proceso de acuerdo con los balances de masa, modelos cinéticos de colección, recuperación en la zona de espuma y recuperación multiespecie. En este caso el simulador entrega otras variables de operación que en algunos casos son de interés como el Bias y el holdup de aire en la zona de colección. Además de los resultados metalúrgicos correspondientes a la concentración de cobre en el concentrado y la recuperación. El primero en especial es muy relevante para esta investigación ya se busca generar un sensor virtual para predecir la ley de cobre en el concentrado.

En la Figura 3-2 se presenta un esquema del simulador, en donde un por un lado se pueden apreciar las variables disponibles en el simulador y que se utilizarán como perturbaciones. Estas variables son utilizadas como *input* para los modelos del simulador, el cual entrega los resultados metalúrgicos, tanto como la recuperación y también la ley de concentrado.

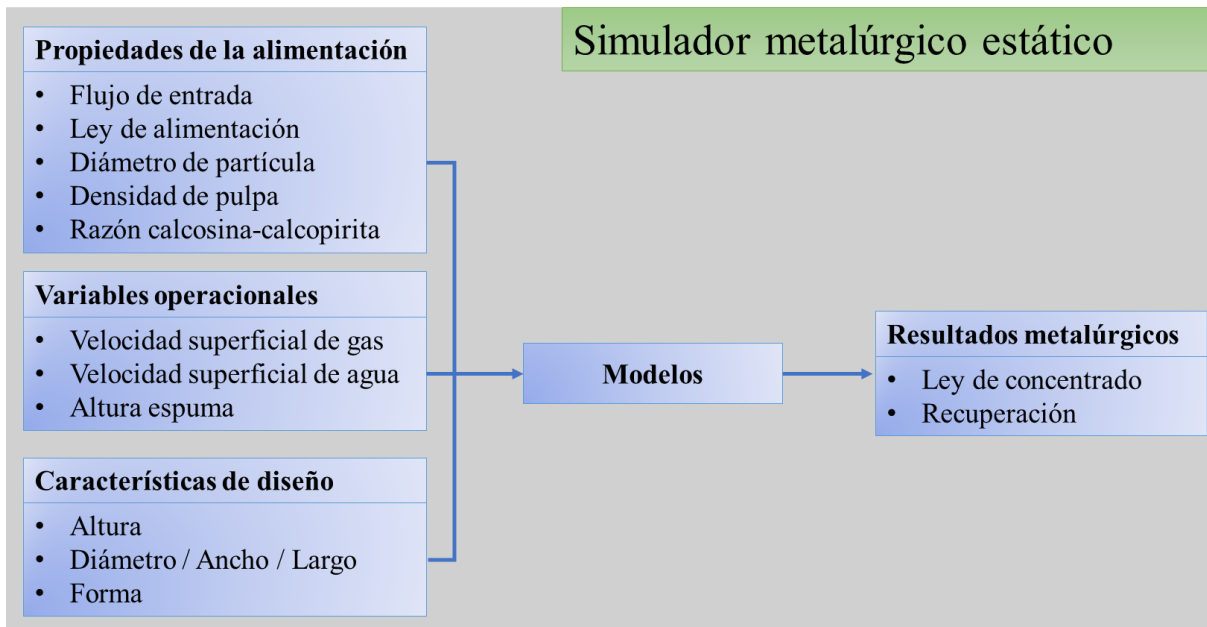


Figura 3-2 Esquema simulador estático utilizado

3.3. Respuesta transiente

Como se explicó anteriormente el simulador está limitado a entregar información que es útil solamente cuando el proceso se encuentra en estado estacionario. Para efectos de la investigación se necesita también información sobre la trayectoria que sigue la variable de interés cuando al sistema modifica sus condiciones de operación respecto de perturbaciones aplicadas en las variables de entrada. Para simular la respuesta transiente se acopló el simulador explicado anteriormente con un simulador fenomenológico dinámico que estudia como varía la ley de concentrado cuando se aplica una perturbación a la columna.

Para el desarrollo de este simulador dinámico se establecieron tres supuestos, el primero es que la respuesta de este sistema corresponde a uno de primer orden, el segundo supuesto es que la constante de tiempo asociada al sistema es conocida y además es función de las variables de operación del sistema, y el tercer supuesto es que el sistema está afectado por retardo por transporte.

3.3.1. Sistema de primer orden

En general un proceso recibe una entrada $u(t)$ para el cual se genera una salida $y(t)$, en este sentido un sistema de primer puede ser representado mediante una ecuación diferencial de primer orden como sigue:

$$\tau \cdot \frac{dy(t)}{dt} + y(t) = k \cdot u(t); \quad y(0) = 0 \quad (\text{Ec. 3-1})$$

En este caso la variable de salida corresponde a la variable que es de interés predecir, es decir la ley de concentrado. Por otra parte, $u(t)$ puede ser cualquiera de las variables del sistema que pueden ser modificadas mediante perturbaciones.

Es común que la resolución de estas ecuaciones se realice con transformadas de Laplace, de este modo tiene en el dominio s una entrada $U(s)$ la cual genera una salida $Y(s)$, la función que relaciona la salida del sistema, respecto de la entrada o perturbación se denomina función de transferencia $G(s)$, lo cual se representa esquemáticamente en la Figura 3-3.

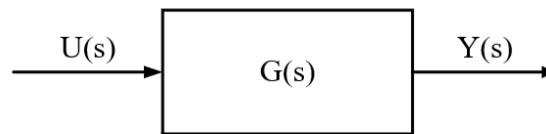


Figura 3-3 Esquema entrada y salida sistema

Luego resolviendo la Ec. 3-1, se tiene la función de transferencia para un sistema de primer orden:

$$G(s) = \frac{Y(s)}{U(s)} = \frac{K_i}{\tau \cdot s + 1} \exp(-t \cdot s) \quad (\text{Ec. 3-2})$$

Donde:

- $Y(s)$: Variable de salida del sistema
- $U(s)$: Variable de entrada al sistema
- K_i : Ganancia estática del sistema
- τ : Constante de tiempo del sistema

- Ganancia estática.

Se denomina ganancia estática de un sistema a la relación de ganancia entre la entrada y la salida del proceso. Es decir, cuando el sistema es sometido a una perturbación del tipo escalón, la variable de salida se estabilizará en un régimen permanente, por tanto, la ganancia estática corresponderá a la relación entre el cambio que sufre la variable de salida y el cambio en la variable de entrada. De lo anterior se puede intuir que la respuesta permanente o respuesta estacionaria se refiere al comportamiento de la salida del proceso cuando el tiempo tiende al infinito como se ve en la siguiente ecuación.

$$K_i = \frac{\Delta Y}{\Delta U} = \frac{Y_{\infty} - Y_i}{U_{\infty} - U_i} \quad (\text{Ec. 3-3})$$

Se ejemplifica la ganancia estática según la Ec. 3-3 para el caso de la velocidad superficial de gas (J_g) como perturbación, sin embargo, puede ser utilizada por cualquiera de las variables de entrada disponibles en simulador.

$$K_{Ley-Jg} = \frac{Ley_{conc_{\infty}} - Ley_{conc_i}}{Jg_{\infty} - Jg_i} \quad (\text{Ec. 3-4})$$

Esta es la forma en el que simulador metalúrgico estático y el dinámico se acoplan, ya que el primero entrega información sobre los estados estacionarios: iniciales y finales. Y el simulador dinámico entregará información sobre la respuesta transiente entre un estado a otro de acuerdo con la ganancia estática según las distintas perturbaciones.

3.3.2. Constante de tiempo

La constante de tiempo de un sistema de primer orden, se define como el tiempo requerido para el sistema alcance el 63% del valor final que tendrá en el régimen permanente. Por tanto, es una muestra de la rapidez del sistema ante una determinada perturbación para alcanzar el estado estacionario. Cuanto menor es la constante de tiempo, más rápida es la respuesta del sistema. Si la constante de tiempo es mayor, el sistema se mueve de forma más lenta en su respuesta transitoria. Además, se dice que la duración de esta respuesta transitoria es aproximadamente 4 veces la constante de tiempo.

Uno de los supuestos es que está constante de tiempo sea una función de variables de la alimentación y/u operacionales, usualmente se puede considerar la constante de tiempo para una columna de flotación como sigue (Yianatos & Vinniet, 2015).

$$\tau = \frac{A_c \cdot h_c \cdot (1 - \varepsilon_g)}{Q_f} \quad (\text{Ec. 3-5})$$

Donde:

- A_c : Área transversal de la columna
- h_c : Altura de la zona de colección
- ε_g : Concentración de gas (Hold-up)
- Q_f : Flujo volumétrico de pulpa en la alimentación.

3.3.3. Retardo por transporte

Como se explicó en el capítulo anterior, el proceso de flotación puede ser representado según la Figura 2-1, como dos procesos en serie, el de colección y seguido por el de separación. Por tanto, una vez que se aplica una perturbación al sistema, se genera un gradiente de concentración en la zona de espuma en dirección vertical, ya que la zona más cercana a la descarga se encontrará a distinta concentración que los agregados más cercanos a la zona de colección. Luego si se midiese la concentración en ambos puntos, se podría notar que están desfasadas por un tiempo t_r .

De esta forma se puede simular la concentración en la salida del equipo considerando un modelo de primer orden con retardo por transporte. Para el caso de una columna de flotación, este tiempo se puede estimar como el tiempo de desplazamiento que le toma al agregado recorrer por completo la zona de espuma.

Suponiendo que la zona de espuma propicia una zona de calma adecuada, el desplazamiento del agregado partícula-burbuja se compone por dos componentes, un movimiento vertical producto de la flotabilidad y otro horizontal ya que al llegar a la superficie se mueve de forma radial hacia las canaletas de concentrado, como se ve la Figura 3-4.

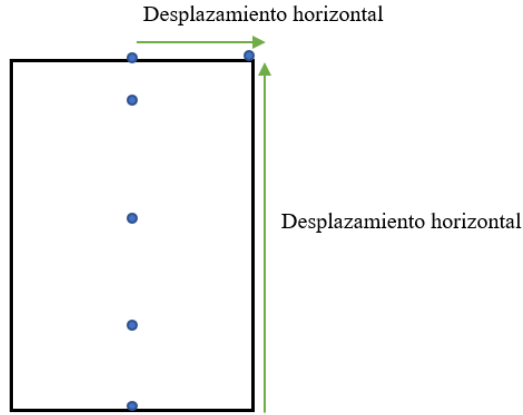


Figura 3-4 Desplazamiento del agregado en la zona de espuma

Luego el tiempo de retardo para una columna se define como:

$$t_r = td_v + td_h \quad (\text{Ec. 3-6})$$

En este sentido el tiempo de desplazamiento, puede ser determinando mediante la velocidad y la distancia recorrida en cada dirección la cual está determinada por las características geométricas de la columna.

El tiempo de desplazamiento td_v del agregado partícula-burbuja dentro de la zona de espuma, se obtiene a partir de la profundidad de espuma y la velocidad vertical, como sigue:

$$td_v = \frac{H_f}{v_{burbuja_v}} \quad (\text{Ec. 3-7})$$

De forma análoga se puede relacionar la velocidad horizontal y la distancia recorrida en esa dirección,

$$td_h = \frac{D_{col}}{2 \cdot v_{burbuja_h}} \quad (\text{Ec. 3-8})$$

Respecto de la velocidad con que se mueven los agregados en cada dirección, se tiene para la componente vertical como (Arrendondo, 2016),

$$v_{burbuja_v} = \frac{J_g}{\varepsilon_{g_{espuma}}} \quad (\text{Ec. 3-9})$$

Según investigaciones el valor del parámetro $\varepsilon_{g_{espuma}}$ que es la concentración de aire en la zona de espuma, se mueve entre un rango de 80-90 % (Arrendondo, 2016). Para efectos de esta investigación se definió tal valor en un 90%.

Para la componente horizontal la velocidad de desplazamiento se define según la siguiente ecuación,

$$v_{burbuja_h} = \left(\frac{J_g - \delta}{2 \cdot H_f \cdot \varepsilon_{g_{espuma}}} \right) \cdot \left(\frac{D_{col}}{2} \right) \quad (\text{Ec. 3-10})$$

Donde δ corresponde a un parámetro de corrección que indica la tasa de rompimiento de burbujas por unidad de volumen y tiempo. Este parámetro se ha determinado según investigaciones tomando un valor de 0,2 cm/s para flujos de aire en torno a 1 cm/s (Arrendondo, 2016).

Desde el punto de vista del diseño de experimentos, es común que se realicen perturbaciones del tipo escalón, por tanto, es esa respuesta en específico la que es de interés para esta investigación. Luego se puede obtener la respuesta escalón de un sistema de primer orden con retardo por transporte mediante la siguiente expresión,

$$y(t) = A \cdot K_i \cdot \left(1 - e^{-\frac{(t-t_r)}{\tau}} \right) \quad (\text{Ec. 3-11})$$

El desarrollo de este simulador dinámico, en conjunto con el simulador estático proponen una alternativa para la generación de datos necesarios para realizar el ajuste del sensor virtual. Si bien puede que el proceso real no se comporte como un sistema de primer orden en todo el rango operacional, para efectos de esta investigación es suficiente ya que los alcances que tiene esta investigación están enfocados a la aplicación de la metodología del ajuste de un sensor virtual, la cual puede ser replicada de la misma forma en caso de contar con datos de proceso reales.

Para complementar la explicación de la simulación de datos, a continuación, en la Figura 3-5 se presenta un esquema que expone como es la integración del simulador estático y como se utiliza en conjunto con el modelo de primer orden para generar datos dinámicos. Comenzando desde dos estados del proceso, para lo cual se calculan los resultados metalúrgicos estáticos, luego con esa información el simulador dinámico determina los parámetros del modelo de primer orden, obteniéndose así la trayectoria de los resultados metalúrgicos de interés.

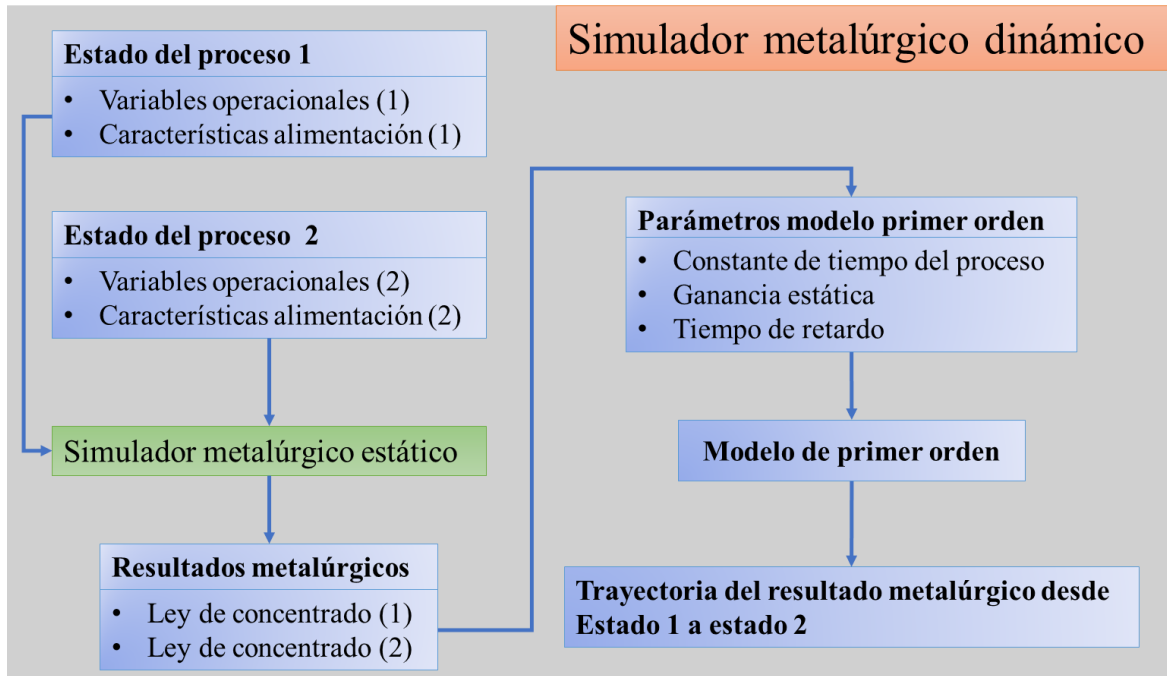


Figura 3-5 Esquema integración del simulador estático y dinámico

3.4. Diseño de experimentos

El objetivo del diseño de experimentos es generar un set de datos con el propósito de imitar la operación de una planta real. Para ello con la utilización del simulador formulado anteriormente se realizan una serie de perturbaciones según los niveles de las variables manipulables, los cuales se definen a continuación.

3.4.1. Selección de variables y sus niveles

Para poder simular los datos experimentales y que estos representen correctamente la mayoría de las zonas operacionales, se decide realizar un diseño experimental con todas las variables disponibles en el simulador. En este caso corresponden a 5 variables que caracterizan las propiedades de la alimentación: flujo de alimentación, ley de cobre, diámetro de partícula, densidad de pulpa y razón calcosina-calcopirita. Por otra parte, se cuenta con tres variables operacionales que se utilizan frecuentemente en el control de columnas, como son velocidades superficiales de aire y agua de lavado, además de la profundidad de espuma.

Se establecieron 5 niveles para cada una de las variables en consideración que cada perturbación tenga una relevancia en la ley de concentrado, pero que también permita generar estados

intermedios según las condiciones de operación. Los niveles de las variables, junto con la nomenclatura están presentados en la Tabla 3-4.

Tabla 3-4 Matriz de variables y sus niveles

Variable	Nomenclatura	Nivel					Unidad
		1	2	3	4	5	
Flujo de alimentación	Q_f	5	7	9	12	15	m3/min
Ley de cobre en alimentación	x_f	5	7	9	12	15	%
Diámetro de partícula	D_p	30	40	50	60	70	μm
Densidad de pulpa	ρ_{pulp}	1,10	1,15	1,20	1,25	1,30	g/L
Razón calcosina/calcopirita	$R_{c/c}$	10	15	25	35	50	%
Velocidad superficial del aire	J_g	1,0	1,5	2,0	2,5	3,0	cm/s
Altura de espuma	H_f	40	55	75	90	100	cm
Velocidad superficial de agua de lavado	J_w	0,0	0,03	0,05	0,08	0,1	cm/s

En cuanto a la forma y dimensiones geométricas estas se mantuvieron constante para todas las simulaciones, en este sentido se definió una columna de forma cilíndrica con las siguientes dimensiones de altura, diámetro y perímetro real, las cuales están presentadas en la Tabla 3-5.

Tabla 3-5 Dimensiones de la columna de flotación simulada

Variable	Valor	Unidad
Altura de la columna	1400	cm
Diámetro de la columna	500	cm
Perímetro real	2000	cm

3.4.2. Matriz de experimentos

Para la generación de los datos experimentales se realizan perturbaciones de forma aleatoria, para ello se elige una variable a la vez y se aumenta o disminuye un nivel. Para aislar los efectos de cada variable se modifica una variable a la vez, además este aumento es de más o menos un nivel según corresponda para que las perturbaciones no sean tan agresivas, es decir que una variable, por ejemplo, no pase desde el nivel 1 al 5. De esta forma se podrá obtener una serie de tiempo con variables operacionales y la ley de concentrado como si se tratase de una planta real.

1. Se establece un estado inicial (1) con las ocho variables disponibles.
2. De forma aleatoria se elige 1 de las variables a perturbar y el sentido que tendrá la perturbación, es decir si se aumenta o disminuye un nivel. En caso de que la variable se encuentre en el nivel 5 y la perturbación sea en la dirección de un aumento, se deberá disminuir un nivel en su lugar ya que no se puede aumentar una variable más allá de los niveles determinados. De forma análoga, si se determina la disminución de una variable en nivel 1, se deberá aumentar un nivel en su lugar.
3. Se calcula la respuesta transiente de la ley de concentrado, utilizando la Ecuación 3.11. Considerando el estado inicial (1) y el estado final (2). La expresión es computada para el espacio de tiempo determinado entre 0 y 10 constantes de tiempo, a intervalos de 1 minuto cada uno.

El tiempo establecido entre cada perturbación se relaciona con la duración de la respuesta transiente, la cual se modela en este caso mediante un modelo de primer orden. Según este modelo se alcanza un 98,2% del valor de la respuesta estacionaria en un tiempo igual a 4 constantes de tiempo. Luego, con el objetivo de capturar el desarrollo completo de la trayectoria transiente, el muestreo se extiende hasta 10 constantes de tiempo.

4. Se almacenan los resultados como serie de tiempo que contiene las ocho variables operacionales y la ley de concentrado de cobre.
5. Se actualiza el estado inicial por el estado dos, desde el cual se realizará la siguiente perturbación.
6. Se repiten los pasos 2 al 5 hasta alcanzar el número de perturbaciones esperadas.

3.4.3. Generación del set de datos

En este caso se realizaron un total de 200 perturbaciones, de acuerdo con el procediendo explicado en el punto anterior, en la Figura 3-6 se puede ver el resultado de la simulación, en la cual se presenta de forma gráfica la variación que tiene la ley de concentrado según los cambios que se van aplicando en las distintas variables operacionales.

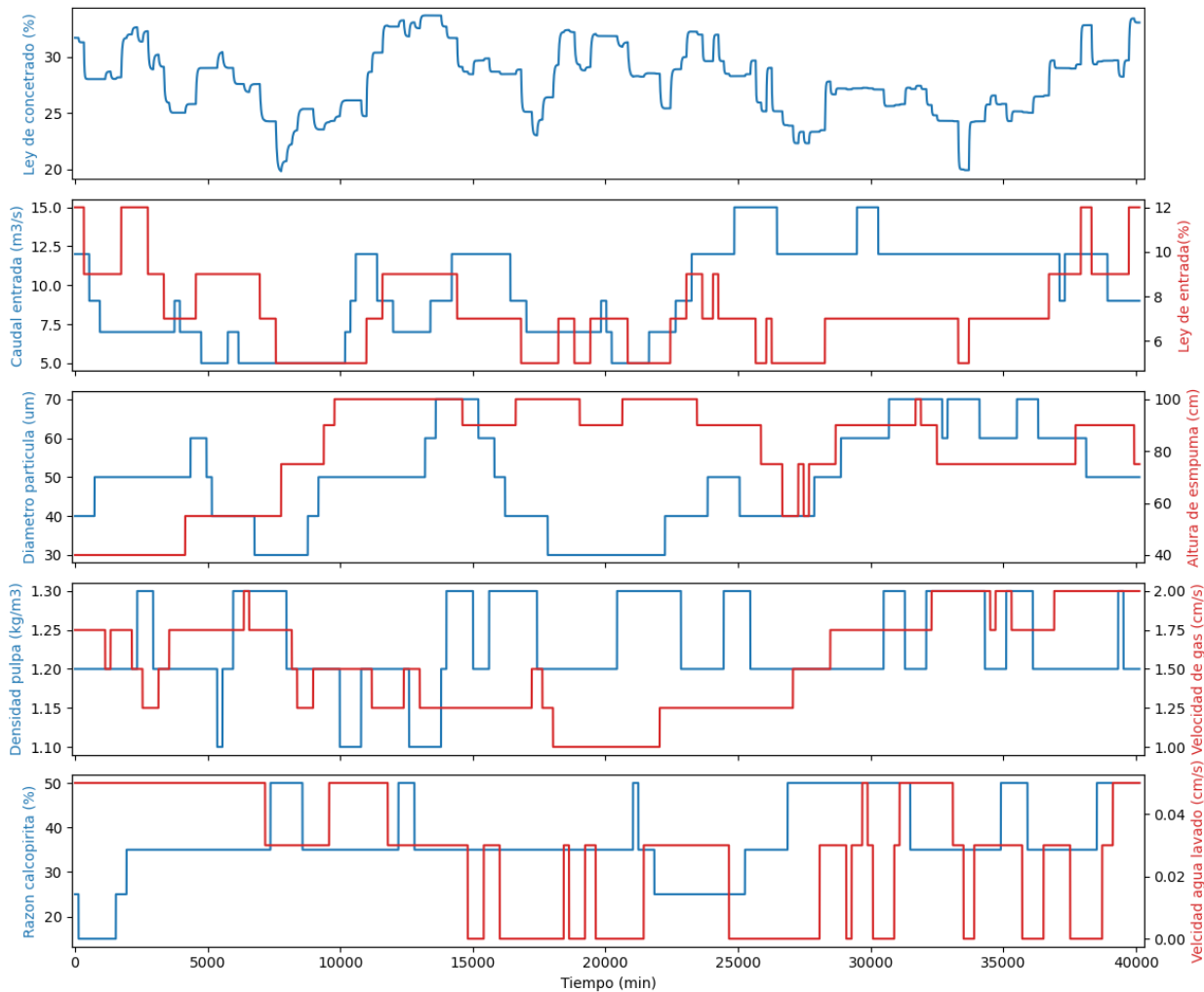


Figura 3-6 Resultados simulación: Ley de concentrado y niveles operacionales

Además, se presenta un pequeño análisis estadístico preliminar en la Tabla 3-6 que entrega información de la distribución de cada una de las variables involucradas en la simulación estableciendo así la zona de operación y ajuste del sensor. La tabla incluye indicadores de tendencia central, como el promedio y mediana, indicadores de posición como los cuartiles y medidas de dispersión como la desviación estándar. Adicional se presenta el índice de correlación de la ley de

concentrado x_c respecto de las variables operacionales, este indicador es una medida de dependencia lineal entre dos variables, el cual va desde -1 a 1, mientras más cercano es la magnitud de este valor a 1 mayor es la dependencia, y el signo determina si la relación existente es directamente o inversamente proporcional.

A primera vista se puede apreciar que, las variables H_f , x_f presentan un valor positivo y significativo, lo que es consistente con la fenomenología explicada, ya que al aumentar estas variables se espera un aumento en la ley del concentrado. Por otra parte, las variables que afectan de forma negativa la ley de concentrado corresponden a ρ_{pulp} , $R_{c/c}$ y J_g . Las variables restantes presentan un bajo índice de correlación por tanto no tienen mayor efecto en los cambios de la ley de concentrado.

Tabla 3-6 Resultados metalúrgicos obtenidos

Variable	Promedio	Desviación estándar	Cuartil 1	Cuartil 2	Cuartil 3	$Corr(x_c)$	Unidad
Q_f	9,5	3,0	7,0	9,0	12,0	-0,10	m3/min
x_f	7,3	1,8	7,0	7,0	9,0	0,68	%
D_p	49,1	12,6	40,0	50,0	60,0	-0,07	μm
ρ_{pulp}	1,2	0,1	1,2	1,2	1,3	-0,29	g/L
$R_{c/c}$	36,7	8,5	15,0	30,0	40,0	-0,26	%
J_g	1,5	0,3	1,2	1,5	2,5	-0,34	cm/s
H_f	80,1	19,4	50,0	90,0	90,0	0,21	cm
J_w	0,03	0,02	0,00	0,05	0,08	0,07	cm/s
x_c	27,9	3,0	25,4	28,2	29,7	1,0	%

Es importante destacar que, aunque estas correlaciones observadas en las simulaciones proporcionan información del proceso, se debe tener cautela al extrapolar directamente estos resultados a situaciones del mundo real.

Esto se debe a que las simulaciones, por su naturaleza modelada, simplifican e idealizan ciertos aspectos del sistema, y las relaciones identificadas pueden depender en gran medida de las suposiciones y parámetros seleccionados durante el proceso de modelado. Para este caso en

específico, el tratamiento del tamaño de partícula como una constante y no mediante una distribución de tamaño como es en la realidad, similar a lo que ocurre con el tiempo de residencia de las partículas, el cual no es el mismo para todas, así como la falta de consideración de factores de mineralogía como la liberación y cambios en la composición producto de la variabilidad intrínseca que existe en la alimentación pueden provocar diferencias con un equipo real. Por lo tanto, se debe reconocer las limitaciones y entender que las simulaciones ofrecen una representación aproximada de la realidad, pudiendo no establecer relaciones exactas.

De todas formas, existiendo diferencias respecto de un equipo real, para efectos de la metodología que es desarrollar un sensor virtual, los datos obtenidos pueden ser utilizados para entrenar el modelo dinámico ya que este debe ser capaz de reconocer las relaciones correspondientes a un conjunto de datos independiente si es de un entorno simulado o real.

3.5. Implementación modelo PLS Dinámico

3.5.1. Variables del modelo PLS y frecuencia de muestreo

Siguiendo con la implementación del sensor virtual, antes se deben construir los bloques de información X e Y utilizando los datos experimentales obtenidos, dicha información se debe disponer de forma conveniente de acuerdo con lo explicado en el capítulo anterior.

En primer lugar, se deben determinar cuáles son las variables que se utilizarán y estarán disponibles para poder realizar la predicción, en este caso son consideradas para el ajuste del sensor las variables de la alimentación y operacionales disponibles con las cuales se generaron los datos. Por otra parte, la variable a predecir u objetivo corresponde a la ley de concentrado, todas estas variables se detallan en la Tabla 3-7.

En algunos casos es posible, según el conocimiento que se cuente del proceso no considerar algunas variables que a priori se conozca que no tienen efecto en la variable a predecir. Sin embargo, si no se conocen del todo las interacciones que puedan tener las variables, es posible utilizar todas las variables disponibles sin problema ya que el algoritmo por si solo es capaz de determinar cuáles son las variables más importantes para construir el modelo y eventualmente descartar variables que no tengan incidencia en la variable objetivo.

Tabla 3-7 Detalle de las variables consideradas en el modelo PLS.

Variables predictoras		Variable Objetivo	
Q_f	Flujo de alimentación	x_c	Ley de cobre en concentrado
x_f	Ley de cobre en alimentación		
D_p	Diámetro de partícula		
ρ_{pulp}	Densidad de pulpa		
$R_{c/c}$	Razón calcosina/calcopirita		
J_g	Velocidad superficial del aire		
H_f	Altura de espuma		
J_w	Velocidad superficial de agua de lavado		

Otro punto importante tiene que ver con la frecuencia de muestreo, ya que, si bien con el uso de las simulaciones es posible obtener valores de la variable objetivo cada un minuto o menos, lo que ocurre en la realidad es que, si una variable es compleja de medir, difícilmente se tendrá una frecuencia de muestreo tan alta. En este sentido para que la implementación sea representativa del proceso real, en el ajuste del sensor se consideraron las mediciones cada 20 minutos, esta es una de las principales limitantes en una implementación real, ya que la frecuencia disponible de la variable a predecir será un factor determinante en la resolución del sensor virtual.

3.5.2. Software SOLO

La implementación del sensor se realizó en la herramienta *PLS Toolbox* el cual es un complemento desarrollado por la empresa *Eigenvector Research Incorporated* (Eigenvector, 2023) esta herramienta está basada en el *software* comercial *Matlab*. La herramienta incorpora una serie de métodos de análisis de datos que son de interés para científicos e ingenieros, entre reconocimientos de patrones, métodos de clusterización y clasificación, como también regresiones.

En este caso se utilizó el módulo de análisis de regresión PLS, el cual su interfaz se puede ver en la Figura 3-7. Al utilizar esta herramienta, la implementación se hace de forma muy intuitiva siguiendo un flujo de trabajo, el cual está marcado con un rectángulo azul ubicado en la parte derecha de la figura presentada, en donde se importan y se pre-procesan los datos experimentales,

además se pueden obtener gráficos con los indicadores estadísticos más importantes, con los que se puede asegurar si el modelo es representativo de los datos, y en consecuencia determinar si es adecuado para la aplicación buscada.

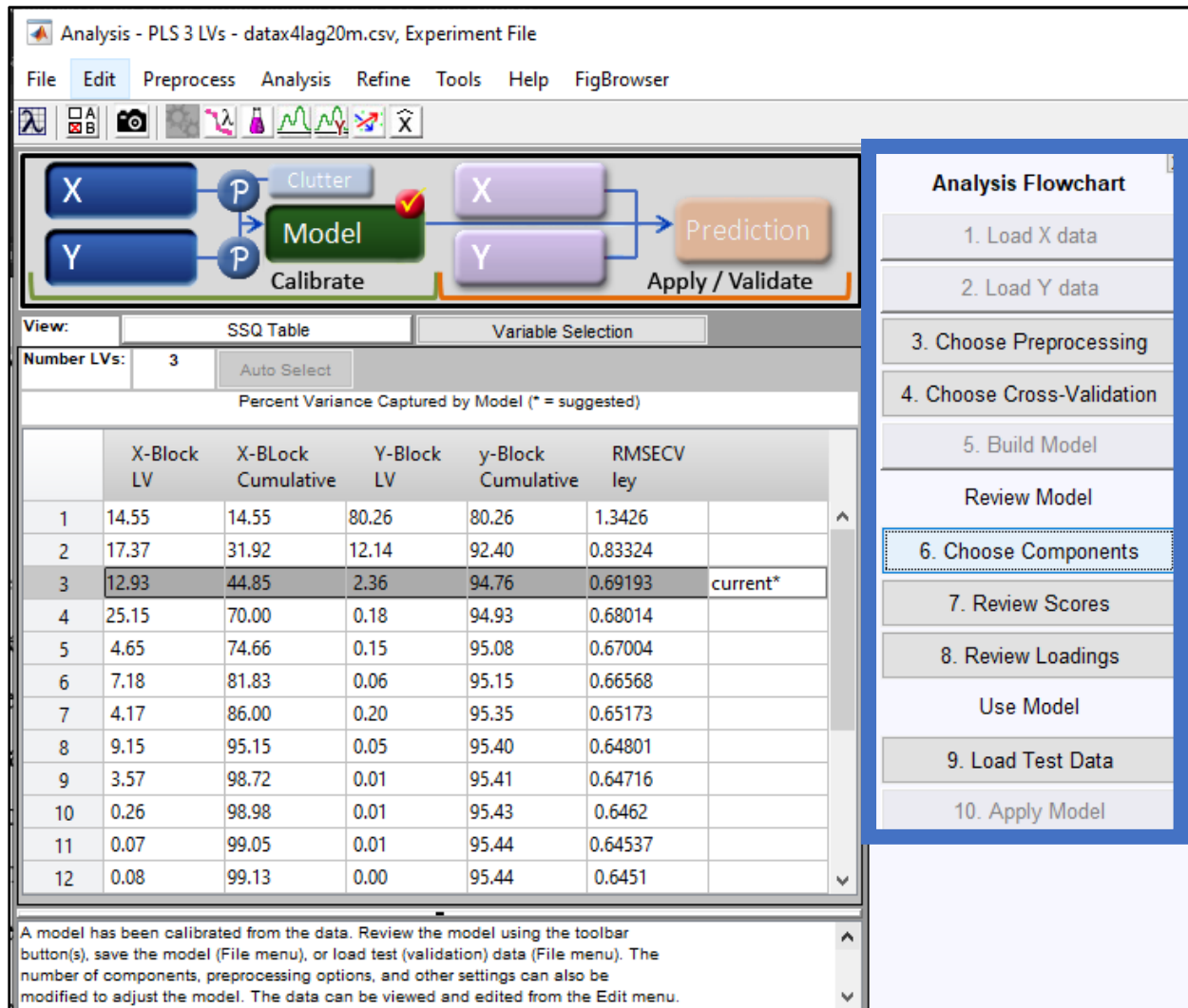


Figura 3-7 Modulo análisis regresión PLS

A continuación, se presentan de forma resumida los pasos para obtener el modelo:

- i. Cargar datos experimentales

El primer paso corresponde a la carga de las matrices X e Y, las cuales contienen las variables regresoras y objetivo respectivamente. La forma en que se dispone la información, debe ser con las observaciones como filas y las variables en las columnas.

ii. Preprocesamiento

El programa tiene disponible una serie de herramientas útiles para el preprocesamiento de datos como: normalización, escalado, filtros, entre otros. En este caso se utilizó únicamente la opción de auto escalado, ya que es una de las condiciones necesarias para aplicar la metodología PLS, lo que hace esta opción es centrar los datos en cero y escalarlos para que su desviación estándar sea la unidad, de esta forma se puede asegurar que las diferentes dimensiones de las medidas no tengan repercusiones en la calidad del modelo.

iii. Validación cruzada

Es común en el marco del análisis de regresiones utilizar la validación cruzada (*Cross-Validation*) esto para evaluar los resultados el modelo obtenido, en este caso para cada una de las iteraciones que realiza el programa se determina un subconjunto de observaciones para validar el modelo y el resto se utiliza para el entrenamiento de dicho modelo. El tamaño de ese subconjunto usualmente corresponde al 10% de los datos totales (Jolliffe, 2002), el cual es el valor utilizado en esta investigación.

iv. Construcción del modelo

Finalmente se selecciona *Build Model* para que el programa compute el algoritmo y determine los modelos, en este caso la herramienta calcula los modelos que consideran desde 1 a 20 componentes principales, entregando información respecto de la bondad de ajuste de cada uno, en la Figura 3-7 se pueden ver indicadores como, la varianza capturada de X e Y, además del error cuadrado medio de validación cruzada. Como información adicional, la herramienta recomienda al usuario un modelo con determinadas componentes principales como modelo preliminar, una vez confirmado el modelo se pueden analizar los distintos indicadores como los estadígrafos T^2 y Q, además se puede obtener y exportar el vector de regresión según la forma obtenida en la Ecuación 2-11.

3.6. Modelo FIR

3.6.1. Bloque X e Y

Según lo expuesto anteriormente, para capturar correctamente la dinámica del proceso se deben incluir valores pasados de las variables en el bloque X, en este caso para el modelo FIR solo se incluyen valores pasados de las variables de entrada.

Para este proceso en específico, a partir de las simulaciones realizadas el valor de la constante de tiempo según la Ec. 3-5, está entre 18 a 25 minutos. Luego, en consideración de la dinámica del proceso se agregaron un total de 4 *lags* o valores pasados de las variables predictoras para obtener una resolución adecuada en la respuesta dinámica del proceso.

La elección de la cantidad de valores pasados a considerar en el modelo tiene relación con la duración de la respuesta transiente del proceso y la propia frecuencia de muestreo de la variable que en este caso es de 20 minutos. La dinámica del proceso indica que la estabilización frente a un cambio se completa en cuatro constantes de tiempo, luego la adición de cuatro los valores pasados en los intervalos de la frecuencia de muestreo puede capturar de forma completa la respuesta.

Evidentemente que una aplicación real, la cantidad de valores dependerá de la frecuencia de muestreo con la que se mide la variable a predecir y también la propia dinámica del proceso, con lo que se debe buscar un equilibrio entre la complejidad de la matriz X y la precisión del sensor.

Siguiendo con la implementación, el bloque X está compuesto por las variables y sus respectivos valores pasados como sigue, en primer lugar, se define el conjunto de observaciones u que contiene las variables de entrada,

$$u(t) = [Q_f(t) \ x_f(t) \ D_p(t) \ \rho_{pulp}(t) \ R_{c/c}(t) \ J_g(t) \ H_f(t) \ J_w(t)] \quad (\text{Ec. 3-12})$$

Finalmente, el bloque X se obtiene al construir la matriz ampliada según la Ecuación 2-18 adicionando los respectivos valores pasados, en este caso cuatro, quedando el bloque de la siguiente forma,

$$X(h = 4) = [u(t) \ u(t - 1) \ u(t - 2) \ u(t - 3) \ u(t - 4)] \quad (\text{Ec. 3-13})$$

Como resultado la dimensionalidad del bloque X es de 2007 observaciones y 40 variables, por otra parte, la dimensionalidad del bloque Y es de 2007 observaciones y 1 variable.

3.6.2. Componentes principales

Una vez que se han construido los bloques de información X e Y correctamente, se deben seguir los pasos explicados en el Capítulo 3.5.2, en donde será necesario determinar la cantidad de componentes principales que determinaran el modelo final. En esta investigación el criterio

utilizado para la elección de estas componentes, tiene relación con la varianza capturada por el bloque X, para el cual se elegirán las componentes principales que expliquen el 80 % de la varianza.

Se debe notar que las componentes se ordenan de mayor a menor según su aporte a la varianza capturada en ambos bloques de información, en este sentido mientras mayor es la cantidad de variables latentes, mayor es la varianza capturada en ambos vectores X e Y, sin embargo, se corre el riesgo de un sobreajuste del modelo provocando una disminución en la capacidad predictiva del modelo lo cual no es recomendable. Luego, es importante destacar que, no solo basta con elegir un modelo según una varianza específica ya que, como cualquier modelo, este se debe ajustar adecuadamente a los datos por medio del error y además que cumplan los límites establecidos según los estadígrafos T^2 y Q.

En la Tabla 3-8 se presenta el resumen estadístico que incluye las varianzas capturadas en el bloque X e Y por cada componente, y también el error cuadrado medio de validación cruzada. En este caso según el criterio utilizado, el 80% de la varianza capturada se alcanza con 6 variables latentes, siendo este el valor seleccionado para la obtención del modelo.

Tabla 3-8 Resumen estadístico de las componentes principales en modelo FIR

CP	Varianza en X	Varianza acumulada de X	Varianza en Y	Varianza acumulada de Y	RMSECV: Ley concentrado
1	14,51	14,51	80,16	80,16	1,3463
2	17,38	31,89	12,18	92,34	0,8366
3	12,90	44,79	2,37	94,71	0,69537
4	25,16	69,95	0,18	94,89	0,6836
5	4,67	74,61	0,15	95,04	0,67347
6	7,54	82,15	0,07	95,11	0,66918
7	4,72	86,87	0,20	95,32	0,65649
8	8,34	95,21	0,06	95,38	0,65241
9	3,51	98,72	0,01	95,39	0,65176
10	0,26	98,98	0,01	95,40	0,65182

3.6.3. Límites estadígrafos T^2 y Q

Una vez que se ha establecido el modelo final se debe analizar su bondad de ajuste según los estadígrafos presentados anteriormente. Usualmente estos indicadores están reportados en unidades que son sensibles al número de variables en las matrices de información, el número de componentes y el tipo de preprocesamiento (Eigenvector, 2023). Luego, para poder realizar una comparación entre distintos modelos se requiere la normalización de dichos valores. Por lo que es común presentar ambos estadígrafos como cantidades reducidas, para obtener los indicadores reducidos se calcula un límite de confiabilidad (establecido a un 95% de significancia), de esta forma se dividen los estadígrafos de cada observación por dicho límite.

Una observación es considerada anormal si su valor está por sobre el 50% del límite, luego otra de las ventajas que se tiene al presentar los estadígrafos reducidos es que resulta más conveniente para determinar las observaciones anormales, ya que cualquier punto que sobrepasa el umbral de 1,5 es considerado anormal.

En la Figura 3-8 se presenta la comparación entre los datos predichos por el modelo y los datos experimentales, además se presenta un mapa de color que relaciona el estadígrafo Q Residuals reducido de cada observación. En este caso se puede observar que la coloración es mayormente azul lo que significa que se encuentran dentro de los rangos esperados del modelo y, por tanto, el modelo explica una mayor correlación, y a priori resultaría ser un ajuste adecuado para los datos.

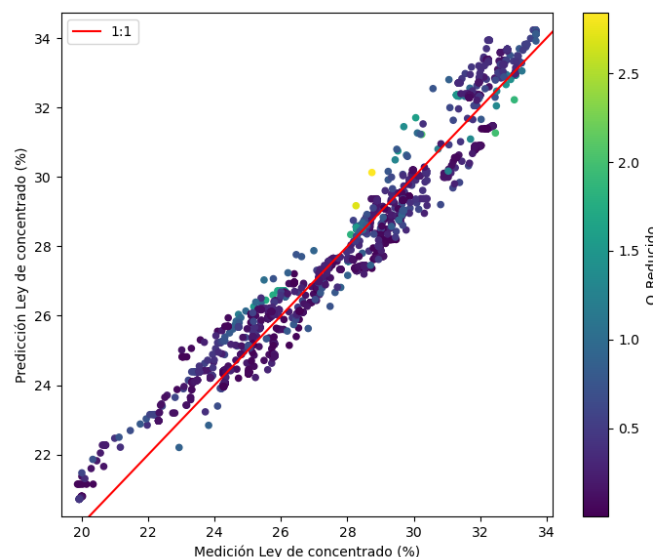


Figura 3-8 Modelo FIR vs Simulador, Q Residuals

En la Figura 3-9 se presenta la comparación entre los datos predichos por el modelo y los datos experimentales, además se presenta un mapa de color que relaciona el estadígrafo *Hotelling's* reducido de cada observación. Para este estadígrafo se puede ver que la coloración del grafico está predominante en torno al 1, sin embargo, se pueden apreciar ciertas zonas en donde el indicador se acerca al umbral especialmente en las zonas en donde el resultado metalúrgico se encuentra cercanos a los valores mínimos y máximos. En este sentido existen algunas observaciones que cumplen lo esperado por el modelo, sin embargo, están entregando resultados anormales.

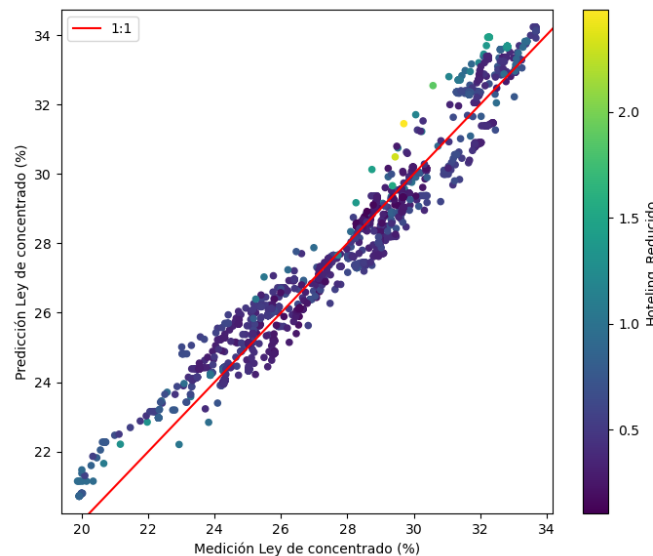


Figura 3-9 Modelo FIR vs Simulador, *Hotelling's* Reducido

Una vez obtenido el modelo, se puede exportar el vector de regresión correspondiente al modelo con 6 componentes principales, el vector es presentado en la Tabla 8-1 ubicada en el Anexo 1.

3.6.4. Análisis de residuos

En la Figura 3-10 se presentan los residuos del modelo generado con 6 componentes principales. Los residuos están calculados como la diferencia entre la predicción del modelo y los datos experimentales, en la figura se pueden apreciar ciertas zonas en las que el modelo no se ajusta a los datos experimentales. En específico para leyes de concentrado menores a 24% el modelo tiende a sobre estimar el valor experimental, caso contrario se ve en la zona correspondiente a 26-28% en donde el modelo en la mayoría de los casos entrega una subestimación. Y finalmente en el resto del rango operacional se ve una alta dispersión de los residuos, especialmente entre 30-32%

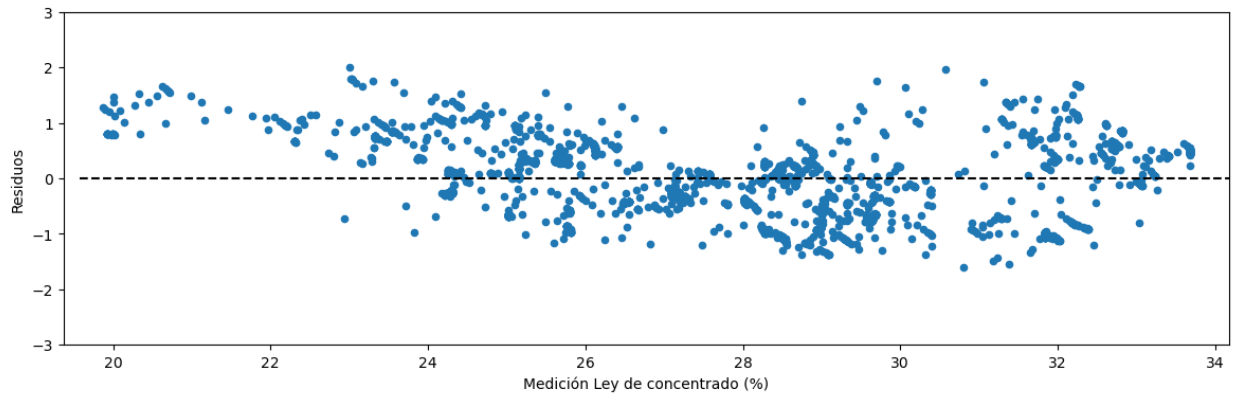


Figura 3-10 Residuos Modelo FIR

3.7. Modelo corregido

A partir del análisis de estadígrafos y residuos, es posible evidenciar que el modelo en algunas zonas de operación no tiene un buen ajuste. Esto en cierto modo es esperado, ya que, si bien la metodología PLS realiza una regresión más eficiente, factores como la complejidad del proceso, su dinámica y las interacciones que puedan existir entre las variables, resulte en que el proceso no pueda ser representado según un modelo lineal en todo su rango de operación.

En este sentido las opciones a seguir pueden ser, por ejemplo, restringir el uso del modelo a zonas en donde el ajuste sea óptimo, otra opción es que se reconozca que el proceso no sigue un modelo lineal determinándose así algún otro tipo de modelo no lineal, y por último, que es la opción que se utilizó en esta investigación es considerar la formulación de varios modelos locales que representen de mejor forma el proceso en cada una de las zonas mejorando así la predicción global. Esta opción resulta útil, ya que se mantiene la simplicidad de una regresión lineal respecto con el manejo e implementación, y además se mantienen las relaciones lineales las cuales son útiles para ser utilizadas en el control automático.

Luego, la principal dificultad con respecto a la aplicación de modelos locales, es determinar las zonas en donde el modelo global tiene una falta de ajuste. Una de las alternativas más intuitivas es enfocar el análisis en base a una condición central de operación y desde ahí realizar un análisis de sensibilidad en donde se aplican perturbaciones, para las cuales se contrasta el valor medido con el valor predicho, similar a un diseño experimental.

Sin embargo, una de las desventajas de esta aproximación es que se debe contar con un alto conocimiento del proceso, ya que en la medida que aumenten las variables involucradas también se deberán realizar más perturbaciones, de esta forma el conocimiento ayuda a determinar cuáles son las variables que más impactan en la predicción. Otra de las desventajas tiene que ver con las interacciones que puedan existir entre variables, en el sentido de que su impacto en los residuos es complejo de determinar, ya que al tratarse de un análisis de sensibilidad modificando de una variable a la vez, el efecto de dicha interacción no se podrá distinguir a simple vista.

En este trabajo para poder determinar las zonas en donde se aplicarán los modelos locales, se propone realizar una clasificación de los residuos utilizando un modelo de árbol de decisión.

3.7.1. Caracterización de los residuos

Para determinar las zonas de los modelos locales, el primer paso es categorizar la variable a clasificar, en este caso el objetivo de los modelos locales es reducir y mantener el error de predicción dentro de una banda aceptable. Luego una vez calculados los errores porcentuales de cada observación, se categoriza cada una de ellas según el detalle presentado en la Tabla 3-9.

Tabla 3-9 Límites banda de aceptación y N ° de observaciones

Error (%)	Error (Categorico)	N ° Observaciones
[0; 2]	0	1132
[2; ∞ [	1	875

Como se puede ver en la tabla, se clasifica la observación como dentro de la banda (0) si su error porcentual es menor al 2%, existiendo un total de 1132 dentro de dicha banda. Por otra parte, si una observación tiene un error mayor al 2%, se considera fuera de la banda de aceptación (1), quedando las 875 observaciones restantes en esta clasificación. Una vez construida la variable según la que se clasificaran las observaciones, es tiempo de aplicar el modelo de clasificación según árboles de decisión.

En este caso se ejemplifica la aplicación de modelos locales con un 2%, sin embargo los límites de la banda para el error pueden ser modificados según el uso estimado para el sensor y/o la precisión esperada para la medición, eventualmente si se requiere aún más precisión se puede definir una banda más pequeña, eso sí para que la clasificación sea óptima posiblemente el nivel de podado

aumente, generando una mayor cantidad de modelos locales, por lo que se debe buscar un equilibrio entre la precisión y la cantidad de modelos locales.

3.7.2. Clasificación de los residuos

En este caso la implementación del árbol de decisión se realizó mediante el lenguaje de programación *Python*, utilizando la librería de código abierto *scikit-learn*, esta librería tiene foco en herramientas de *machine learning*, en la cual están disponibles diversos algoritmos de clasificación, clusterización y regresión, entre ellos el método de árboles de decisión (Scikit-learn, 2023). Se aplica el modelo de clasificación de árbol de decisión según el paquete de código abierto *scikit-learn* y su modelo *DecisionTreeClassifier*, en este caso en la Figura 3-11 se presenta el árbol de decisión con una profundidad máxima o podado a nivel 2.

En este caso se puede ver que el modelo de clasificación resulta en cuatro nodos terminales, correspondientes a los nodos número 2, 3, 5 y 6. Además en cada uno de los nodos se puede apreciar la clase que predomina según las reglas jerárquicas que se han establecido, es decir, si la observación presenta un error dentro o fuera de la banda de aceptación.

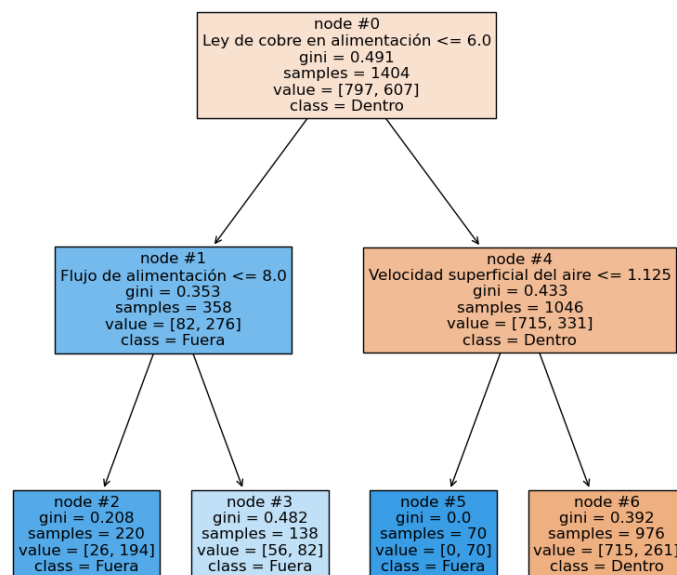


Figura 3-11 Modelo árbol de decisión: Errores modelo FIR

Para la formulación de modelos locales, se utilizarán los nodos terminales como referencia para generar el subconjunto de observaciones. La idea es generar un modelo local específico para los nodos terminales en donde el árbol de clasificación determine que esas observaciones están fuera de la banda, esto con el fin de mejorar la predicción en dicha zona operacional.

En el caso de los nodos terminales donde la clasificación indique que están dentro de la banda de aceptación, existen dos opciones, la primera es, en consideración de que el modelo global ya entrega una predicción adecuada es seguir utilizando ese modelo, la otra opción es generar un modelo local que considere todas las observaciones restantes que no fueron consideradas en los otros modelos locales, siendo esto lo que se realizó en este trabajo, además esta aproximación genera un mejor modelo para aquellas observaciones.

Para obtener el set de reglas que define cada uno de los subconjuntos, se debe tener en cuenta la convención que existe para interpretar el diagrama, en este sentido, cuando se realiza la división del nodo según la regla divisoria, aparecen dos ramas, el nodo generado que se conecta con la rama izquierda contiene las observaciones que cumplen dicha regla, caso contrario con el nodo que se conecta con la rama derecha, que agrupa las observaciones que no cumplen la regla. De esta forma, desde el nodo principal se avanza hasta los nodos terminales agregando cada una de las reglas según corresponda. A partir del árbol de decisión presentando Figura 3-11 se extraen las reglas que definen los subconjuntos según el procedimiento explicado anteriormente, obteniéndose así las zonas de aplicación de los modelos locales que son presentadas en la Tabla 3-10.

El modelo local generado por el nodo 5 se desestimó, ya que el vector de regresión no presentaba coherencia con la fenomenología de proceso, esto explicado por la baja cantidad de observaciones del subconjunto, no habiendo suficiente información para modelar correctamente dicha zona operacional de forma local.

Tabla 3-10 Modelos locales: Subconjuntos de aplicación y observaciones

Modelo	Subconjuntos	N ° Observaciones
1	$(x_f \leq 6,0) \& (Q_f \leq 8,0)$	313
2	$(x_f \leq 6,0) \& (Q_f > 8,0)$	180
3	Todas las observaciones, menos los subconjuntos 1 y 2.	1514

3.7.3. Construcción de modelos locales

Una vez que se han determinado las zonas de operación para los modelos locales, con cada uno de los subconjuntos se realiza el mismo procedimiento presentado desde los Capítulos 3.6.1 hasta el 3.6.3. Cabe destacar que se mantiene la forma del modelo, por tanto, se deben incluir también valores pasados de las variables como quedó determinado en la Ecuación 3-13.

En la Tabla 3-11, se presenta la elección de las componentes principales para los distintos modelos locales. Según el criterio utilizado anteriormente, los modelos explican el 80% de la varianza en X entre las 5 y 6 componentes principales, similar al modelo global el cual se determinó con 6 componentes.

Por efectos de espacio, para cada modelo local solo se muestran los indicadores como la varianza acumulada y RMSECV, el detalle para cada modelo en torno al comportamiento de aquellos indicadores en función de las componentes principales se presenta en el Anexo 2.

A primera vista una de las observaciones más interesantes que entregan los modelos locales, es la disminución que tiene el error de calibración/validación, ya que disminuye notablemente en comparación al modelo global presentando anteriormente, el cual presenta un valor de 0,66918. Este valor disminuye en casi la mitad en algunos modelos locales, en este sentido en se debería esperar un mejor ajuste en toda la zona operacional.

Tabla 3-11 Resumen de CP, Varianza acumulada de X e Y, RMSECV en cada modelo local

Modelo	CP	Varianza acumulada de X	Varianza acumulada de Y	RMSECV: Ley concentrado
1	6	81,85	98,21	0,35461
2	5	86,69	98,05	0,41167
3	5	80,02	97,72	0,39669

3.7.4. Límites estadígrafos T^2 y Q

En la Figura 3-12 se presenta la comparación entre los datos predichos por cada uno de los modelos locales y los respectivos datos experimentales, además se presenta un mapa de color que relaciona el estadígrafo *Q Residuals* reducido de cada observación.

En la figura se puede observar que para el caso de los modelos 1 y 2, la gran mayoría de las observaciones, se encuentran dentro del límite establecido en cada modelo, predominando la coloración azul en dichos modelos. Para el caso del modelo 3, si bien se pueden apreciar observaciones que superan el límite, son casos puntuales y sin presentar desviaciones importantes desde el límite. Con lo anterior se asegura que los datos cumplen lo esperado por cada uno de los modelos locales.

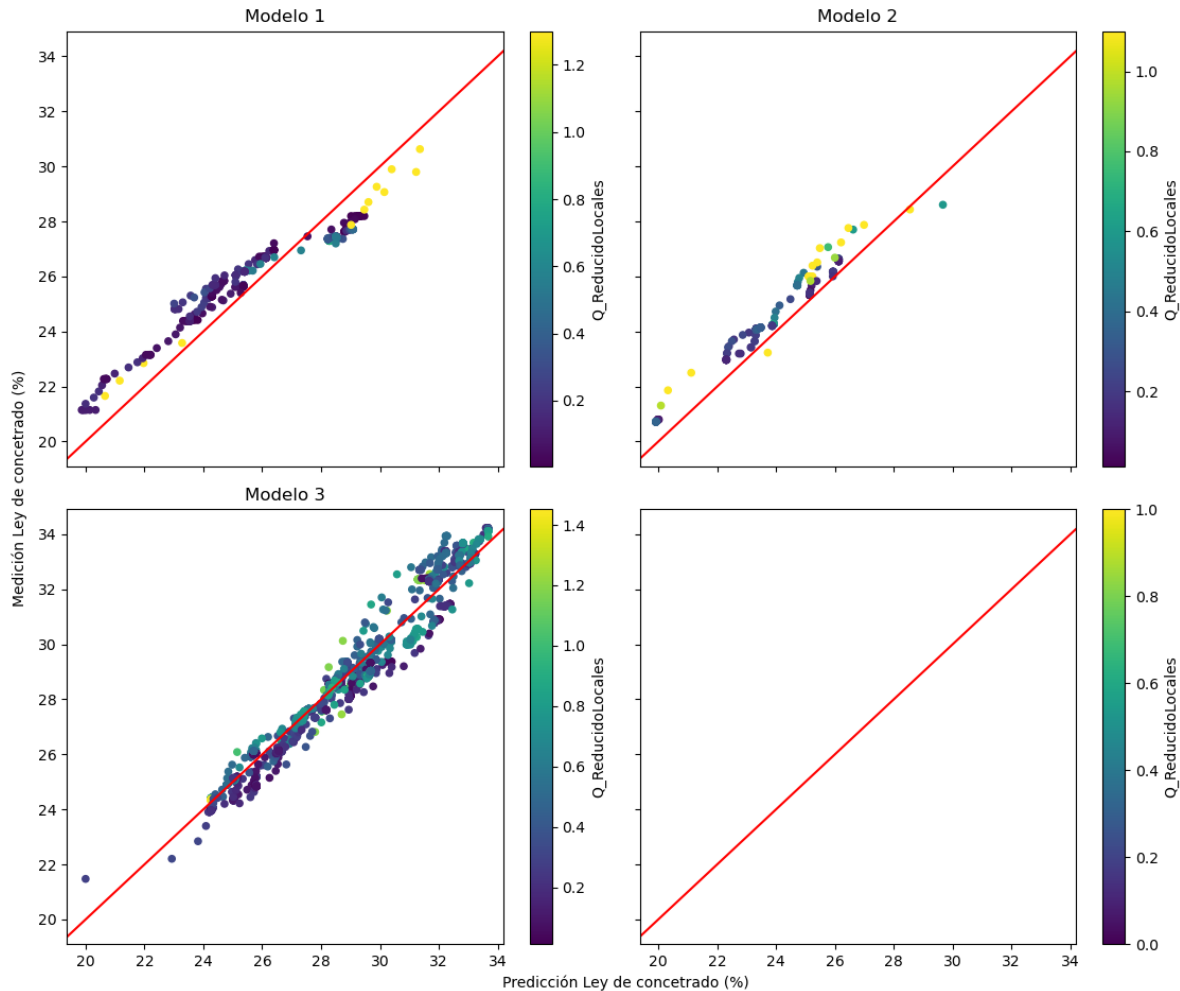


Figura 3-12 Modelos Locales vs Simulador, Q Residuals

De forma análoga en la Figura 3-13 se presenta la comparación entre los datos predichos por cada uno de los modelos locales y los respectivos datos experimentales, además se presenta un mapa de color que relaciona el estadígrafo *Hottelling's* reducido de cada observación.

Desde la figura se puede apreciar que para el caso de los modelos 1 y 3, el indicador Hottelling's está dentro de los límites establecidos para cada modelo, no habiendo observaciones anormales. En el caso del modelo 2, se puede apreciar que existen algunas observaciones que aumentan la variación, pero están presentes en baja cantidad y no se alejan en exceso desde el límite.

Respecto a este indicador, se tiene entonces que los tres modelos locales se encuentran en rangos aceptables, sin zonas en las que predominen observaciones anormales. En este sentido, las observaciones están cumpliendo con lo esperado en el modelo.

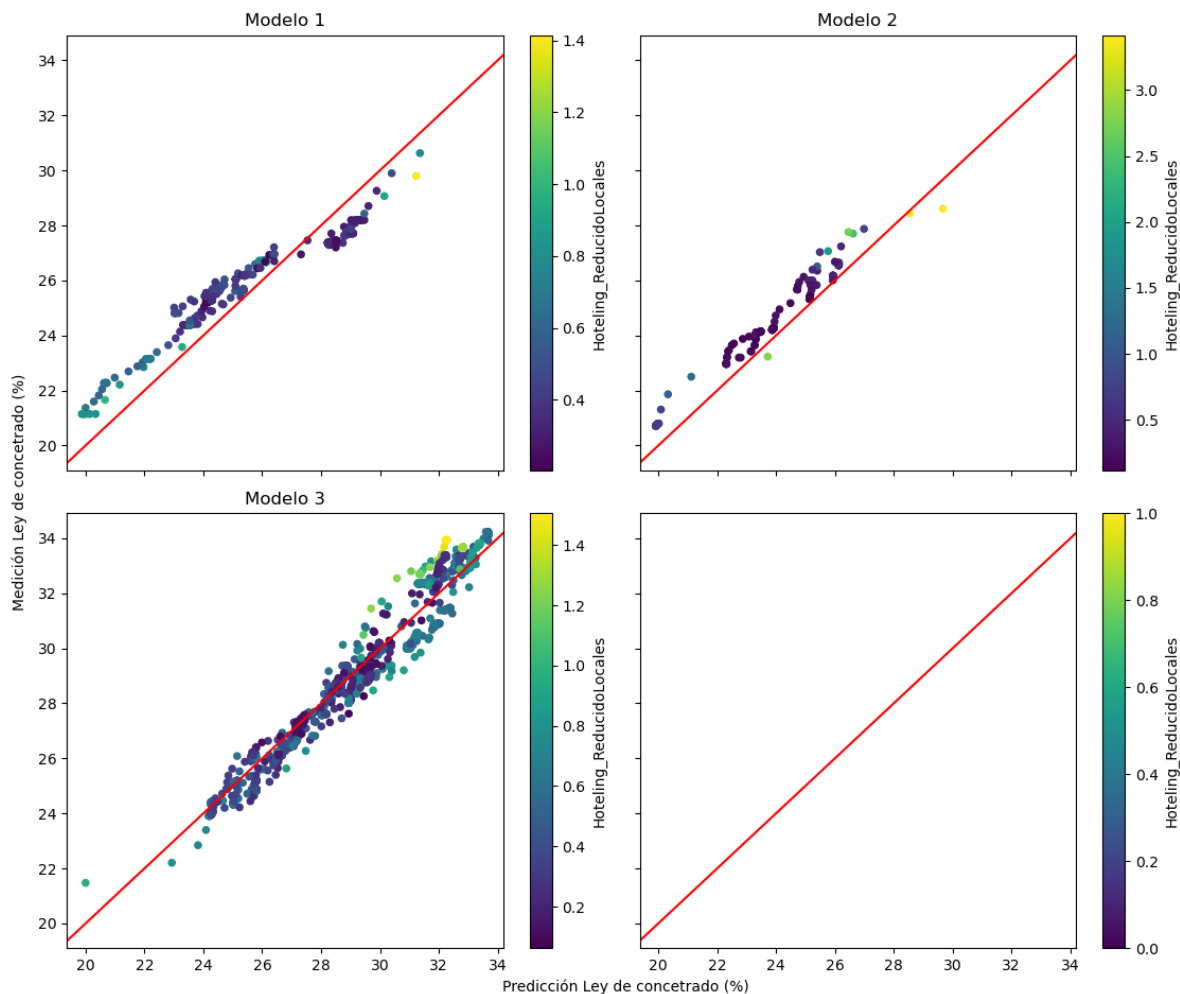


Figura 3-13 Modelos Locales vs Simulador, Hotelling's Reducido

Una vez obtenido los modelos locales, se pueden exportar los vectores de regresión correspondientes, dichos vectores están presentados desde la Tabla 8-2 a la Tabla 8-4 ubicadas en el Anexo 1.

3.7.5. Análisis de residuos

En la Figura 3-14 se presentan los residuos según los cuatro modelos locales determinados. Los residuos están calculados como la diferencia entre la predicción del modelo y los datos experimentales, en la figura se pueden apreciar mejoras notables respecto de la aplicación con un modelo global, ya que se tiene una distribución de los residuos mucho más compacta y en torno al cero, por otra parte, no se aprecian zonas en donde exista algún comportamiento marcado en el que el modelo para ciertas zonas realice subestimaciones o sobreestimaciones, como ocurría en el caso del modelo global.

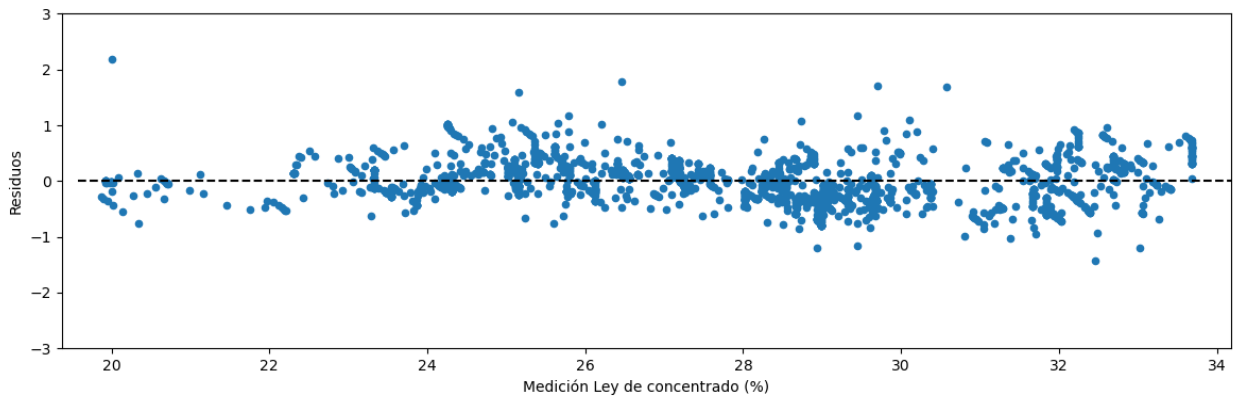


Figura 3-14 Residuos Modelos Locales

3.8. Modelo ARX

3.8.1. Bloque X e Y

A continuación, se presenta la construcción del modelo autorregresivo (ARX) que, a diferencia de los modelos presentados anteriormente, además de incluir en el bloque X valores pasados de las variables de entrada, incluye también valores pasados de la variable objetivo, en este caso la ley de concentrado.

Para la construcción del modelo se sigue el mismo flujo que en los otros modelos, en primer lugar, se deben incluir los respectivos valores pasados en el bloque X, en este sentido se define el vector u que contiene las variables de entrada,

$$u(t) = [Q_f(t) \ x_f(t) \ D_p(t) \ \rho_{pulp}(t) \ R_{c/c}(t) \ J_g(t) \ H_f(t) \ J_w(t)]. \quad (\text{Ec. 3-14})$$

Para el caso de los modelos con una forma autorregresiva, la cantidad de valores pasados usualmente es mucho menor que para el modelo FIR, en este caso el valor adecuado es de 1 valor pasado, para determinar dicha cantidad se ajustaron modelos con un mayor número de *lags*, no encontrándose una mejora significativa en el error de predicción, y por tanto no justificando el aumento de la complejidad del modelo, quedando finalmente el bloque X como sigue,

$$X(h = 1) = [u(t) \ u(t - 1) \ x_c(t - 1)] \quad (\text{Ec. 3-15})$$

Notar que incluye el vector $x_c(t - 1)$ el cual representa 1 *lag* para la concentración de salida (variable objetivo). Como resultado la dimensionalidad del bloque X es 2010 observaciones y 17 variables, por otra parte, la dimensionalidad del bloque Y es de 2010 observaciones y 1 variable.

3.8.2. Componentes principales

Una vez construido los bloques de información se procede a construir el modelo en el software SOLO, en la Tabla 3-12 se presenta el resumen estadístico de los indicadores del modelo ARX, en este caso se puede ver el que 80% de la varianza acumulada en el bloque X se alcanza con 7 componentes que es el valor elegido para el modelo.

Tabla 3-12 Resumen estadístico de las componentes principales en modelo ARX

CP	Varianza en X	Varianza acumulada de X	Varianza en Y	Varianza acumulada de Y	RMSECV: Ley concentrado
1	17,59	17,59	91,40	91,40	0,88689
2	17,88	35,46	5,93	97,33	0,49388
3	13,14	48,61	0,95	98,28	0,39678
4	23,26	71,87	0,10	98,38	0,38463
5	4,28	76,14	0,23	98,61	0,35649
6	3,37	79,51	0,58	99,20	0,27257
7	8,53	88,04	0,19	99,39	0,23756
8	7,80	95,84	0,01	99,41	0,23473
9	3,70	99,54	0,01	99,41	0,23376
10	0,13	99,67	0,06	99,47	0,22346

3.8.3. Límites estadígrafos T^2 y Q

En la Figura 3-15 se presenta la comparación entre los datos predichos por el modelo ARX y los datos experimentales, además se presenta un mapa de color que relaciona el estadígrafo Q *Residuals* reducido de cada observación.

En este caso se observa que la gran mayoría de las observaciones se encuentran dentro del límite establecido para el modelo, con una coloración de color azul predominante, habiendo pocas observaciones anormales y que no se disparan del límite, estableciéndose así que los datos cumplen lo esperado por el modelo.

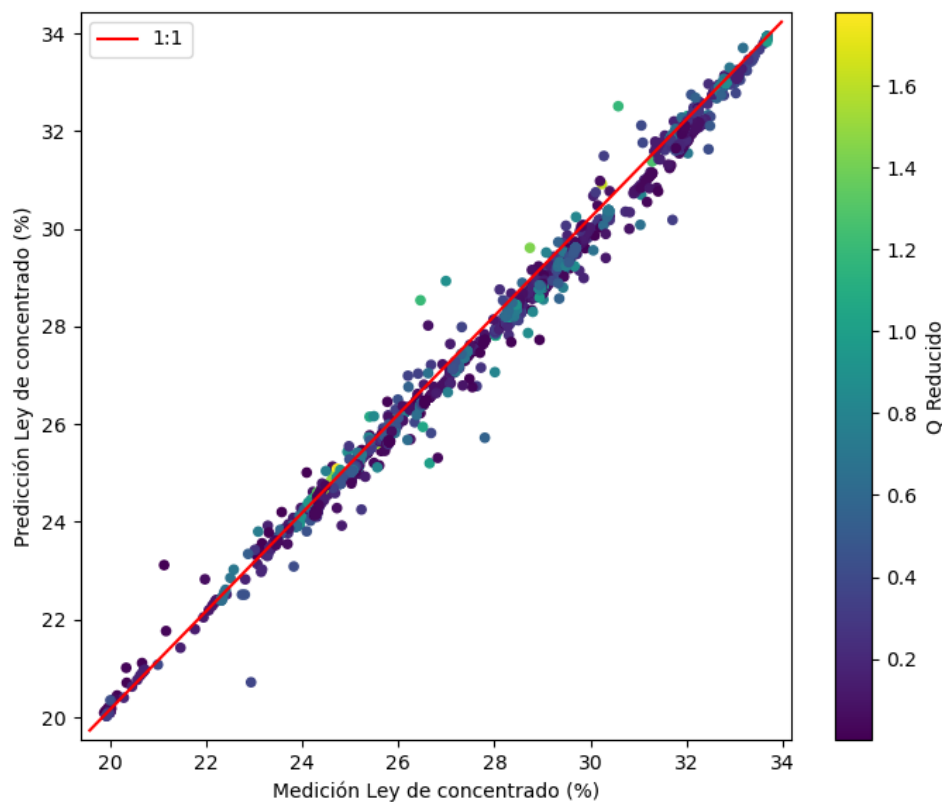


Figura 3-15 Modelo ARX vs Simulador, Q *Residuals*

En la Figura 3-16 se presenta la comparación entre los datos predichos por el modelo ARX y los datos experimentales, además se presenta un mapa de color que relaciona el estadígrafo *Hotelling's* reducido de cada observación, en la figura se puede apreciar, al igual que para el indicador anterior, que las observaciones cumplen en su mayoría con los límites establecidos, por tanto, se asegura que las observaciones están cumpliendo lo esperado por el modelo.

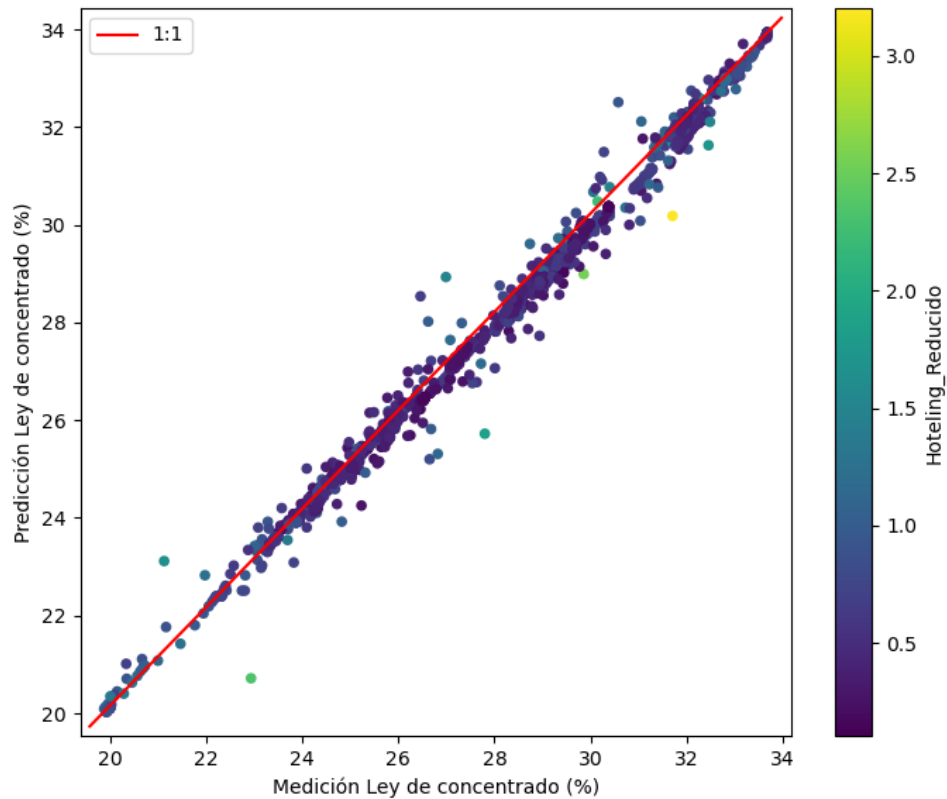


Figura 3-16 Modelo ARX vs Simulador, Hotelling's Reducido

Una vez obtenido el modelo, se puede exportar el vector de regresión correspondiente al modelo con 7 componentes principales, el vector es presentado en la Tabla 8-5 ubicada en el Anexo 1.

3.8.4. Análisis de residuos

En la Figura 3-17 se presentan los residuos según el modelo ARX determinado por 7 componentes principales, los residuos están calculados como la diferencia entre la predicción del modelo y los datos experimentales, en la figura se puede apreciar una distribución de residuos sin comportamientos anormales, acercándose a una distribución normal y compacta en torno al cero, lo cual es esperado ya que el modelo generado con 7 componentes, obtuvo un error de calibración/validación de 0,23756.

Esta mejora en el ajuste, es explicada por la adición de los valores pasados correspondientes a la variable objetivo en el entrenamiento del modelo, de esta forma cada una de las predicciones se realiza a partir del valor anterior real de la variable, disminuyendo así el error.

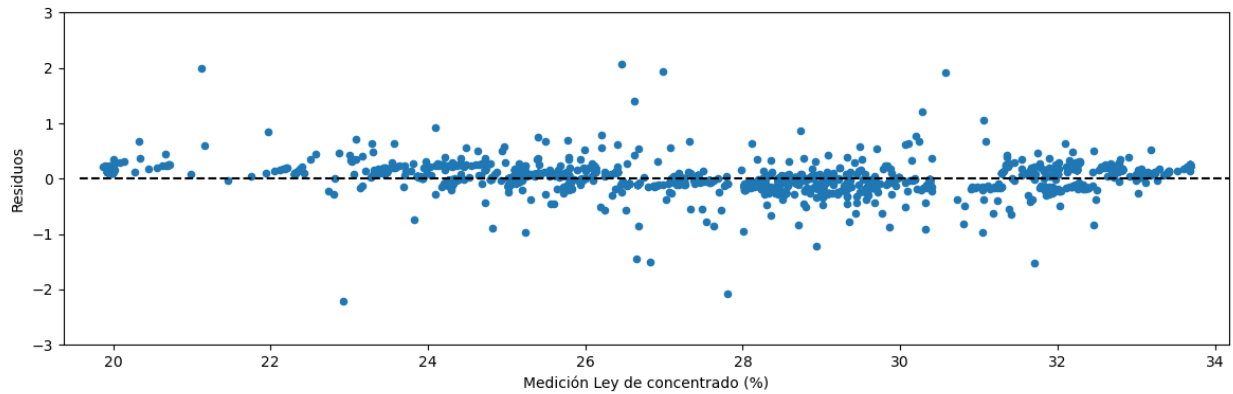


Figura 3-17 Residuos Modelo ARX

3.8.5. Predicción recursiva

Como se comentó, el modelo ARX en su formulación y entrenamiento incluye los valores pasados de la variable objetivo para predecir el estado actual, por lo tanto, se podrá utilizar en la medida que se cuente con algún otro tipo de medición alternativa de la variable como *input* al modelo.

Sin embargo, si en algún proceso se considera la aplicación de un sensor virtual, probablemente es porque no se cuenta con la facilidad o disponibilidad en la medición de una variable en específico, en este sentido, si se quisiera predecir en tiempo real utilizando este modelo, no se contará con el valor real de la variable como *input* para realizar las predicciones.

Para efectos del control de proceso y obtención de métricas, lo deseado es que se pueda contar con una predicción en tiempo real. Como alternativa para utilizar el modelo ARX y generar predicciones sin contar con información anterior exacta de la variable objetivo, lo que se puede realizar es utilizar los valores predichos como *inputs* del modelo, y de esta forma calcular la predicción de forma recursiva, utilizando en cada una de las iteraciones el valor entregado por la predicción anterior de forma sucesiva.

Expuesto lo anterior, en la Figura 3-18 se pueden ver los residuos del modelo ARX calculando la predicción de forma recursiva utilizando los valores predichos anteriormente, se puede apreciar claramente como aumenta la dispersión de los residuos en comparación al gráfico anterior y además son evidentes residuos que están correlacionados. Dichas condicionantes son explicadas por la propagación del error de una predicción a la otra, que se produce al calcular de forma recursiva las predicciones según lo explicado anteriormente.

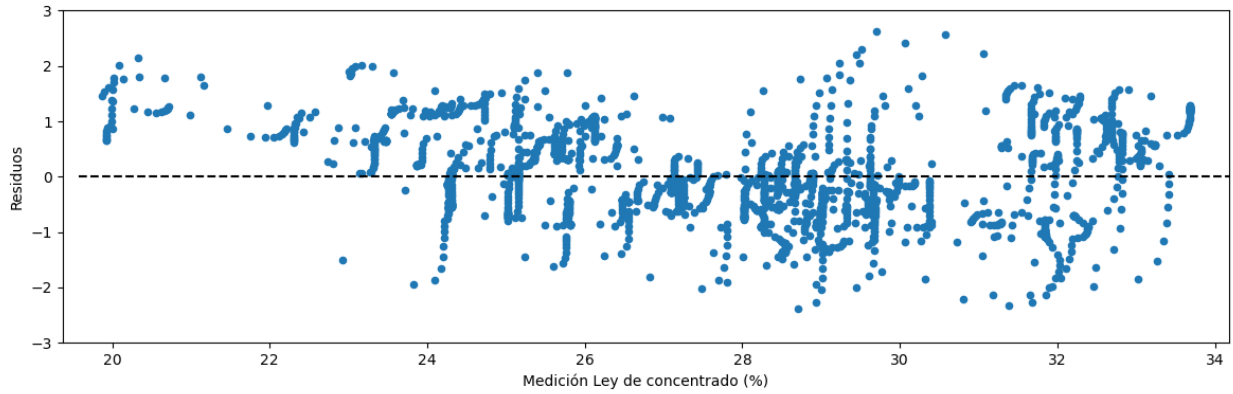


Figura 3-18 Residuos Modelo ARX-Recursivo

3.8.6. Convergencia del modelo

De acuerdo con el punto anterior, las predicciones se ven afectadas fuertemente por la propagación del error desde una predicción a la otra. Una de las formas para mitigar dicha desventaja en el uso del modelo es, por ejemplo, que cada cierto número de predicciones, se reinicie el sensor desde el último valor conocido o más cercano de la ley de concentrado que se tenga disponible.

A continuación, se evalúa el efecto que tiene la proximidad al valor real de $x_c(t - 1)$ que es utilizado como *input* en la predicción, en relación a la convergencia hacia el valor final predicho. En la Figura 3-19 se puede apreciar que en la medida que el valor pasado utilizado como *input* se acerque a el valor real, más rápido alcanzará una predicción adecuada.

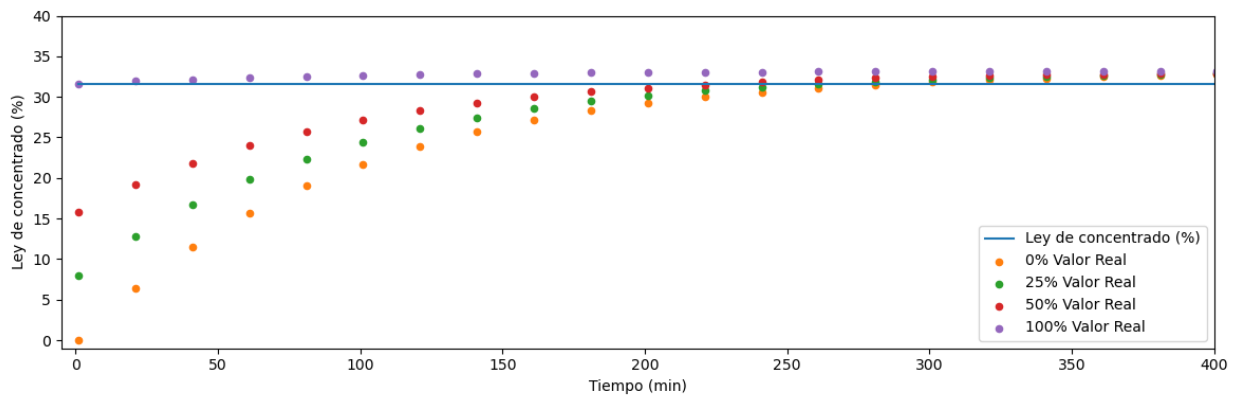


Figura 3-19 Efecto del valor inicial en la convergencia del modelo ARX

Luego, por ejemplo, se puede utilizar el modelo para predecir la ley de concentrado, y de forma paralela realizar una contrastación en el laboratorio cada 10 o 20 predicciones, y reiniciar el sensor

utilizando como *input* el valor obtenido por el laboratorio. Si bien es cierto, que esta opción aún requiere la utilización de mediciones en el laboratorio, la frecuencia de muestro disminuye notablemente, siendo de todas formas una mejora.

Otra alternativa consiste en complementar el sensor con un simulador o modelo estático. Para lo cual, en estados estacionarios conocidos del proceso, se determina la concentración de salida mediante el simulador o modelo estático, y se utiliza este valor como referencia para reiniciar el sensor.

4. DISCUSIÓN DE RESULTADOS

4.1. Resumen de modelos

El resumen de indicadores relevantes correspondientes al ajuste de cada modelo se presenta a continuación en la Tabla 4-1. Se puede ver que el modelo con el menor error corresponde al modelo ARX, lo cual es esperado, ya que, al incluir valores pasados de la variable objetivo, la predicción se hace desde el valor anterior conocido disminuyendo así el error, sin embargo, si se busca la aplicación de un sensor virtual muchas veces no se conoce el valor anterior de la variable, por tanto, su aplicación no sería factible.

Luego, para poder ser utilizado como sensor, la alternativa que existe es calcular de forma recursiva utilizando el valor predicho anteriormente en cada iteración, sin embargo, su desempeño se ve afectado negativamente producto de la propagación del error desde una predicción a la otra, para mitigar dicha desventaja lo que se propone es reiniciar la predicción cada cierto número de iteraciones con el fin de disminuir la propagación del error al mínimo.

En la medida que el valor utilizado como *input* para la predicción sea más cercano al valor real, la predicción será más precisa, de esta forma, por ejemplo, se pueden utilizar mediciones en laboratorio o desde una condición estacionaria conocida del proceso utilizando un simulador estático de este.

Tabla 4-1 Indicadores de los modelos obtenidos

Modelo	Varianza X	Varianza Y	RMSE	Error Máximo
FIR	82,15	95,11	0,6807	2,01
FIR Locales	81,85 86,69 80,02	98,21 98,05 97,72	0,4138	2,18
ARX	88,04	99,39	0,2737	2,07
ARX-Recursivo	88,04	99,39	0,8314	2,62

En comparación, para los modelos del tipo FIR, evidentemente existe una mejora en la utilización de los modelos locales, y en este sentido la clasificación de residuos por medio de árboles de decisión resulta en una herramienta interesante para mejorar el modelo global a partir de modelos locales, ya que en cierta forma permite una clusterización de los datos que son más parecidos,

generando modelos específicos para cada zona de operación, sin embargo, se debe tener en consideración de no generar una cantidad de modelos locales excesiva, ya que esto no permitiría la generalización del modelo y el desempeño podría verse afectado en algunos escenarios operacionales producto de un sobreajuste.

Respecto de la varianza acumulada en los tres modelos se alcanza un valor superior al 80% producto del criterio de selección de las variables latentes, y además para el caso de la varianza acumulada en el bloque Y en todos los modelos alcanza un valor superior al 95%, por lo que resultan en un ajuste adecuado.

En relación a la interpretabilidad de los modelos, se analizan los vectores de regresión, los cuales están presentados en el Anexo 1, y se puede evidenciar que existe una coherencia respecto de la fenomenología y dinámica del proceso, ya que la magnitud de los coeficientes asociados a los valores pasados va disminuyendo en la medida que sean *lags* más antiguas, eso por parte de la dinámica, respecto de la fenomenología los coeficientes con mayor relevancia son los que se relacionan con las variables de: ley cobre de entrada y la velocidad superficial de gas, siendo consistente con la fenomenología esperada para este caso, ya que son dos de las variables más importantes que gobiernan el proceso.

Desde el punto de vista del manejo de los modelos, el modelo ARX tiene una clara ventaja, ya que, al necesitar de una menor cantidad de valores pasados, su dimensionalidad es mucho menor, lo que es un punto importante en una aplicación real, ya que a diferencia de las simulaciones existirán muchas más variables que podrían ser consideradas para el modelo, por tanto, en una aplicación el número de variables en el bloque X podría llegar a convertirse en una limitante.

4.2. Distribución de residuos

A continuación, en la Figura 4-1 se presenta la distribución de errores para los cuatro modelos generados, se reafirma la mejora respecto de la utilización de los modelos locales para el caso del tipo FIR, ya que se aprecia como la distribución se vuelve más centrada en cero, sin algún tipo de sesgo evidente, cumpliendo las suposiciones de una distribución normal.

En el caso del modelo ARX tiene una distribución de residuos prácticamente centrada en cero, muy pronunciada con pocos valores que se desvíen desde el cero, caso contrario con la predicción ARX calculada de forma recursiva, en la cual se la dispersión aumenta de forma considerable, por tanto,

disminuyendo el desempeño del modelo. Respecto de los errores máximos absolutos en los cuatro modelos no sobrepasan el umbral del 2 (% Cobre) a excepción de la predicción calculada con el modelo ARX de forma recursiva.

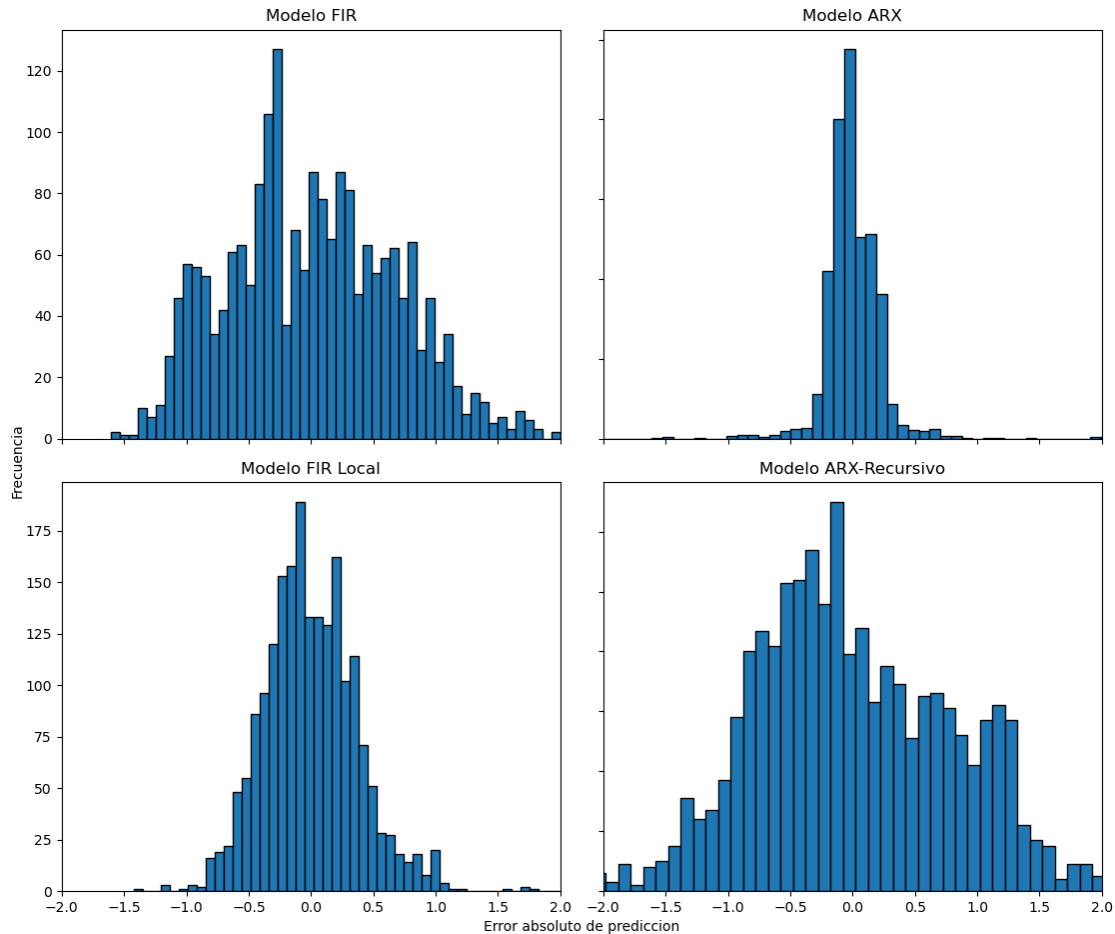


Figura 4-1 Distribución de errores para los modelos generados

Otro criterio para verificar la normalidad de los residuos es la construcción de los gráficos P-P o probabilidad-probabilidad, la idea de estos gráficos es comparar que tan similares son dos distribuciones. Para ello se grafica la probabilidad acumulada de ambas distribuciones en un gráfico de dispersión, luego si ambas distribuciones tienen características similares los datos deberían trazar una línea diagonal o identidad. En este sentido para verificar el criterio de normalidad, se compara una distribución normal teórica versus la distribución de errores obtenidos en los distintos modelos, obteniéndose la Figura 4-2 que se presenta a continuación.

En relación a la normalidad de los residuos se puede apreciar que el modelo FIR global presenta una pequeña desviación la cual mejora con la aplicación de los modelos locales. En el caso del modelo ARX calculado de forma recursiva se aleja la normalidad producto de la propagación de los errores desde una perturbación a la otra. Mientras que los modelos FIR Locales y ARX se aproximan a una distribución normal.

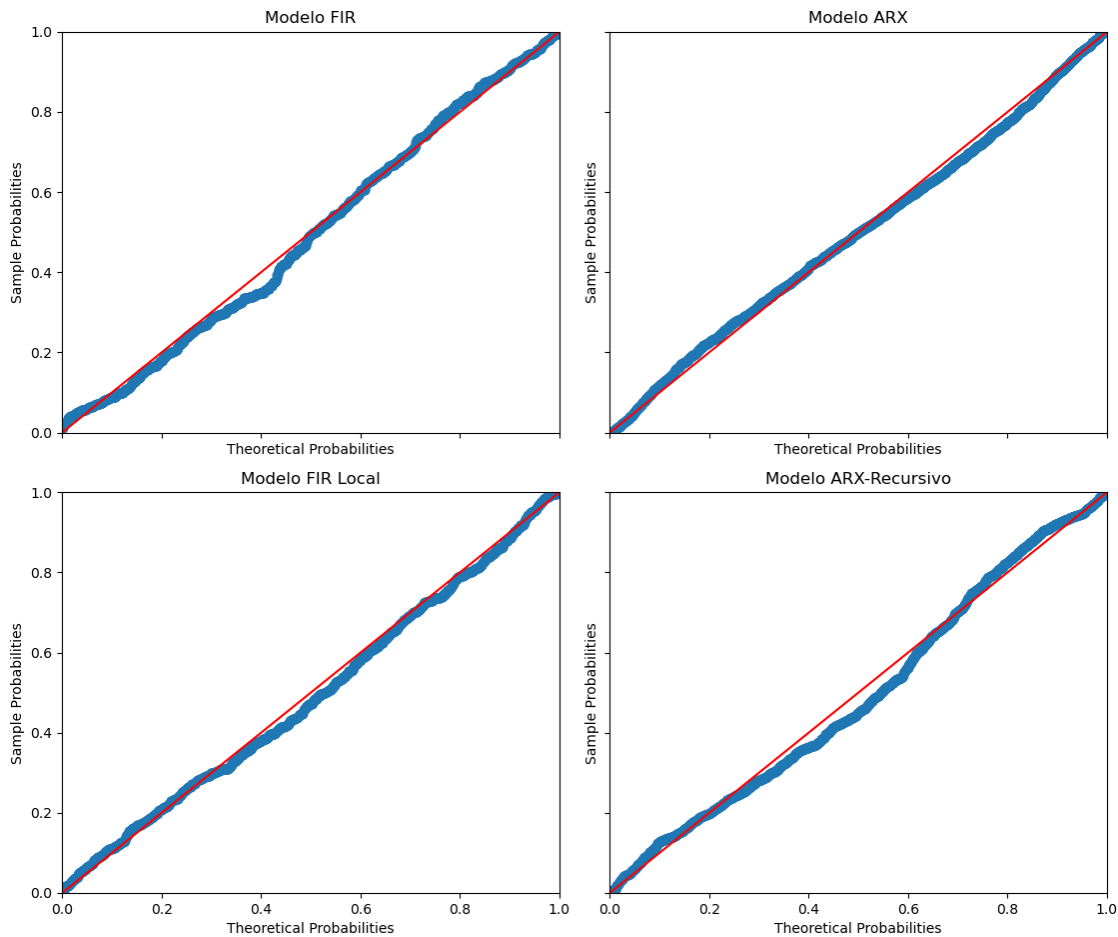


Figura 4-2 Grafios P-P : Verificación normalidad de residuos

4.3. Análisis de propagación del error

En una planta real es común enfrentarse a diferencias en la medición de variables debido a condiciones ambientales cambiantes, dispositivos de medición imperfectos o a desgastes naturales. Estos errores introducen cierta incertidumbre a los datos. Dado que esta información se utiliza para la predicción del sensor, los errores se propagan hacia las predicciones, afectando negativamente la calidad del sensor virtual.

Un análisis para evaluar la robustez de los modelos obtenidos es el análisis de propagación de errores. Este análisis no solo permite comprender cómo se comportará el sensor frente a la incertidumbre asociada a sus entradas, sino que también nos proporciona información sobre las variables que tienen el mayor impacto. Con este conocimiento, se pueden desarrollar estrategias para minimizar las fuentes de error en las técnicas de medición, mejorando así la predicción del sensor. Para evaluar la propagación, se realiza un análisis de sensibilidad, en donde a las variables involucradas en el sensor virtual se le adicionan distintos valores como error de medición, en este caso 1%, 5%, 10%, 20%, 40%. En este sentido, de una variable a la vez se va verificando el efecto que tienen los distintos errores de medición y se revisa su impacto en la ley de concentrado. Este análisis de sensibilidad se realizó utilizando el nivel 3 del diseño experimental como caso base.

En la Figura 4-3 se presenta los resultados del análisis de propagación para el modelo del tipo FIR, en este caso se presentan las variables que más importancia tienen respecto de la propagación del error en la variable objetivo. Se puede ver que en la medida que aumenta el error de medición aumenta también la variación en la variable objetivo. Predomina la importancia en la medición de la ley de entrada, seguidos por la medición de la densidad y la velocidad superficial de gas. Los resultados tienen que ver con la formulación del modelo, en este caso lineal, por tanto, la propagación del error es directamente proporcional al coeficiente asociado a cada variable, siendo estas la más significativas en el vector de regresión.

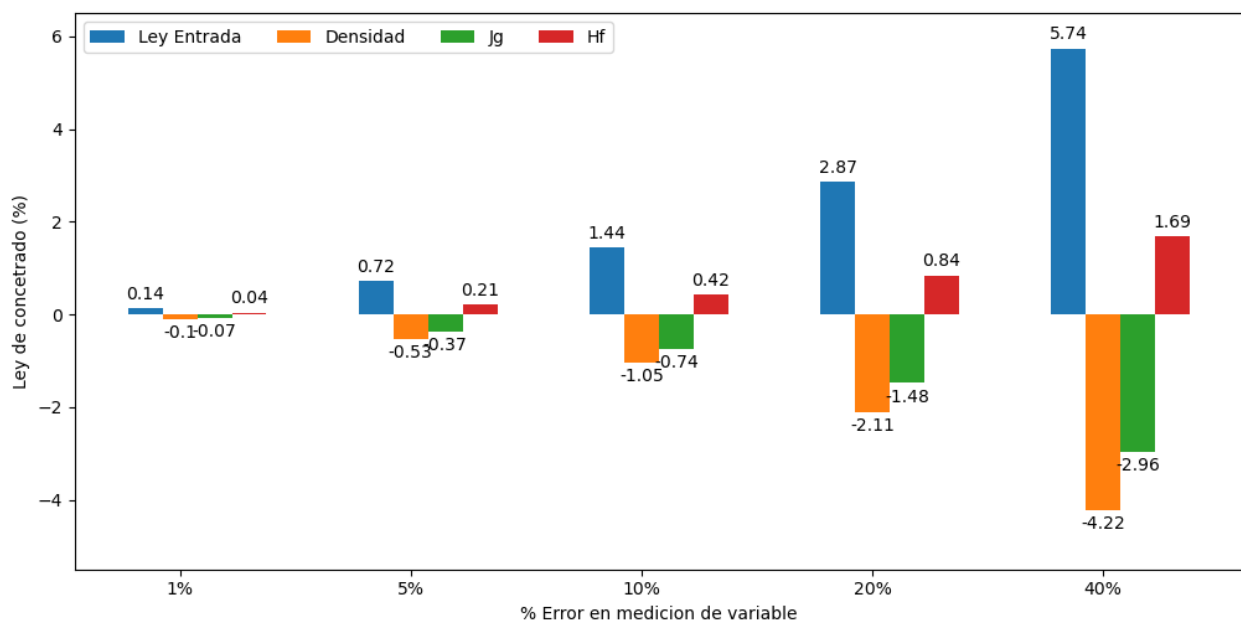


Figura 4-3 Análisis propagación del error de medición en Ley de concentrado modelo FIR

En la Figura 4-4 se presenta los resultados gráficos del análisis de propagación para el modelo del tipo ARX, en este caso se presentan las variables que más importancia tienen respecto de la propagación del error en la variable objetivo. Al igual que en el caso anterior un aumento en el porcentaje de error que se agrega a las mediciones, aumenta la variación en la ley del concentrado.

A diferencia del modelo FIR, aquí se ve como predomina de forma muy marcada la medición de la ley de concentrado del periodo anterior, lo que es esperado producto de la forma autorregresiva del modelo, ya que al incluir una menor cantidad de *lags* en su formulación, la importancia de la variable regresiva es mayor debido a que esta incluye los efectos cumulativos anteriores de proceso aumentando así la sensibilidad del sensor respecto a dicha variable. Esto es consistente con el análisis que se hizo respecto de la convergencia del modelo, que indicaba que mientras más cercano es el input utilizado más rápido se ajustaba la predicción.

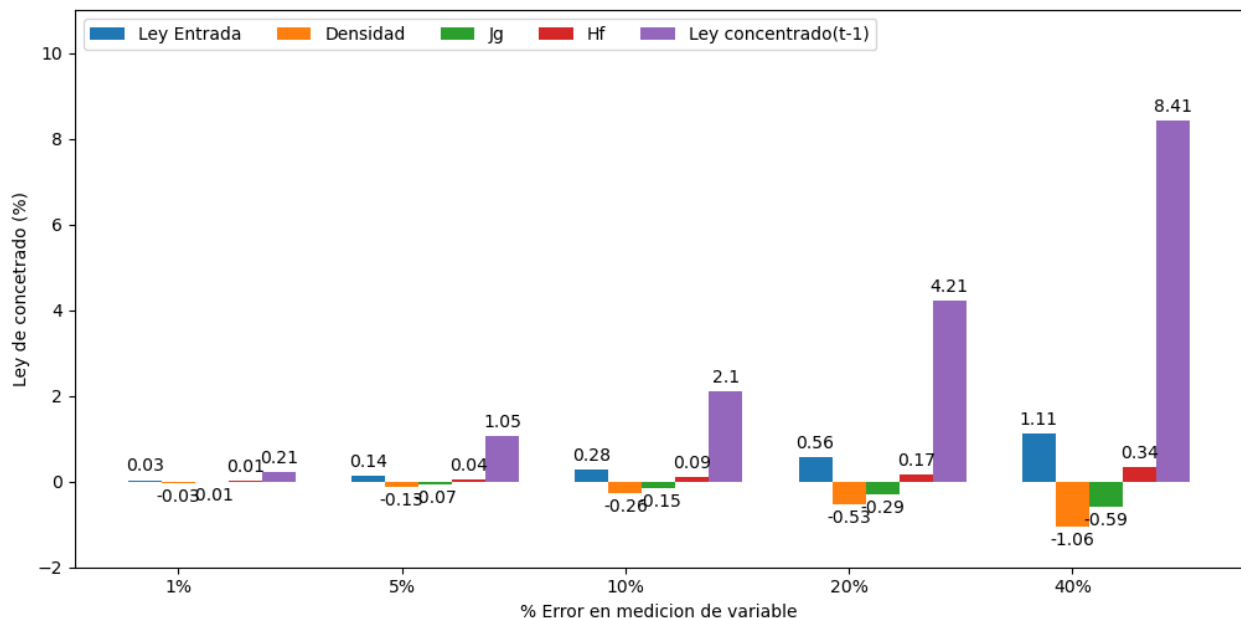


Figura 4-4 Análisis propagación del error de medición en Ley de concentrado modelo ARX

4.4. Consolidado

Para finalizar la implementación la Figura 4-5 muestra el resumen de forma gráfica respecto del comportamiento de cada modelo versus el valor experimental. La figura se realiza presentando el dato experimental y la predicción para esa observación por cada modelo simultáneamente, en este sentido es posible visualizar cada ajuste y poder tener una comparación de los modelos.

En la figura se aprecian desviaciones respecto de los valores experimentales, dichas desviaciones se pueden atribuir principalmente a dos factores. El primero tiene que ver con la formulación utilizada para los modelos, en este caso lineal para un proceso que no tiene dicho comportamiento, por tanto, en la linealización se produce una pérdida de información relevante ya que el modelo solamente es capaz de representar relaciones lineales. Es por eso que en el desarrollo de modelos locales FIR, es posible reducir los residuos, ya que al acotar las áreas de operación se puede asimilar un comportamiento más lineal del proceso en esas áreas específicas.

En segundo lugar, las desviaciones se explican también por la naturaleza dinámica del modelo, que intenta representar la componente dinámica del proceso mediante la inclusión de valores pasados o *lags*. La representación de la trayectoria se basa en la actualización de los valores pasados, lo que resulta en una aproximación que no se traduce en una curva continua, sino más bien en diferencias finitas. Esta aproximación discreta puede generar discrepancias entre los datos experimentales y las predicciones del modelo aumentando el error de predicción.

Luego la adición de valores pasados debe estar respaldada por la propia dinámica del proceso, y con la frecuencia de muestreo disponible, ya que para poder modelar la respuesta transiente se deben incluir la suficiente información anterior del proceso, en este caso para el modelo FIR la inclusión de 4 valores pasados, espaciados en 20 minutos permiten capturar la respuesta del proceso que fue modelada según un modelo de primer orden, dicha respuesta tiene una duración aproximada de 4 constantes de tiempo (aproximadamente 80 minutos).

En cuanto al modelo ARX la inclusión del valor pasado de la concentración de salida en los datos de entrenamiento permite disminuir la cantidad de *lags* desde 4 a 1, ya que gran parte de los efectos del estado anterior del proceso están explicados por el valor anterior de la propia variable a predecir.

Finalmente, dependiendo de la calidad de la información disponible para reiniciar el sensor y cómo se comporta propagación del error, la aplicación de modelo ARX con reinicio o recalibración cada cierto número de iteraciones corresponde al modelo más representativo ya que al incluir 1 *lag* su complejidad y manejo es menor. En caso de no conseguir estabilidad con el ARX-Recursivo o de no contar con información disponible de la variable objetivo, el modelo más adecuado corresponde a FIR-Locales.

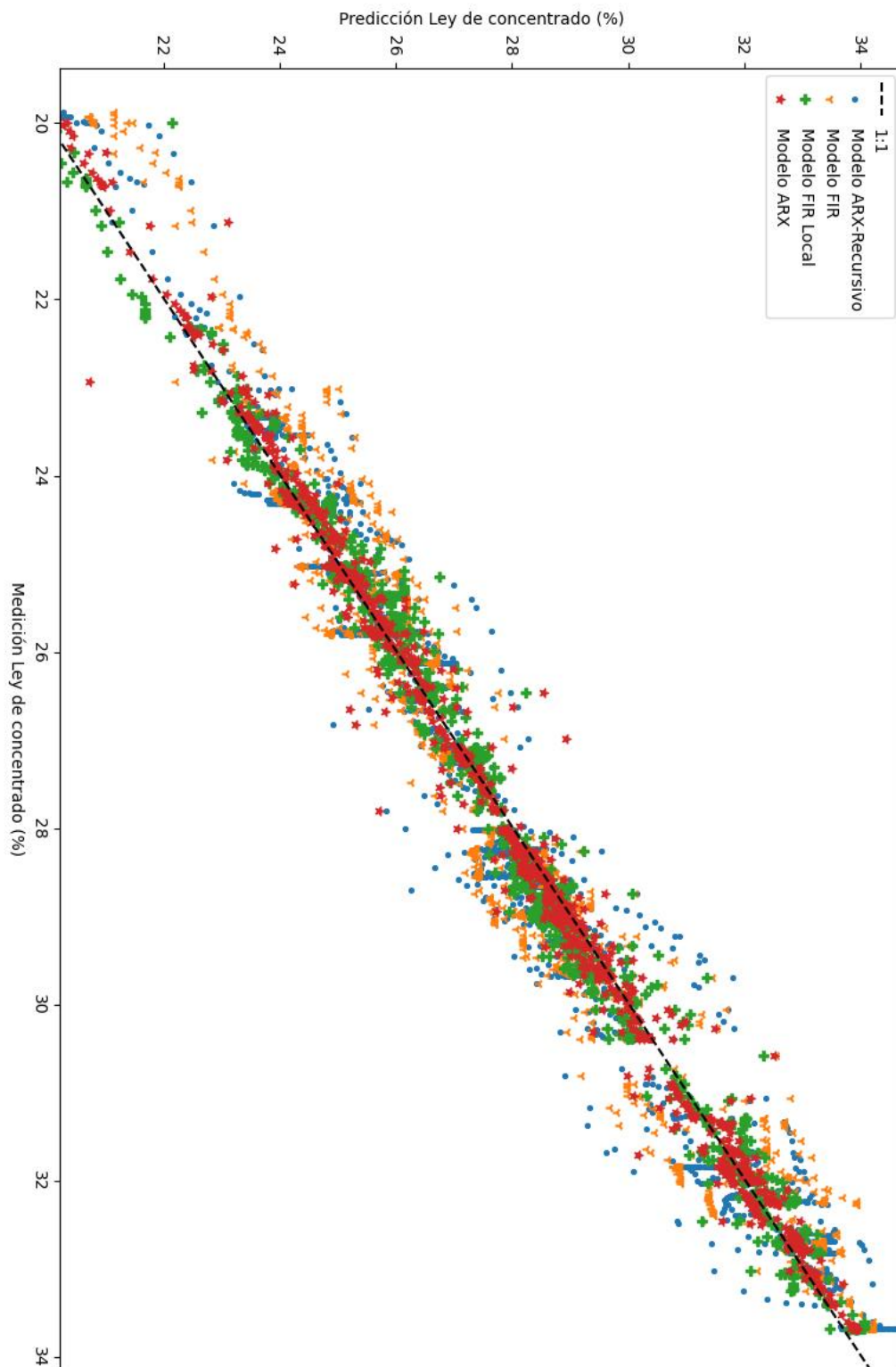


Figura 4-5 Comparación datos experimentales vs predicción de modelos

5. CONCLUSIONES

Mediante la metodología presentada en esta investigación, se desarrolló un sensor virtual dinámico para predecir la ley de concentrado de cobre en una columna de flotación, esto mediante el entrenamiento y evaluación de un modelo predictivo utilizando métodos de proyección, en específico el método PLS Dinámico el cual considera valores pasados o *lags* de las variables durante el ajuste del modelo.

En este sentido se obtiene de forma satisfactoria un set de observaciones para entrenar el modelo PLS, esto mediante la integración de un simulador fenomenológico estático con un modelo de primer orden que permitió simular de forma dinámica una columna de flotación. Junto con un diseño experimental de 8 variables y 5 niveles, lograron generar un set de 2011 observaciones que son consistentes con la fenomenología del proceso, que si bien presentan algunas desviaciones son explicadas por simplificaciones en los modelos en que se basa el simulador.

Se obtienen modelos que difieren en la inclusión de *lags* de la variable objetivo en el bloque X, el primera caso del tipo FIR, que no incluye, se abordó con un primer modelo que abarca una zona de operación global, con un error de 0,6807. La segunda formulación incluye tres modelos locales, en donde las zonas fueron determinados por la clasificación de los residuos del modelo global mediante árboles de decisión, reduciendo significativamente el error a 0,4138. Esta aproximación resulta en una disminución del error en casi la mitad, en este sentido la clasificación vía árboles de decisión es un aporte es una alternativa útil para evaluar en que zonas el modelo debe mejorar.

Además, se entrena un modelo ARX que, en este caso si incorpora valores pasados de la variable objetivo, logrando un error de 0,2737. Sin embargo, la limitación radica en la falta frecuente del valor anterior de la variable objetivo, lo que dificulta su uso como sensor virtual. Se propone una solución recursiva, aunque esto aumenta el error (0,8314) debido a la propagación del error entre predicciones sucesivas. La sugerencia para mitigar este problema implica reiniciar el sensor periódicamente con valores conocidos, como los obtenidos mediante técnicas de laboratorio o simulaciones, reduciendo así la propagación del error.

La varianza capturada en los modelos para el bloque X supera el 80%, y además respecto el bloque Y supera el 95%. En relación a la coherencia de los vectores de regresión, las variables que más influencia tienen, son las relacionadas con las variables de: ley de entrada y velocidad superficial

de gas, siendo consistente con la fenomenología del proceso. Adicionalmente se revisan los residuos de los modelos, no habiendo desviaciones importantes de la normalidad y en cuanto a errores máximos no superan el 2% Cu, por tanto, según lo expuesto anteriormente la metodología PLS, con la inclusión de valores pasados permitió la construcción de los modelos correctamente.

Se realizó un análisis de sensibilidad frente a errores de medición, con niveles de error desde 1% a 40%, encontrándose para el modelo FIR que la variable que más influye en la propagación del error corresponde la ley de entrada variando desde 0,14% a 5,74%. Para el modelo ARX la variable con mayor influencia corresponde a la ley de concentrado en el periodo anterior, variando desde 0,21% a 8,41% este análisis permitió validar la robustez de los modelos, indicando que los modelos pueden verse afectados negativamente por incertidumbres en la medición y se deben establecer estrategias para mitigar los errores de medición en dichas variables.

Los modelos más relevantes para ser utilizados como sensor virtual corresponden a los modelos locales FIR y también al modelo ARX utilizado de forma recursiva. Finalmente, el desarrollo de esta investigación valida la aplicación de la metodología utilizando datos experimentales generados por un simulador, y por tanto los resultados obtenidos en esta investigación indican que existe una oportunidad para probar dicha metodología utilizando datos correspondientes un entorno de prueba real y que en caso de validarse con datos reales podría suponer de una mejora en la información disponible en planta y una toma de decisiones más acertada a los cambios de una planta.

6. RECOMENDACIONES

Esta sección tiene como objetivo entregar algunas recomendaciones que tienen que ver con la aplicación de la metodología en un sitio de prueba con datos reales. Desde el punto de vista de la metodología lo más crítico tiene que ver con la calidad de información utilizada para el entrenamiento del modelo PLS.

En este sentido, las industrias miden y almacenan una gran cantidad de datos del proceso para sus sistemas de optimización y control, sin embargo, a diferencia de las simulaciones, esta adquisición de datos no está exenta de problemas que deben ser considerados a la hora de la aplicación de un sensor virtual, como lo son la frecuencia de muestreo, pérdida de información y datos anormales.

Respecto de la frecuencia de muestreo, es común que las variables más fáciles de medir tengan una mayor frecuencia de muestreo en comparación a las variables que son difíciles de medir, por tanto, se hace necesario sincronizar la frecuencia de muestreo de las variables.

En este sentido las opciones son dos, la primera es realizar un *down-sampling* de las variables con mayor frecuencia, excluyendo las observaciones que no tengan un valor medido en la variable objetivo. Y la segunda opción es considerar la realización de una interpolación o predicción intermedia según algún tipo de modelo, sin embargo, la dificultad es estimar un modelo que represente correctamente los datos sin aumentar la incertidumbre del sensor.

Dicho lo anterior, la aproximación más simple es el *down-sampling*, pero se debe considerar que la frecuencia con que se mide la variable objetivo determinará la ventana de predicción del modelo, por tanto, la respuesta del sensor y evidentemente tendrá un impacto en el desempeño del este, luego, se recomienda que, durante la adquisición de los datos de entrenamiento, en lo posible aumentar de forma temporal la frecuencia de muestreo para mejorar la calidad de este.

Las observaciones anormales son un punto muy importante a considerar también, es necesario que observaciones que se desvíen en gran proporción de las mediciones típicas sean abordadas, ya que la presencia de ellas en el entrenamiento del modelo deteriorará la calidad de la predicción de este.

Las mediciones anormales usualmente son causadas, por ejemplo, por el malfuncionamiento de sensores, errores de comunicación o degradación de sensores, en este sentido es vital que las rutinas

de calibración y mantención de los sensores funcionen correctamente en la durante adquisición de datos para el entrenamiento.

Aun habiendo métodos y procedimientos para evitar la aparición de mediciones anormales se debe establecer un criterio para detectarlas y por consiguiente tratarlas. En la literatura existen algunos métodos basados en técnicas de estadística, en donde una de las más más utilizadas y simples corresponde a la regla de las 3 desviaciones estándar, sin embargo, este método presenta limitantes que no son recomendables para la aplicación, ya que se trata de un análisis univariable que no considera relaciones entre las variables, además la presencia de observaciones anormales tiende a incrementar los límites, causando una disminución en la cantidad de datos anormales detectados y lo más importante es que no es aplicable en presencia de datos dinámicos.

En este sentido, un criterio más robusto para determinar las observaciones anormales corresponde a los estadígrafos calculados en la metodología de componentes principales, correspondientes a *Q-Residuals* y *Hottelling's T^2* , como fue explicado anteriormente se deben analizar los límites de cada uno de los estadígrafos, con lo cual se pueden determinar dos tipos de anomalías, el primer caso corresponde a una falla operacional, este tipo de falla se reconoce cuando el indicador T^2 de la observación está sobre dicho límite, el segundo tipo de anomalía tiene que ver con la falla en la medición de un sensor, este tipo de anomalía se presenta cuando la observación supera el límite establecido para el estadígrafo Q.

Una vez determinadas las mediciones anormales, se debe decidir que realizar con ellas, en este sentido existen algunas opciones, la primera tiene que ver con la eliminación de dicha muestra desde el set de entrenamiento, sin embargo, si los datos anormales acumulan una gran cantidad, puede haber una pérdida de información importante. Otra opción, es reemplazar la medición anormal con algún *input*, como por ejemplo podría ser con el promedio o la mediana de los datos existentes, o también la aplicación de un simulador de la planta.

Por otro lado, en relación con las variables a considerar en el modelo, aunque la metodología PLS posee la capacidad de identificar las variables más relevantes para la predicción, es recomendable realizar un primer filtro de dichas variables. Debido a que, en entornos industriales reales, la cantidad de variables medidas puede alcanzar decenas o incluso cientos, muchas de las cuales pueden carecer de relevancia para la predicción de la ley de concentrado.

A pesar de la gran cantidad de variables medidas, todavía existen variables que se miden de forma muy esporádica o que directamente no se miden por motivos similares a esta aplicación, ya sea la inexistencia de un instrumento robusto o por métodos de medición complejos.

La omisión de estas variables no medidas que podrían tener influencia en la ley de concentrado, pueden resultar en un aumento significativo en el error de predicción. La razón es que la falta de información sobre estas variables durante el entrenamiento del modelo impide que el modelo predictivo identifique las relaciones e interacciones respecto de la variable que se desea predecir.

Asimismo, es el caso de variables que, aunque se midan, ofrecen una descripción incompleta de la magnitud. Un ejemplo es la distribución de tamaño de partícula, que a menudo se resume en un solo valor. Este enfoque simplificado no puede representar completamente la complejidad de la distribución y de esta forma afecta la precisión del modelo.

Efectos similares ocurren con otras características de la alimentación, como la liberación, composición mineralógica, y la presencia o cantidad de reactivos utilizados, variables que influyen significativamente en el comportamiento del proceso. Luego mientras mejor estén medidas y caracterizadas las variables, mayores son las posibilidades de implementar un sensor virtual de forma exitosa.

7. REFERENCIAS

Acosta, S., 2007. *Monitoreo en línea de la operación de una columna de flotación a escala piloto usando métodos de proyección*, Valparaíso: Tesis Magíster Ingeniería Química, Universidad Técnica Federico Santa María.

Arrendondo, C., 2016. *Análisis estructural y dinámico de espuma para la evaluación de mejoras tecnológicas del lavado de espuma-mineralizada en columnas de flotación*, Santiago: Tesis Ingeniería Civil de Minas, Universidad de Chile.

Bergh, L. & Acosta, S., 2009. On-Line Fault Detection on a Pilot Flotation Column Using Linear PCA Models. *International Symposium on Process Systems Engineering*, Issue 10, pp. 1437-1442.

Bergh, L. & Yianatos, J., 2003. Flotation column automation: state of the art. *Control Engineering Practice*, Issue 11, pp. 67-72.

Breiman, L., 1984. *Classification and Regression Trees*. 1st Edition ed. New York: Routledge.

Cayulef, E., 2017. *Determinación de costos para análisis de elementos de interés en muestras de minerales.*, Valparaíso: Tesis Técnico Universitario en Minería y Metalurgia, Universidad Técnica Federico Santa María.

COCHILCO, 2017. *Sulfuros primarios: Desafíos y oportunidades*, Santiago: Comisión Chilena del Cobre.

COCHILCO, 2021. *Anuario de estadísticas del cobre y otros minerales*, Santiago: Comisión Chilena del Cobre.

De Jong, S., 1993. SIMPLS : An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, Volumen 18, pp. 251-263.

Dong, Y. & Qin, S., 2015. Dynamic-Inner Partial Least Squares for Dynamic Data Modeling. *IFAC-PapersOnLine*, 48(8), pp. 117-122.

Dudzic, M., Vaculik, V. & Miletic, I., 2000. On-line applications of multivariate statistics at Dofasco. *IFAC Proceedings Volumes*, 33(22), pp. 425-430.

Eigenvector, 2023. *Advanced Chemometrics Software for use with MATLAB®*. [En línea]
Available at: <https://eigenvector.com/software/pls-toolbox/>
[Último acceso: Julio 2023].

Esposito, F., Malerba, D., Semeraro, G. & Kay, J., 1997. A comparative analysis of methods for pruning decision trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5), pp. 476-491.

Garcia-Muñoz, S., Kourti, T. & MacGregor, J., 2004. Model Predictive Monitoring for Batch Processes. *Ind. Eng. Chem. Res.*, 43(18), pp. 5929-5941.

Hodouin, D., Jämsä-Jounela, S., Carvalho, M. & Bergh, L., 2000. State of the art and challenges in mineral processing control. *IFAC Proceedings Volumes*, 33(22), pp. 49-60.

Jolliffe, I., 2002. Principal component analysis. En: New York: Springer, pp. 1-3.

Kaspar, M. & Harmon, W., 1992. Chemometric methods for process monitoring and high-performance controller design. *American Institute of Chemical Engineers*, 38(10), pp. 1593-1608.

Kaspar, M. & Harmon, W., 1993. Dynamic PLS modelling for process control. *Chemical Engineering Science*, 48(20), pp. 3447-3461.

Kotsiantis, S., 2013. Decision trees: a recent overview. *Artificial Intelligence Review*, Volumen 39, pp. 261-283.

Ku, W., Storer, R. & Georgakis, C., 1995. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 30(1), pp. 179-196.

Li, G., Qin, S. & Zhou, D., 2014. A New Method of Dynamic Latent-Variable Modeling for Process Monitoring. *IEEE Transactions on Industrial Electronics*, 61(11), pp. 6438-6445.

Lin, B., Recke, B. & J, K. K., 2007. A systematic approach for soft sensor development. *Computers & Chemical Engineering*, 31(5), pp. 419-425.

MacGregor, J., 1994. Statistical Process Control of Multivariate Processes. *IFAC Proceedings Volumes*, 27(2), pp. 427-437.

MacGregor, J. & Kourti, T., 1995. Statistical process control of multivariate processes. *Control Eng. Practice*, 3(3), pp. 403-414.

MacGregor, J. y otros, 2007. Multivariate methods for the analysis of data-bases, process monitoring and control in the material processing industries. *IFAC Proceedings Volumes*, 40(11), pp. 193-198.

Metso Outotec, 2023. *Analizador Courier® 5X/6X SL*. [En línea]

Available at: <https://www.metso.com/es/portafolio/analizador-courier-5x-6x-sl/>

[Último acceso: Mayo 2023].

Mingers, J., 1989. An Empirical Comparison of Pruning Methods for Decision Tree Induction. *Machine Learning*, Volumen 4, pp. 227-243.

Moolman, D., Eksteen, J., Aldrich, C. & Van Deventer, J., 1996. The significance of flotation froth appearance for machine vision control. *International Journal of Mineral Processing*, pp. 135-158.

Narayanan, H., Sokolov, M., Butté, A. & Morbidelli, M., 2019. Decision Tree-PLS (DT-PLS) algorithm for the development of process: Specific local predictions models. *Biotechnology Progress*, 35(4), pp. 13-24.

Qin, S. & McAvoy, T., 1996. Nonlinear FIR modeling via a neural net PLS approach. *Computers & Chemical Engineering*, 20(2), pp. 147-159.

Scikit-learn, 2023. *Decision Tree*. [En línea]

Available at: <https://scikit-learn.org/stable/modules/tree.html>

[Último acceso: Julio 2023].

Shean, B. & Cilliers, J., 2011. A review of froth flotation control. *International Journal of Mineral Processing*, Issue 100, pp. 57-71.

Somvanshi, M., Chavan, P., Tambade, S. & Shinde, S. V., 2016. A review of machine learning techniques using decision tree and support vector machine. *International Conference on Computing Communication Control and automation*, pp. 1-7.

Suknovic, M., Delibasic, B. & Jovanoci, M., 2012. Reusable components in decision tree induction algorithms. *Computation Statics*, Issue 27, pp. 127-148.

- Wang, D., Liu, J. & Srinivasan, R., 2010. Data-Driven Soft Sensor Approach for Quality. *IEEE Transactions on Industrial Informatics*, 6(1), pp. 11-17.
- Wills, B. & Finch, J., 2016. *Will's Mineral Processing Technology*. Octava ed. Oxford: Butterworth-Heinemann.
- Wold, S., Ruhe, A., Wold, H. & Dunn, I. W. J., 1984. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM Journal on Scientific and Statistical Computing*, 5(3), pp. 735-743.
- Yianatos, J., Carrasco, C., Bergh, L. & Vinnett, L. & T. C., 2012. Modelling and simulation of rougher flotation circuits. *International Journal of Mineral Processing*, Volumen 112–113, pp. 63-70.
- Yianatos, J. & Diaz, F., 2011. Hydrodynamic Characterization of Industrial Flotation Machines Using Radioisotopes.. En: *Radioisotopes - Applications in Physical Sciences*.. s.l.:s.n., pp. 391-416.
- Yianatos, J. & Vinnett, L., 2015. *Flotación de minerales: Fundamentos, tecnología y aplicación*. Valparaíso: USM.

8. ANEXOS

Anexo 1 : Vectores de regresión

Tabla 8-1 Vector de regresión Modelo FIR

Variable	Coeficiente asociado a cada valor pasado $X(t - h)$				
	4	3	2	1	0
Q_f	-0,0002	-0,0023	-0,0045	-0,0071	-0,0099
x_f	0,2551	0,2773	0,3189	0,3656	0,3785
D_p	0,0003	-0,0002	-0,0007	-0,0012	-0,0019
ρ_{pulp}	-1,2402	-1,4724	-1,8787	-2,2269	-1,9640
$R_{c/c}$	-0,0019	-0,0010	-0,0005	-0,0003	-0,0008
J_g	-0,9708	-0,9788	-0,9954	-1,0058	-0,9774
H_f	0,0114	0,0112	0,0112	0,0113	0,0112
J_w	-0,1170	0,2921	0,6577	0,9291	0,9504
Intercepto	30,4899				

Tabla 8-2 Vector de regresión Modelo Local 1

Variable	Coeficiente asociado a cada valor pasado $X(t - h)$				
	4	3	2	1	0
Q_f	-0,054	-0,0283	-0,0077	0,0096	0,0247
x_f	0,3716	0,3022	0,3401	0,52	0,0
D_p	-0,0026	-0,0026	-0,0032	-0,0046	-0,0066
ρ_{pulp}	-2,1658	-2,1549	-2,176	-2,0722	-1,6291
$R_{c/c}$	0,0013	0,003	0,0041	0,0045	0,0043
J_g	-1,7135	-1,7664	-1,8431	-1,866	-1,7362
H_f	0,0108	0,0105	0,0106	0,0108	0,0108
J_w	0,771	1,7631	2,711	3,5936	4,3497
Intercepto	37,4674				

Tabla 8-3 Vector de regresión Modelo Local 2

Variable	Coeficiente asociado a cada valor pasado $X(t - h)$				
	4	3	2	1	0
Q_f	0,0058	0,0079	0,0101	0,0127	0,0158
x_f	0,085	0,1579	0,3869	0,9569	0
D_p	-0,0066	-0,0063	-0,0062	-0,0062	-0,0062
ρ_{pulp}	-1,6672	-1,8864	-2,1641	-2,4657	-2,6643
$R_{c/c}$	-0,0064	-0,0054	-0,0049	-0,0048	-0,0051
J_g	-0,726	-0,7556	-0,7916	-0,8277	-0,8444
H_f	0,0087	0,0094	0,0107	0,0122	0,0125
J_w	-0,5566	-0,1102	0,3774	0,9103	1,4289
Intercepto	32,4766				

Tabla 8-4 Vector de regresión Modelo Local 3

Variable	Coeficiente asociado a cada valor pasado $X(t - h)$				
	4	3	2	1	0
Q_f	-0,0159	-0,0153	-0,015	-0,0151	-0,0154
x_f	0,2312	0,2455	0,2655	0,2847	0,2849
D_p	-0,0025	-0,0026	-0,0027	-0,0028	-0,003
ρ_{pulp}	-2,3633	-2,3863	-2,5208	-2,6885	-2,6966
$R_{c/c}$	0,0003	0,0005	0,0005	0,0005	0,0002
J_g	-0,8165	-0,8218	-0,832	-0,8425	-0,8461
H_f	0,0106	0,0105	0,0105	0,0105	0,0104
J_w	-0,4807	-0,4016	-0,3859	-0,4521	-0,6537
Intercepto	37,6987				

Tabla 8-5 Vector de regresión Modelo ARX

Variable	Coeficiente asociado a cada valor pasado $X(t - h)$	
	1	0
Q_f	-0,0042	-0,0125
x_f	0,1518	0,1579
D_p	0,0018	0,0003
ρ_{pulp}	-1,2686	-0,9350
$R_{c/c}$	0,0006	0,0006
J_g	-0,5568	-0,4234
H_f	0,0059	0,0055
J_w	-0,1792	0,5094
x_c	0,8088	-
Intercepto	6,3486	

Anexo 2 : Construcción modelos locales

Tabla 8-6 Resumen estadístico de las componentes principales en modelo Local 1

CP	Varianza en X	Varianza acumulada de X	Varianza en Y	Varianza acumulada de Y	RMSECV: Ley concentrado
1	35,96	35,96	73,21	73,21	1,3082
2	8,97	44,93	23,71	96,92	0,4424
3	23,02	67,95	0,50	97,42	0,40208
4	2,85	70,81	0,62	98,04	0,38655
5	4,03	74,84	0,13	98,18	0,35446
6	7,01	81,85	0,03	98,21	0,35461
7	5,59	87,44	0,02	98,22	0,35609
8	1,98	89,42	0,05	98,27	0,35062
9	6,60	96,02	0,01	98,28	0,35371
10	0,50	96,52	0,02	98,30	0,36013

Tabla 8-7 Resumen estadístico de las componentes principales en modelo Local 2

CP	Varianza en X	Varianza acumulada de X	Varianza en Y	Varianza acumulada de Y	RMSECV: Ley concentrado
1	24,09	24,09	92,94	92,94	0,57376
2	23,24	47,33	2,84	95,78	0,52017
3	21,09	68,42	1,09	96,87	0,47726
4	12,16	80,58	0,58	97,45	0,44638
5	6,10	86,69	0,59	98,05	0,41167
6	5,01	91,69	0,43	98,47	0,4113
7	2,73	94,43	0,23	98,71	0,377
8	1,69	96,11	0,18	98,88	0,38894
9	0,40	96,51	0,11	99,00	0,38563
10	1,22	97,73	0,02	99,01	0,37379

Tabla 8-8 Resumen estadístico de las componentes principales en modelo Local 3

CP	Varianza en X	Varianza acumulada de X	Varianza en Y	Varianza acumulada de Y	RMSECV: Ley concentrado
1	20,96	20,96	80,54	80,54	1,1563
2	19,34	40,30	13,41	93,96	0,64473
3	19,78	60,09	2,79	96,74	0,47372
4	12,67	72,75	0,90	97,64	0,40336
5	7,26	80,02	0,08	97,72	0,39669
6	6,15	86,17	0,06	97,78	0,39162
7	6,40	92,57	0,07	97,86	0,3856
8	2,19	94,76	0,16	98,02	0,37131
9	3,84	98,60	0,01	98,03	0,37024
10	0,24	98,85	0,02	98,05	0,37009