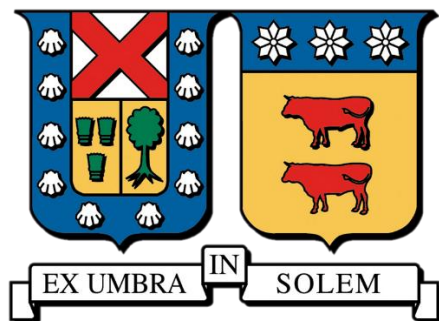


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INGENIERÍA QUÍMICA Y AMBIENTAL
VALPARAÍSO – CHILE



**DESARROLLO DE UN SENSOR VIRTUAL DINÁMICO DE
CONCENTRACIONES EN UNA CELDA DE FLOTACIÓN ROUGHER
PILOTO, UTILIZANDO LA METODOLOGÍA PLS**

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

LEONARDO BLADIMIR AVILEZ BOZO

PROFESOR GUÍA: DR. LUIS BERGH O.

PROFESOR CO-REFERENTE: DR. CLAUDIO ACUÑA P.

Valparaíso 2026



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: Desarrollo de un sensor virtual dinámico de concentraciones en una celda de flotación rougher piloto, utilizando la metodología PLS

Nombre del candidato(a): Leonardo Bladimir Avilez Bozo

Carrera / Grado: Ingeniería civil química

Campus: Casa central Valparaíso Departamento: Ingeniería química y ambiental

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Luis Bergh Olivares, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (marcar una opción):

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

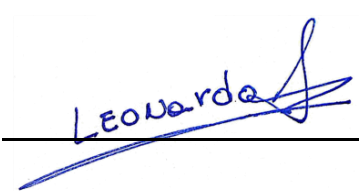
Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 24-07-2026

Firma: 

Estudiante o Candidato(a):

Fecha: 23/07/2026

Firma: 

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Resumen

El presente trabajo tiene como objetivo el desarrollo de un sensor virtual dinámico para estimar la ley de concentrado en una celda de flotación, considerando la dificultad de medir esta variable en tiempo real mediante sensores físicos. Para ello, se implementaron modelos matemáticos capaces de inferir el valor de esta variable a partir de otras más accesibles, utilizando la metodología de regresión por mínimos cuadrados parciales (PLS) aplicada en estructuras dinámicas tipo FIR y ARX. El conjunto de variables consideradas incluye seis entradas principales: flujo de alimentación (Q_F), velocidad superficial del aire (J_g), altura de espuma (H_f), porcentaje de sólidos (S_f), porcentaje de gruesos (G_f) y ley de alimentación (X_f). Estas variables fueron obtenidas desde un entorno de simulación calibrado para representar el comportamiento estacionario de la celda de flotación.

Dado que el simulador opera en estado estacionario, se incorporó un modelo dinámico adicional asumiendo un sistema de primer orden con retardo por transporte despreciable. Se generaron 729 observaciones a partir de una matriz experimental completa, las cuales fueron posteriormente depuradas según criterios de recuperación de cobre y la selección de una ruta de perturbación representativa, conservándose finalmente 323 observaciones válidas. Este conjunto se utilizó para entrenar los modelos mediante el *software* SOLO, empleando auto escalado y validación cruzada para la calibración. Se construyeron y evaluaron modelos FIR (global y locales) y ARX (ideal y no ideal). El modelo FIR global presentó un error (RMSE) de 1,37, mientras que los modelos locales lograron reducirlo a 1,22, demostrando mayor precisión en zonas específicas del dominio operativo. El modelo ARX ideal, que incorpora retroalimentación de la variable de salida, obtuvo el menor error general (0,94), aunque su aplicabilidad real es limitada debido a la necesidad de conocer valores reales pasados. En contraposición, el modelo ARX no ideal, alimentado con predicciones pasadas, mostró el mayor error (1,56), aunque con la ventaja de ser completamente autónomo. Se identificó que este último requiere ser reiniciado cada 15,8 horas para evitar acumulación de errores. Los modelos FIR locales y ARX ideal mostraron mejor comportamiento, con residuos más simétricos y centrados. Se efectuó también un análisis de sensibilidad, revelando que la variable J_g tiene mayor impacto en el modelo FIR, mientras que en el modelo ARX, los valores pasados de la propia salida son los más influyentes. Esto sugiere la necesidad de contar con sensores precisos para estas variables si se busca implementar el sensor virtual en planta. En conclusión, los modelos FIR locales y ARX ideal ofrecen el mejor balance entre precisión y robustez para estimar la ley de concentrado. Sin embargo, la elección del modelo depende del contexto operativo: mientras que los modelos FIR locales son preferibles con datos de alta calidad, el modelo ARX no ideal representa una opción más viable en entornos donde se prioriza la autonomía del sensor y se toleran mayores niveles de error.

Abstract

This work aims to develop a dynamic virtual sensor to estimate the concentrate grade in a flotation cell, given the difficulty of measuring this variable in real time using physical sensors. To this end, mathematical models were implemented capable of inferring the value of this variable from other, more accessible ones, using the partial least squares (PLS) regression methodology applied to dynamic FIR- and ARX-type structures. The set of variables considered includes six main inputs: feed flow rate (Q_F), superficial air velocity (J_g), froth height (H_f), percent solids (S_f), percent coarse particles (G_f), and feed grade (X_f). These variables were obtained from a simulation environment calibrated to represent the steady-state behavior of the flotation cell.

Since the simulator operates in steady state, an additional dynamic model was incorporated, assuming a first-order system with negligible transport delay. A total of 729 observations were generated from a full experimental matrix, which were subsequently refined according to copper recovery criteria and the selection of a representative perturbation route, ultimately retaining 323 valid observations. This dataset was used to train the models using SOLO software, applying autoscaling and cross-validation for calibration. FIR models (global and local) and ARX models (ideal and non-ideal) were built and evaluated. The global FIR model showed an error (RMSE) of 1.37, while the local models reduced it to 1.22, demonstrating greater accuracy in specific regions of the operating domain. The ideal ARX model, which incorporates feedback from the output variable, achieved the lowest overall error (0.94), although its real-world applicability is limited by the need to know past real values. In contrast, the non-ideal ARX model, fed with past predictions, showed the highest error (1.56), though with the advantage of being fully autonomous. This latter model was found to require resetting every 15.8 hours to prevent error accumulation. The local FIR models and the ideal ARX model showed the best performance, with more symmetric and centered residuals. A sensitivity analysis was also carried out, revealing that variable J_g has the greatest impact on the FIR model, while in the ARX model, past values of the output itself are the most influential. This suggests the need for accurate sensors for these variables if the virtual sensor is to be implemented on an industrial plant. In conclusion, the local FIR models and the ideal ARX model offer the best balance between accuracy and robustness for estimating concentrate grade. However, model choice depends on the operational context: local FIR models are preferable when high-quality data is available, while the non-ideal ARX model represents a more viable option in settings where sensor autonomy is prioritized and higher error levels are tolerable.

Índice de contenidos

Resumen.....	i
Abstract.....	ii
Índice de tablas.....	vi
Índice de figuras	vii
Nomenclatura	ix
1. Introducción.....	1
1.1. Contexto.....	1
1.2. Situación minera en Chile.....	3
1.3. Sistemas de control en flotación	4
2. Objetivos	6
2.1. Objetivo principal	6
2.2. Objetivos específicos	6
3. Estado del arte	7
3.1. Proceso de flotación.....	7
3.1.1. Fenomenología	7
3.1.2. Equipos.....	8
3.1.3. Variables del proceso	11
3.1.4. Resultados metalúrgicos	15
3.2. Métodos de proyección.....	16
3.2.1. Análisis de componentes principales (PCA)	18
3.2.2. Partial least square (PLS)	23
3.2.3. Partial least square dinámico (PLS)	26
4. Equipo experimental y metodología	33
4.1. Preparación del laboratorio y sistema de control.....	33
4.1.1. Descripción general del equipo a utilizar	34
4.1.2. Programación software de control.....	35
4.1.3. Configuración HMI y calibración de variables	35
4.2. Código de simulación del proceso de flotación	37
4.2.1. Variables de proceso en el código de simulación	38
4.3. Generación de respuesta transiente en los datos	41
4.3.1. Sistema de primer orden.....	42
4.4. Diseño de experimentos.....	48
4.4.1. Selección de variables y sus niveles.....	49
4.4.2. Matriz de experimentos	50
4.4.3. Set de datos.....	52

4.4.4.	Modelo PLS dinámico	55
4.5.	Modelo FIR.....	59
4.5.1.	Bloques de datos X e Y	59
4.5.2.	Número de componentes del modelo	61
4.5.3.	Estadígrafos Q y T^2	62
4.5.4.	Residuos modelo FIR	66
4.6.	Modelos FIR locales.....	68
4.6.1.	Generación de modelos locales	71
4.6.2.	Estadígrafos Q y T^2	73
4.6.3.	Residuos modelos FIR locales.....	77
4.7.	Modelo ARX.....	79
4.7.1.	Bloques de datos X e Y	79
4.7.2.	Número de componentes del modelo	80
4.7.3.	Estadígrafos Q y T^2	81
4.7.4.	Residuos modelo ARX	84
4.7.5.	Modelo ARX no ideal.....	86
4.7.6.	Determinación tiempo de reinicio del sensor	89
5.	Análisis y discusión de resultados	93
5.1.	Resumen modelos generados.....	93
5.2.	Distribución y normalidad de los residuos.....	95
5.3.	Análisis de propagación del error	99
5.4.	Comparación del comportamiento predictivo de los modelos.....	103
6.	Conclusiones	106
7.	Recomendaciones	107
8.	Referencias	109
9.	Anexos	113
9.1.	Anexo 1: Coeficientes de regresión de los modelos	113
9.2.	Anexo 2: Modelos locales.....	115
9.3.	Anexo 3: Relevancia de variables en modelos estimados según VIP scores	116
9.4.	Anexo 4: Gráficos de calibración y rangos operativos de variables clave.....	116
9.5.	Anexo digital 1.....	118

Índice de tablas

Tabla 1. Características de la celda de flotación.....	34
Tabla 2. Variables de alimentación.....	39
Tabla 3. Variables de operación.....	40
Tabla 4. Variables geométricas.....	40
Tabla 5. Matriz de variables y sus niveles.....	49
Tabla 6. Variables fijas de entrada.....	50
Tabla 7. Resultados preliminares de la investigación.....	54
Tabla 8. Sets de datos a utilizar en el método PLS.....	56
Tabla 9. Resumen estadístico modelo FIR.....	62
Tabla 10. Clasificación de las observaciones según el error porcentual.	69
Tabla 11. Modelos locales.....	71
Tabla 12. Resumen estadístico modelos FIR locales.	72
Tabla 13. Resumen estadístico modelo ARX.....	81
Tabla 14. Resumen estadístico de los modelos obtenidos.....	94
Tabla 15. Observación utilizada en el análisis.	100
Tabla 16. Matriz de coeficientes de regresión del modelo FIR global.....	113
Tabla 17. Matriz de coeficientes de regresión del modelo FIR local 1.....	113
Tabla 18. Matriz de coeficientes de regresión del modelo FIR local 2.....	114
Tabla 19. Matriz de coeficientes de regresión del modelo ARX.....	114
Tabla 20. Resumen estadístico del modelo FIR local 1.....	115
Tabla 21. Resumen estadístico del modelo FIR local 2.....	115
Tabla 22. Relevancia de variables según VIP scores en el modelo FIR.....	116
Tabla 23. Relevancia de variables según VIP scores en el modelo ARX.....	116

Índice de figuras

Figura 1. Representación de la zona de colección y zona de espuma	8
Figura 2. Esquema de celda mecánica de aire forzado (Yianatos Bernardino, 2005)	10
Figura 3. Esquema de celda mecánica auto-aspirante (Yianatos Bernardino, 2005)	10
Figura 4. Esquema de la Columna de Flotación (Wills & Napier-Munn, 2006).....	11
Figura 5. Sistema de celdas de flotación rougher.	33
Figura 6. Esquema de celdas de flotación rougher piloto (Durán, 2014).....	34
Figura 7. Código de simulación flotación en celda rougher.....	37
Figura 8. Código de simulación en estado estacionario	41
Figura 9. Esquema de sistema con entrada y salida	42
Figura 10. Zonas de transporte de espuma en celda de flotación piloto.....	46
Figura 11. Esquema de simulador estático y dinámico integrados.....	48
Figura 12. Set de datos generado a partir de simulador (dinámico más estacionario).	53
Figura 13. Interfaz de análisis PLS programa SOLO.....	57
Figura 14. Simulador vs modelo FIR, Hotelling's T^2 reducido.	64
Figura 15. Simulador vs modelo FIR, Q residual reducido.....	66
Figura 16. Residuos modelo FIR.....	68
Figura 17. Modelos 1, 2 y 3 vs simulador, Hotelling's T^2 reducido.	75
Figura 18. Modelos 1, 2 y 3 vs simulador, Q residual reducido.....	77
Figura 19. Residuos modelos locales FIR.	79
Figura 20. Simulador vs modelo ARX, Hotelling's T^2 reducido.....	83
Figura 21. Simulador vs modelo ARX, Q residual reducido.....	84
Figura 22. Residuos modelo ARX.....	86
Figura 23. Residuos modelo ARX no ideal.....	87
Figura 24. Residuos modelo ARX no ideal temporales.	89
Figura 25. Comparación de RMSE (Ventana deslizante).....	92
Figura 26. Distribución de errores (residuos) para cada modelo evaluado.	97
Figura 27. P-P plots de residuos: evaluación de la normalidad en los modelos.....	99
Figura 28. Propagación del error de medición en la ley de concentrado modelo FIR.	101
Figura 29. Propagación del error de medición en la ley de concentrado modelo ARX.	103
Figura 30. Comparación del comportamiento predictivo de los modelos.....	105
Figura 31. Calibración bomba de alimentación.....	116
Figura 32. Calibración de nivel sensor ultrasónico	117
Figura 33. Calibración sensor flujo de aire.	117

Figura 34. Relación flujo de aire con porcentaje de apertura de válvula.	118
Figura 35. Calibración de relación entre sensor y válvula de aire.....	118

Nomenclatura

X_c : Ley de mineral valioso en el concentrado

R : Recuperación de cobre

C : Flujo másico de concentrado

F : Flujo másico de alimentación

X_f : Ley de mineral valioso en la alimentación

X : Matriz de datos con m observaciones y n variables (variables independientes)

S : matriz de covarianza o correlación

$\vec{v}$: Vector propio (*eigenvector*)

λ : Valor propio (*eigenvalue*)

P : Matriz de cargas (*loadings*) de X

V : Matriz de cargas (*loadings*) de Y

T : Matriz de puntuaciones (*scores*) de X

U : Matriz de puntuaciones (*scores*) de Y

T^2 : Estadígrafo Hotelling's T^2

Q : Estadígrafo Q *residuals*

Y : Matriz de datos con m observaciones y p variables (variables dependientes)

k : Número de variables de salida de los modelos (componentes principales)

E : Matriz de errores

H : Matriz de errores

B : Matriz de los coeficientes de regresión

w : Vector de *loadings* de la matriz X

c : Vector de *loadings* de la matriz Y
 W : Matriz *weights* del modelo PLS
 $u(t)$: Vector de variables independientes
 b_t : Coeficiente de regresión para variables independientes
 a_t : Coeficiente de regresión para variables dependientes pasadas
 l : Número de retrasos (*lags*)
 X_{ARX} : Matriz de predictores modelo ARX
 X_{FIR} : Matriz de predictores modelo FIR
 Y_{ARX} : Matriz de salida modelo ARX
 Y_{FIR} : Matriz de salida modelo FIR
 K y g : Ganancia estática del sistema
 $U(s)$: Variable de entrada al sistema
 $Y(s)$: Variable de salida del sistema
 τ : Constante de tiempo del sistema
 t_d : Retardo por transporte
 $G(s)$: Función de transferencia
 A : Amplitud
 A_c : Área de la sección transversal de la celda
 H_c : Altura de la zona de colección
 ε_g : Concentración de aire (Hold up)
 Q_F : Flujo volumétrico de la pulpa en la alimentación
 t_{fv} : Tiempo de ascenso vertical de la espuma

t_{fh} : Tiempo de desplazamiento horizontal de la espuma

v_{fv} : Velocidad de ascenso vertical de la espuma

ε_f : Concentración de aire en la fase de espuma

Q_F : Flujo de alimentación de pulpa

S_f : Porcentaje de solidos

X_f : Ley de cobre en alimentación

G_f : Porcentaje de gruesos en alimentación

J_g : Velocidad superficial del aire

H_f : Altura de espuma

1. Introducción

1.1. Contexto

La corteza y el manto terrestre están formados por material mineral, es decir, por rocas de distinta naturaleza y estado físico, según las condiciones fisicoquímicas presentes en la Tierra. Las rocas se definen de manera menos precisa: son masas sólidas de materia mineral que forman parte de nuestro planeta, y algunas están compuestas casi por completo de un solo mineral (Wills & Napier-Munn, 2006).

La reactividad de los metales con su entorno, particularmente con oxígeno, azufre y dióxido de carbono, determina su forma en la corteza terrestre y como depósitos en el lecho de los océanos. Metales como la plata, el cobre y el mercurio se encuentran tanto en forma nativa como en sulfuros, carbonatos y cloratos. Los metales más reactivos siempre se presentan en compuestos, como los óxidos o sulfuros de hierro, y los óxidos y silicatos de aluminio y berilio. Los compuestos naturales de estos elementos se conocen como minerales.

Los minerales se definen como sustancias inorgánicas con composición química y estructura atómica definidas. Sin embargo, el término "mineral" se usa a menudo de manera más amplia para incluir cualquier recurso económico extraído de la tierra. Así, el carbón, la tiza, las arcillas y el granito no se consideran minerales en sentido estricto, aunque popularmente se les clasifica como tales. Estos materiales son en realidad rocas, que no son homogéneas en composición química ni física y consisten en una variedad de minerales que forman gran parte de la corteza terrestre.

Los yacimientos minerales, tanto metalíferos como no metálicos, son concentraciones de sustancias útiles que generalmente están distribuidas de manera escasa en la corteza terrestre. Los yacimientos metalíferos son concentraciones extremas de metales que originalmente estaban dispersos. Los metales de interés suelen estar químicamente unidos a otros elementos formando las menas minerales, que se mezclan con minerales no metálicos o roca, denominados ganga. La combinación de menas minerales y ganga constituye la mena, que generalmente se presenta en forma de rocas (Napier-Munn, Morrell, Morrison, & Kojovic, 1996).

Las menas minerales son los minerales que se pueden utilizar para extraer uno o más metales. En su mayoría, son minerales metálicos como la calcopirita, que se extrae para obtener cobre. Algunos son minerales no metálicos, como la malaquita, la bauxita o la cerusita, que contienen cobre, aluminio y

plomo, respectivamente. Un mismo metal puede extraerse de diferentes minerales; por ejemplo, el cobre se encuentra en la calcosina, bornita, calcopirita, cuprita, cobre nativo y malaquita. Asimismo, una sola mena mineral puede contener más de un metal; por ejemplo, la estannita contiene estaño y cobre (Herrmann & Zappettini, 2014).

Por lo tanto, el beneficio de minerales metálicos se entiende como el conjunto de operaciones mecánicas, químicas o biológicas destinadas a obtener ciertos elementos metálicos contenidos en algunas rocas de la corteza terrestre, donde se movilizan, transforman y consumen grandes cantidades de materia. Estas operaciones o procesos usualmente se clasifican como:

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

Los procesos de liberación y conminución de minerales metálicos son cruciales en el beneficio de minerales, enfocándose en separar los minerales valiosos de la ganga para facilitar su concentración y recuperación. La reducción del tamaño de las rocas que contienen minerales valiosos permite liberar estos minerales de la matriz rocosa no valiosa, asegurando una separación física eficiente. Este paso es fundamental para preparar los minerales en etapas posteriores, maximizando así la eficiencia del proceso y garantizando la viabilidad económica y técnica de la explotación minera (Wills & Napier-Munn, 2006).

Por otro lado, los procesos de separación o concentración de minerales metálicos, como la flotación, la lixiviación y la separación magnética, entre otros, se emplean para concentrar los minerales valiosos y separarlos de la ganga utilizando propiedades físicas y químicas específicas. La flotación aprovecha diferencias en la capacidad de los minerales para adherirse a burbujas, la lixiviación disuelve selectivamente los minerales deseados mediante soluciones químicas, y la separación magnética se basa en las propiedades magnéticas de los minerales.

El objetivo principal de estos procesos es mejorar la concentración de los minerales valiosos, facilitando su refinamiento y uso posterior. La elección del método de separación adecuado maximiza la eficiencia de la recuperación y minimiza los costos operativos, asegurando la sostenibilidad económica de los proyectos mineros.

En Chile, los procesos de separación o concentración de minerales de cobre están influidos por el tipo de mena presente. Las menas oxidadas, como la malaquita y la cuprita, se encuentran cerca de la

superficie y se tratan principalmente con lixiviación ácida debido a su alta solubilidad en soluciones ácidas. En contraste, las menas sulfuradas, como la calcopirita, bornita y calcosina, que se encuentran en depósitos más profundos, se tratan principalmente mediante flotación, aprovechando sus propiedades sulfídicas para la concentración eficiente de cobre.

1.2. Situación minera en Chile

En lo que respecta a la minería en Chile, este sector se posiciona como uno de los pilares fundamentales que sustentan la economía del país. Desde el año 2003 hasta el 2022, la participación de la minería en el PIB ha fluctuado entre un 8% y un 14%. En 2022, esta participación alcanzó un valor significativo del 14,2%, consolidándose como un motor económico crucial (Ministerio de Minería, 2023). De este porcentaje, el 10,9% corresponde al aporte de la minería del cobre, mientras que el resto proviene de otras actividades mineras. Esto subraya la importancia del cobre en la economía chilena y su contribución al liderazgo del sector minero. En 2022, la minería se destacó como la actividad económica más importante a nivel nacional, superando a otros sectores como los servicios personales, el comercio y los servicios financieros. Esta preeminencia destaca la centralidad de la minería en el desarrollo económico de Chile (Ministerio de Minería, 2023).

Las exportaciones de cobre en Chile (2003-2022) se han mantenido prácticamente constantes, siempre superando las 5.000 kilotoneladas métricas de cobre fino. Estas exportaciones consisten principalmente en productos como concentrado de cobre, cátodos y cobre blíster. Sin embargo, se está observando un cambio significativo en la composición del mineral extraído: los sulfuros de cobre son cada vez más abundantes, mientras que los óxidos son cada vez más escasos. Según proyecciones de COCHILCO, el procesamiento de óxidos pasará de representar el 30,8% en 2015 a solo el 12% en 2027 (Ministerio de Minería, 2023).

Este cambio en la disponibilidad de minerales impacta directamente los métodos de procesamiento utilizados. Aunque los sulfuros de cobre pueden ser procesados mediante lixiviación, este método presenta ciertas dificultades. Los sulfuros tienen una cinética de reacción más lenta y tienden a formar una capa pasiva que impide que el agente lixivante reaccione eficientemente con el mineral. Como resultado, los sulfuros de cobre deben ser tratados de manera tradicional, mediante flotación y pirometalurgia (Villela, Kutscher, Castillo, & Cantallopts, 2017). Este tratamiento es significativamente diferente del procesamiento de los óxidos de cobre, que se realiza principalmente a través de hidrometalurgia. Así, la industria debe adaptarse a estas diferencias en los procesos para mantener la eficiencia en la producción y exportación de cobre.

1.3. Sistemas de control en flotación

El uso de sistemas de control en procesos de flotación es fundamental en la industria minera, especialmente en el procesamiento de minerales. El objetivo del control de procesos de concentración de minerales, como es el caso de la flotación, es reducir la variabilidad en objetivos metalúrgicos como la recuperación y la ley de concentrado, variabilidad ocasionada naturalmente por el mineral alimentado.

El reducir esta variabilidad genera mejoras en la productividad, garantizando mayores recuperaciones para la misma ley, o mejores leyes de concentrado para la misma recuperación. Adicionalmente, logra indicadores operacionales más consistentes, permite reducir los costos y mejores tomas de decisiones (Yianatos & Vinnett, 2015).

Con el objetivo de lograr controlar y supervisar el proceso de flotación es que se utilizan modelos. Los modelos de micro o macro escala se utilizan para lograr describir la cinética de flotación, por un lado, los modelos a microescala tratan de determinar las relaciones causa-efecto entre las variables, lo cual es muy complejo de realizar debido a la interacción entre parámetros físicos y químicos. Por otro lado, los modelos a macro escala se encargan de analizar el proceso global de flotación a través de expresiones matemáticas (Miettinen, Ralston, & Fornasiero, 2010).

Los modelos a macro escala se dividen en modelos fenomenológicos y empíricos. Los métodos empíricos relacionan las variables de entrada y salida con la recuperación de mineral, mediante expresiones matemáticas y modelos estadísticos, que determinan la dependencia o independencia de los parámetros analizados. Los modelos fenomenológicos, en cambio, consideran parámetros cinéticos o los mecanismos de transformación que ocurren al interior del proceso. Adicionalmente, los modelos pueden incluir las siguientes consideraciones (Gharai & Venugopal, 2016):

- Modelos en estado estacionario o modelos dinámicos (transientes)
- Modelos determinísticos o estocásticos
- Modelos causales (modelos de entrada y salida) o modelos no causales (asociadas a relaciones entre las variables de proceso, como pueden ser las restricciones)
- Modelos lineales o no lineales
- Modelos basados en ecuaciones matemáticas o en reglas difusas

Es por este motivo que, en el ámbito de la minería, la calidad de los datos o la información no es solo una buena práctica, sino un pilar fundamental que sustenta la eficiencia operativa, la rentabilidad y la seguridad. Las condiciones adversas de la flotación minera, como la naturaleza corrosiva y abrasiva

de los minerales y los reactivos, imponen desafíos significativos a la instrumentación de medición, lo que puede comprometer la fiabilidad y precisión de los datos generados (Hodouin, Jämsä-Jounela, Carvalho, & Bergh, 2001). Una baja calidad de la información puede llevar a decisiones operativas subóptimas, ineficiencias en el uso de recursos y, en última instancia, a una disminución de la productividad (Napier-Munn, Morrell, Morrison, & Kojovic, 1996). Por lo tanto, la inversión en instrumentación robusta, el mantenimiento adecuado y la implementación de sistemas de gestión de datos confiables son esenciales para asegurar la integridad de los datos que alimentan los modelos de proceso y los sistemas de control avanzado, permitiendo una optimización real del proceso de flotación (Hodouin, Jämsä-Jounela, Carvalho, & Bergh, 2001). La falta de conocimiento del proceso, especialmente su comportamiento transiente y la relación entre las variables del proceso, provocan una variabilidad en el proceso, lo que hace que el control sea de baja calidad y que, de forma parcial, anulen el uso de estrategias de control más complejas (Yianatos Bernardino, 2005).

El monitoreo en línea de la flotación puede incluir:

- Flujo volumétrico y la densidad de la pulpa de alimentación
- Flujo de aire
- Nivel de pulpa
- Dosificación de reactivos
- pH
- Medición de leyes en las corrientes de alimentación, concentrado y colas

Monitoreo que es uno de los más importantes avances en el control de la flotación, ya que permite la automatización del proceso basado en la medición directa de un objetivo metalúrgico o en el uso de sistemas de control supervisor experto (Yianatos Bernardino, 2005). Por otra parte, las variables más comúnmente manipuladas al operar una planta son:

- Nivel de pulpa
- Flujo de aire
- Dosificación de reactivos
- Flujo de agua de lavado

Es gracias a la manipulación de estas variables que, por ejemplo, en columnas de flotación industrial se ha logrado un control efectivo mediante el uso de un sistema de control supervisor, el cual coordina la manipulación de las variables con el fin de obtener una ley de concentrado en un rango delimitado (Yianatos, Bergh, & Aguilera, 2003).

En la industria minera, contar con mediciones precisas, exactas y de fácil disponibilidad es de vital importancia. Estas mediciones son fundamentales para el diseño y funcionamiento de sistemas de control efectivos, así como para obtener resultados metalúrgicos consistentes y óptimos, especialmente en lo que respecta a la ley del concentrado. Una adecuada gestión de estos datos permite optimizar los procesos y garantizar la eficiencia y calidad del producto final.

En el presente trabajo se busca predecir la ley de concentrado de cobre en una celda de flotación piloto tipo *rougher*, mediante la construcción de un sensor virtual dinámico. Este sensor se desarrollará utilizando metodologías de proyección, específicamente mediante la técnica de regresión por mínimos cuadrados parciales (PLS, por sus siglas en inglés).

Para llevar a cabo esto, se cuenta con un código de simulación en estado estacionario que será acoplado a un simulador dinámico. Esta integración permitirá generar las observaciones necesarias para la calibración y evaluación del sensor virtual propuesto, asegurando su validez frente al escenario operacional planteado.

2. Objetivos

2.1. Objetivo principal

Desarrollar un sensor virtual dinámico, que pueda predecir la concentración de cobre en línea, en una celda de flotación *rougher* piloto, utilizando la metodología PLS.

2.2. Objetivos específicos

- ❖ Diseñar experimentos para diferentes condiciones de la alimentación al proceso y de las variables operacionales disponibles en el código de simulación en estado estacionario.
- ❖ Generar datos, usando el código de simulación, que relacionen las variables independientes del proceso con el objetivo o resultado metalúrgico de interés (ley de concentrado de cobre).
- ❖ 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 posibles.

3. Estado del arte

3.1. Proceso de flotación

3.1.1. Fenomenología

La flotación de minerales es un proceso de separación fisicoquímico fundamental en la industria minera. Su principal objetivo es concentrar minerales valiosos, como el cobre, separándolos de la ganga, es decir, del material sin valor económico. Esta separación se basa en las diferencias en las propiedades superficiales de los minerales, particularmente en su grado de hidrofobicidad. El proceso de flotación es complejo y multifacético, ya que implica interacciones tanto a nivel molecular como macroscópico. En términos generales, el mineral de interés se modifica para volverse hidrofóbico, de modo que pueda adherirse a las burbujas de aire introducidas en la pulpa. Por otro lado, el material no deseado permanece hidrofílico, lo que significa que es mojado por el agua y no se adhiere a las burbujas, permitiendo su separación. La selectividad del proceso, es decir, la capacidad para separar eficientemente el mineral valioso de la ganga se logra mediante la adición de colectores. Estos son reactivos químicos que se adsorben selectivamente sobre la superficie del mineral de interés, modificando su humectabilidad. La adsorción del colector puede ser de naturaleza física o química, y resulta esencial para inducir la hidrofobicidad necesaria para la flotación exitosa del mineral (Wills & Finch, 2015).

Además de la hidrofobicidad inducida, la interacción entre las burbujas de aire y las partículas minerales constituye un pilar fundamental en el proceso de flotación. Para que el proceso sea efectivo, es esencial que las partículas hidrofóbicas colisionen con las burbujas y se adhieran de manera estable a su superficie. Esta adhesión implica la delgada película de agua que separa la partícula de la burbuja, un fenómeno controlado por la energía superficial de los componentes involucrados y por la acción de ciertos reactivos, como los espumantes. Los espumantes son reactivos químicos que reducen la tensión superficial del agua, permitiendo la formación de burbujas más estables y la generación de una espuma mineralizada robusta sobre la superficie de la pulpa. Sin una espuma lo suficientemente estable, las partículas adheridas a las burbujas no podrían ser transportadas y recuperadas eficientemente (Fuerstenau, Jameson, & Yoon, 2007). Las partículas que se adhieren a las burbujas de aire y que son capaces de ser colectadas se denominan flotables, pasando a la fase de espuma y, posteriormente, al concentrado. Mientras que, las partículas que no son flotables abandonan el sistema por la corriente de cola o relave.

Por otro lado, la cinética de flotación, que describe la velocidad a la que los minerales se recuperan, está influenciada por múltiples factores operacionales y de diseño. Entre los más relevantes se encuentran el tamaño de las partículas, el flujo de aire, el diseño específico de la celda de flotación, la concentración de los reactivos (espumantes y colectores) y el tiempo de contacto entre partículas y burbujas (Lynch, Johnson, Manlapig, & Thorne, 1981).

A la hora de analizar el proceso de flotación, se consideran dos zonas con características muy distintas entre ellas: en primer lugar, la zona de colección (pulpa) y, en segundo lugar, la zona de limpieza (espuma), lo que se puede visualizar por medio de la Figura 1.

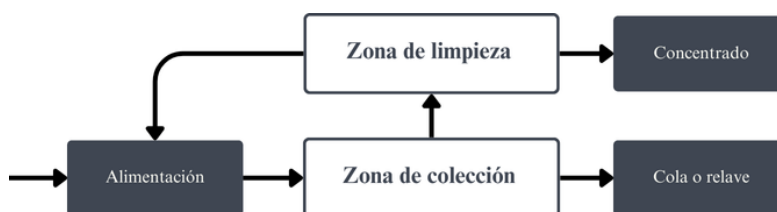


Figura 1. Representación de la zona de colección y zona de espuma

En la zona de colección, ocurre el primer encuentro entre las partículas minerales, que descienden en la pulpa, y las burbujas de aire, que ascienden desde el fondo de la celda. La eficiencia en la recolección y recuperación de las partículas está determinada por la frecuencia y efectividad de las colisiones y posteriores adhesiones que permiten la formación del agregado burbuja-partícula. Por su parte, la zona de limpieza corresponde al trayecto que recorren las burbujas cargadas con mineral a través del lecho de espuma, hasta alcanzar el rebalse, donde finalmente se recupera el concentrado (Yianatos Bernardino, 2005).

3.1.2. Equipos

La técnica de flotación, crucial en la industria minera para la concentración de minerales, ha experimentado una notable evolución en sus equipos a lo largo de las décadas. Hasta mediados de los años 70, el panorama tecnológico estaba dominado por las celdas con agitación mecánica y auto aspirantes (Yianatos Bernardino, 2005). No obstante, a partir de los años 80, se produjo un cambio significativo con la incorporación masiva de celdas de gran tamaño, tanto auto aspirantes como de aire forzado. Paralelamente a esta tendencia, se introdujeron también las celdas neumáticas de flotación, diversificando las opciones disponibles para la industria (Finch, 1995).

A pesar de estos avances y el desarrollo de una vasta gama de equipos a lo largo de aproximadamente cien años, es importante señalar que el diseño fundamental de las celdas de flotación mecánica, ampliamente utilizadas en aplicaciones industriales, no ha experimentado cambios significativos (Yianatos Bernardino, 2005).

El objetivo principal de los equipos de flotación es la separación y concentración de minerales valiosos de la ganga, aprovechando las diferencias inherentes en las propiedades superficiales de las partículas minerales. Este proceso se fundamenta en la generación de burbujas de aire que se dispersan eficientemente en la pulpa mineral, una mezcla homogénea de agua y minerales finamente triturados (Wills & Napier-Munn, 2006).

Las partículas de minerales valiosos, después de ser acondicionadas con reactivos específicos, exhiben propiedades hidrofóbicas, lo que les permite adherirse selectivamente a las burbujas de aire. Una vez cargadas con estas partículas, las burbujas ascienden a la superficie del equipo de flotación, formando una capa de espuma mineralizada. Este fenómeno posibilita la recolección de las partículas flotadas como concentrado, mientras que el material no flotado, que carece de afinidad por las burbujas, se hunde y se descarga como relave (Fuerstenau, Jameson, & Yoon, 2007).

Los equipos de flotación son esenciales en el procesamiento de minerales, diseñados para optimizar tanto la recuperación como la pureza del concentrado. Esto implica maximizar la cantidad de minerales valiosos recuperados y, a su vez, minimizar la presencia de ganga en el concentrado final (Kelly & Spottiswood, 1982). Dentro de la amplia gama de estos equipos, destacan dos categorías principales: las celdas mecánicas y las celdas neumáticas.

- Celdas mecánicas

Las celdas mecánicas de flotación son, sin duda, las más comunes en la industria minera. Su diseño incorpora un impulsor mecánico cuya función es agitar la pulpa (una mezcla de mineral finamente molido y agua) y dispersar eficientemente el aire en burbujas diminutas. Esta agitación mecánica genera una zona de alta turbulencia, lo que facilita la crucial adhesión de las partículas de mineral a las burbujas de aire (Wills & Napier-Munn, 2006).

Es importante destacar que, dentro de las celdas mecánicas, los diseños de gran tamaño que se incorporaron masivamente a partir de los años 80 incluyen tanto las variantes auto aspirantes como las de aire forzado. Las celdas auto aspirantes utilizan la acción del impulsor para succionar aire, mientras que las de aire forzado dependen de un soplador externo para inyectar aire presurizado en la pulpa (Fuerstenau, Jameson, & Yoon, 2007).

La eficacia de estas celdas radica en su alta eficiencia de agitación, que asegura una mezcla homogénea de la pulpa y el aire, incrementando así la probabilidad de contacto entre las partículas de mineral y las burbujas. Además, las celdas mecánicas ofrecen un control preciso sobre el flujo de aire y la velocidad del impulsor, permitiendo una optimización del proceso de flotación adaptada a las características de cada mineral (Fuerstenau, Jameson, & Yoon, 2007). No obstante, es importante señalar que el impulsor mecánico conlleva un consumo energético considerable y que las partes móviles, como el impulsor y el estator, requieren mantenimiento regular (Wills & Napier-Munn, 2006).

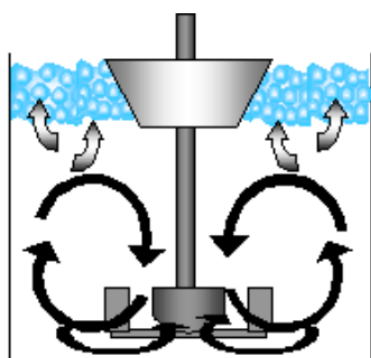


*Figura 2. Esquema de celda mecánica de aire forzado
(Yianatos Bernardino, 2005)*



*Figura 3. Esquema de celda mecánica auto-aspirante
(Yianatos Bernardino, 2005)*

- Celdas Neumáticas

En contraste, las celdas neumáticas de flotación, a menudo referidas como celdas de columna, se distinguen por la ausencia de partes móviles mecánicas para la agitación. En su lugar, estas celdas emplean un flujo de aire presurizado para generar burbujas finas que se mezclan con la pulpa. Estructuralmente, son más altas y delgadas que sus contrapartes mecánicas, y su funcionamiento se basa en la inyección de aire a través de difusores ubicados en la base de la columna (Wills & Napier-Munn, 2006).

Una ventaja clave de las celdas neumáticas es su menor consumo energético, directamente relacionado con la ausencia de componentes mecánicos de agitación. La falta de agitación mecánica puede propiciar un entorno de flotación más tranquilo, lo que a su vez puede resultar en una mayor selectividad y una mejor calidad del concentrado (Finch & Dobby, 1991). Sin embargo, estas celdas pueden enfrentar desafíos al procesar pulpas con altos contenidos de sólidos debido a la menor agitación. Además, el control de la distribución de burbujas y el flujo de pulpa puede ser más complejo en comparación con las celdas mecánicas (Wills & Napier-Munn, 2006).

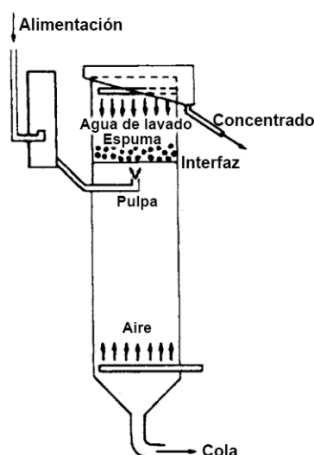


Figura 4. Esquema de la Columna de Flotación (Wills & Napier-Munn, 2006).

La elección entre celdas mecánicas y neumáticas depende de varios factores, como el tipo de mineral, la capacidad de procesamiento requerida y las condiciones específicas del proceso.

3.1.3. Variables del proceso

La optimización y el control de las celdas de flotación, en particular de las celdas mecánicas objeto de esta investigación, dependen de una gestión precisa de diversas variables operativas. Estas resultan fundamentales para alcanzar los objetivos de recuperación de minerales valiosos y asegurar la pureza del concentrado. A continuación, se presentan las variables más relevantes, junto con aquellas utilizadas en el desarrollo de este trabajo.

- Tiempo de residencia

El tiempo de residencia es un factor fundamental que afecta directamente la ley y, de manera más significativa, la recuperación del mineral flotado. Su ajuste se logra principalmente modificando el flujo de la pulpa, la concentración de sólidos en la alimentación y la altura de la zona de colección (Yianatos Bernardino, 2005). El tiempo de residencia de las partículas sólidas en una celda de flotación bien agitada es similar al del líquido, lo que indica una baja segregación de partículas. La medición y modelado de la distribución del tiempo de residencia (RTD) son cruciales para entender la hidrodinámica y optimizar el diseño y operación de las celdas industriales (Diaz & Yianatos, 2010). La manipulación de esta variable es vital para asegurar que las partículas tengan el contacto suficiente con las burbujas para la flotación, sin exceder el tiempo que podría llevar a la carga excesiva de ganga.

- Reactivos

La adición estratégica de reactivos es indispensable para un control efectivo de la selectividad y la estabilidad de la espuma en el proceso de flotación. Los reactivos más comunes utilizados incluyen:

- ❖ Colector: Su función principal es hacer hidrofóbicas las partículas del mineral de interés, permitiendo su adhesión selectiva a las burbujas de aire (Gaudin, 1922)
- ❖ Espumante: Es un reactivo que desempeña un papel crucial en el proceso de flotación, ya que reduce la tensión superficial del agua. Esta reducción permite la formación de burbujas más pequeñas y estables, las cuales son esenciales para el transporte del concentrado en la fase de espuma. Gracias a este reactivo, la espuma se mantiene estabilizada, lo que mejora la eficiencia del proceso de separación de minerales (Fuerstenau, Jameson, & Yoon, 2007).
- ❖ Regulador de pH: Generalmente, la cal se utiliza para ajustar el pH de la pulpa. Este control es vital, ya que el pH influye en la eficacia de los demás reactivos y en el comportamiento de las superficies minerales, optimizando la interacción burbuja-partícula (Wills & Finch, 2015).

La dosificación de estos reactivos se ajusta, ya sea manual o automáticamente, en diferentes puntos del circuito de procesamiento, desde la molienda húmeda hasta las propias celdas de flotación, basándose en los tonelajes y flujos de alimentación (Kelly & Spottiswood, 1982).

- Tamaño de partícula

La eficiencia de la recuperación de minerales por flotación en celdas mecánicas es óptima en un rango intermedio de tamaño de partícula, que usualmente se encuentra entre 50 y 100 μm . La recuperación tiende a disminuir tanto para partículas más finas como para las más gruesas (Wills & Finch, 2015):

- ❖ Partículas finas o ultrafinas (menores a 20 μm): Estas partículas presentan los mayores desafíos en la flotación. Su baja inercia provoca que se muevan con el fluido, dificultando su colisión y adhesión a las burbujas de aire. Su pequeña superficie de contacto en relación con su masa disminuye la probabilidad de captura efectiva.
- ❖ Partículas gruesas: Sufren de un menor grado de liberación, lo que significa que a menudo contienen ganga incrustada. Además, tienen un tiempo de residencia reducido en la celda y una menor eficiencia de colección. Su problema más crítico es la propensión a la ruptura del agregado burbuja-partícula debido a la alta turbulencia presente en la celda (Wills & Napier-Munn, 2006).

- Flujo de aire

El flujo de aire es una de las variables más importantes para controlar el proceso de flotación, particularmente en celdas de columna, debido a su gran influencia en la recuperación del mineral. Dentro de los límites de estabilidad del equipo de flotación, la recuperación del mineral flotado generalmente aumenta al incrementar el flujo de aire hasta alcanzar un valor máximo. Este aumento se debe a la mayor área superficial de las burbujas que se introducen en el equipo de flotación (Fuerstenau, Jameson, & Yoon, 2007). Sin embargo, un incremento excesivo del flujo de aire puede ser perjudicial, causando mayor turbulencia, arrastre de material no deseado (ganga) y la posible pérdida de la interfase aire-pulpa. Para condiciones típicas de operación en celdas y columnas, la velocidad superficial del aire varía entre 1 y 2 [cm/s] (Yianatos Bernardino, 2005).

- Concentración de aire (Hold-up)

El Hold-up de aire representa la fracción volumétrica de aire que está contenida en una sección específica del equipo de flotación. Comúnmente, se determina en la zona de colección de la celda y es una variable que depende del flujo de aire, el tamaño de las burbujas, la densidad de la pulpa, la carga de sólidos en las burbujas y la velocidad de descenso de la pulpa. La medición del hold-up de aire puede realizarse utilizando sensores de presión, siendo un indicador clave del rendimiento de la aireación y la hidrodinámica de la celda (Fuerstenau, Jameson, & Yoon, 2007).

- Tamaño de burbujas

El tamaño medio de las burbujas y su distribución juegan un papel crucial en la flotación, impactando directamente la eficiencia de la colección y el transporte de partículas. Las burbujas pequeñas, por ejemplo, pueden mejorar la cinética de colección y el transporte de sólidos por volumen de aire, lo que se traduce en una mayor eficiencia del proceso (Fuerstenau, Jameson, & Yoon, 2007). Sin embargo, burbujas excesivamente pequeñas pueden ascender a una velocidad más lenta que la velocidad de descenso de la pulpa, lo que podría resultar en la pérdida de partículas hidrófobas hacia el flujo de relave.

Es fundamental identificar un tamaño medio óptimo de burbujas que maximice la eficiencia de la flotación. Este tamaño ideal varía en función del tamaño promedio de las partículas presentes en la pulpa y puede ajustarse mediante las variables operativas del sistema de aireación y la adición de espumantes. El diámetro y la distribución del tamaño de las burbujas son directamente influenciados por el tipo de generador de burbujas, su modo de operación y mantenimiento, el flujo de aire introducido y la cantidad de reactivo espumante añadido (Fuerstenau, Jameson, & Yoon, 2007).

- Profundidad de espuma

La profundidad de espuma en las celdas de flotación es una variable operativa crítica para asegurar la eficiencia y el rendimiento general del proceso. Se refiere a la altura de la capa de espuma donde se acumula el concentrado. Su manejo adecuado es crucial para la selectividad y la recuperación de minerales valiosos. Una profundidad de espuma bien controlada permite un tiempo de residencia suficiente para que la ganga atrapada pueda drenar y regresar a la pulpa, mejorando así la calidad del concentrado (ley). Por otro lado, una profundidad insuficiente puede llevar a una pérdida de recuperación, mientras que una excesiva puede afectar negativamente la ley al arrastrar más agua y finos de ganga. Además, controlar esta variable es clave para mantener una ley de concentrado óptima, lo que mejora la calidad y el valor económico del producto final (Wills & Napier-Munn, 2006). Niveles consistentes de espuma también contribuyen a la estabilidad operativa, previniendo fluctuaciones indeseadas en la recuperación y la calidad del concentrado, y asegurando un flujo constante de concentrado hacia las etapas posteriores del proceso.

- Flujo de Alimentación de Pulpa

El flujo de alimentación de pulpa (Q_F) corresponde a la tasa de material (mineral y agua) que ingresa al circuito de flotación por unidad de tiempo. Es un factor determinante en la capacidad de producción de la planta y su control resulta crítico, ya que afecta directamente el tiempo de residencia de las partículas en las celdas. Un flujo elevado puede reducir dicho tiempo por debajo del necesario para que se produzcan las interacciones entre partículas y burbujas, lo que compromete la recuperación (R) y la ley de concentrado (X_c), disminuyendo así la eficiencia del proceso (Wills & Finch, 2015).

- Porcentaje de Sólidos

El porcentaje de sólidos (S_f) representa la fracción de material sólido respecto al peso total de la pulpa. Esta variable incide en la reología del sistema, modificando su viscosidad y densidad. Un porcentaje elevado aumenta la viscosidad, dificultando la dispersión de burbujas y la movilidad de partículas, lo que reduce la probabilidad de colisiones eficientes. En cambio, un valor demasiado bajo diluye el mineral y disminuye la capacidad de las celdas. Por ello, la selección del porcentaje de sólidos debe balancear consumo de agua, energía y cinética de flotación, de manera de asegurar un desempeño adecuado del proceso (Wills & Finch, 2015).

- Ley de Alimentación

La ley de alimentación (X_f) describe la concentración de mineral valioso presente en el material que ingresa al circuito o, en este caso, a la celda de flotación. Es una variable clave desde el punto de vista metalúrgico, ya que constituye la base para calcular la recuperación y la producción final de metal. Sin embargo, sus fluctuaciones representan un desafío para el control del proceso, puesto que, al tratarse de una condición no manipulable, deben ajustarse otras variables, como la dosificación de reactivos, para mantener el rendimiento esperado (Wills & Napier-Munn, 2006).

3.1.4. Resultados metalúrgicos

La efectividad de la operación de una celda de flotación se evalúa principalmente a través de sus resultados metalúrgicos, los cuales cuantifican la eficiencia con la que el mineral valioso es separado de la ganga. Estos resultados son el reflejo directo de cómo las variables operativas controladas influyen en el proceso. Las métricas fundamentales incluyen la recuperación y la ley del concentrado del mineral de interés.

- Ley del Concentrado

La ley del concentrado (X_c) es una medida de la pureza del mineral valioso en el producto final obtenido. Se expresa como el porcentaje en masa del metal o mineral de interés en el concentrado (Wills & Napier-Munn, 2006). Una alta ley implica que hay menos material no deseado (ganga) mezclado con el mineral valioso, lo que reduce los costos de transporte y los procesos metalúrgicos posteriores (Wills & Napier-Munn, 2006). La ley es fuertemente influenciada por la selectividad de los reactivos químicos, el control de la profundidad de espuma (que permite un drenaje efectivo de la ganga) y la minimización del arrastre mecánico de partículas estériles (Fuerstenau, Jameson, & Yoon, 2007). Un equilibrio entre ley y recuperación es a menudo el objetivo, ya que mejorar una puede ir en detrimento de la otra.

- Recuperación

La recuperación (R) se refiere al porcentaje del mineral valioso presente en la alimentación que logra ser transferido al concentrado final. Es una medida de la eficiencia con la que el mineral deseado es colectado. Una alta recuperación indica que se está perdiendo menos mineral valioso en los relaves. Los factores que optimizan la eficiencia de la flotación, como el tiempo de residencia adecuado, el tamaño óptimo de partícula y la correcta dispersión de burbujas, son cruciales para maximizar la recuperación (Wills & Napier-Munn, 2006).

Sin embargo, es fundamental comprender que la búsqueda de una recuperación excesivamente alta se logra a expensas de la ley del concentrado. Para recuperar hasta la última partícula valiosa, las condiciones de flotación se vuelven menos selectivas, arrastrando consigo una mayor cantidad de ganga. Esto significa que, al intentar recuperar más mineral, la pureza del concentrado disminuye. Por lo tanto, el objetivo es encontrar un equilibrio óptimo entre la recuperación y la ley que maximice el valor económico global del proceso, considerando los costos de transporte, los procesos de refinación posteriores y las penalizaciones por baja ley (Wills & Napier-Munn, 2006).

Este porcentaje puede ser calculado por medio de la siguiente ecuación:

$$R (\%) = \frac{C * X_c}{F * X_f} * 100 \quad (3.1)$$

Donde:

- C : Flujo másico de concentrado
- F : Flujo másico de alimentación
- X_c : Ley de mineral valioso en el concentrado
- X_f : Ley de mineral valioso en la alimentación

Además de la recuperación y la ley del concentrado, existen otros indicadores relevantes para evaluar el rendimiento de una celda o circuito de flotación. Uno de ellos es la ley del mineral valioso en el relave (t), la cual constituye una herramienta clave para identificar ineficiencias y detectar oportunidades de mejora en el proceso (Wills & Napier-Munn, 2006).

3.2. Métodos de proyección

En el panorama actual de la ciencia y la industria, es común enfrentar procesos que generan cientos o miles de variables, con registros cada vez más frecuentes. Esta avalancha de datos a menudo se presenta de forma compleja y difícil de interpretar. A medida que los procesos se vuelven más intrincados, los conjuntos de datos exhiben un elevado número de variables y, frecuentemente, multicolinealidad, es decir, fuertes correlaciones entre las variables que describen el proceso (Jolliffe, 2002). Esta complejidad puede obstaculizar la visualización de la estructura de los datos, generar inestabilidad en los modelos estadísticos y aumentar el riesgo de sobreajuste.

Para manejar esta complejidad, se han desarrollado los métodos de proyección. Su concepto central es transformar el conjunto original de variables de alta dimensión en un nuevo conjunto de menor

dimensión, que capture y preserve la información más relevante del sistema. En esencia, estos métodos buscan las "sombras" o "proyecciones" más informativas de los datos en un espacio más manejable, minimizando la pérdida de información del conjunto original (Abdi & Williams, 2010).

En un contexto multivariado, donde las variables originales definen los ejes de un espacio de alta dimensión, los métodos de proyección crean nuevos ejes. Estos nuevos ejes son combinaciones lineales de las variables originales y dan como resultado nuevas variables (o componentes) que no solo reducen la dimensionalidad, sino que también resuelven el problema de la multicolinealidad, ya que están diseñadas para no estar correlacionadas entre sí (Jolliffe, 2002).

La aplicación de estos métodos ofrece ventajas sustanciales, lo que explica su adopción generalizada en diversas disciplinas:

- **Simplificación de la Complejidad (Reducción de Dimensionalidad):** Al trabajar con un número reducido de variables transformadas, la interpretación de la estructura de los datos se simplifica notablemente. Esta reducción no solo permite visualizar conjuntos de datos complejos mediante gráficos bidimensionales o tridimensionales, sino que también disminuye significativamente la carga computacional en las etapas posteriores del análisis. Este beneficio es crucial para la exploración de datos y la preparación de los mismos para algoritmos analíticos avanzados (Jolliffe, 2002).
- **Estabilidad y Robustez en el Modelado:** La multicolinealidad es un problema conocido en técnicas estadísticas como la regresión lineal, ya que puede generar coeficientes inestables y difíciles de interpretar. Los métodos de proyección, al generar componentes ortogonales o con menor correlación, mitigan este problema, dando lugar a modelos más estables y generalizables que se desempeñan mejor con datos nuevos (Frank & Friedman, 1993).

Gracias a su eficacia y versatilidad, los métodos de proyección se han convertido en herramientas fundamentales en disciplinas donde el análisis de datos multivariados es una necesidad constante, incluyendo la quimiometría, la sensometría y la biología, entre otras. Dentro de esta familia, dos de los métodos más influyentes y utilizados son el Análisis de Componentes Principales (PCA) y la Regresión de Mínimos Cuadrados Parciales (PLS) (Wold, Sjostrom, & Eriksson, 2001). Aunque ambos comparten el objetivo general de proyectar los datos en un espacio de menor dimensión conservando la mayor cantidad de información posible, difieren en sus objetivos específicos y en la forma en que construyen sus componentes, lo que los hace más apropiados para diferentes escenarios de análisis y modelado.

3.2.1. Análisis de componentes principales (PCA)

El Análisis de Componentes Principales (PCA) es un método cardinal en el análisis de datos multivariados, a menudo referido como "la madre de todos los métodos de análisis multivariados". Se clasifica como una técnica de aprendizaje no supervisado y es ampliamente utilizada en la fase exploratoria del análisis de datos (Jolliffe, 2002). Su objetivo principal es reducir la dimensionalidad de un conjunto de datos que contiene un gran número de variables interrelacionadas, logrando conservar la mayor cantidad de información posible, medida a través de la varianza presente en los datos originales (Abdi & Williams, 2010).

Esta reducción de la dimensión se consigue transformando el conjunto original de variables en un nuevo conjunto de variables, denominadas componentes principales, que no están correlacionadas entre sí (Jolliffe, 2002). Cada componente principal es una combinación lineal de todas las variables originales del conjunto de datos (Wold, Esbensen, & Geladi, 1987). Estas nuevas componentes pueden ser utilizadas posteriormente en métodos de aprendizaje supervisado, como la Regresión de Componentes Principales (PCR) o la Regresión de Mínimos Cuadrados Parciales (PLS) (Martens & Næs, 1992).

PCA también funciona como una potente herramienta de visualización para el análisis exploratorio de datos (Jolliffe, 2002). Al proyectar los datos en representaciones bidimensionales o tridimensionales utilizando las componentes principales, es posible identificar patrones, agrupamientos, tendencias o *outliers* que serían imperceptibles en el espacio original de alta dimensión (Jolliffe, 2002).

Es importante señalar que la aplicación de PCA pierde sentido si todas las variables en el conjunto de datos original son completamente independientes u ortogonales entre sí. Esto se debe a que el método busca y explota la redundancia y correlación entre las variables para lograr la reducción de dimensionalidad; si no existe tal interdependencia, no hay información que comprimir de manera significativa (Jolliffe, 2002).

Las componentes principales se construyen siguiendo un orden de importancia decreciente en términos de la cantidad de información (variabilidad) que cada una puede retener (Jolliffe, 2002). Así, la primera componente principal siempre capturará la mayor proporción de la varianza total de los datos originales, siendo la más importante en términos de información. Las componentes subsiguientes capturarán la varianza restante de forma decreciente, siendo la última componente la que menos información retenga. Este ordenamiento jerárquico es crucial para la reducción de

dimensionalidad, ya que permite seleccionar un número reducido de componentes que explican la mayor parte de la variabilidad del sistema sin perder información crítica (Abdi & Williams, 2010).

La validez y la interpretación de los resultados de PCA se sustentan en una serie de supuestos inherentes al sistema original:

- **Linealidad:** PCA asume que las relaciones entre las variables originales y las componentes principales son lineales. Esto significa que las componentes se expresan como combinaciones lineales de las variables de entrada. Si las relaciones subyacentes en los datos son intrínsecamente no lineales, PCA podría no ser la técnica más eficiente para capturar toda la estructura compleja de los datos (Jolliffe, 2002). En estos casos, pueden considerarse métodos no lineales o bien restringir el rango operativo de ciertas variables para aproximar un comportamiento más lineal, facilitando así una representación más eficiente mediante PCA; por ejemplo, acotando variables como la velocidad superficial del aire o la altura de espuma en procesos de flotación.
- **Varianza como medida de información:** El método define la "información" primariamente como la varianza de los datos. Su objetivo es maximizar la varianza explicada por cada componente principal. Esto implica que las variables con mayor variabilidad o dispersión en los datos son las que más contribuyen a las primeras componentes principales. Es fundamental considerar que una alta varianza no siempre equivale a "información útil"; PCA permite separar la variabilidad estructurada del proceso del ruido aleatorio, concentrando la información relevante en las primeras componentes principales.

Antes de aplicar PCA, es fundamental realizar un preprocesamiento adecuado del conjunto de datos, por dos razones principales:

- **Sensibilidad a valores atípicos (*outliers*):** PCA es susceptible a la influencia de datos extremos, los cuales pueden distorsionar significativamente la dirección de las componentes principales. No obstante, una vez construido el modelo, PCA permite identificar observaciones atípicas mediante el análisis de estadígrafos como Hotelling's T^2 y el Q *residuals*, los cuales facilitan la detección de desviaciones respecto al comportamiento normal. Por ello, resulta necesario considerar procedimientos de detección y tratamiento de *outliers* de forma previa al análisis (Jolliffe, 2002).
- **Estandarización de variables:** Dado que las componentes principales se construyen a partir de combinaciones lineales de las variables originales, sus escalas influyen directamente en el resultado. Si las variables están expresadas en distintas unidades (por ejemplo, millones

frente a unidades), aquellas con mayor varianza dominarán el análisis. Para evitar este sesgo, se realiza una estandarización: se sustrae la media y se divide por la desviación estándar de cada variable, asegurando así que todas tengan media 0 y desviación estándar 1 (Jolliffe, 2002).

El núcleo de este método se basa en el algebra lineal, por ende, para lograr obtener los componentes principales y reducir la dimensionalidad del conjunto de datos original se realiza una descomposición por valores y vectores propios (eigendecomposición) de la matriz de covarianza (Jolliffe, 2002). Con los datos ya preprocesados, el siguiente paso es cuantificar las relaciones lineales entre todas las parejas de variables. Esto se logra por medio de la matriz de covarianza o correlación, donde en esta matriz los elementos varían entre -1 y 1, indicando la fuerza y la dirección de la relación lineal entre cada par de variables.

Para una matriz o conjunto de datos X estandarizada con m observaciones y n variables la matriz, la matriz covarianza o correlación se calcula como:

$$Cov(X) = \frac{X^T X}{m - 1} \quad (3.2)$$

Donde X^T es la transpuesta de la matriz de datos. Esta matriz $S(Cov(x))$ cuadrada de $n \times n$ dimensiones contiene:

- En la diagonal de la matriz (S_{jj}): La varianza de cada variable. En el caso de preprocesar los datos los valores en la diagonal serán iguales a 1
- Fuera de la diagonal (S_{ij}): La covarianza o correlación entre la variable i y la variable j

Esta matriz contiene toda la información sobre la variabilidad y las correlaciones entre las variables originales que el método PCA intentara simplificar. Posteriormente, por medio de esta matriz de correlación se realiza la descomposición por valores y vectores propios, donde el objetivo es encontrar las direcciones (vectores propios) en el espacio de los datos originales donde la varianza es máxima (valores propios) (Abdi & Williams, 2010).

Para una matriz cuadrada S (matriz de covarianza/correlación), un vector propio corresponde a un vector no nulo que, cuando se multiplica por S , solo presenta un cambio en su escala (se alarga o acorta) pero nunca cambia su dirección. El factor de escala corresponde al valor propio (λ) asociado. Esto se describe por la ecuación:

$$S * \vec{v} = \lambda * \vec{v} \quad (3.3)$$

Al resolver esta ecuación para S (matriz de correlación/covarianza) nos proporciona un conjunto de "n" valores propios ($\lambda_1, \lambda_2, \dots, \lambda_n$) y sus respectivos "n" vectores propios ($\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$).

Los vectores propios o también conocidos como los vectores de “carga” (*loadings*) son las direcciones de los componentes principales. Cada vector de *loadings* es ortogonal (perpendicular) a los demás, lo cual garantiza que los componentes principales resultantes no se encuentren correlacionados, además cada uno de estos vectores indican el peso de cada variable original a la hora de construir cada componente principal (Jolliffe, 2002). Estos vectores forman la base del nuevo espacio de menor dimensionalidad. Es por esto por lo que la ecuación anterior se puede expresar como:

$$S * \vec{p}_i = \lambda_i * \vec{p}_i \quad (3.4)$$

Donde $\vec{p}$ son los vectores de *loadings*, y donde los valores propios (λ) representan la cantidad de varianza que cada componente principal explica. Por ende, un valor propio más grande indica que la componente principal asociada captura una mayor proporción de la variabilidad (información) total de los datos. Con los valores y vectores propios ya calculados, se procede a ordenar los pares (valor propio, vector propio) de mayor a menor valor propio como se indicó previamente. Por lo tanto, la componente principal asociada al valor propio más grande se convierte en la primera componente principal (PC_1). Esta es la dirección que explica la mayor varianza en los datos. La segunda componente principal (PC_2) es la dirección ortogonal que explica la mayor varianza restante, y así sucesivamente hasta llegar a un total de "n" componentes principales, por ende, se debe aplicar un criterio para determinar cuántos componentes principales se deben utilizar.

La suma de todos los valores propios es igual a la varianza total presente en la matriz de covarianza/correlación (Jolliffe, 2002). Para reducir la dimensionalidad, se realiza una selección de un subconjunto de los primeros "k" componentes principales que retengan la mayor parte posible de la información. Para determinar el número de componentes principales, los métodos más comunes corresponden a:

- **Varianza acumulada explicada:** Elegir "k" componentes que expliquen un porcentaje acumulado deseado de la varianza total (normalmente 80 o 90%) (Abdi & Williams, 2010).

- **Criterio de Kaiser:** Mantener únicamente componentes principales cuyos valores propios sean mayores a 1. Valores menores a 1 implican que la componente principal explica menos varianza (información) que una sola variable original preprocesada (Kaiser, 1960).

Finalmente, ya con el número de componentes principales determinado a partir de alguno de los métodos de selección, se procede a realizar la proyección de los datos originales (preprocesados) al nuevo espacio creado por los componentes principales. Para poder lograr esto se crea una matriz P de $n \times k$ dimensiones, donde cada columna corresponde a los k vectores propios seleccionados y ordenados según sus valores propios decrecientes. Posteriormente se realiza el cálculo de las puntuaciones (*scores*) que corresponden a las coordenadas de cada observación en el nuevo espacio de los componentes principales. Se obtienen multiplicando la matriz de datos original preprocesada ($X_{estandarizado}$) por la matriz de cargas (*loadings*) P :

$$T = X_{estandarizado} * P \quad (3.5)$$

Donde T corresponde a la matriz de “puntuaciones” y tendrá dimensiones $m \times k$. Cada columna de T representa una componente principal ($PC_1, PC_2, \dots, PC_k$), y cada fila de T contiene las “nuevas coordenadas” de cada observación en este espacio reducido de k dimensiones, con $k < n$.

Además de los valores propios, cargas (*loadings*) y puntuaciones (*scores*) que describen la estructura inherente de los datos, el Análisis de Componentes Principales (PCA) genera dos estadígrafos cruciales para el monitoreo de procesos y la detección de anomalías: el estadígrafo Hotelling's T^2 y el Q residuals (o Error Cuadrático de Predicción, SPE) (Montgomery, Peck, & Vining, 2012). Estos estadígrafos permiten evaluar la calidad del ajuste de cada observación al modelo PCA y determinar si un punto de datos se comporta de manera esperada o si representa una desviación o un *outlier*.

- Estadígrafo Hotelling's T^2

El estadígrafo T^2 es una medida de la variación de cada muestra dentro del espacio del modelo PCA generado. Indica que tan lejos se encuentra una observación/muestra del centro del modelo PCA ($scores = 0$) (Jolliffe, 2002). Para cada observación en el nuevo espacio, el estadígrafo T^2 se calcula utilizando sus puntuaciones (*scores*) en las k componentes principales seleccionadas y los valores propios asociados a cada componente, calculando así el estadígrafo como es expresa en la siguiente ecuación:

$$T_i^2 = \sum_{j=1}^k \frac{t_{ij}^2}{\lambda_j} \quad (3.6)$$

Donde t_{ij} corresponde a la puntuación (*score*) de la observación i en la componente principal j , T_i^2 corresponde al valor del estadígrafo de la observación i y λ_j corresponde al valor propio de la componente principal j .

- Estadígrafo Q *residuals*

El estadígrafo Q residuales, también conocido como Error Cuadrático de Predicción (SPE), indica que tan bien se ajusta cada observación al modelo (Jackson, 2003). Es una medida de la diferencia, o residual, entre una observación en el espacio original y la observación proyectada en el espacio nuevo generado por el método PCA. Cuantifica cuanto de la varianza (información) de una observación no llega a ser explicada por las k componentes principales retenidas. Para cada observación i , el estadígrafo se calcula como la suma de los cuadrados de los errores generados al momento de reconstruir esta observación del nuevo espacio k dimensional reducido al espacio original, ósea cuantificar cuanta información se ha perdido para esta observación, lo que se puede observar en la siguiente ecuación:

$$Q_i = \sum_{j=1}^n (x_{ij} - \hat{x}_{ij})^2 \quad (3.7)$$

Donde x_{ij} es el valor muestreado de la variable j para la observación i (espacio original), y $\hat{x}_{ij}$ corresponde al valor reconstruido de esa variable utilizando el modelo PCA con k componentes seleccionados. En esencia, es la suma de los cuadrados de los elementos del vector residual para cada observación.

3.2.2. Partial least square (PLS)

La regresión de mínimos cuadrados parciales (PLS), es una técnica multivariada que combina características del análisis de componentes principales (PCA) con la regresión múltiple. A diferencia del método PCA, que es una técnica de aprendizaje no supervisado centrada únicamente en la varianza de un conjunto de datos, PLS es una técnica supervisada diseñada específicamente para modelar la

relación entre un conjunto de variables independientes, X , y uno o más conjuntos de variables dependientes, Y (Wold, Sjostrom, & Eriksson, 2001).

El objetivo de esta metodología es construir un modelo predictivo robusto y explicativo, que es especialmente útil cuando el número de variables independientes (X) es muy grande, existe una fuerte correlación entre ellas o incluso cuando las observaciones totales son menores al número de variables utilizadas en el proceso. PLS se considera como un método robusto ante la incertidumbre de las mediciones ya que, PLS no presupone que los predictores sean fijos, sino que estos predictores se miden con cierto error, a diferencia de otros métodos como la regresión múltiple (Geladi & Kowalski, 1986).

Dentro de la regresión que realiza este método, se distinguen dos casos, si las variables dependientes (Y) se encuentran formadas por una o más variables. En el caso de solo presentar una única variable dependiente, se hablará de una regresión PLS1, que será cuando el proceso solo cuente con una única variable de respuesta y .

En el otro caso, donde la regresión PLS presente más de una variable dependiente (matriz Y) se denominará PLS2. El objetivo en ambos casos sigue siendo el mismo, crear modelos de calibración para la predicción de una o más variables dependientes a partir de las variables independientes. Tal y como ocurría en la metodología PCA, el espacio n dimensional original para los datos X se proyectará en un número reducido de componentes, pero adicionalmente este método PLS toma el p espacio dimensional original para los datos dependientes (Y) y los proyectará en un número reducido de componentes al igual que con las variables independientes. Con los componentes x e y relacionados por pares de máxima covarianza de sus *scores* (Frank & Friedman, 1993).

PLS es un método que busca maximizar una función objetivo bajo ciertas restricciones, siendo la función objetivo la covarianza entre los *scores* de x y los *scores* (puntuaciones) de y , con la restricción de ortogonalidad entre estas puntuaciones dependientes y las independientes.

Para la regresión centrándose en el caso más general que es el PLS2, se tiene que X es una matriz con m observaciones y n variables independientes ($m \times n$), mientras que Y es una matriz que al igual que X presenta m observaciones, pero con p variables dependientes. Conteniendo así n variables predictoras y p variables respuesta. De manera similar a lo realizado en la metodología PCA las matrices deben ser preprocesadas para luego trabajar con ellas, dado que, al igual que con PCA, los componentes generados en PLS se construyen a partir de las combinaciones lineales de las variables de origen, por ende, sus escalas influyen directamente en el resultado, por lo tanto, para evitar este

sesgo se realiza un preprocesamiento a los datos (Eriksson, y otros, 2006). Con las matrices X e Y preprocesadas, el objetivo del método es encontrar una relación lineal entre las variables dependientes e independientes de la forma:

$$Y = X * B + E \quad (3.8)$$

Donde E corresponde a la matriz de errores y B la matriz con los coeficientes de regresión de tamaño $n \times p$. Donde las matrices X e Y se modelan por variables latentes de acuerdo con modelos de regresión:

$$X = T * P^T + E_X \quad (3.9)$$

$$Y = U * V^T + E_Y \quad (3.10)$$

Donde las matrices E_X y E_Y son matrices de error. T y U son matrices de puntuación (*scores*); P y V son matrices de carga (*loadings*). Tanto las matrices de *scores* como las de *loadings* tendrán tantas columnas como número de componentes seleccionados, número de componentes que, de manera similar a lo que ocurre en PCA debe ser menor o igual que el valor mínimo entre el número de observaciones, el número de variables dependientes e independientes.

Como se indicó previamente los *scores* tanto en X como en Y corresponden a combinaciones lineales de estas variables. Donde los vectores t_j , u_j , p_j y v_j denotan la j -ésima componente seleccionada para $j = 1, \dots, k$. Donde los *scores* de las variables independientes y dependientes están conectados por una relación lineal interna.

$$u_j = d_j * t_j + h_j \quad (3.11)$$

Donde h_j corresponde al vector error y d_j el parámetro de regresión. Con lo que, de manera general esta relación entre las matrices T y U sería:

$$U = T * D + H \quad (3.12)$$

Con H como la matriz de errores y D una matriz solamente con elementos en su diagonal. Tendría sentido seguir utilizando los vectores de *loadings* en las matrices previamente indicadas (P y V), pero por razones técnicas del modelo, se utilizarán otros tipos de vectores de *loadings* que se denominarán

w y c , siendo w el vector de *loadings* de las variables independientes y c el vector de las variables dependientes, de tal forma que:

$$t = X * w \quad (3.13)$$

$$u = Y * c \quad (3.14)$$

Donde se encuentran bajo la restricción de requerir ortogonalidad para los vectores de *loadings*, así como necesitar que presenten una norma unitaria. Por ende, tendremos un problema de maximización que puede describirse como la maximización de:

$$t^T * u = (X * w)^T * Y * c = w^T * X^T * Y * c \quad (3.15)$$

Todo lo previamente indicado se calcula mediante distintos algoritmos matemáticos (kernel, NIPALS, SIMPLS, entre otros). Generalmente, la primera componente se determina de manera similar en todos ellos. Sin embargo, las diferencias entre los algoritmos radican en la forma en que se calculan los componentes restantes seleccionados en el método. Por ejemplo, en el caso del algoritmo SIMPLS tendremos que la matriz de los coeficientes de regresión (B) es (Jong, 1993):

$$B = WT^TY \quad (3.16)$$

Una vez determinado el modelo por medio del algoritmo, se pueden utilizar los estadígrafos previamente indicados (T^2 y Q), donde podremos observar que tan bien se ajustan las observaciones al modelo (Q *residuals*) y también determinar la varianza de cada observación dentro del nuevo espacio dimensional generado por el método PLS (Hotelling's T^2). Lo cual va acompañado de los típicos análisis realizados a regresiones como lo puede ser el análisis de residuos.

3.2.3. Partial least square dinámico (PLS)

Los procesos industriales, y particularmente aquellos presentes en la industria minera, son inherentemente dinámicos. Esto significa que las salidas o respuestas del sistema en un momento determinado no dependen únicamente de las condiciones actuales, sino también de las condiciones pasadas. Por ejemplo, en una planta de flotación de minerales, una modificación en la dosificación de reactivos como el colector o el espumante no produce un efecto inmediato sobre la ley de concentrado del mineral de interés. Existen retardos significativos debido al tiempo de residencia de

la pulpa en las celdas, a la cinética de las reacciones involucradas y al tiempo de transporte dentro del circuito (MacGregor & Kourti, 1995).

La regresión de mínimos cuadrados parciales dinámica (PLS dinámico) surge como una extensión de la metodología PLS tradicional, precisamente para enfrentar estas complejidades temporales. Mientras que el enfoque tradicional asume relaciones en estado estacionario entre variables independientes (X) y dependientes (Y), el PLS dinámico incorpora explícitamente la estructura temporal y los retardos del sistema en el modelo (MacGregor & Kourti, 1995).

La principal diferencia entre PLS tradicional y dinámico radica en la forma en que manejan la dependencia temporal de los datos. El PLS tradicional predice cada observación de Y usando únicamente la observación correspondiente en X, ignorando los efectos acumulados de entradas y salidas pasadas. En procesos con dinámicas relevantes, esto puede derivar en modelos inexactos y una comprensión deficiente del sistema.

Por otro lado, los métodos de regresión PLS dinámica amplían las capacidades de la metodología PLS tradicional al incorporar explícitamente información temporal en las matrices de entrada. Esto se logra mediante la inclusión, en la matriz de variables predictoras (X), no solo de los valores actuales de las variables del proceso, sino también de sus valores pasados (*lags*).

La expansión de la matriz de entrada con variables retardadas permite que el modelo capture de manera más efectiva la dinámica del sistema. En particular, facilita la representación de la inercia del proceso, los tiempos de retardo característicos de las operaciones unitarias, y el modo en que las perturbaciones en las variables de entrada se propagan y afectan a las variables de salida a lo largo del tiempo (Wold, Sjostrom, & Eriksson, 2001).

El desarrollo de los métodos de regresión PLS dinámicos no representó un salto repentino, sino una evolución natural impulsada por la necesidad de aplicar herramientas multivariadas a sistemas con comportamiento dinámico. Esta evolución tiene sus raíces en el trabajo pionero de Herman Wold en la década de 1960, quien sentó las bases del análisis multivariante. Con el tiempo, surgió la necesidad de adaptar estas técnicas para abordar datos caracterizados por autocorrelación y estructuras temporales, particularmente en el contexto de procesos industriales. Los primeros estudios en la aplicación de métodos quimiométricos para el monitoreo y el diseño de controladores de alto rendimiento fueron fundamentales para esta evolución (Kaspar & Ray, 1992).

Se reconoció que ignorar la estructura temporal de los datos podía conducir a errores significativos en el modelado y en la interpretación de los resultados. En respuesta a este desafío, se desarrolló la

metodología PLS dinámica, la cual integra explícitamente modelos de series de tiempo (como los modelos ARX y FIR) en la etapa de construcción de las matrices de datos utilizadas por el algoritmo PLS (Kaspar & Ray, 1993).

Gracias a la integración de estructuras de series de tiempo, el PLS dinámico permite capturar y modelar directamente los efectos dinámicos presentes en los procesos industriales, superando así las limitaciones del PLS tradicional, que no está diseñado para representar de manera explícita las relaciones temporales entre variables (Ronen, Sanders, Tan, Mort, & Doyle, 2009).

Estos avances fueron fundamentales para la evolución del PLS, transformándolo de una herramienta estática de análisis multivariado en una metodología robusta aplicable al monitoreo estadístico de procesos (MSPC) y al control predictivo en tiempo real. Su implementación ha sido especialmente valiosa en industrias con procesos complejos y dinámicos, como la minería.

Para incorporar la dinámica temporal en las matrices de entrada del algoritmo PLS, se recurre a representaciones basadas en modelos de series de tiempo. Entre las más utilizadas se encuentran los modelos Autorregresivos con Entradas Exógenas (ARX) y los modelos de Respuesta a Impulso Finito (FIR) (Ljung, 1999). Estas estructuras definen cómo las salidas actuales dependen de entradas y salidas pasadas, permitiendo transformar un problema inherentemente dinámico en uno estático mediante la expansión de la matriz de predictores X . De esta forma, es posible capturar las relaciones temporales de manera implícita dentro del marco del PLS.

- Modelo Autorregresivo con Entradas Exógenas (ARX por sus siglas en ingles)

El modelo ARX es una representación versátil y ampliamente utilizada para describir sistemas dinámicos lineales (Ljung, 1999). Este modelo expresa la salida actual de un sistema como una combinación lineal de sus propias salidas pasadas (componente autorregresiva) y de las entradas pasadas (componente exógena). Esta formulación resulta especialmente intuitiva en el contexto de procesos industriales, donde el estado actual de una variable (por ejemplo, el pH en una celda de flotación) suele ser una continuación de su estado anterior, influenciada por las acciones tomadas en el sistema.

Por ejemplo, imaginemos que se desea predecir la ley de concentrado de cobre ($y(t)$) en el instante de tiempo t . El modelo ARX diría que esta ley no depende únicamente de las variables independientes en un pasado reciente, sino que también de cuál fue la ley de concentrado en el momento anterior $y(t - 1)$ o hace dos momentos $y(t - 2)$. La inclusión de $y(t - 1)$, $y(t - 2)$, ..., $y(t - l)$, permite capturar la “memoria” o inercia del proceso.

Para una sola variable dependiente $y(t)$ y una única variable de entrada o independiente $u(t)$, el modelo ARX se puede expresar como:

$$y(t) = f(u(t-1), \dots, u(t-l)) + f(y(t-1), \dots, y(t-l)) + Error \quad (3.17)$$

Mas formalmente:

$$y(t) = b_{t-1}u(t-1) + \dots + b_{t-l}u(t-l) + a_{t-1}y(t-1) + \dots + a_{t-l}y(t-l) + e(t) \quad (3.18)$$

Donde:

- $y(t)$: Valor de la salida en el instante de tiempo t
- $u(t-l)$: Valores de la variable independiente u en momentos pasados, considerando que presenta un total de l lags o retrasos en el modelo
- $y(t-l)$: Valores de la variable de salida, considerando un total de l lags o pasos atrás
- b_{t-1} y a_{t-1} : Coeficientes que el modelo va a generar

Considerando que el algoritmo opera sobre la base de la metodología PLS tradicional, es necesario construir una nueva matriz de predictores, denominada X_{ARX} , para que pueda capturar adecuadamente la dinámica del sistema.

Esta matriz incluye columnas correspondientes a cada variable independiente con sus respectivos retardos o lags definidos. Además, incorpora columnas adicionales para los lags de las variables dependientes o de salida. De esta forma, se representa explícitamente la estructura temporal del proceso. Cada fila de X_{ARX} corresponde a una observación del proceso en un instante de tiempo t , mientras que las columnas contienen los valores previos de sus variables independientes junto con los valores pasados de las variables de salida.

$$X_{ARX} = \begin{bmatrix} u_1(t_1-1) & \dots & u_1(t_1-l) & y_1(t_1-1) & \dots & y_1(t_1-l) \\ u_1(t_2-1) & \dots & u_1(t_2-l) & y_1(t_2-1) & \dots & y_1(t_2-l) \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ \vdots & \dots & \vdots & \vdots & \ddots & \vdots \\ u_1(t_m-1) & \dots & u_1(t_m-l) & y_1(t_m-1) & \dots & y_1(t_m-l) \end{bmatrix} \quad (3.19)$$

Al aplicar esta nueva matriz de predictores X_{ARX} en el algoritmo PLS, se obtiene como resultado la matriz de salida Y_{ARX} , la cual contiene los valores estimados de las variables dependientes que se desean predecir a partir de la información contenida en X_{ARX} .

$$Y_{ARX} = \begin{bmatrix} y_1(t_1) \\ y_1(t_2) \\ \vdots \\ y_1(t_m) \end{bmatrix} \quad (3.20)$$

En caso de presentar más variables de salida, la matriz Y_{ARX} presentará más columnas, al igual que en el caso de la matriz X_{ARX} . Las matrices deben preprocesarse (centrado y escalado) para posteriormente utilizarse en el algoritmo PLS tradicional.

- Modelo de Respuesta a Impulso Finito (FIR por sus siglas en ingles)

El modelo FIR (*Finite Impulse Response*) se centra en describir cómo la salida actual de un proceso corresponde a una suma ponderada de las entradas pasadas hasta un cierto punto en el tiempo (Ljung, 1999). De manera similar al modelo ARX, utiliza una cantidad definida de datos pasados para predecir adecuadamente una variable dependiente o de salida; sin embargo, en el caso del FIR, solo considera las variables independientes o de entrada con un número determinado de retardos o *lags*. Este modelo es especialmente útil para entender de forma directa el efecto que una perturbación puntual en la entrada tiene sobre la salida a lo largo del tiempo.

Por ejemplo, si en una celda de flotación la ley de alimentación del mineral de interés $u(t)$ cambia, la ley de concentrado del mineral en la celda $y(t)$ no reaccionara de manera instantánea. Un modelo FIR indica que la ley de concentrado actual $y(t)$ es una suma del efecto de la ley de alimentación de hace 1 minuto, la ley de alimentación de hace 2 minutos, y así sucesivamente, hasta un cierto número de retrasos. Este enfoque es útil cuando hay un retardo de tiempo claro y un efecto que se “desvanece” con el tiempo.

Para una sola variable de salida $y(t)$ y una única variable de entrada o independiente $u(t)$, el modelo FIR se puede expresar como:

$$y(t) = f(u(t-1), \dots, u(t-l)) + Error \quad (3.21)$$

Mas formalmente:

$$y(t) = b_{t-1}u(t-1) + \dots + b_{t-l}u(t-l) + e(t) \quad (3.22)$$

Donde:

- $y(t)$: Valor de la salida en el instante de tiempo t
- $u(t-l)$: Valores de la variable independiente u en momentos pasados, considerando que presenta un total de l lags o retrasos en el modelo
- b_{t-1} : Coeficientes que el modelo va a generar a partir de la perturbación o impulso generado

Similar al caso del modelo ARX, y considerando que el algoritmo funciona con base en la metodología PLS tradicional, la nueva matriz de predictores X_{FIR} incluye columnas correspondientes a cada variable independiente con sus retardos o lags definidos. Cada fila de esta matriz representa una observación del proceso en un instante de tiempo t . Las columnas de X_{FIR} contienen los valores históricos de las variables independientes (retardos o lags), mientras que la matriz de salida Y_{FIR} es idéntica a la que se utiliza en el modelo ARX.

$$X_{FIR} = \begin{bmatrix} u_1(t_1-1) & \dots & u_1(t_1-l) & u_2(t_1-1) & \dots & u_2(t_1-l) \\ u_1(t_2-1) & \dots & u_1(t_2-l) & u_2(t_2-1) & \dots & u_2(t_2-l) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ u_1(t_m-1) & \dots & u_1(t_m-l) & u_2(t_m-1) & \dots & u_2(t_m-l) \end{bmatrix} \quad (3.23)$$

Para cada variable de entrada u_i , se incluyen en la matriz de predictores sus valores pasados ($u_i(t-1), u_i(t-2), \dots, u_i(t-l)$), considerando l retrasos o lags en el proceso. A partir de esta matriz de predictores obtendremos la matriz de salida (Ecuación 3.24).

$$Y_{FIR} = \begin{bmatrix} y_1(t_1) \\ y_1(t_2) \\ \vdots \\ y_1(t_m) \end{bmatrix} \quad (3.24)$$

Similar al modelo ARX, el modelo FIR requiere un preprocesamiento previo de los datos con los cuales se va a trabajar. Los métodos PLS dinámicos, tanto FIR como ARX, al incorporar explícitamente la dinámica del proceso de esta manera, se convierten en herramientas indispensables para la construcción de modelos predictivos precisos en la industria. Estos métodos sientan las bases

para el desarrollo de sistemas de control avanzado y monitoreo estadístico de procesos (MSPC) más robustos, contribuyendo en última instancia a una operación más eficiente y optimizada (MacGregor & Kourti, 1995).

4. Equipo experimental y metodología

El objetivo principal de esta metodología es desarrollar un sensor virtual dinámico que permita estimar las concentraciones de salida en una celda de flotación *rougher* piloto. Este proceso se basa en el *Partial Least Square* (PLS), una técnica adecuada para manejar conjuntos de datos complejos y altamente correlacionados como los que se encuentran en el proceso de flotación.

4.1. Preparación del laboratorio y sistema de control

En primer lugar, debido a que esta metodología necesita de equipo experimental para poder establecer la región de operación de las distintas variables involucradas en el proceso, se va a utilizar una de las celdas disponibles en el laboratorio de control, celda en la que se debe realizar una calibración y configuración del sistema de control que presenta (Figura 5). Esto implica verificar y calibrar los sensores, medidores y controladores instalados en la celda de flotación *rougher* piloto. Es crucial asegurar que todos los equipos estén correctamente configurados para proporcionar mediciones precisas y confiables de las variables operativas de la celda de flotación.



Figura 5. Sistema de celdas de flotación rougher.

En este caso, dado que se utilizará solo una celda de flotación *rougher* del sistema completo, se simplificará la preparación del sistema de control. Esto facilitará una atención detallada y garantizará que la configuración inicial sea precisa y efectiva antes de proceder con las siguientes etapas de la investigación.

4.1.1. Descripción general del equipo a utilizar

El sistema analizado simula el funcionamiento de una celda de flotación *rougher* (Figura 5), etapa clave en la separación de minerales valiosos mediante la adhesión de partículas a burbujas de aire generadas en presencia de reactivos químicos. Las celdas de flotación del sistema *rougher* del laboratorio son idénticas en diseño. Sus dimensiones se presentan en el diagrama de la Figura 6, mientras que sus características principales se detallan en la Tabla 1.

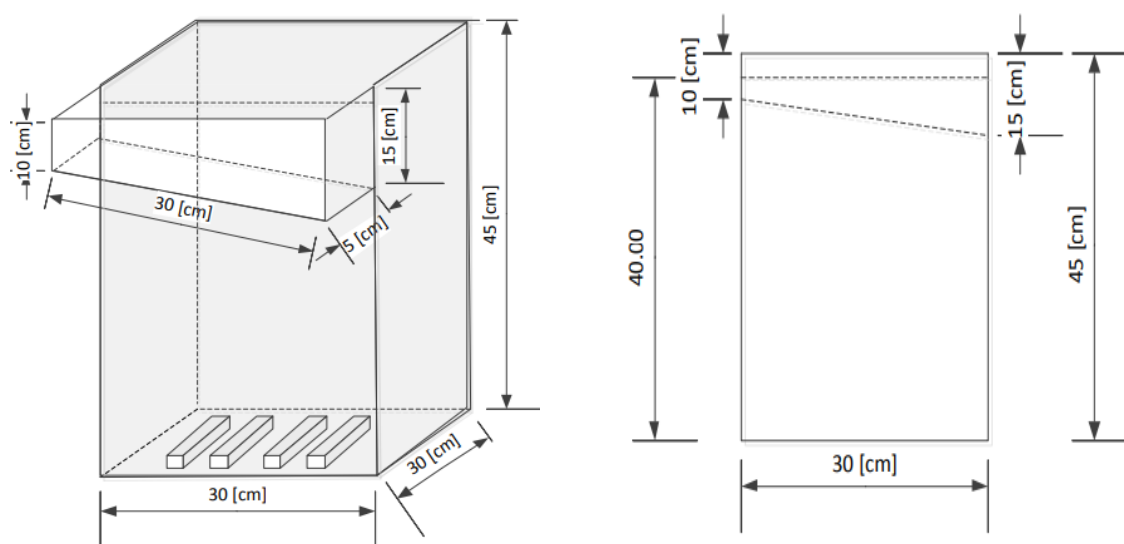


Figura 6. Esquema de celdas de flotación rougher piloto (Durán, 2014).

Tabla 1. Características de la celda de flotación

Alto	45	[cm]
Ancho	30	[cm]
Largo	30	[cm]
Número de difusores	4	-
Material de celda	Acrílico	-
Sistema de agitación	Forzado	-
Tipo de rebalse	Periférico	-
Tipo de difusor	Piedra porosa	-

La alimentación a la celda se realiza mediante una bomba peristáltica que extrae agua desde un estanque de suministro. Como se observa en la Figura 5, la celda cuenta con un agitador que permite

regular el tamaño de las burbujas durante la operación, funcionando a una velocidad fija de 564 RPM, con control únicamente de encendido y apagado.

Para las labores de medición y control, el sistema dispone de distintos sensores:

- **Flujo de aire:** medido por un medidor de flujo másico, que registra con precisión la cantidad de aire suministrada a la celda.
- **Nivel en la celda:** medido por un sensor capacitivo y un sensor ultrasónico. No obstante, dado que el sensor capacitivo no entrega lecturas confiables en niveles altos, en este trabajo se emplea exclusivamente el sensor ultrasónico, capaz de medir en todo el rango de altura de la celda y, por lo tanto, ideal para un monitoreo continuo y preciso.

En cuanto a los actuadores, la celda cuenta con dos válvulas tipo *pinch* destinadas a regular el flujo de aire y el nivel. Ambas operan mediante un transductor I/P (*current to pressure*), que convierte una señal eléctrica de 4–20 [mA] en una señal neumática de 3–15 [psig], ajustando la variable de proceso según las necesidades de operación.

4.1.2. Programación software de control

La configuración del sistema de control para la primera celda inicia con la conexión eléctrica entre instrumentos y PLC, considerando la naturaleza y dirección de las señales. Estas pueden ir del PLC hacia los equipos (control, activación o desactivación) o de los instrumentos hacia el PLC (envío de datos). Posteriormente, se realiza la programación por bloques unitarios, configurando actuadores (válvulas y bomba) y sensores (ultrasónico y capacitivo), aunque para el control de nivel se emplea únicamente el ultrasónico. También se agrupan los elementos del flujo de aire y se implementan los bloques PID para regular nivel y flujo. La programación del *Ladder* define direcciones tipo %R para todas las señales registradas en KepServer con su respectiva descripción. Luego, en el programa SMC (*System Management Console*) se cargan las etiquetas (*tags*) para habilitar su uso en otros equipos y en la creación de la interfaz HMI (*Human Machine Interface*).

4.1.3. Configuración HMI y calibración de variables

Con las variables cargadas en el nuevo equipo, se diseña la interfaz HMI en el programa HMI InTouch, incorporando los *tags* de nivel, flujo de aire, parámetros de control PID, encendido/apagado de equipos y *setpoints*. Entre ellos destacan las tres variables clave de este trabajo: flujo de alimentación de pulpa, velocidad superficial del aire y altura de espuma, que junto con variables

virtuales conformarán las entradas del simulador descrito en secciones posteriores para el desarrollo del sensor virtual.

Los registros de estas variables y de otros instrumentos se presentan inicialmente en formato “número de máquina” (0 a 32.767), lo que dificulta su control directo. Por ello, es necesaria su conversión a unidades de ingeniería mediante una calibración minuciosa de cada instrumento, actuador y equipo, aplicando metodologías específicas en cada caso. En lo que sigue, se describen la calibración y rangos de operación de las tres variables que constituyen el foco de este procedimiento.

Calibración de la bomba (Flujo de alimentación de pulpa): La calibración se realizó mediante el método de aforo volumétrico, que permite establecer la relación entre un número de máquina específico y el flujo de agua en litros por minuto. Para ello, se ajustó el número de máquina en el programa Proficy, que controla el flujo de alimentación de la bomba, y se efectuaron múltiples mediciones variando dicho valor. Esto permitió generar una función, obtenida a partir de un gráfico que relaciona el flujo (en unidades de ingeniería) con el número de máquina entregado por el sistema (Figura 31, Anexo 4). Una vez incorporada la calibración en el HMI, se obtuvo una lectura más clara y precisa del flujo de alimentación de la bomba. De este modo, se determinó que el rango operativo de la bomba (y, por ende, del flujo de alimentación de pulpa) se sitúa entre 0 y 4,4 [L/min]. Estos límites representan las capacidades máximas y mínimas de operación de la celda, aunque el rango utilizado en la simulación será más acotado y se encontrará contenido dentro de dichos valores.

Calibración del sensor de nivel (altura de espuma): La calibración del nivel de la celda (altura del líquido) se llevó a cabo utilizando instrumentos de medición en centímetros, como regla o cinta métrica. A través del programa Proficy se fijó un número de máquina en el sensor ultrasónico mediante la apertura o cierre de la válvula de control de nivel. Una vez estabilizada la lectura, se midió la altura real del líquido, permitiendo establecer la relación entre el número de máquina y la altura en centímetros.

Esta relación se expresa mediante una función que permite convertir el número de máquina en una medida de altura de espuma. Dicha función corresponde a una recta con pendiente igual a la del nivel, pero con signo opuesto, generando una simetría respecto al eje Y (Figura 32, Anexo 4). Así, los límites máximos y mínimos para la altura de espuma son 21,3 y 0 [cm], respectivamente. Sin embargo, al igual que con el flujo de alimentación de pulpa, estos límites corresponden a valores teóricos máximos y mínimos definidos por las características del equipo, no a los utilizados en la operación real de la celda piloto, cuyo rango efectivo de operación se encuentra dentro de dichos valores y será especificado en secciones posteriores.

Calibración del flujo de aire (velocidad superficial del aire): La calibración se realizó utilizando un rotámetro para medir el flujo en litros por minuto. A través del *software Proficy*, se ajustó el porcentaje de apertura de la válvula de aire, registrando simultáneamente los números de máquina del sensor y de la válvula. Aunque ambos comparten el mismo rango (0 a 32.767), sus valores no coincidieron, por lo que fue necesario establecer una relación con el flujo real medido en $[\text{cm}^3/\text{s}]$. Estas correspondencias se presentan en las Figura 33 y Figura 34 (Anexo 4), mientras que la Figura 35 (Anexo 4) muestra el gráfico final que cuantifica dicha relación. A partir de esta calibración y considerando el área de la celda, se determinó que la velocidad superficial del aire varía teóricamente entre 0 y 1,8 $[\text{cm/s}]$. No obstante, estos valores corresponden a limitaciones de los equipos, siendo los rangos efectivos de operación más acotados, los cuales se detallan posteriormente.

Con este procedimiento se completó la calibración de la velocidad superficial del aire, junto con la ya realizada para la altura de espuma y el flujo de alimentación de pulpa, estableciendo así los límites máximos y mínimos de las tres variables clave que serán empleadas como entradas para la construcción del sensor virtual.

4.2. Código de simulación del proceso de flotación

El código de simulación utilizado para simular el proceso de flotación en una única celda *rougher* se encarga de recibir variables de entrada para posteriormente entregar la recuperación de cobre, la ley de cobre en el concentrado y la ley de colas. Originalmente el código simulaba el sistema de flotación *rougher* presentado en la Figura 5, considerando las 3 celdas presentes en el sistema. Este código de Matlab consiste en una función que solicita variables de alimentación para poder otorgar resultados en estado estacionario de la operación de esta única celda (Figura 7).

```
function [recuperacion1, LeyCobreConcentrado, LeyCobreColas]=Simuladormemoria(flujoVolumetricoPulpaAlimentacion,porcentajeSolidosAlimentacion)

% Ley_Tail es la ley de la cola 3

areaCelda      = 0.09*100*100 ; % cm^2

% Parametros de la ecuacion de climper

P              = [0.001197 0.000977 ; 0.00209 0.001790 ; 0.001898 0.001525];
Rmax(:,1,1)    = [56.58 ; 94.74 ; 98.2];
Rmax(:,2,1)    = [27.92 ; 78.95 ; 86.76];

% Calculo pulpa alimentacion (habria que revisar unidades)

densidadSolido = 3.8; %habria que ver a que hace referencia esta densidad, que minerales, etc.
densidadAgua   = 1; %nada que agregar esta bien

porcentajeAguaAlimentacion(1) = 100-porcentajeSolidosAlimentacion(1); %el calculo es correcto

% Formula comun para determinar densidad de pulpa (bien calculado)

densidadPulpaAlimentacion(1) = ((porcentajeSolidosAlimentacion(1)/100)/densidadSolido+(porcentajeAguaAlimentacion(1)/100)/densidadAgua);

flujoPulpaAlimentacion(1) = flujoVolumetricoPulpaAlimentacion(1)*densidadPulpaAlimentacion(1);

flujoSolidosAlimentacion(1) = flujoPulpaAlimentacion(1)*(porcentajeSolidosAlimentacion(1)/100);
```

Figura 7. Código de simulación flotación en celda rougher

Debido a que el código originalmente fue creado para este sistema de celdas, las modificaciones realizadas fueron mínimas, centradas únicamente en pasar de simular 3 celdas a una única celda. La clara y obvia limitante del código es el hecho de que los resultados metalúrgicos que entrega corresponden a estados estacionarios del proceso de flotación, por ende, para poder simular el proceso de forma dinámica, se añadirá al código un modelo de primer orden que me permita obtener la trayectoria que se sigue desde un valor de ley de cobre en el concentrado a otro, generado a partir de la perturbación de una de las variables de entrada que requiere el código de simulación.

Otro elemento a resaltar es que el código caracteriza una simulación de un proceso de flotación *rougher* que en la alimentación contiene Pirita (FeS_2), Calcopirita (CuFeS_2), Molibdenita (MoS_2) y con una densidad del sólido igual a 3,8 [kg/L].

4.2.1. Variables de proceso en el código de simulación

4.2.1.1. Variables de entrada en el código

Las variables consideradas se dividen en dos grupos principales: aquellas que caracterizan el flujo de alimentación que ingresa a la celda y las que describen las condiciones en las que esta opera. A estas se suma una tercera categoría, correspondiente a las variables que definen las características propias de la celda simulada en el código.

- Variables de alimentación

Se encargan de describir y detallar las características que presenta la corriente que se alimenta a la celda de flotación *rougher*. En este caso el código considera propiedades como el porcentaje de sólidos en la alimentación, el flujo de alimentación en la celda, el contenido de cobre, fierro y molibdeno en la corriente, y también el porcentaje de gruesos que contiene la alimentación.

En el caso de las variables de alimentación, es importante destacar que el laboratorio no trabaja con un flujo real de pulpa ni opera directamente con minerales. Por ello, resulta necesario diferenciar entre las variables que pueden ser modificadas de manera efectiva mediante el *software* de control y aquellas que no. En este contexto, la única variable que puede ajustarse de forma real corresponde al flujo de alimentación, asumiéndose que este es equivalente al flujo de pulpa. En contraste, las demás variables previamente descritas en este apartado se consideran “virtuales”, ya que no es posible determinar sus rangos de valores directamente en el laboratorio. Para estas variables se establecieron rangos máximos y mínimos definidos a partir de los valores generales con los que operan plantas de flotación, los cuales delimitan su comportamiento dentro del modelo. Todas estas variables, junto con sus respectivos rangos de operación, se presentan en la Tabla 2.

Tabla 2. Variables de alimentación.

Variable	Valor mínimo	Valor máximo	Unidad
Flujo de alimentación	3	4	[L/min]
Porcentaje de solidos	20	40	(%)
Ley de cobre en alimentación	0,3	1	(%)
Ley de fierro en alimentación	1,5		(%)
Ley de molibdeno en alimentación	-		(%)
Porcentaje de gruesos en alimentación	5	20	(%)

En el caso del flujo de alimentación, sus límites operativos se establecieron a partir de las calibraciones realizadas y de la lógica de funcionamiento de una celda de flotación. Estos valores garantizan que el proceso opere dentro de un rango coherente con condiciones reales, asegurando resultados representativos para los objetivos de esta investigación.

Respecto a las variables virtuales, se definió una ley de cobre en la alimentación entre 0,3 y 1 %, en concordancia con la realidad actual de los yacimientos (Ndoro & Witika, 2017). Para el porcentaje de sólidos, se adoptó un rango entre 20 y 40 %, tomando como referencia el valor típico de 30 %, utilizado en celdas de flotación para evitar problemas de viscosidad excesiva en la pulpa (Ormeño Puentes, 2024). En cuanto a otras leyes, el molibdeno no se consideró en este análisis debido a su baja concentración y al enfoque centrado en el cobre de este trabajo. Para la ley de fierro, se adoptó un valor fijo de 1,5 %, consistente con trabajos previos realizados con el mismo código de simulación (Bravo Rain, 2017).

Finalmente, variables adicionales como la densidad del sólido (3,8 [kg/L]) y el tamaño de partícula (45 [μm] para finos y 255 [μm] como máximo) fueron fijadas directamente por el simulador, sin necesidad de ajustes adicionales.

- Variables de operación

Estas corresponden a las variables que permiten definir en qué condiciones opera la celda de flotación. Las variables utilizadas en este caso corresponden a la velocidad superficial del aire y la altura de espuma en la celda. Los rangos de operación de estas variables se detallan en la Tabla 3.

Tabla 3. Variables de operación.

Variable	Valor mínimo	Valor máximo	Unidad
Velocidad superficial del aire	0,5	1	[cm/s]
Altura de espuma	1	4	[cm]

Estos valores son considerados a partir de las calibraciones realizadas junto con trabajos realizados en el mismo sistema de celdas, donde se definen valores en los cuales la celda de laboratorio realmente opera y que son los reflejados en la tabla previa. Evidentemente, al tratarse de mediciones de laboratorio, estas mediciones son con una alimentación ausente de sólidos y manteniendo constante la concentración de espumante, con una velocidad de agitación constante (Zambrano Ríos, 2024).

- Variables geométricas

Estas variables están directamente relacionadas con la celda piloto con la cual se trabaja y que fue descrita previamente, siendo una celda rectangular que en la Tabla 4 se detallan sus dimensiones junto con el área superficial y volumen. En este caso el código ya considera estos valores para los cálculos internos que debe realizar, motivo por el cual no se requiere que se ingresen estas variables.

Tabla 4. Variables geométricas.

Variable	Valor	Unidad
Altura de la celda	45	[cm]
Largo de la celda	30	[cm]
Ancho de la celda	30	[cm]
Área superficial de la celda	900	[cm ²]
Volumen de la celda	40.500	[cm ³]
Tipo de celda	Rectangular	-

4.2.1.2. Variables de salida del código

El código a partir de las variables que caracterizan la alimentación y las variables de operación, junto con las dimensiones de la celda piloto, permite generar una serie de resultados metalúrgicos del proceso a partir de los balances de materia, el modelo cinético de la zona de colección, el modelo de recuperación de la zona de espuma y la distribución de tamaños de los minerales. El código aparte de entregar la ley de cobre en el concentrado, indica la recuperación de cobre en esta corriente, junto con

la ley de cobre en las colas, siendo los más especiales y útiles la ley de cobre y la recuperación de cobre en el concentrado, el primero debido a que se busca generar un sensor virtual para predecir sus valores en la celda piloto y el segundo porque posteriormente será utilizado como un filtro para todos los resultados obtenidos al utilizar este código de simulación.

En la Figura 8 se puede apreciar una representación por medio de un esquema de lo que realiza este código de simulación de Matlab, en este se detallan aquellas variables que serán utilizadas para perturbar el sistema y aquellas que se mantendrán fijas e imperturbables. Variables de entrada para las cuales el código entregara los resultados metalúrgicos previamente mencionados.

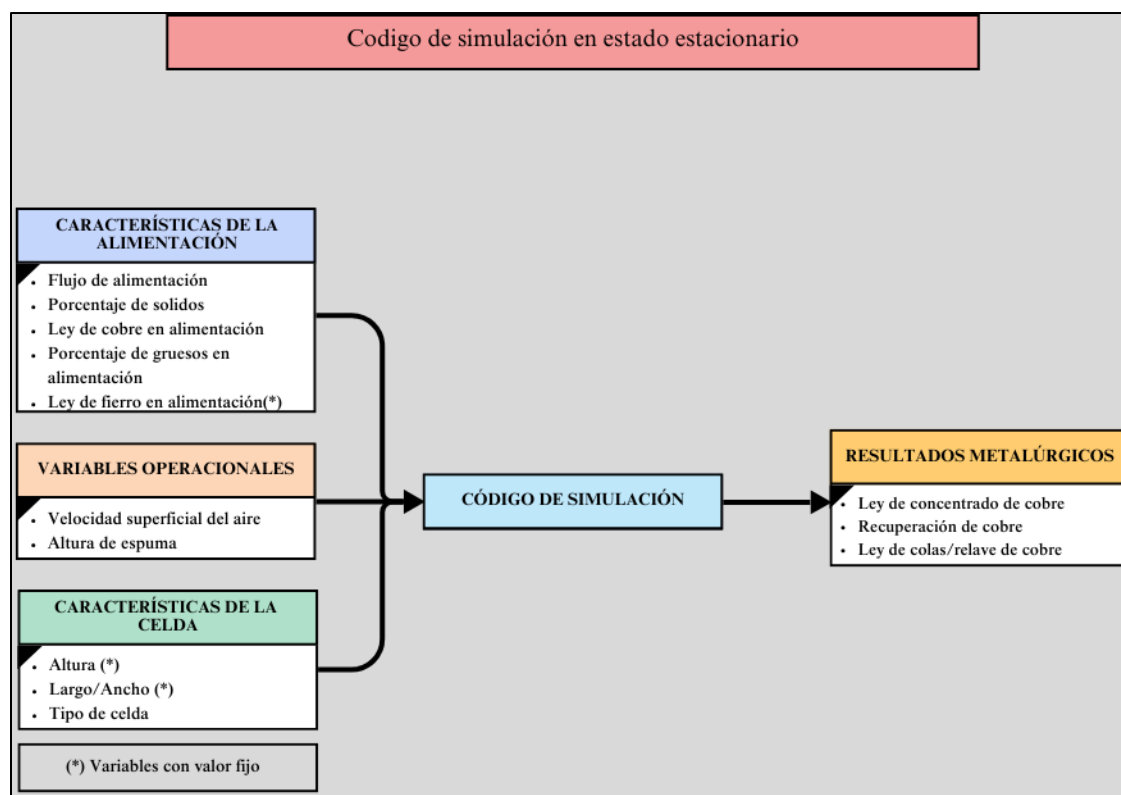


Figura 8. Código de simulación en estado estacionario

4.3. Generación de respuesta transiente en los datos

Tal como fue indicado previamente el código de simulación presenta la limitante de entregar sus resultados metalúrgicos cuando el proceso se encuentra en estado estacionario, motivo por el cual para la investigación se hace esencial agregar un sistema o modelo dinámico que permita predecir como cambiara la ley de concentrado en el tiempo a partir de las perturbaciones realizadas en las variables de entrada, y que de esa forma permita construir la trayectoria que sigue esta variable de interés en un tiempo determinado. De esta forma se lograría simular el comportamiento dinámico de

la ley de concentrado de cobre de acorde a las perturbaciones realizadas en la operación de la celda o las características que presenta la alimentación. Por ende, al código de simulación previamente explicado y detallado en la Figura 7, se le debe acoplar un simulador del comportamiento dinámico de la variable de interés, lo que permite detallar como actúa la ley de concentrado a partir de una perturbación específica.

Para lograr esta simulación del comportamiento dinámico de la variable de interés se establece como supuesto que la respuesta del proceso analizado se comporta como un sistema de primer orden, adicionalmente se considera que el tiempo de residencia de este sistema es conocido y es función de las variables de entrada al sistema (geométricas, operación y alimentación), adicionalmente en este sistema de primer orden se supone que existe un retardo por transporte.

4.3.1. Sistema de primer orden

De forma general un proceso recibe una cierta entrada $u(t)$ y termina generando una salida $y(t)$. Si llevamos estas señales al dominio de Laplace tendremos ahora una entrada $U(s)$ que genera una salida $Y(s)$, con una función que relaciona esta entrada con la salida, denominada función de transferencia $G(s)$, en el dominio s . De modo que, esquemáticamente tendremos (Figura 9):

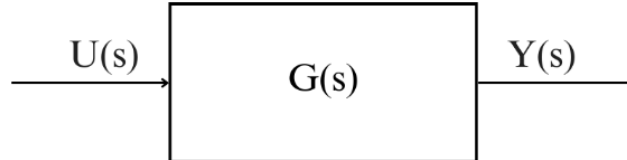


Figura 9. Esquema de sistema con entrada y salida

Se denominan sistemas de primer orden a aquellos en los que el sistema puede ser representado mediante una ecuación de primer orden, donde se reducen al siguiente formato:

$$\tau * \frac{dy(t)}{dt} + y(t) = k * u(t) \quad (4.1)$$

Donde $y(t)$, que corresponde a la variable salida, sería la variable de interés a predecir en esta investigación, es decir, la ley de concentrado de cobre. Mientras que, $u(t)$ corresponde a cualquiera de las variables de entrada presentadas previamente y que pueden ser modificadas mediante perturbaciones (variables fijas excluidas). Tomando las transformadas de Laplace se puede resolver

la ecuación previa, donde obtendremos finalmente la función de transferencia para un sistema de primer orden:

$$Y(s) = G(s) * U(s) \quad (4.2)$$

$$G(s) = \frac{Y(s)}{U(s)} = \frac{K}{\tau * s + 1} \quad (4.3)$$

$G(s)$ corresponde a una función de transferencia para un sistema de primer orden sin retardo por transporte. Por ende, por motivos de la investigación y los supuestos previamente expuestos, la función de transferencia sería:

$$G(s) = \frac{K}{\tau * s + 1} * \exp(-t_d * s) \quad (4.4)$$

Donde:

- $Y(s)$: Variable de salida del sistema
- $U(s)$: Variable de entrada al sistema
- K : Ganancia estática del sistema
- τ : Constante de tiempo del sistema
- t_d : Retardo por transporte

4.3.1.1. *Ganancia estática del sistema*

Se denomina ganancia estática de un sistema a la relación de ganancia entre la entrada y la salida del proceso. Es decir, cuando la entrada es constante (escalón) y la salida se estabiliza (régimen permanente), la razón del cambio en la salida provocado por la perturbación entre el salto escalón o cambio generado en la entrada (perturbación) nos da lo conocido como ganancia estática del sistema (Ecuación 4.5).

$$K = \frac{\Delta Y}{\Delta U} = \frac{Y_{\infty} - Y_i}{U_{\infty} - U_i} \quad (4.5)$$

De la ecuación anterior se puede deducir que la respuesta permanente o respuesta estacionaria corresponde al comportamiento de la salida del proceso o sistema cuando el tiempo tiende a infinito. En el marco de esta investigación, la perturbación o cambio introducido en la entrada puede estar

asociada a cualquiera de las variables de entrada no fijas consideradas en el proceso simulado. El efecto de dicha perturbación en la entrada se refleja en la salida como una variación en la variable de interés principal, es decir, la ley de concentrado. A modo de ejemplo, en la Ecuación 4.6 se presenta el caso particular en el que se genera una perturbación en la altura de espuma de la celda piloto, ilustrando cómo dicho cambio afecta directamente la respuesta del sistema.

$$K_{Ley_{Conc}-H_f} = \frac{Ley_{Conc_{\infty}} - Ley_{Conc_i}}{H_{f_{\infty}} - H_{f_i}} \quad (4.6)$$

Esta es la demostración de cómo es posible acoplar el simulador del comportamiento dinámico de la variable de interés (modelo de primer orden) y el código de simulación en estado estacionario, ya que el segundo se encarga de generar los resultados metalúrgicos a partir de las variables de entrada en el sistema, en un estado inicial (previo a la perturbación) y un estado final (después de la perturbación). Con lo que el simulador dinámico puede construir la trayectoria y, por ende, la respuesta transiente que sigue la variable de salida entre un estado y otro, a partir la variable de entrada perturbada y su magnitud.

4.3.1.2. *Constante de tiempo del sistema*

La constante de tiempo de un sistema de primer orden se define como el tiempo necesario para que el sistema alcance el 63,2 % del valor final o respuesta estacionaria, es decir, la salida cuando el tiempo tiende a infinito. Este parámetro representa la velocidad de respuesta del sistema frente a una perturbación en la entrada. En consecuencia, una constante de tiempo pequeña implica una respuesta rápida, mientras que una constante de tiempo grande refleja un sistema más lento en su dinámica transiente. La respuesta transiente corresponde al comportamiento dinámico del sistema desde el estado inicial hasta alcanzar el estado estacionario. En sistemas de primer orden, se considera que la duración de esta etapa equivale aproximadamente a cuatro veces la constante de tiempo.

Tal como se indicó previamente, en este trabajo se asume que la constante de tiempo depende de las variables de entrada del sistema. En una celda de flotación piloto, el tiempo de residencia puede modificarse principalmente a través de variaciones en el flujo, la concentración de sólidos de la alimentación y la altura de la zona de colección (Yianatos Bernardino, 2005). Este tiempo puede estimarse a partir de la siguiente ecuación:

$$\tau = \frac{A_c * H_c * (1 - \varepsilon_g)}{Q_F} \quad (4.7)$$

Donde:

- A_c : Área de la sección transversal de la celda
- H_c : Altura de la zona de colección
- ε_g : Concentración de aire (Hold up)
- Q_F : Flujo volumétrico de la pulpa en la alimentación

En el caso de la concentración de aire en una celda mecánica industrial, su valor típicamente varía entre un 5% y un 15% (Fuerstenau, Jameson, & Yoon, 2007). Para los fines de esta investigación, se utilizará una concentración del 10%, asumiéndose que este valor se mantiene constante durante todo el trabajo realizado.

4.3.1.3. Retardo por transporte

Como se observó previamente en la Figura 1, el proceso de flotación de forma simplificada puede ser dividido en dos zonas principales: la zona de colección, en el cual las partículas minerales se adhieren a burbujas dentro de la pulpa, y la zona de espuma, donde estas burbujas ascienden y se desplazan hacia el borde de la celda para su extracción. La zona de espuma en particular actúa como una etapa de retención temporal, por ende, existe un tiempo característico de permanencia antes de que sean evacuadas como concentrado las partículas.

Una vez que se aplica una perturbación al sistema, como un cambio en la alimentación o en el flujo superficial de aire, se genera un gradiente de concentración vertical en la zona de espuma. Esto se debe a que las burbujas recién ingresadas a la zona de espuma desde la zona de colección contienen una composición diferente a las que ya se encuentran próximas a la descarga. Por tanto, si se midiera la concentración en dos puntos de la zona de espuma, uno cercano al ingreso desde la pulpa y otro cerca de la canaleta de recolección, se observaría un desfase temporal entre ambas señales. Este desfase se interpreta como un retardo por transporte, ya que refleja el tiempo necesario para que los agregados recorran la zona de espuma desde su entrada hasta su salida (Rojas Santibañez, 2023).

Con esto se puede simular la ley de concentrado de cobre obtenido en la celda de flotación *rougher* piloto aplicando un modelo de primer orden con retardo por transporte, utilizando ya los tres supuestos previamente planteados. Este tiempo para un equipo de flotación se considera como el tiempo que le toma al agregado desplazarse desde la entrada hasta la salida de la zona de espuma. El transporte de esta espuma desde su entrada hacia su salida considera dos componentes: el tiempo que tarda la espuma en desplazarse verticalmente desde la base de la zona de espuma (interfaz pulpa-espuma) hasta el nivel del borde del canal de concentrado y el tiempo que tarda la espuma en

desplazarse horizontalmente hasta el borde de la celda de flotación, tal como se puede ver en la Figura 10.

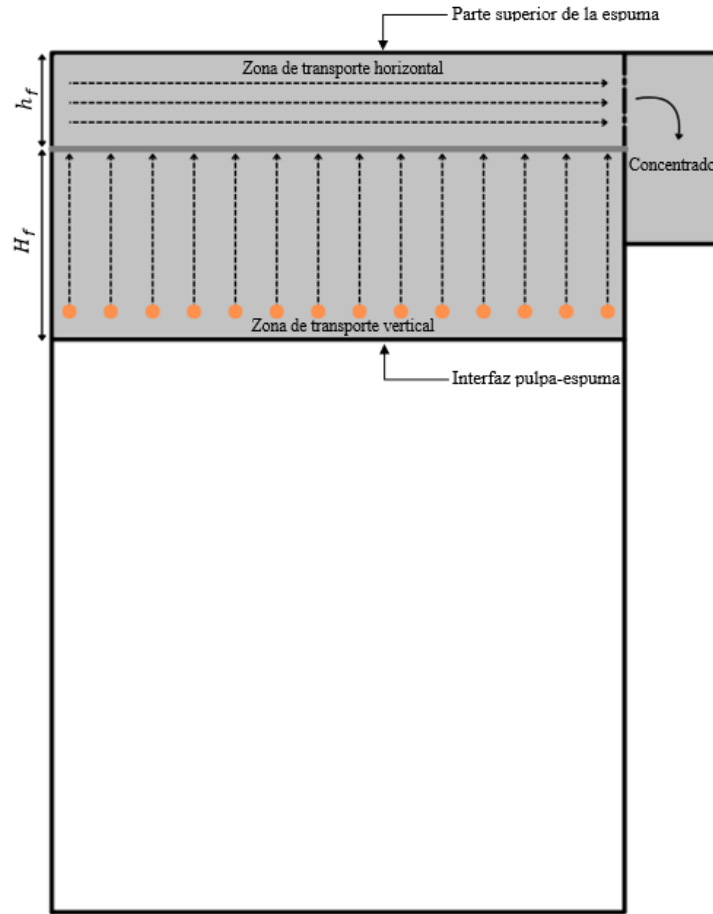


Figura 10. Zonas de transporte de espuma en celda de flotación piloto

Este modelo a utilizar supone que las burbujas de aire y las partículas adheridas que entran en la fase de espuma ascienden verticalmente hasta alcanzar la altura del borde, en este caso 40 [cm] según se demuestra en la Figura 6. En este modelo el tiempo de retardo generado por la fase de espuma en la celda piloto puede ser expresado como:

$$t_d = t_{fv} + t_{fh} \quad (4.8)$$

En el caso de la componente vertical, el tiempo se determina a partir de la velocidad de ascenso vertical de la espuma y la altura de la espuma:

$$t_{fv} = \frac{H_f}{v_{fv}} = \frac{H_f * \varepsilon_f}{J_g} \quad (4.9)$$

Donde:

- H_f : Altura de la espuma comprendida entre la interfaz pulpa-espuma y el borde de descarga de la canaleta de concentrado (Figura 10)
- v_{fv} : Velocidad de ascenso vertical de la espuma
- J_g : Velocidad superficial del aire
- ε_f : Concentración de aire en la fase de espuma

Para efectos de esta investigación, como se está trabajando en una celda de flotación piloto de pequeña escala de un laboratorio, se realiza el supuesto de que el componente horizontal del transporte de espuma es insignificante en comparación con el componente vertical anteriormente indicado, por ende, el tiempo total de retardo o el tiempo total medio corresponde a lo indicado en la Ecuación 4.10 (Zheng, Franzidis, & Manlapig, 2004).

$$t_d = t_{fv} = \frac{H_f * \varepsilon_f}{J_g} \quad (4.10)$$

En el caso de la concentración de aire en la fase de espuma (ε_f), en una celda de flotación convencional, sus valores típicos oscilan entre el 97% y el 80%, para efectos de esta investigación se trabajará con un valor de 90% (Schwarz & Alexander, 2006).

Con todo lo indicado previamente, para poder obtener la respuesta transiente del sistema al realizar una perturbación en las variables de entrada se utiliza la siguiente expresión:

$$y(t) = A_i * K_i * \left(1 - e^{-\frac{(t-t_d)}{\tau}}\right) \quad (4.11)$$

Al momento de realizar los experimentos, se aplicaron múltiples perturbaciones de tipo escalón a cada una de las variables de entrada. En este contexto, la expresión previamente presentada adquiere una relevancia fundamental, ya que permite determinar el comportamiento de la variable de interés (ley de concentrado de cobre) a partir de datos generados por un código de simulación en estado estacionario.

El acoplamiento entre el simulador estático (código en Matlab) y el simulador dinámico (sistema de primer orden) posibilita la generación de datos que sirven de base para la construcción y ajuste del sensor virtual aplicado a la celda piloto. Si bien el proceso real puede no presentar un comportamiento idéntico al descrito en todo su rango operacional, esta aproximación resulta suficiente para los objetivos de la investigación. En efecto, el alcance de este trabajo se centra en el desarrollo metodológico de un sensor virtual, el cual puede ser replicado de igual manera en caso de disponer de datos de un proceso real.

Como complemento a lo explicado previamente, en la Figura 11 se presenta un esquema que ilustra el acoplamiento entre el código de simulación estático y el simulador dinámico de primer orden. Dicho esquema muestra cómo, a partir de los resultados metalúrgicos estacionarios obtenidos para un estado inicial y otro final, generado por la perturbación de una variable de entrada en estado estacionario, el simulador dinámico de primer orden estima los parámetros del modelo (t_d , τ y K). Con esta información es posible determinar la trayectoria transiente que sigue la variable de interés desde el estado inicial hasta alcanzar el estado final.

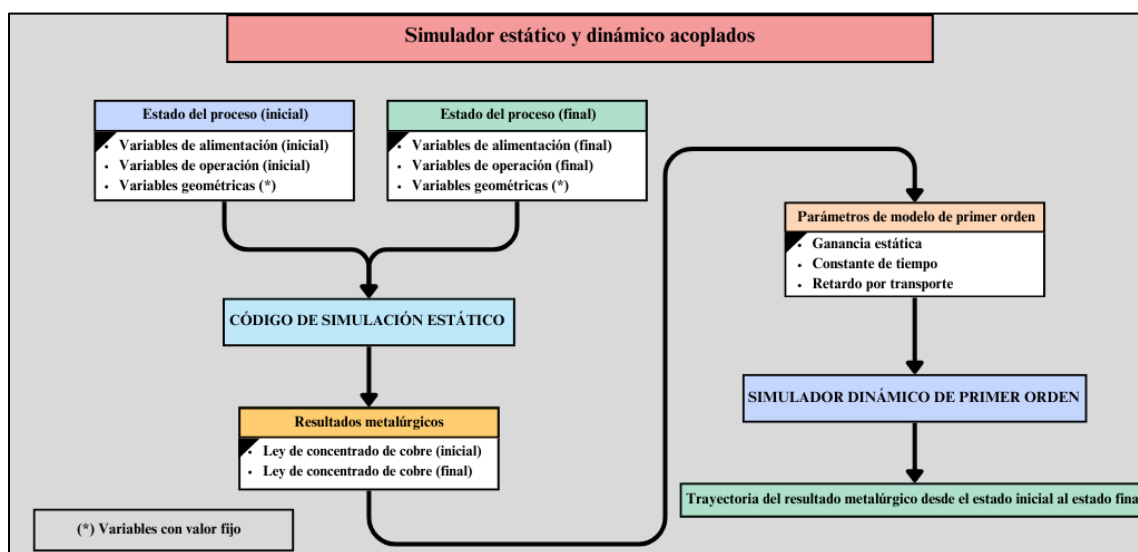


Figura 11. Esquema de simulador estático y dinámico integrados

4.4. Diseño de experimentos

El propósito de esta sección es generar un conjunto de datos que imite de manera realista la operación de la celda de flotación piloto. Para llevar a cabo esto, se utiliza el simulador integrado, el cual permitirá, a partir de una serie de perturbaciones, describir la trayectoria que sigue la variable de interés entre cada estado y su respectiva perturbación. Estas perturbaciones se aplicarán de acuerdo con las variables disponibles previamente indicadas, así como con los niveles que estas presenten.

A continuación, se detallan las variables cuyos valores pueden ser manipulados en la investigación.

4.4.1. Selección de variables y sus niveles

Para asegurar que los datos experimentales representen la mayoría de las zonas de operacionales, se realiza un diseño de experimentos que incluya cada una de las variables previamente indicadas en el código de simulación y que presenten la opción de perturbarla. En este caso, tal como se indicó previamente, estas variables son la ley de cobre en la alimentación, el flujo volumétrico de pulpa alimentado en la celda, el porcentaje de solidos en la alimentación, el porcentaje de gruesos en la alimentación. Adicionalmente se agregan las variables que modifican la operación de la celda, en este caso el flujo superficial de aire y la altura o profundidad de espuma en la celda, lo que termina combinando, las variables definidas a partir de los datos experimentales del sistema de control en la celda de flotación piloto, con las variables definidas de manera virtual en el código de simulación.

En total, se considerarán seis variables que caracterizan tanto la alimentación como la operación de la celda piloto en esta investigación. A estas variables se establecieron tres niveles para cada una de las variables, de modo que cada perturbación tenga un impacto visible en la ley de concentrado y que permita generar estados intermedios en las condiciones de operación, lo cual disminuirá la incertidumbre del sensor virtual que se busca desarrollar. Los niveles de cada una de las variables y su respectiva nomenclatura se presentan a continuación en la Tabla 5.

Tabla 5. Matriz de variables y sus niveles

Variable	Nomenclatura	Nivel			Unidad
		(1)	(2)	(3)	
Flujo de alimentación de pulpa	Q_F	3	3,5	4	[L/min]
Porcentaje de solidos	S_f	20	30	40	(%)
Ley de cobre en alimentación	X_f	0,3	0,65	1	(%)
Porcentaje de gruesos en alimentación	G_f	5	12,5	20	(%)
Velocidad superficial del aire	J_g	0,5	0,75	1	[cm/s]
Altura de espuma	H_f	1	2,5	4	[cm]

Tal como se indicó anteriormente, las dimensiones físicas de la celda piloto utilizada en la investigación (altura, largo, ancho y volumen), así como su forma geométrica, se mantienen constantes en todas las simulaciones realizadas con el código. Además de estas características

estructurales, se consideran otras variables asociadas a la alimentación, las cuales también se mantendrán fijas a lo largo de toda la investigación.

Todas estas variables de valor constante se resumen en la Tabla 6. Por otro lado, las especificaciones geométricas y constructivas de la celda estudiada se derivan de lo mostrado en la Figura 6.

Tabla 6. Variables fijas de entrada

Variable	Valor	Unidad
Altura de la celda	45	[cm]
Largo de la celda	30	[cm]
Ancho de la celda	30	[cm]
Área superficial de la celda	900	[cm ²]
Volumen de la celda	40.500	[cm ³]
Tipo de celda	Rectangular	-
Ley de fierro en alimentación	1,5	(%)
Ley de molibdeno en alimentación	-	(%)
Densidad del sólido	3,8	[kg/L]
Diámetro de finos de partícula en alimentación	45	[μ m]
Diámetro máximo de partícula en alimentación	255	[μ m]

4.4.2. Matriz de experimentos

Para generar los datos experimentales utilizados en esta investigación, se realizaron perturbaciones aleatorias con el objetivo de cubrir la totalidad de combinaciones posibles definidas por la matriz de experimentos. Dado que el código de simulación que entrega los resultados metalúrgicos está implementado en un entorno como MATLAB, fue posible generar todas las combinaciones de estados de las variables de entrada de forma rápida y eficiente. En este caso, considerando seis variables de entrada, cada una con tres niveles, se obtuvieron un total de $3^6 = 729$ combinaciones únicas u observaciones.

Además de la ley de concentrado de cobre, el modelo entrega otros resultados metalúrgicos relevantes, como la ley de colas y la recuperación de cobre. No obstante, dada la gran cantidad de combinaciones generadas, muchas de las observaciones presentan recuperaciones poco coherentes desde el punto de vista operativo. Por ejemplo, valores de recuperación cercanos al 80% no reflejan condiciones realistas en una celda de flotación mecánica operando de forma aislada.

Con el fin de asegurar la coherencia y representatividad de los datos, se aplicó un filtro basado en la recuperación de cobre, restringiendo las observaciones a un rango considerado operacionalmente aceptable. Se establecieron límites entre un mínimo de 52,5% y un máximo de 72,5% de recuperación, valores que representan condiciones normales de operación y que permiten reducir la incertidumbre en los resultados posteriores. Tras aplicar este filtro, se conservaron 398 observaciones de las 729 originales.

En este contexto, y siguiendo el marco de construcción de la matriz de experimentos, fue necesario revisar en detalle los supuestos considerados en la formulación de la Ecuación 4.11, particularmente en lo relativo al efecto del retardo por transporte (t_d). Este parámetro, si bien es relevante en ciertos sistemas dinámicos, debía ser evaluado para determinar su influencia real en el comportamiento del sistema estudiado. Para ello, se comparó el valor de retardo más alto dentro del subconjunto de observaciones filtradas con la constante de tiempo (τ) más baja registrada en ese mismo conjunto. El análisis mostró que, incluso en el escenario donde el retardo por transporte podría ejercer su mayor impacto en la Ecuación 4.11, este valor apenas representa un 1,65% de la constante de tiempo. Esta relación tan reducida indica que su efecto es prácticamente despreciable.

En consecuencia, se decidió no incorporar este término en el desarrollo del modelo, considerando un tiempo de retardo igual a cero en la construcción de la matriz de experimentos. Esta simplificación no compromete la validez del análisis posterior y, al mismo tiempo, permite enfocar el estudio en los factores que realmente condicionan la respuesta del sistema.

A partir de las observaciones filtradas y de la no consideración del retardo en el modelo, se construyó una ruta de perturbación aleatoria diseñada para aislar el efecto de cada variable de entrada individualmente y obtener la respuesta transiente del sistema. Tras este procedimiento, se conservaron 323 observaciones de las 398 ya filtradas, configurando así una dinámica más representativa del comportamiento real de una celda de flotación individual. Los pasos a seguir fueron los siguientes:

- 1) **Selección de un estado inicial (estado 1):** Se elige aleatoriamente una combinación de valores de las seis variables de entrada disponibles.
- 2) **Perturbación aleatoria de una variable:** Se selecciona una de las seis variables de manera aleatoria. Esta puede aumentar o disminuir uno o dos niveles, siempre respetando los límites impuestos (niveles 1 a 3). Por ejemplo:
 - Si la variable está en nivel 3, solo puede disminuir
 - Si está en nivel 1, solo puede aumentar

- 3) **Cálculo de la respuesta transiente:** Se estima la evolución de la ley de concentrado al cambiar del estado inicial (1) al nuevo estado (2), utilizando la Ecuación 4.11. Esta ecuación se evalúa en un intervalo de tiempo que va desde 0 hasta 4τ , donde τ es la constante de tiempo más grande entre las observaciones filtradas. Para este estudio, el valor máximo encontrado fue $\tau_{m\acute{a}x} = 10,53$ minutos, coherente con los tiempos de residencia típicos en este tipo de celdas.

La evaluación de la respuesta transiente se realiza en intervalos de $\tau = 10,53$ minutos, garantizando así una cantidad uniforme de observaciones en cada transición.

- 4) **Cobertura de la respuesta estacionaria:** Este intervalo de 4τ permite alcanzar al menos un 98,2% del valor de la ley de concentrado en estado estacionario, lo cual es suficiente para considerar que se ha alcanzado el nuevo equilibrio. En los casos en que la constante de tiempo es menor a 10,53 minutos, este porcentaje será incluso superior.
- 5) **Registro de resultados:** Los datos generados se almacenan como una serie de tiempo, incluyendo:
- Las seis variables de entrada (alimentación y operacionales)
 - Los tres resultados metalúrgicos (ley de colas, ley de concentrado y recuperación de cobre)
- 6) **Actualización del estado inicial:** El estado final (2) se convierte en el nuevo estado inicial (1), y se repite el proceso desde el paso 2.

4.4.3. Set de datos

Una vez aplicadas todas las perturbaciones posibles sobre las observaciones filtradas, según el procedimiento descrito en la sección anterior, se generaron resultados que integran tanto el comportamiento estático como el dinámico del sistema. Estos resultados se presentan en la Figura 12, donde se muestra la evolución de las variables de entrada y la ley de concentrado como respuesta a las perturbaciones realizadas. El gráfico permite analizar cómo varía el sistema frente a cambios controlados en una sola variable, reflejando la respuesta dinámica estimada a partir de la Ecuación 4.11. Finalmente, los datos sin procesar que dieron origen a estos resultados se encuentran disponibles en formato Excel en el Anexo digital 1.

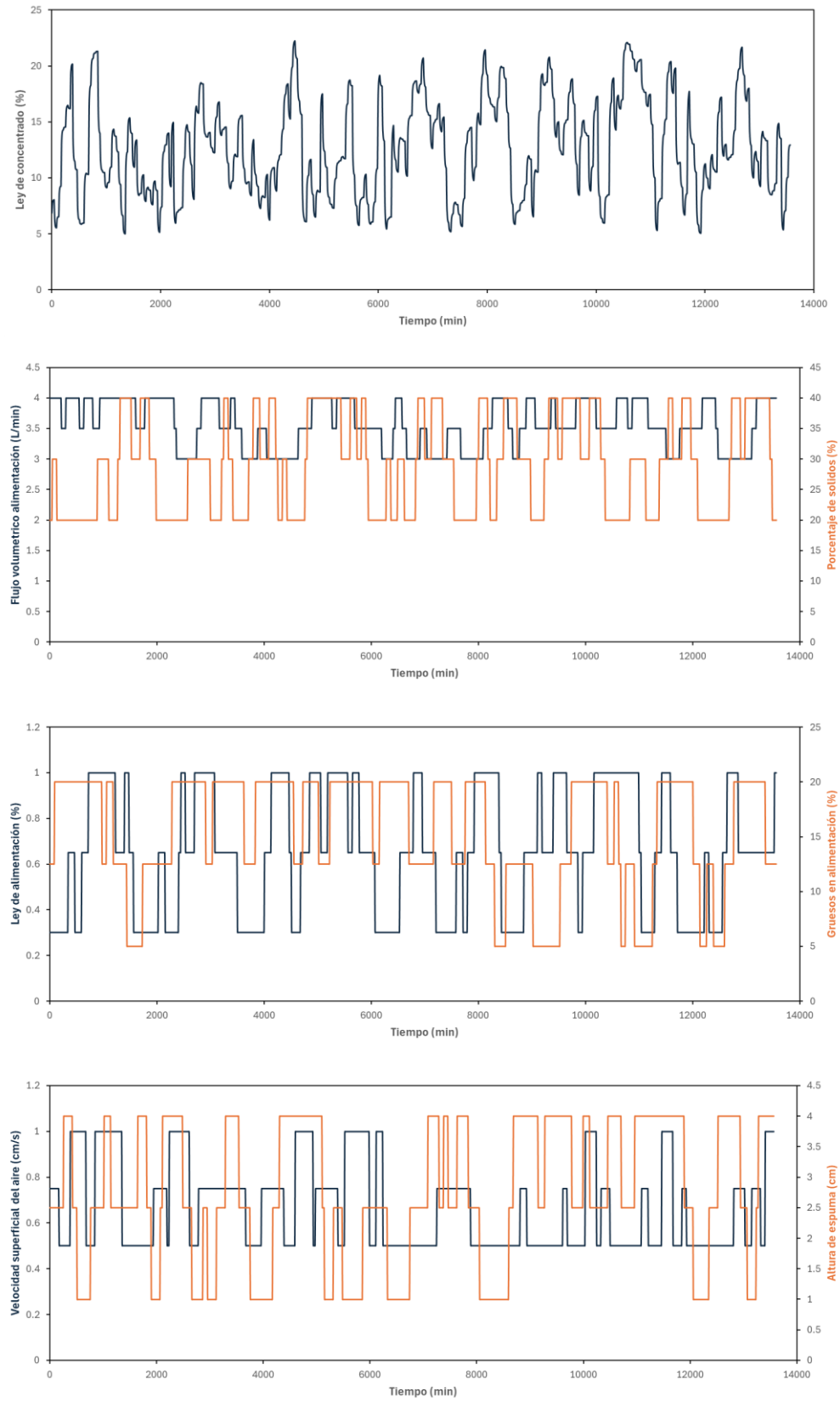


Figura 12. Set de datos generado a partir de simulador (dinámico más estacionario).

Junto con esta visualización, se presenta un análisis preliminar del conjunto de datos dinámicos en la Tabla 7. Este análisis incluye indicadores estadísticos básicos como la media y la desviación estándar de cada variable, además del índice de correlación lineal entre la ley de concentrado (X_c) y las variables de entrada. Este índice sirve como medida de dependencia lineal entre dos variables, con valores que oscilan entre -1 y 1. Cuanto mayor sea la magnitud del valor, mayor será la dependencia entre ambas variables. El signo, por su parte, indica si la relación es directamente proporcional (positivo) o inversamente proporcional (negativo).

Tabla 7. Resultados preliminares de la investigación.

Variable	Nomenclatura	Promedio	Desviación estándar	Índice de correlación	Unidades
Flujo de alimentación de pulpa	Q_f	3,56	0,40	-0,12	[L/min]
Porcentaje de solidos	S_f	28,94	8,20	-0,34	(%)
Ley de cobre en alimentación	X_f	0,65	0,28	0,55	(%)
Porcentaje de gruesos en alimentación	G_f	15,57	5,32	-0,06	(%)
Velocidad superficial del aire	J_g	0,67	0,20	-0,57	[cm/s]
Altura de espuma	H_f	2,74	1,17	0,11	[cm]
Ley de cobre en concentrado	X_c	12,65	4,34	1,00	(%)

Del análisis de estos coeficientes se observa que la ley de cobre en alimentación (X_f) presenta la mayor correlación positiva con la ley de concentrado (X_c), con un valor de 0,55. Este resultado indica una relación directamente proporcional moderadamente fuerte, lo que sugiere que aumentos en la ley de alimentación tienden a reflejarse en un incremento en la ley de concentrado. Por otro lado, la velocidad superficial del aire (J_g) muestra una correlación negativa significativa de -0,57, lo que indica que valores más altos de esta variable se asocian con una disminución en la ley de cobre en el concentrado. En contraste, el porcentaje de sólidos (S_f) presenta una correlación negativa débil con X_c , mientras que el porcentaje de gruesos (G_f), el flujo de alimentación (Q_f) y la altura de espuma (H_f) exhiben valores de correlación muy bajos, cercanos a cero. Esto sugiere que, en el rango de

operación considerado, perturbaciones en estas tres variables tienen un impacto limitado sobre la ley de concentrado, al menos desde una perspectiva lineal.

4.4.4. Modelo PLS dinámico

4.4.4.1. *Variables de entrada y de salida del modelo*

Como fue explicado en puntos previos, las variables de entrada utilizadas en el desarrollo del modelo PLS dinámico en esta investigación corresponden a las variables de alimentación y variables de operación, detalladas en la Tabla 5. En cuanto a las variables de salida del modelo, para efectos de esta investigación se considerará únicamente la ley de concentrado de cobre, ya que corresponde a la variable de interés en la construcción del sensor virtual. Por lo tanto, al contar con una única variable de respuesta, el modelo generado será un modelo PLS1 dinámico, es decir, un modelo que trabaja con múltiples variables predictoras, pero solo una variable de salida. En este caso, se utilizarán seis variables predictoras y una única variable de respuesta.

Con base en lo anterior, se definen los bloques o matrices de datos necesarios: la matriz de variables predictoras (X) y la matriz de respuesta (Y). Para ser utilizadas en el programa que permite realizar la regresión PLS, esta información debe organizarse de acuerdo con lo explicado previamente en la sección teórica: cada fila representa una observación o muestra, y cada columna corresponde a una de las variables predictoras presentes en el sistema (ver Tabla 8). Cabe destacar que, a pesar de los valores de los índices de correlación presentados en la Tabla 7, todas las variables de entrada serán incluidas como predictoras. Esto se debe a que, aunque algunas de ellas presenten baja o nula correlación con la variable de interés, se desconoce el efecto combinado o las interacciones que podrían tener en el sistema. Por esta razón, se incluyen todas las variables, considerando que el propio método PLS determinará posteriormente cuáles tienen una incidencia significativa en el modelo y descartará aquellas que considere poco relevantes.

Respecto a la frecuencia de las observaciones o muestras, se ha definido que, con el fin de otorgar mayor realismo al proceso modelado, esta será de 10,53 minutos, valor previamente determinado como representativo del sistema. Así, entre cada perturbación existirán un total de cuatro muestras: el estado inicial, el estado final y dos mediciones intermedias. Esta frecuencia se considera adecuada dada la dificultad de obtener mediciones frecuentes de ciertas variables en un entorno real, por lo que se estima que un intervalo de 10,53 minutos refleja adecuadamente la dinámica de un proceso de flotación en celdas mecánicas.

Tabla 8. Sets de datos a utilizar en el método PLS.

Matriz de variables predictoras (X)	Matriz de variables de respuesta (Y)
Flujo de alimentación de pulpa (Q_F)	Ley de cobre en concentrado (X_C)
Porcentaje de solidos (S_f)	
Ley de cobre en alimentación (X_f)	
Porcentaje de gruesos en alimentación (G_f)	
Velocidad superficial del aire (J_g)	
Altura de espuma (H_f)	

4.4.4.2. Programa SOLO

El desarrollo del sensor fue realizado mediante el *software* SOLO, desarrollado por Eigenvector Research, Inc., una herramienta especializada en el análisis de datos multivariados, quimiometría y aprendizaje automático (Eigenvector Research Inc., s.f.). Este programa está diseñado para asistir tanto a investigadores como a profesionales en la extracción de información significativa y en la construcción de modelos predictivos a partir de conjuntos de datos complejos, especialmente aquellos que provienen de experimentos o procesos industriales.

SOLO permite trabajar con grandes volúmenes de datos que incluyen múltiples variables, facilitando la identificación de relaciones y estructuras ocultas en los datos. Asimismo, permite reducir la complejidad de estos conjuntos para hacerlos más interpretables, identificar las variables que realmente influyen en un proceso y, en consecuencia, tomar decisiones mejor fundamentadas a partir del análisis.

Entre sus funcionalidades principales se incluyen métodos estadísticos como el Análisis de Componentes Principales (PCA), utilizado para la exploración de datos y la reducción de dimensionalidad; y Mínimos Cuadrados Parciales (PLS), aplicado en el modelado de regresión. Además, cuenta con algoritmos de clasificación como PLS-DA y SIMCA, que permiten identificar patrones y clasificar observaciones dentro de los datos. El *software* también incorpora potentes herramientas de preprocesamiento de datos, fundamentales para la limpieza y normalización de la información, así como avanzadas capacidades de visualización, que permiten interpretar gráficamente los resultados obtenidos.

Para efectos de esta investigación, se utilizará el análisis de Mínimos Cuadrados Parciales (PLS) disponible en este *software*, lo cual se puede observar en la Figura 13. Este análisis sigue constantemente la misma estructura lógica e intuitiva. La estructura lógica que se sigue es:

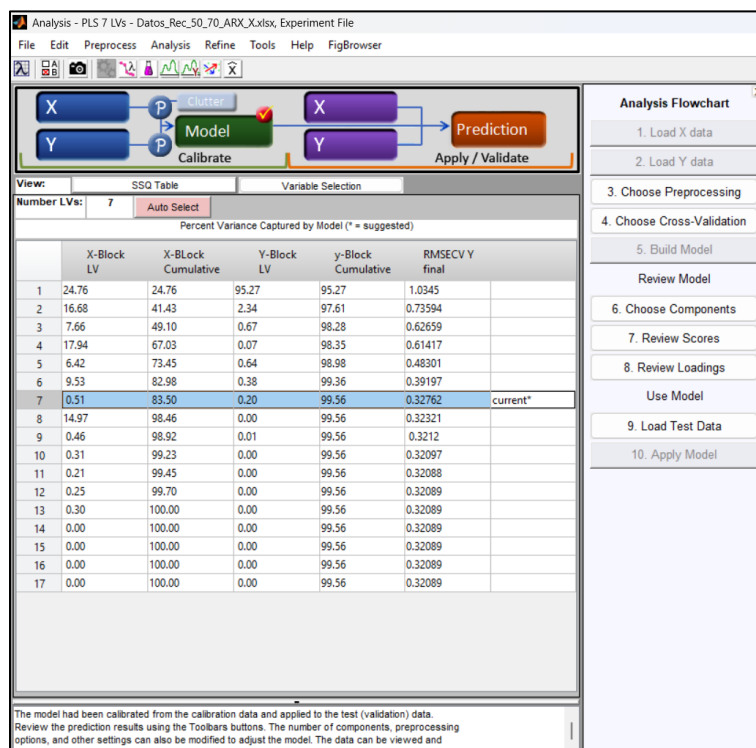


Figura 13. Interfaz de análisis PLS programa SOLO.

- Cargar datos experimentales:** El proceso comienza con la carga de la matriz de variables predictoras (X), que contiene la información de entrada del sistema, seguida por la matriz de variables de respuesta (Y), que representa los resultados o valores que se desean predecir. Cada matriz, como se informó previamente, contiene en cada fila una de las muestras realizadas y en sus columnas las variables consideradas en el modelo.
- Preprocesamiento de los datos:** El *software* cuenta con una serie de herramientas para realizar un preprocesamiento a los datos cargados, como el auto escalado o la transformación de los datos, entre otros. Como fue mencionado en secciones previas, para efectos de esta investigación únicamente se aplicará un auto escalado a los datos, el cual consiste en hacer que todos los datos presenten una media de cero y una desviación estándar igual a la unidad, y que corresponde a un requisito obligatorio para aplicar adecuadamente la metodología PLS.
- Validación cruzada:** SOLO incorpora técnicas de validación cruzada que son esenciales para evaluar la robustez y la capacidad predictiva de los modelos. En la validación cruzada, el conjunto de datos se divide en subconjuntos. El modelo se entrena en una parte de los datos (conjunto de

calibración o entrenamiento) y se prueba en la parte restante (conjunto de validación o prueba). Esto ayuda a obtener una estimación más fiable del rendimiento del modelo con datos nuevos y a evitar el sobreajuste. Como esta investigación considera grandes números de observaciones y, por ende, el conjunto de datos es bastante grande, comúnmente para estos casos se utiliza una proporción del 70% para la calibración y solo un 30% para la validación del modelo, lo cual se encuentra dentro de los rangos típicos de proporciones para cada subconjunto (Hastie, Tibshirani, & Friedman, 2009).

- d) **Construcción del modelo:** Luego de completar los tres pasos previos, se selecciona la opción “*Build Model*” en el diagrama de flujo del análisis PLS. Al hacerlo, el *software* ejecuta el algoritmo interno para la construcción del modelo. Una vez construido, es posible visualizar, mediante una tabla (Figura 13), la varianza o información que cada componente retiene o captura, así como el porcentaje acumulado de esta varianza tanto para las matrices X como Y. Adicionalmente, en la última columna de la tabla se presenta el error cuadrático medio obtenido mediante validación cruzada. Finalmente, el *software* entrega una recomendación sobre el número óptimo de componentes que debería incluir el modelo, la cual puede ser modificada según el método o criterio de selección empleado.
- e) **Selección del número de componentes:** La opción de selección del número de componentes permite desplegar diversos gráficos e información complementaria que facilitan al usuario la determinación del número óptimo de componentes a utilizar en el modelo. Tal como se expuso en la sección teórica de esta investigación, esta elección depende enteramente del criterio adoptado por el usuario, siendo todos ellos válidos desde el punto de vista metodológico. Entre los criterios más comunes se encuentran el de varianza acumulada explicada y el criterio de Kaiser, que selecciona aquellos componentes con valores propios mayores a uno. Una vez definido el número de componentes, es posible realizar diversos análisis adicionales, tales como la interpretación de los *scores* y *loadings*, la evaluación del comportamiento de las predicciones frente a los valores medidos, el análisis de residuos y la revisión de los estadígrafos Q y T^2 . Finalmente, el programa SOLO permite extraer los coeficientes del modelo en forma de un vector. Este vector, en conjunto con los valores pasados de las variables de entrada y de salida (en el caso del modelo ARX), permiten estimar el estado actual de la variable de interés.

4.5. Modelo FIR

4.5.1. Bloques de datos X e Y

Tal como se explicó previamente, el modelo FIR (*Finite Impulse Response*) se enfoca en describir cómo la salida actual de un proceso corresponde a una suma ponderada de las variables de entrada pasadas hasta un punto de tiempo definido. Esto lo convierte en una herramienta particularmente útil para analizar el efecto que una perturbación puntual en una variable de entrada ejerce sobre la salida del sistema a lo largo del tiempo.

En este estudio, los datos obtenidos de la simulación bajo la ruta de perturbación aleatoria consideran tiempos de residencia en la celda que fluctúan entre 7,29 y 10,53 minutos, calculados según la Ecuación 4.7 y dependientes de las variables de entrada. No obstante, para mantener un marco de análisis estandarizado, se definió un intervalo fijo de 10,53 minutos entre observaciones, independiente del tiempo de residencia instantáneo. Asimismo, entre perturbaciones se estableció un tiempo equivalente a cuatro veces la constante de tiempo más alta ($4 \times 10,53 = 42,12$ minutos), asegurando que cada respuesta dinámica sea comparable y completa dentro del mismo horizonte temporal.

Las observaciones obtenidas a partir de la ruta de perturbación aleatoria evidencian que los efectos asociados a cada perturbación, expresados como $g_i * A_i$ o equivalentemente $K_i * A_i$ (considerando $K_i = g_i$), varían de manera significativa según la variable de entrada modificada, ya sea de alimentación u operacional. Esto provoca modificaciones de distinta magnitud en la ley de concentrado, lo que permite evidenciar empíricamente el impacto de cada variable de entrada en la salida del sistema. Junto con la variación en los tiempos de residencia, este comportamiento confirma la naturaleza no lineal del proceso, la cual se busca representar posteriormente mediante la combinación lineal de variables pasadas a través de las metodologías PLS-FIR y PLS-ARX. Para obtener la respuesta transiente del sistema frente a una perturbación, se utiliza la siguiente expresión:

$$y(t) = A_i * g_i * \left(1 - e^{-\frac{(t)}{\tau}}\right) \quad (4.12)$$

En cada perturbación realizada se observan no solo distintos tiempos de residencia, sino también diferentes efectos $g_i * A_i$ o $K_i * A_i$, lo que refleja las diversas formas en que cada variable de entrada modifica la respuesta del sistema. En particular, el flujo de alimentación de pulpa (Q_F) presenta efectos ($g_i * A_i$) entre -0,34 % y 0,27 %; el porcentaje de sólidos (S_f) entre -3,42 % y 3,44 %; la ley

de cobre en la alimentación (X_f) entre $-5,29\%$ y $5,42\%$; el porcentaje de gruesos (G_f) entre $-1,41\%$ y $1,38\%$; la velocidad superficial del aire (J_g) entre $-9,35\%$ y $9,18\%$; y la altura de espuma (H_f) entre $-1,30\%$ y $2,23\%$.

Estos rangos reflejan el impacto diferenciado de cada variable en la ley de concentrado, mostrando cómo la salida se ve modificada de manera particular según la variable intervenida. De este modo, cada perturbación genera trayectorias transientes únicas, cuya forma depende directamente de la variable modificada y de su influencia sobre la variable de interés del sistema.

Por otra parte, la normalización de los intervalos temporales tiene como objetivo simplificar y estandarizar el desarrollo de los modelos considerados en este trabajo. En concreto, se garantiza que la respuesta dinámica del sistema frente a una perturbación se capture completamente dentro del intervalo de 42,12 minutos, lo cual equivale a 4 veces la constante de tiempo más alta utilizada ($\tau_{m\acute{a}x} = 10,53$ minutos). Bajo este esquema, se asegura que como mínimo se alcance el 98,2% de la respuesta final ante una perturbación, incluso en el peor de los casos ($\tau_{m\acute{a}x} = 10,53$ minutos). En aquellos escenarios donde el tiempo de residencia sea menor, se alcanzarán porcentajes incluso superiores de la respuesta final.

Con base en lo anterior, se definió de manera preliminar un total de 4 *lags* o retrasos para el modelo FIR, lo cual permite capturar adecuadamente la dinámica del sistema desde el estado inicial hasta el estado final inducido por una perturbación.

Posteriormente, se evaluaron modelos adicionales con 2, 3, 5, 6, 7 y 8 *lags* con el propósito de validar dicha elección inicial. Esta evaluación se realizó utilizando el *software* SOLO, empleando como criterio principal el error medio cuadrático de validación cruzada (RMSECV), métrica que permite cuantificar la capacidad predictiva del modelo sobre datos no utilizados en la etapa de calibración y que es entregada directamente por el *software* para cada configuración analizada.

Los resultados obtenidos indicaron que el modelo más adecuado corresponde al que utiliza un total de 6 *lags*, ya que la incorporación de un mayor número de retardos (7 y 8 *lags*) no generó reducciones significativas del RMSECV, evidenciando una ganancia marginal en el desempeño predictivo. En contraste, una menor cantidad de *lags*, incluyendo el caso preliminar de 4 *lags*, provocó un incremento apreciable del RMSECV, reflejando una pérdida considerable de precisión en la estimación de la ley de concentrado de cobre.

Una vez definido el número de retardos a utilizar, se procede a establecer la composición de la matriz o bloque de datos X, la cual, para este modelo, considera únicamente valores pasados de las variables

de entrada. En este contexto, cada una de las observaciones $u(t)$ generadas incluye las siguientes variables de entrada:

$$u(t_1) = [Q_f(t_1), S_f(t_1), X_f(t_1), G_f(t_1), J_g(t_1), H_f(t_1)] \quad (4.13)$$

Finalmente, se establece que la matriz de datos X, tal como se describe en la Ecuación 3.23, se construye a partir de los valores pasados de las variables de entrada. En este estudio, se consideraron un total de 6 retardos (*lags*), con un intervalo fijo de $\tau_{m\acute{a}x} = 10,53$ minutos entre cada observaci3n. Esto da lugar a la siguiente estructura matricial para X:

$$X_{FIR} = \begin{bmatrix} u(t_1 - \tau_{m\acute{a}x}) & u(t_1 - 2\tau_{m\acute{a}x}) & u(t_1 - 3\tau_{m\acute{a}x}) & \cdots & u(t_1 - 6\tau_{m\acute{a}x}) \\ u(t_2 - \tau_{m\acute{a}x}) & u(t_2 - 2\tau_{m\acute{a}x}) & u(t_2 - 3\tau_{m\acute{a}x}) & \cdots & u(t_2 - 6\tau_{m\acute{a}x}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u(t_m - \tau_{m\acute{a}x}) & u(t_m - 2\tau_{m\acute{a}x}) & u(t_m - 3\tau_{m\acute{a}x}) & \cdots & u(t_m - 6\tau_{m\acute{a}x}) \end{bmatrix} \quad (4.14)$$

Por su parte, la matriz de datos Y considerada para este modelo es id3ntica a la presentada en la Ecuaci3n 3.24, correspondiendo a la 3nica variable de salida: la ley de concentrado de cobre. En cuanto a la matriz X, esta posee un tama1o de 1.283 x 36, es decir, contiene 1.283 observaciones y un total de 36 variables. En el caso de la matriz Y, se cuenta con las mismas 1.283 observaciones y una 3nica variable. Una vez ingresadas las matrices X e Y en el *software*, se procede a realizar el preprocesamiento (auto escalado) y la validaci3n cruzada con el fin de construir el modelo.

4.5.2. N3mero de componentes del modelo

Una vez que se han ingresado los bloques X e Y en el *software*, se procede a seguir los pasos descritos previamente en la Secci3n 4.4.4.2. En esta etapa, es necesario determinar la cantidad de componentes principales que formar3n parte del modelo FIR final. En este trabajo, la selecci3n del n3mero de componentes se basa exclusivamente en el criterio de varianza acumulada explicada. Se opta por el n3mero m3nimo de componentes que permitan explicar al menos el 80% de la varianza total de los datos.

No obstante, aunque este criterio es la base principal para la selecci3n, no resulta suficiente por s3 solo. Es imprescindible adem3s verificar que los datos se mantengan dentro de l3mites aceptables definidos por los estad3grafos Q y T^2 , los cuales permiten identificar y eliminar posibles observaciones an3malas que podr3an afectar la calidad del ajuste del modelo.

Una vez considerados los criterios previamente mencionados, se determina que el número óptimo de componentes principales a utilizar en el modelo FIR es igual a 7, los cuales explican en conjunto un 80,08% de la varianza total. Esta información se encuentra contenida en la tabla resumen generada por el *software* al momento de construir el modelo, tal como se muestra en el ejemplo ilustrado en la Figura 13. Dicha tabla presenta detalles relevantes como la varianza explicada por cada componente en los bloques X e Y, su correspondiente acumulado, así como la Raíz del Error Cuadrático Medio de la Validación Cruzada (RMSECV).

Toda esta información específica para el modelo FIR desarrollado en este trabajo se presenta en la Tabla 9.

Tabla 9. Resumen estadístico modelo FIR.

Componente principal	Varianza bloque X	Varianza bloque X acumulativa	Varianza bloque Y	Varianza bloque Y acumulativa	RMSECV Ley de concentrado
1	15,91	15,91	76,66	76,66	2,1237
2	17,79	33,69	5,33	81,99	1,871
3	10,89	44,59	2,52	84,52	1,739
4	11,83	56,42	1,28	85,79	1,6713
5	3,61	60,03	2,89	88,68	1,5049
6	16,01	76,04	0,23	88,91	1,4902
7	4,04	80,08	1,67	90,59	1,3869
8	14,61	94,69	0,34	90,93	1,3638
9	0,82	95,51	0,34	91,27	1,3419
10	0,35	95,86	0,32	91,59	1,3218
11	0,94	96,80	0,02	91,61	1,3204

4.5.3. Estadígrafos Q y T^2

Tal como se indicó previamente, los estadígrafos Q y T^2 son utilizados para verificar que el conjunto de datos empleado no contenga observaciones atípicas. Estas se identifican cuando una observación presenta un valor superior al límite definido para alguno de estos estadígrafos. Una vez detectadas, dichas observaciones se eliminan del conjunto, permitiendo así un filtrado efectivo de los datos y la construcción de un modelo más robusto, en este caso, un modelo FIR. Cabe destacar que tanto el

estadígrafo Q como el estadígrafo de Hotelling T^2 dependen directamente del número de componentes principales seleccionados, así como del preprocesamiento aplicado a los datos, siendo en este estudio el auto escalado. Por este motivo, y con el fin de facilitar el análisis, la comparación entre observaciones y la detección de posibles datos anómalos, se opta por normalizar ambos estadígrafos utilizando una escala reducida. Esta normalización permite interpretar de manera más clara los resultados y establecer criterios uniformes para la detección de valores atípicos.

La normalización se lleva a cabo estableciendo un límite de confianza para cada uno de los estadígrafos. En este trabajo, se adopta un nivel de confianza del 95% (normalmente varían desde 95 a 99%), valor que además es el establecido por defecto por el *software* utilizado. A partir de esto, se calcula el valor reducido dividiendo los estadígrafos Q y T^2 obtenidos para cada observación por sus respectivos límites de confianza. Este procedimiento permite interpretar ambos indicadores en una escala unificada y sin unidades, lo que facilita la identificación de las observaciones que excedan los límites esperados.

En la escala reducida utilizada para los estadígrafos Q y T^2 , el límite corresponde a la unidad. En este contexto, una observación es considerada atípica o anómala cuando presenta un valor que excede en un 50% dicho umbral, es decir, cuando alcanza valores iguales o superiores a 1,5 (valor crítico). Este tipo de observaciones, al encontrarse significativamente alejadas del comportamiento general del conjunto de datos, se interpretan como no representativas del sistema.

A pesar de que el conjunto de datos utilizado para la generación de los modelos proviene de un entorno de simulación, la presencia de datos anómalos o no representativos continúa siendo un aspecto relevante a considerar. Esto se debe a que la inclusión de observaciones atípicas en el subconjunto de calibración puede comprometer la robustez del modelo, distorsionando la representación de la dinámica general del sistema y limitando su capacidad de generalización frente a nuevas entradas.

Por ello, se llevó a cabo un proceso de filtrado previo a la construcción del modelo FIR final, descartando aquellas observaciones que se desvían significativamente de los límites establecidos para los estadígrafos Q y T^2 . Esto busca garantizar que la estructura final del modelo no se vea afectada por datos anómalos que no reflejan adecuadamente el comportamiento esperado del sistema, manteniendo así su capacidad predictiva.

En el caso del modelo FIR final desarrollado, se generaron dos gráficos comparativos que permiten evaluar visualmente el comportamiento de la ley de concentrado estimada frente a su valor real, obtenido directamente del simulador. Estos gráficos, además de comparar visualmente ambas señales,

incorporan un mapa de color basado en los valores de los estadígrafos reducidos, permitiendo identificar de manera gráfica y clara la calidad de cada observación posterior al proceso de filtrado.

En primer lugar, se presenta la Figura 14, la cual incorpora el estadígrafo Hotelling's T^2 . En esta figura, cada observación está coloreada de acuerdo con el valor reducido de dicho estadígrafo. Se observa que los valores predominantes del estadígrafo T^2 reducido a lo largo del conjunto de datos se encuentran aproximadamente en el rango de 0,2 a 0,8, lo cual se refleja en la predominancia de tonalidades azules, celestes y verdes claros en la coloración del gráfico. Exceptuando algunas zonas muy puntuales, prácticamente todas las observaciones se mantienen por debajo del valor umbral de 1, y ninguna supera el valor crítico establecido de 1,5.

Adicionalmente, no se identifican patrones definidos que indiquen aumentos progresivos en los valores de T^2 hacia los extremos o el centro del gráfico. Las pocas observaciones con valores cercanos al umbral crítico, representadas mediante tonalidades amarillas, se localizan en regiones aleatorias del dominio y no conforman un patrón discernible. Estas observaciones son escasas y presentan valores de T^2 que, en ningún caso, superan 1,3.

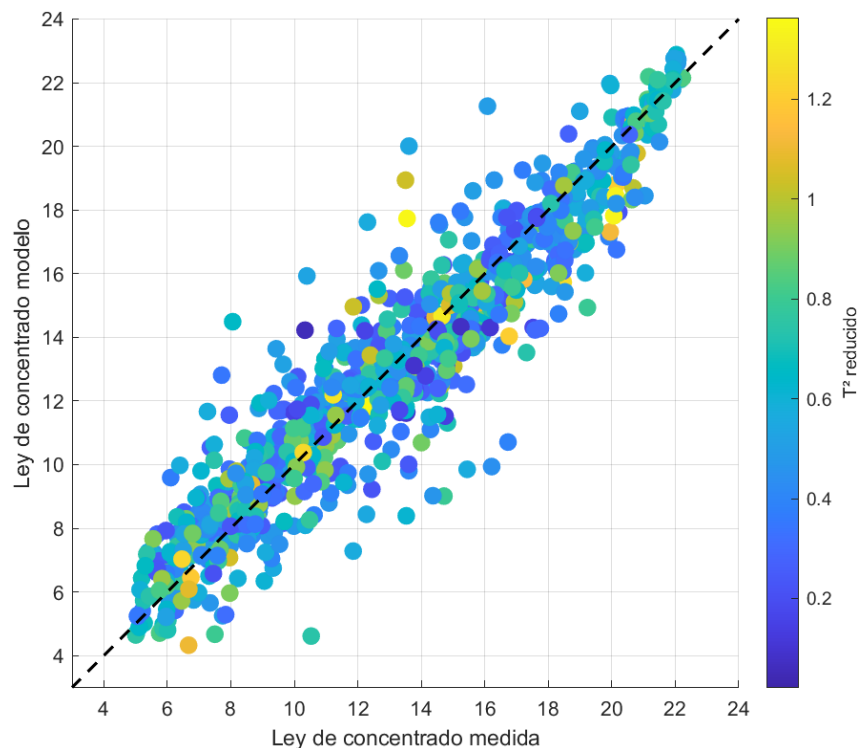


Figura 14. Simulador vs modelo FIR, Hotelling's T^2 reducido.

Por otro lado, se presenta la Figura 15, la cual incorpora el estadígrafo Q residual reducido, en contraste con la Figura 14, donde se utilizó el estadígrafo Hotelling's T^2 reducido. En esta figura se

observa una distribución de color considerablemente más uniforme, predominando tonalidades que van desde el azul oscuro hasta el azul claro en prácticamente la totalidad de las observaciones. Esta coloración sugiere que los valores del estadígrafo Q reducido se encuentran, en su mayoría, en un rango aproximado entre 0 y 0,6, es decir, muy por debajo del umbral de valor 1.

La presencia de estas tonalidades indica que las observaciones han sido, en general, bien representadas por el modelo FIR, mostrando niveles bajos de error residual. Si bien es posible identificar puntos o grupos de observaciones con valores mayores o cercanos a 1 (umbral), evidenciados por las tonalidades amarillas y verdes claras presentes en la figura, ninguna de ellas supera el valor o umbral crítico de 1,5. Estas corresponden a observaciones o agrupaciones muy puntuales que no permiten identificar un patrón claro que sugiera la causa de dichos incrementos.

Adicionalmente, se aprecia que la ley de concentrado estimada por el modelo presenta una clara tendencia a alinearse con la diagonal de igualdad, lo que indica un comportamiento general acertado en la predicción. No obstante, se observa una dispersión considerable en todo el rango de operación del sensor, con un leve ensanchamiento en el extremo inferior del dominio (leyes más bajas), el cual tiende a disminuir progresivamente a medida que se avanza hacia el extremo superior. Esta variación en la dispersión, si bien es perceptible, no compromete de manera significativa la calidad general del ajuste. Una proporción importante de las observaciones se alinea adecuadamente con la diagonal de igualdad, lo que refuerza la validez del modelo. Sin embargo, también se identifican estimaciones que no logran ajustarse de forma óptima a los valores obtenidos mediante simulación, presentando desviaciones que varían desde leves hasta considerablemente marcadas.

Este comportamiento sugiere que el modelo, o sensor virtual, ofrece un desempeño predictivo aceptable dentro del dominio simulado. No obstante, es importante tener presente que en determinadas condiciones operacionales las estimaciones generadas podrían no alcanzar los niveles deseados de precisión y exactitud, lo cual debe considerarse al momento de su implementación práctica.

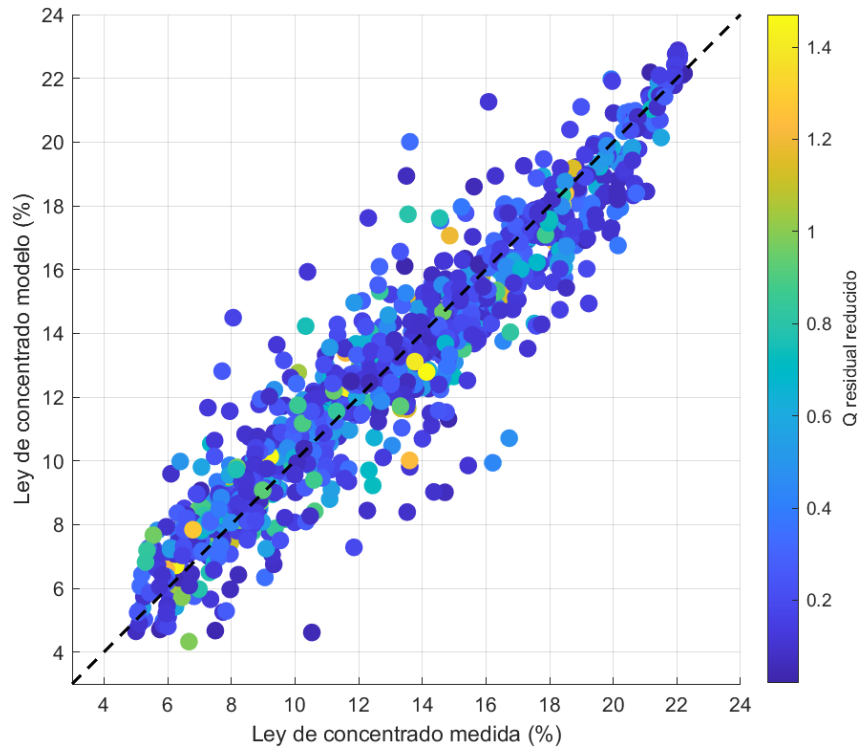


Figura 15. Simulador vs modelo FIR, Q residual reducido.

Una vez completadas todas las etapas previas y obtenido el modelo FIR definitivo, se procede a la extracción del vector de regresión. En este caso, se trata de un modelo construido con 7 componentes principales e incorpora 6 valores pasados o *lags* en su estructura. Este vector de regresión representa el conjunto de coeficientes asociados a cada una de las variables de entrada o independientes, definiendo la relación lineal entre las entradas pasadas del sistema (6 retardos) y la variable de salida modelada. En total, el modelo utiliza 36 variables, correspondientes a los valores pasados de las variables de entrada seleccionadas. Cada uno de estos términos posee un coeficiente de regresión específico, cuya magnitud y signo determinan su contribución al valor estimado de la ley de concentrado. El vector de regresión completo se presenta en la Tabla 16, disponible en el Anexo 1, donde se muestran todos los coeficientes asociados a las variables utilizadas en el modelo FIR final.

4.5.4. Residuos modelo FIR

En esta investigación, y específicamente para el modelo FIR desarrollado, los residuos se definen como la diferencia entre la predicción de la ley de concentrado de cobre realizada por el modelo y los valores experimentales, los cuales fueron obtenidos a partir del simulador. Esta métrica permite evaluar directamente el error de estimación cometido en cada observación. El *software* de simulación empleado distingue entre los datos utilizados para la calibración del modelo y aquellos reservados

para su validación. La Figura 16 presenta los residuos obtenidos en la modelación FIR diferenciando visualmente ambos subconjuntos, siendo esta distinción esencial, ya que los datos de validación no se incluyen en el ajuste del modelo y, por tanto, ofrecen una medida objetiva de su capacidad predictiva frente a datos nuevos.

Con el fin de complementar el análisis, se incorporan en el gráfico los percentiles 2,5 % y 97,5 %, representados mediante líneas segmentadas de color naranja y verde, respectivamente. Estos percentiles, calculados únicamente a partir del subconjunto de validación, constituyen un apoyo visual clave para evaluar la dispersión y variabilidad de los residuos de este subconjunto.

El análisis de los residuos del subconjunto de validación revela varios patrones. Entre 5 % y 13 % de ley de concentrado medida (o simulada), el modelo presenta una tendencia a sobreestimar los valores reales. En contraste, en el intervalo entre 13 % y 17 %, el modelo presenta una tendencia a infraestimar los valores, generando predicciones sistemáticamente menores que las medidas.

A lo largo de casi todo el rango de operación, se detecta una alta dispersión de residuos, con una varianza de 2,03 y diferencias que alcanzan hasta un 7 % en el subconjunto analizado. Esta dispersión se hace más evidente gracias a la inclusión de los percentiles 2,5 % y 97,5 %, y tiende a disminuir progresivamente a medida que el sensor opera en regiones con una mayor ley de cobre en el concentrado. Si bien existen tramos con mayor variabilidad, los residuos no presentan un sesgo sostenido y tienden a mantenerse centrados en torno al cero, lo que indica una distribución equilibrada entre sobreestimaciones e infraestimaciones.

En el tramo superior del rango de operación, específicamente entre 17 % y 22 %, se observan algunas predicciones que tienden a infraestimar los valores reales en el subconjunto de validación. No obstante, en términos generales, se aprecia una tendencia a residuos bajos, con observaciones más centradas en torno al cero y con desviaciones menores hacia la infraestimación. Este comportamiento refleja un ajuste particularmente adecuado del modelo en esta región (especialmente entre 21 % y 22 %), donde los errores de predicción resultan más acotados en comparación con los otros tramos descritos.

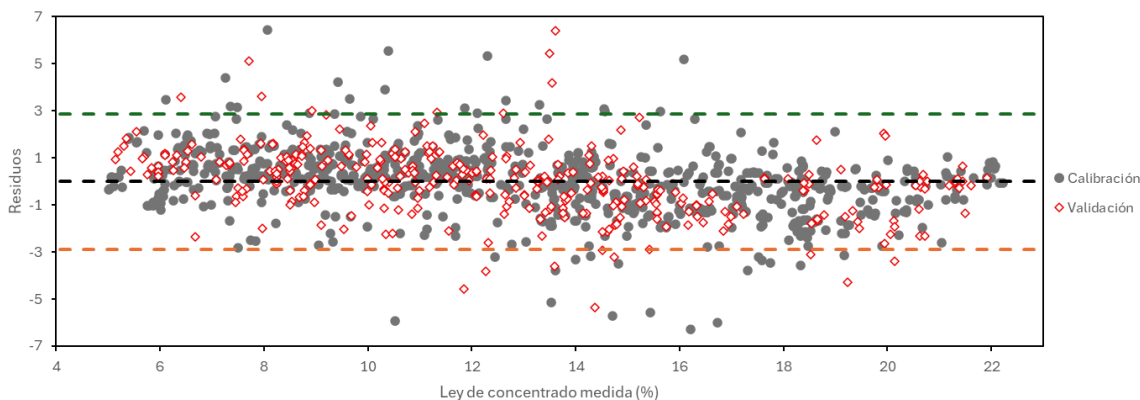


Figura 16. Residuos modelo FIR.

4.6. Modelos FIR locales

A partir del análisis de los residuos y de los valores obtenidos para los estadígrafos Q y T^2 reducidos sobre los datos de validación, presentados en las Secciones 4.5.3 y 4.5.4, se puede evaluar de manera objetiva la capacidad predictiva del modelo FIR global. Los resultados evidencian que el modelo, en su rol de sensor virtual, presenta limitaciones en ciertas zonas del rango operativo, lo cual se manifiesta particularmente en el comportamiento de los residuos.

Estas discrepancias indican que el modelo FIR global no logra ajustarse con igual precisión en todas las regiones del dominio, comprometiendo la calidad de sus estimaciones. Este comportamiento, sin embargo, resulta esperable, ya que tanto el tiempo de residencia (τ) como la ganancia (K_i o g_i) del sistema fluctúan constantemente en función de las perturbaciones inducidas por cambios en las condiciones operativas de la celda o en las características de la pulpa procesada. Tales variaciones afectan de manera directa la dinámica del sistema, generando cambios en el comportamiento de la variable de interés (ley de concentrado) que el modelo no logra capturar completamente bajo una estructura global única.

Si bien la metodología PLS-FIR ofrece ventajas destacables, como la capacidad para manejar multicolinealidad y su tolerancia al ruido en las variables de entrada, su estructura lineal y dependiente únicamente de las entradas pasadas limita su capacidad para representar fenómenos dinámicos más complejos. En particular, al no incorporar retroalimentación de la salida, el modelo es incapaz de capturar relaciones que dependen de la historia del sistema o de efectos no lineales, lo que explica su rendimiento desigual en ciertas zonas del dominio operativo.

En respuesta a estas limitaciones, se propone la implementación de modelos FIR locales como estrategia de mejora. Esta aproximación permite segmentar el espacio operativo en regiones más homogéneas desde el punto de vista dinámico, ajustando modelos específicos a cada una. Con ello, se busca representar de manera más precisa el comportamiento del sistema en zonas donde el modelo global muestra un desempeño deficiente, incrementando así la capacidad predictiva del sensor virtual.

Para construir estos modelos locales, se realiza en primera instancia una categorización de las observaciones, con base en la magnitud de sus errores de predicción. Esta clasificación es esencial, ya que el propósito principal de dividir el dominio en subconjuntos es precisamente reducir el error de estimación y asegurar que dicho error se mantenga dentro de rangos considerados aceptables para la operación del sensor.

En esta investigación, se establece como criterio de aceptabilidad un error porcentual igual o inferior al 5%, lo que representa una tolerancia razonable para las aplicaciones del modelo. Por otro lado, errores porcentuales superiores al 5% se consideran indicativos de desviaciones significativas entre los valores estimados por el modelo y los valores de referencia obtenidos desde el simulador, y por lo tanto requieren una atención especial en el proceso de modelado local.

Una vez calculados los errores porcentuales para cada observación, se procede a clasificarlas según este criterio binario: observaciones dentro del rango aceptable y observaciones con desviaciones elevadas. Esta categorización constituye la base para la definición de los modelos locales y se resume en la Tabla 10.

Tabla 10. Clasificación de las observaciones según el error porcentual.

Error porcentual	Categoría	Número de observaciones
[0;5]	Aceptable	546
]5; ∞ [	No aceptable	737

De un total de 1.283 observaciones, 546 presentan errores porcentuales iguales o inferiores al 5%, lo que representa un 42,6% del conjunto total de datos. Estas observaciones se consideran adecuadamente representadas por el modelo global, ya que sus errores se encuentran dentro del rango definido como aceptable. Por otro lado, 737 observaciones (57,4%) presentan errores superiores al 5%, clasificándose como casos con desviaciones elevadas. Este volumen considerable de datos con errores significativos indica que existe una porción muy importante del espacio operativo en el que el modelo FIR no logra un ajuste adecuado.

Una vez clasificadas las observaciones según su error porcentual, se procede a una segunda etapa de segmentación, esta vez basada en una de las variables de entrada del modelo FIR. El objetivo de esta nueva segmentación es agrupar observaciones con comportamientos similares, permitiendo desarrollar modelos locales que se ajusten de manera más específica a las condiciones particulares de cada grupo.

En el caso de las observaciones que presentan errores porcentuales aceptables, se determina que continúen utilizando el modelo global FIR, dado que este ya proporciona una predicción adecuada. Si bien se consideró como prueba la posibilidad de realizar una modelación local únicamente con estas observaciones, los resultados fueron inconsistentes: algunas predicciones mejoraron en precisión, pero otras presentaron errores porcentuales que superaban ampliamente el umbral establecido. Por esta razón, se opta por mantener el modelo global para este subconjunto, priorizando la coherencia en el desempeño general y asegurando que todas las observaciones se mantengan bajo el límite de error permitido, en lugar de generar un modelo focalizado que presente alta variabilidad en los resultados.

En consecuencia, la segunda etapa de clasificación se aplica exclusivamente a las observaciones con errores porcentuales no aceptables, es decir, aquellas superiores al 5%. Para este conjunto, se propone la construcción de dos modelos locales adicionales. La definición de estos modelos requiere segmentar este subconjunto en función de una variable que tenga un impacto significativo en el modelo global, de modo que la segmentación sea coherente con la dinámica subyacente del sistema. Para ello, se recurre al análisis de importancia de variables proporcionado por el *software* de modelación SOLO, seleccionando únicamente aquellas variables cuya contribución al modelo FIR sea considerable, ya sea de forma positiva o negativa.

Este análisis identifica que las tres variables más relevantes para el modelo FIR global, ordenadas de mayor a menor impacto, son: la velocidad superficial del aire (J_g), la ley de cobre en la alimentación (X_f) y el porcentaje de sólidos en la alimentación (S_f). El orden completo de importancia puede consultarse en la Tabla 22 del Anexo 3, donde se presenta el ranking completo de variables.

Con esta información, se selecciona la velocidad superficial del aire (J_g) como criterio de partición para el subconjunto de observaciones con errores porcentuales no aceptables. La división se realiza en función de su valor, estableciendo un umbral en $J_g = 0,75$ [cm/s], de acuerdo con los niveles definidos en la matriz experimental utilizada. En consecuencia, se generan dos modelos locales: el primero considera exclusivamente observaciones con valores de J_g mayores o iguales a 0,75 [cm/s] (es decir, 0,75 y 1 [cm/s]), mientras que el segundo se restringe a valores J_g menores a 0,75 [cm/s],

lo cual corresponde a valores iguales a 0,5 [cm/s]. Esta clasificación permite capturar con mayor precisión la influencia de esta variable clave en zonas donde el modelo global no logra representar adecuadamente la dinámica del sistema.

Con base en el procedimiento descrito, se han definido tres subconjuntos distintos, cada uno asociado a un modelo local específico. Cada uno de estos modelos posee sus respectivas zonas de operación y coeficientes de regresión particulares. La Tabla 11 presenta los modelos locales numerados junto con sus subconjuntos correspondientes, además de detallar la cantidad de observaciones contenidas en cada uno.

Tabla 11. Modelos locales.

Modelo	Subconjunto	Número de observaciones
1	$]5; \infty [$ y $J_g < 0,75$	361
2	$]5; \infty [$ y $J_g \geq 0,75$	376
3 (Modelo global FIR)	$[0;5]$	546

4.6.1. Generación de modelos locales

Una vez definidos los modelos locales junto con sus respectivos rangos de operación, se procede a aplicar el mismo procedimiento metodológico a cada uno de los subconjuntos identificados, con excepción de aquel que continuará utilizando el modelo global. Este procedimiento corresponde al descrito previamente en las secciones 4.5.1 a 4.5.3, el cual establece la estructura del modelo FIR y el enfoque de modelación utilizado. Dado que se mantiene la misma configuración estructural (específicamente, el uso de seis retardos o *lags*), en cada modelo local se incorporan seis valores pasados de las variables de entrada, tal como se definió en la Ecuación 4.14.

Cada uno de los modelos locales generados sigue el mismo criterio de partición de datos utilizado en el modelo FIR global: el 70% de las observaciones se emplea para la etapa de calibración y el 30% restante para la validación del modelo. La Tabla 12 presenta un resumen de los aspectos más relevantes del desempeño de cada modelo local. En ella se indica, para cada modelo, la cantidad de componentes principales seleccionados, los cuales se determinan siguiendo el mismo criterio utilizado para el modelo global FIR: elegir el número mínimo de componentes que permita explicar al menos el 80% de la varianza total.

Dado que las tablas de resumen estadístico completas de los modelos ocupan una gran cantidad de espacio, en la Tabla 12 se incluyen únicamente los elementos más representativos. Entre estos se

destacan la varianza acumulada explicada en los bloques X e Y, así como el error cuadrático medio de la validación cruzada (RMSECV) para los modelos locales 1 y 2. En el caso del modelo 3, se presentan los valores correspondientes al modelo global, ya que este se mantiene para el subconjunto de observaciones con errores porcentuales aceptables. La información detallada sobre el comportamiento de la varianza explicada en los bloques X e Y, junto con los valores del RMSECV en función del número de componentes utilizados, para los modelos locales 1 y 2, se encuentra disponible en las Tabla 20 y Tabla 21, incluidas en el Anexo 2.

Tabla 12. Resumen estadístico modelos FIR locales.

Modelo	Componentes principales	Varianza bloque X acumulativa	Varianza bloque Y acumulativa	RMSECV Ley de concentrado
1	7	85,03	81,57	1,7899
2	7	80,35	72,05	1,7341
3 (Modelo global FIR)	7	80.08	90.59	1.3869

Si bien el criterio adoptado en este trabajo para utilizar el modelo 3 (modelo global FIR) se basa en que la observación presente un error porcentual igual o inferior al 5 %, las condiciones operacionales bajo las cuales se cumple este criterio difieren de manera relevante respecto de aquellas asociadas a errores superiores a dicho umbral.

En particular, las observaciones para las cuales se aplica preferentemente el modelo 3 se caracterizan por operar con niveles altos de ley de cobre en la alimentación, correspondientes principalmente al nivel superior definido (1 %). Estas condiciones se asocian, además, a niveles bajos de porcentaje de sólidos en la alimentación, concentrándose mayoritariamente en el nivel inferior (20 %). Asimismo, estas configuraciones tienden a excluir escenarios de operación con niveles altos de altura de espuma (4 cm), privilegiando el uso del nivel medio (2,5 [cm]). No obstante, bajo combinaciones específicas del resto de las variables de entrada, también se admiten condiciones extremas dentro de los niveles discretos definidos.

Adicionalmente, se observa que el modelo global FIR se aplica mayoritariamente en condiciones de baja velocidad superficial del aire, correspondiente al nivel inferior definido (0,5 [cm/s]). En conjunto, estas combinaciones de niveles describen las condiciones operacionales típicas en las que el modelo global presenta un desempeño satisfactorio, en contraste con aquellas configuraciones en

las que se observan errores porcentuales superiores al umbral establecido y resulta más conveniente la aplicación de modelos locales.

Cabe destacar que esta caracterización no implica que el modelo global FIR no pueda ser utilizado fuera de dichas combinaciones típicas de niveles. El análisis se presenta únicamente con el objetivo de identificar las configuraciones del espacio operacional en las que el modelo global ofrece, de manera recurrente, un desempeño adecuado, y distinguirlas de aquellas en las que la aplicación de modelos locales permite mejorar la precisión en la estimación de la ley de cobre en el concentrado.

4.6.2. Estadígrafos Q y T^2

Para el desarrollo de los modelos FIR locales, se aplicaron los mismos procedimientos metodológicos y suposiciones que en el caso del modelo FIR global, con la diferencia de que ahora dichos procedimientos se ejecutan de forma independiente para cada uno de los modelos generados. En consecuencia, los estadígrafos Q y T^2 se utilizaron nuevamente para verificar que los subconjuntos de datos empleados en cada modelo local no contuvieran observaciones atípicas. Los límites definidos cumplen la misma función que en el modelo global: detectar y eliminar datos fuera de rango, pero esta vez aplicados a cada subconjunto específico, lo que permite construir modelos más robustos y filtrar de forma efectiva la información.

Tal como en el modelo FIR global, se emplearon los estadígrafos en su forma reducida, estableciendo un límite estándar de valor 1 y un umbral crítico de 1,5 para ambos casos (Q y T^2 reducidos). Una vez filtrados los datos de cada subconjunto, se generaron gráficos comparativos que permiten evaluar visualmente el comportamiento de la ley de concentrado estimada por cada modelo respecto a los valores reales (simulados).

La Figura 17 presenta el estadígrafo Hotelling's T^2 aplicado a cada uno de los modelos locales generados (modelos 1 y 2), así como al modelo 3, el cual corresponde al modelo global FIR utilizado para su respectivo subconjunto de observaciones.

En el Modelo 1, la mayoría de las observaciones presenta valores de T^2 reducido entre 0,3 y 0,8, lo cual se refleja en la predominancia de tonalidades azules y verde claro en el gráfico. Este comportamiento indica que, en general, las observaciones no superan el umbral establecido (valor igual a 1), y mucho menos el umbral crítico de 1,5. Si bien se identifican algunas observaciones con valores entre 0,9 y 1,1 (representadas por tonalidades entre amarillo y anaranjado), estas son escasas, se presentan de forma aislada y ninguna de ellas supera el umbral crítico definido.

Adicionalmente, aunque una cantidad considerable de observaciones se alinea adecuadamente en torno a la diagonal de igualdad (línea segmentada), el modelo evidencia una dispersión persistente a lo largo de prácticamente todo el rango de operación del sensor local. Esta dispersión se manifiesta en numerosos puntos que se alejan moderadamente de la diagonal, especialmente en el tramo central y en el extremo superior del dominio. Sin embargo, no se observa una variación significativa en el nivel de dispersión según la zona de operación, ya que esta se mantiene relativamente constante a lo largo de todo el rango, con excepción de algunas observaciones aisladas. A pesar de este comportamiento, se logra evidenciar una mejora clara en el ajuste de los datos dentro de la zona especificada, la cual anteriormente presentaba errores porcentuales superiores al 5 % entre las estimaciones del modelo y los valores obtenidos experimentalmente.

En el Modelo 2, los valores del estadígrafo T^2 reducido se concentran prácticamente en su totalidad entre 0,2 y 0,8, lo cual se evidencia en la predominancia de tonalidades azules y verdes claros en el gráfico. Este comportamiento sugiere un ajuste aceptable del modelo, ya que la gran mayoría de las observaciones se ubica por debajo del umbral establecido (valor igual a 1). Si bien algunos puntos muy aislados alcanzan valores entre 0,9 y 1,4, representados por tonalidades que van del verde claro al amarillo, estos no constituyen una acumulación significativa, sino más bien casos puntuales que, además, no superan el umbral crítico impuesto (1,5).

Similar al modelo local 1, no se logra evidenciar un patrón claro en los valores del estadígrafo T^2 reducido, ya que no se observan incrementos localizados o progresivos en zonas específicas del rango de operación. Del mismo modo, se aprecia que una parte considerable de las observaciones se alinea adecuadamente en torno a la diagonal de igualdad (línea segmentada). No obstante, el modelo presenta una dispersión persistente a lo largo de toda la zona de operación, la cual, a diferencia del modelo local anterior, tiende a mantenerse constante o incluso incrementarse hacia el extremo superior del rango. En particular, en la zona comprendida entre 13 % y 17 % de ley de concentrado medida, se observa una combinación de mayor dispersión y una tendencia clara del modelo a subestimar los valores experimentales. A pesar de lo anterior, se logra evidenciar una mejora en el ajuste dentro de esta región, considerando que anteriormente estas observaciones presentaban errores porcentuales superiores al 5 %.

Finalmente, en el Modelo 3, correspondiente al modelo FIR global, se observa el comportamiento más homogéneo de los tres modelos evaluados. Los valores del estadígrafo T^2 reducido se mantienen dentro de un rango estrecho, entre 0,2 y 0,8, con una clara predominancia de tonalidades azules, celestes y verde claro en el gráfico. Esto refleja un ajuste altamente estable, con valores ampliamente por debajo del umbral establecido.

Se logran identificar algunas observaciones con valores de T^2 entre 0,9 y 1,3, representadas por tonalidades que van del verde claro al amarillo. No obstante, estas corresponden a casos puntuales, sin una acumulación significativa ni una agrupación evidente, por lo que no permiten establecer un patrón asociado a zonas específicas del dominio operacional. En otras palabras, no se observan aumentos o disminuciones progresivas del estadígrafo T^2 reducido en regiones particulares del sensor. Evidentemente, al tratarse del modelo construido a partir de observaciones con errores porcentuales inferiores al 5 %, las estimaciones se encuentran altamente alineadas a lo largo de la diagonal de igualdad (línea segmentada), presentando una dispersión mínima. Cabe destacar, sin embargo, que esta dispersión muestra una leve tendencia a incrementarse a medida que el modelo se aproxima al extremo superior del rango de operación, es decir, cuando trabaja con leyes de concentrado más elevadas.

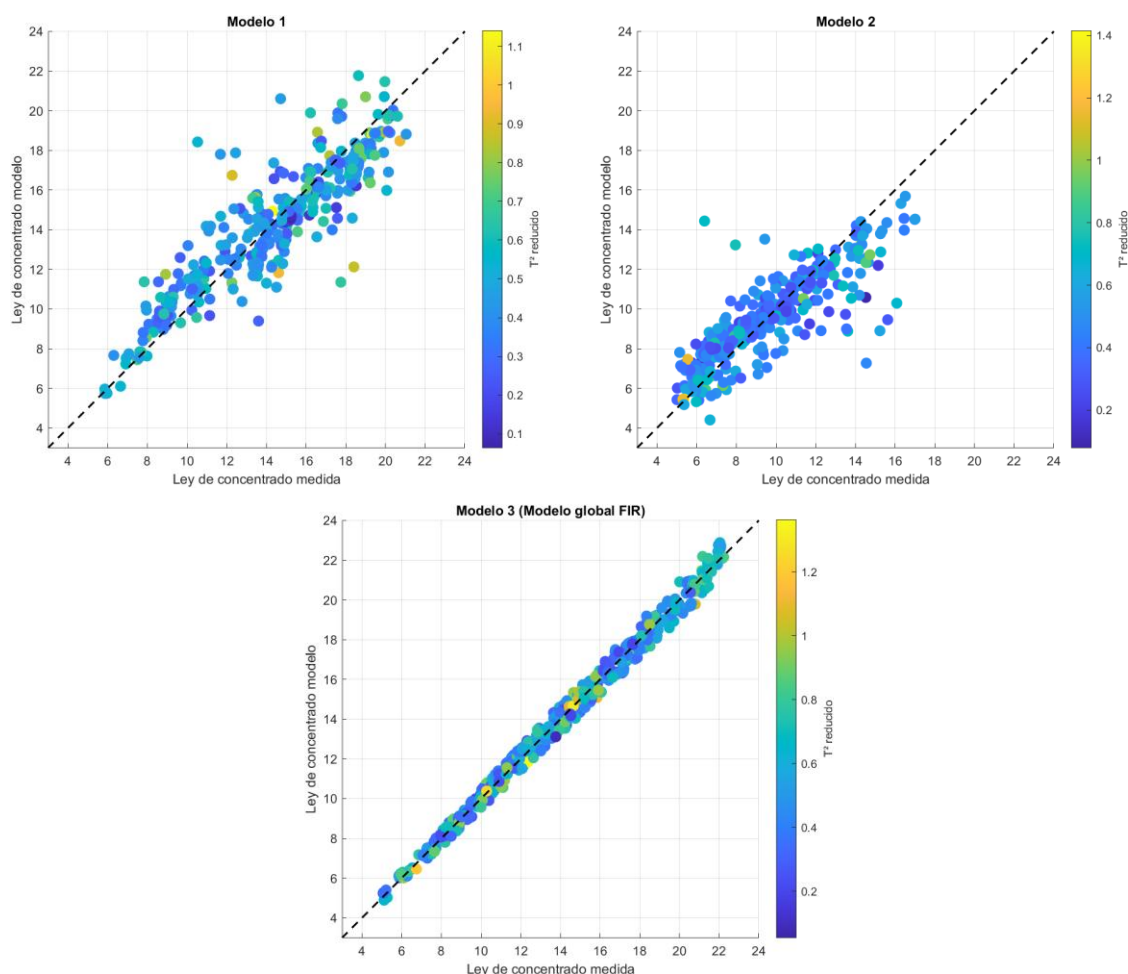


Figura 17. Modelos 1, 2 y 3 vs simulador, Hotelling's T^2 reducido.

De forma análoga al análisis anterior, en la Figura 18 se incorpora el estadígrafo Q residual reducido, en contraste con los gráficos previos donde se utilizó el estadígrafo Hotelling's T^2 reducido.

En el Modelo 1, la mayoría de las observaciones presenta valores del estadígrafo Q residual reducido entre 0,2 y 0,8, lo cual se refleja en la predominancia de tonalidades azules, celestes y verde claro en el gráfico. Se observa que, dentro de la zona de operación del sensor, principalmente entre 14 % y 19 % de ley de concentrado, se concentran la mayor parte de las observaciones que sobrepasan el umbral de 1, alcanzando valores de entre 1,0 y 1,4, representados por tonalidades que van desde el verde claro hasta el amarillo. Sin embargo, el gráfico evidencia que estas observaciones no superan el umbral crítico establecido (1,5) y que, aunque se concentran en una zona específica del dominio, estas observaciones aisladas no constituyen un comportamiento generalizado del modelo.

En el Modelo 2, los valores se concentran principalmente entre 0,1 y 0,8, como lo indica la predominancia de tonalidades azules, celestes y verdes claros en el gráfico. A diferencia del Modelo 1, se observa una menor presencia de observaciones cercanas al umbral crítico, y una cantidad aún más reducida de casos que alcanzan valores superiores al umbral (valor igual a 1). Estas observaciones no se agrupan en una zona específica del dominio, sino que aparecen de forma dispersa y sin una acumulación significativa. Si bien es posible identificar algunas observaciones puntuales con valores entre 1,2 y 1,4, estas corresponden a casos aislados, ya que la mayoría de los puntos se mantiene claramente por debajo del umbral establecido.

Por su parte, el Modelo 3, correspondiente al modelo FIR global, muestra valores del estadígrafo reducido que se mantienen mayoritariamente en un rango entre 0 y 0,6, con una clara predominancia de tonalidades azules y celestes. Esta distribución evidencia un ajuste altamente estable y sin presencia significativa de errores residuales elevados. Solo se identifican, de manera aislada, algunas observaciones que presentan valores del estadígrafo entre 0,8 y 1,4, representadas por tonalidades que van desde el verde suave hasta el amarillo. Estas observaciones no forman agrupaciones relevantes ni una concentración específica dentro del dominio operativo, apareciendo de forma dispersa y sin superar el umbral crítico de 1,5, de hecho, solo algunas de ellas apenas sobrepasan el umbral de 1.

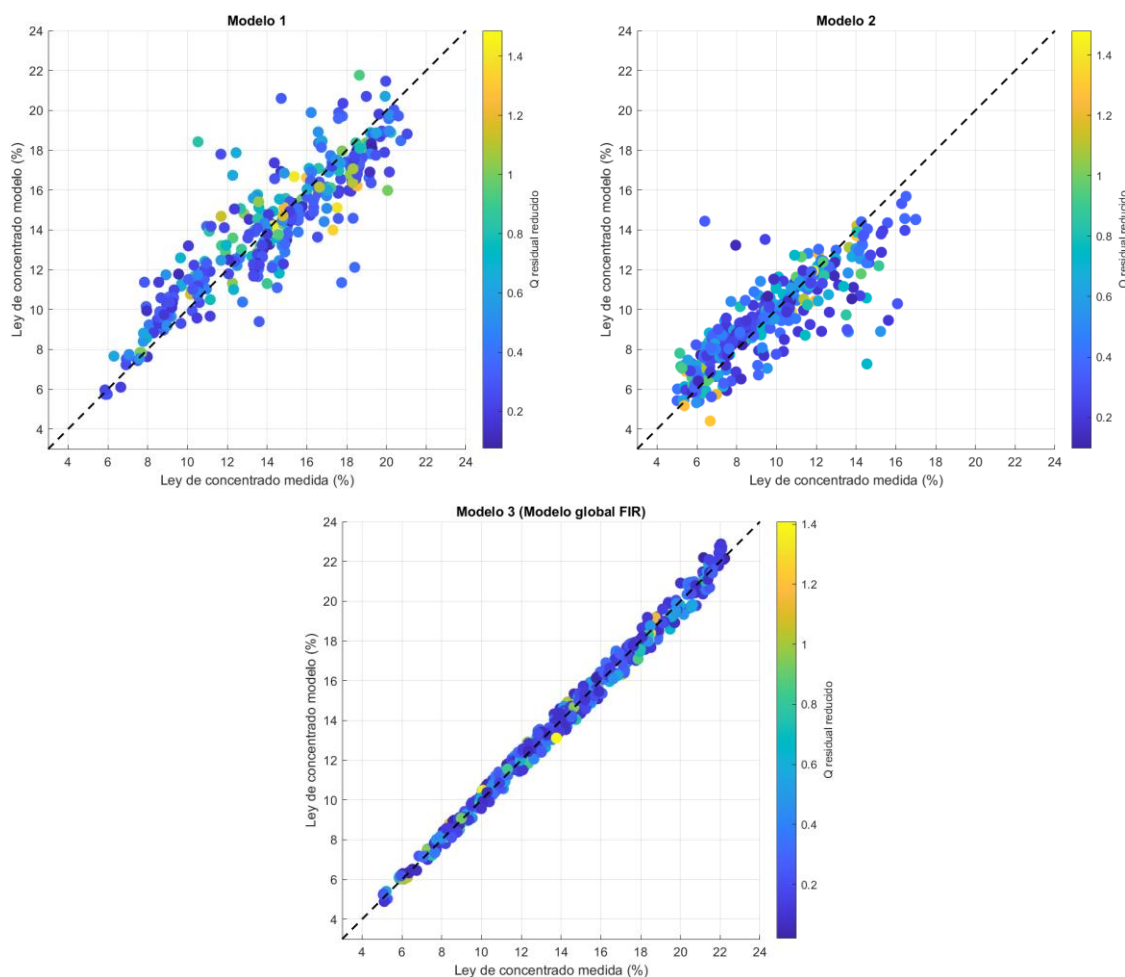


Figura 18. Modelos 1, 2 y 3 vs simulador, Q residual reducido.

Una vez completadas todas las etapas anteriores y obtenidos los modelos FIR locales, se procede a la extracción de los vectores de regresión correspondientes. Los vectores completos para cada modelo se presentan en las Tabla 17 y Tabla 18, disponibles en el Anexo 1, donde se detallan los coeficientes asociados a las variables utilizadas en los modelos locales 1 y 2, respectivamente.

4.6.3. Residuos modelos FIR locales

Similar a lo realizado para el modelo FIR global en la Sección 4.5.4, los residuos se definen como la diferencia entre la predicción de la ley de concentrado de cobre realizada por el modelo y los valores experimentales, permitiendo evaluar de forma directa el error de estimación asociado a cada observación. Dado que el *software* de simulación utilizado permite distinguir entre los datos de calibración y los datos de validación, la Figura 19 presenta los residuos obtenidos para los modelos FIR locales (Modelos 1 y 2) y para el modelo FIR global (Modelo 3), diferenciando visualmente ambos subconjuntos. Esta distinción resulta fundamental, ya que los datos de validación no participan en el ajuste del modelo y, por tanto, proporcionan una medida objetiva de su capacidad predictiva,

simulando su desempeño frente a datos nuevos o no vistos. Al igual que en el caso del modelo global, con el fin de complementar el análisis, se incorporan en el gráfico los percentiles 2,5 % y 97,5 %, representados mediante líneas segmentadas de color naranja y verde, respectivamente.

El análisis de los residuos del subconjunto de validación, apoyado en los percentiles 2,5 % y 97,5 %, evidencia diferencias relevantes entre los enfoques global y local. En particular, los modelos FIR locales presentan residuos más concentrados y mejor centrados en torno a cero a lo largo de prácticamente todo el rango de operación, con una dispersión notablemente menor en comparación con el modelo global. Esta reducción se refleja en una varianza de 1,47 para todo el subconjunto de validación, frente a 2,03 en el modelo FIR global, lo que corresponde a una disminución de 0,56 (27,6 %). Esta mejora también se aprecia al comparar la ubicación de los percentiles para ambos modelos, observándose una clara reducción en la dispersión general y un mejor centrado de los datos respecto al cero.

Este comportamiento es particularmente evidente en la región intermedia del dominio (10 % a 18 % de ley de concentrado), donde el modelo FIR global tiende a mostrar residuos más dispersos y asimétricos, generando zonas con notorias sobreestimaciones o subestimaciones. En cambio, los modelos locales mantienen un patrón más controlado y simétrico, lo que se traduce en una mayor precisión en sus estimaciones.

Si bien en los modelos locales aún se identifican zonas con ligeras tendencias a la sobreestimación o subestimación, la magnitud de estas desviaciones es considerablemente menor que en el modelo global. Esto es particularmente notorio en el rango 5 % a 10 % de ley de concentrado, que en el modelo global presentaba una marcada tendencia a la sobreestimación y que, en los modelos locales, muestra un comportamiento mucho más controlado, con residuos significativamente más próximos a cero.

En conjunto, estos resultados demuestran que la aplicación de modelos locales permite reducir de manera significativa la variabilidad de los residuos y mejorar la calidad de las predicciones. Por lo tanto, este enfoque representa una estrategia eficaz para incrementar la precisión y la confiabilidad de los modelos FIR, especialmente en escenarios donde se requiere una capacidad predictiva robusta frente a condiciones operacionales variables.

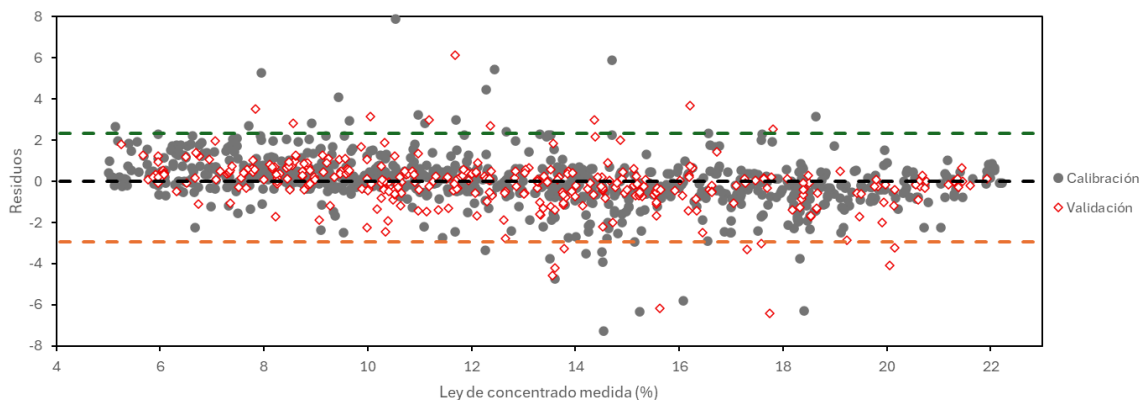


Figura 19. Residuos modelos locales FIR.

4.7. Modelo ARX

4.7.1. Bloques de datos X e Y

La construcción del modelo ARX se realiza siguiendo exactamente los mismos procedimientos descritos en la Sección 4.4.4.2, lo que implica una estructura metodológica similar a la del modelo FIR, tanto en la conformación del bloque o matriz X como del bloque Y. La principal diferencia radica en que el modelo ARX incorpora una componente autorregresiva, es decir, expresa la salida actual como una combinación lineal de sus propias salidas pasadas, junto con las entradas pasadas, mientras que el modelo FIR considera exclusivamente los valores pasados de las variables de entrada. Evidentemente, se utiliza el mismo conjunto de datos aplicado en la modelación FIR, así como las mismas consideraciones y supuestos en relación con las observaciones, conforme a lo descrito en la Sección 4.5.1.

En cuanto a la selección del número de valores pasados (*lags*) en modelos ARX desarrollados mediante PLS, la práctica común sugiere trabajar con entre 1 y 5 *lags*. En este caso, se evaluaron modelos con configuraciones que van desde 1 a 6 valores pasados utilizando el *software* SOLO, empleando como criterio principal el error medio cuadrático de validación cruzada (RMSECV) métrica que es entregada directamente por el software para cada configuración evaluada.

El uso del RMSECV resulta particularmente adecuado en este contexto, ya que permite evaluar la capacidad predictiva del modelo sobre datos no utilizados en la calibración, reduciendo el riesgo de sobreajuste asociado al incremento del número de retardos. De este modo, el criterio adoptado busca fue identificar el modelo más adecuado, entendiendo como tal aquel en el que el aumento del número

de *lags* no produce reducciones significativas del RMSECV, y donde una reducción adicional de *lags* conlleva una pérdida apreciable en la capacidad predictiva de la ley de concentrado.

Tras analizar todos los casos evaluados, se determina que la cantidad de *lags* necesarios para el modelo corresponde a 2. Con este valor definido, se procede a establecer la composición de la matriz o bloque de datos X. Para este modelo, al igual que en el caso del modelo FIR, cada una de las observaciones $u(t)$ se compone de las siguientes variables de entrada:

$$u(t_1) = [Q_F(t_1), S_f(t_1), X_f(t_1), G_f(t_1), J_g(t_1), H_f(t_1)] \quad (4.15)$$

Finalmente, se establece que la matriz de datos X, al corresponder a un modelo ARX, se construye según lo especificado en la Ecuación 3.19, incorporando tanto los valores pasados de las variables de entrada como los de la variable de salida. En este caso, se consideraron un total de 2 retardos (*lags*), con un intervalo fijo de $\tau_{m\acute{a}x} = 10,53$ minutos entre cada observación. Esta configuración da lugar a la siguiente estructura matricial para X:

$$X_{ARX} = \begin{bmatrix} u(t_1 - \tau_{m\acute{a}x}) & u(t_1 - 2\tau_{m\acute{a}x}) & y(t_1 - \tau_{m\acute{a}x}) & y(t_1 - 2\tau_{m\acute{a}x}) \\ u(t_2 - \tau_{m\acute{a}x}) & u(t_2 - 2\tau_{m\acute{a}x}) & y(t_2 - \tau_{m\acute{a}x}) & y(t_2 - 2\tau_{m\acute{a}x}) \\ \vdots & \vdots & \vdots & \vdots \\ u(t_m - \tau_{m\acute{a}x}) & u(t_m - 2\tau_{m\acute{a}x}) & y(t_m - \tau_{m\acute{a}x}) & y(t_m - 2\tau_{m\acute{a}x}) \end{bmatrix} \quad (4.16)$$

En cuanto a la matriz de datos Y utilizada para este modelo, su estructura es idéntica a la empleada en el modelo FIR, con la única diferencia en la cantidad de filas, la cual varía debido a la distinta cantidad de *lags* considerados en cada caso (el modelo ARX presenta cuatro filas más que el FIR). Esta matriz contiene como única variable de salida la ley de concentrado de cobre. Respecto a la matriz de datos X, esta posee un tamaño de 1.287 filas por 14 columnas, lo que corresponde a 1.287 observaciones y un total de 14 variables derivadas de los valores pasados de las entradas y salidas del sistema. Por su parte, la matriz Y presenta también 1.287 observaciones y una única variable de salida.

Una vez definidas ambas matrices, se procede a ingresarlas en el *software* de modelación, donde se aplica el preprocesamiento mediante auto escalado. Posteriormente, se realiza la validación cruzada con el fin de construir el modelo ARX final.

4.7.2. Número de componentes del modelo

Una vez ingresados los bloques X e Y en el *software* de modelación, se procede a seguir el mismo procedimiento aplicado al modelo FIR descrito en la Sección 4.5.2, utilizando como criterio principal

la varianza acumulada explicada para la selección del número de componentes principales. Al igual que en el modelo FIR, se exige que los componentes seleccionados expliquen al menos el 80% de la varianza total. Si bien este criterio constituye la base para determinar el número de componentes, es fundamental complementarlo con la verificación del cumplimiento de los límites aceptables en los estadígrafos Q y T^2 , específicamente en relación con el umbral crítico que permite filtrar observaciones atípicas.

Considerando estos criterios, se determina que el número óptimo de componentes principales para el modelo ARX es seis. Estos seis componentes explican un 80,23 % de la varianza total del sistema. Esta información se encuentra resumida en la Tabla 13, generada por el *software* durante la construcción del modelo ARX. Dicha tabla presenta detalles relevantes, como la varianza explicada por cada componente en los bloques X e Y, sus respectivos valores acumulados, así como la Raíz del Error Cuadrático Medio de la Validación Cruzada (RMSECV).

Tabla 13. Resumen estadístico modelo ARX.

Componente principal	Varianza bloque X	Varianza bloque X acumulativa	Varianza bloque Y	Varianza bloque Y acumulativa	RMSECV Ley de concentrado
1	25,42	25,42	90,41	90,41	1,3408
2	15,65	41,07	1,41	91,93	1,2409
3	10,87	51,94	0,67	92,50	1,1895
4	11,94	63,88	0,68	93,18	1,1328
5	3,52	67,40	1,58	94,77	0,99672
6	15,04	82,44	0,03	94,79	0,99278
7	10,65	93,08	0,12	94,91	0,97674
8	5,28	98,36	0,22	95,13	0,96193
9	0,35	98,71	0,05	95,18	0,95586
10	0,12	98,83	0,09	95,27	0,94748
11	0,20	99,03	0,04	95,31	0,94273

4.7.3. Estadígrafos Q y T^2

Similar a lo realizado con los modelos desarrollados mediante la metodología PLS-FIR, en el caso del modelo ARX se emplearon los estadígrafos Q y T^2 en su forma reducida para verificar que el conjunto de datos utilizado no presentara observaciones atípicas. Se definieron límites específicos

con el objetivo de detectar y eliminar datos fuera del umbral crítico establecido, permitiendo así obtener un modelo más robusto y garantizar un filtrado adecuado de la información. Tal como en los modelos anteriores, se fijó un límite estándar de 1 y un umbral crítico de 1,5 para ambos estadígrafos (Q y T^2 reducidos). Una vez filtrados los datos y generado el modelo ARX final, se elaboraron gráficos comparativos que permiten evaluar visualmente el comportamiento de la ley de concentrado estimada respecto a los valores reales (simulados). En estas representaciones, la codificación de color se utiliza para reflejar los valores de los estadígrafos, facilitando la interpretación conjunta del ajuste del modelo, la calidad de las predicciones y la distribución del estadígrafo correspondiente.

La Figura 20 presenta el estadígrafo Hotelling's T^2 reducido aplicado al modelo ARX, mediante un mapa de color donde cada observación está coloreada de acuerdo con su valor. En esta figura, la mayoría de las observaciones presenta valores de T^2 reducido comprendidos entre 0,2 y 0,8, lo que se refleja en la predominancia de tonos azul oscuro a verde suave en el gráfico. Este comportamiento sugiere que, en su gran mayoría, las observaciones no superan el umbral establecido (valor igual a 1) y, con mayor razón, se mantienen muy por debajo del umbral crítico de 1,5. Aunque se identifican de forma muy puntual algunas observaciones con valores entre 1,0 y 1,3 (representadas por tonos entre amarillo pálido e intenso), estas se encuentran dispersas, sin formar agrupaciones significativas ni asociarse a un rango específico de la ley de concentrado. En términos generales, no se aprecia un incremento progresivo del estadígrafo T^2 reducido hacia los extremos del dominio, lo que indica una distribución estable de este indicador estadístico a lo largo del rango de operación.

Respecto a la ley de concentrado estimada por el modelo, se observa una clara alineación con la diagonal de igualdad (línea segmentada), lo que evidencia un desempeño predictivo sólido. La mayor concentración de observaciones se encuentra próxima a dicha diagonal, por lo que la dispersión observada en algunos puntos no representa el comportamiento predominante. Esta dispersión se extiende aproximadamente desde leyes de concentrado de 6 % hasta 21 %, y tiende a disminuir o mantenerse constante a medida que se avanza hacia valores más altos, lo que refuerza la solidez del modelo en la parte superior del rango operativo.

En conjunto, las desviaciones presentadas por este modelo son más leves y menos frecuentes en comparación con los modelos previamente analizados, lo que refleja un comportamiento predictivo notablemente más estable. Este desempeño posiciona al modelo como una herramienta adecuada para ser utilizada como sensor virtual, dado su alto nivel de precisión y exactitud en la estimación de la ley de concentrado.

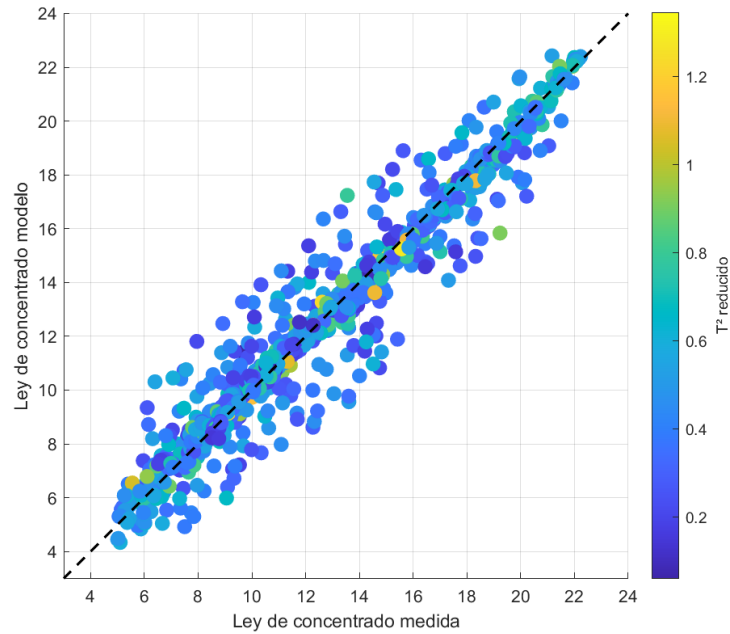


Figura 20. Simulador vs modelo ARX, Hotelling's T^2 reducido.

De forma análoga al análisis anterior, en la Figura 21 se presenta el comportamiento del estadígrafo Q residual reducido, en contraste con el gráfico previo donde se utilizó el estadígrafo Hotelling's T^2 reducido.

En este caso, la mayoría de las observaciones presenta valores de Q residual reducido comprendidos entre 0 y 0,6, lo que se refleja en la predominancia de tonalidades azules y celestes en el gráfico. Esta distribución sugiere un ajuste adecuado del modelo a lo largo de prácticamente todo el dominio de la ley de concentrado, indicando que la mayoría de las observaciones no supera el umbral establecido para este estadígrafo. Se identifican, de forma puntual, algunas observaciones con valores de Q residual reducido entre 1,0 y 1,2, representadas por tonalidades amarillas. No obstante, al igual que en los modelos analizados anteriormente, estas corresponden a casos aislados, sin evidenciar acumulaciones significativas ni una asociación clara con zonas específicas del dominio operativo del sensor. Además, es importante destacar que dichas observaciones no sobrepasan ni se acercan al umbral crítico establecido (1,5).

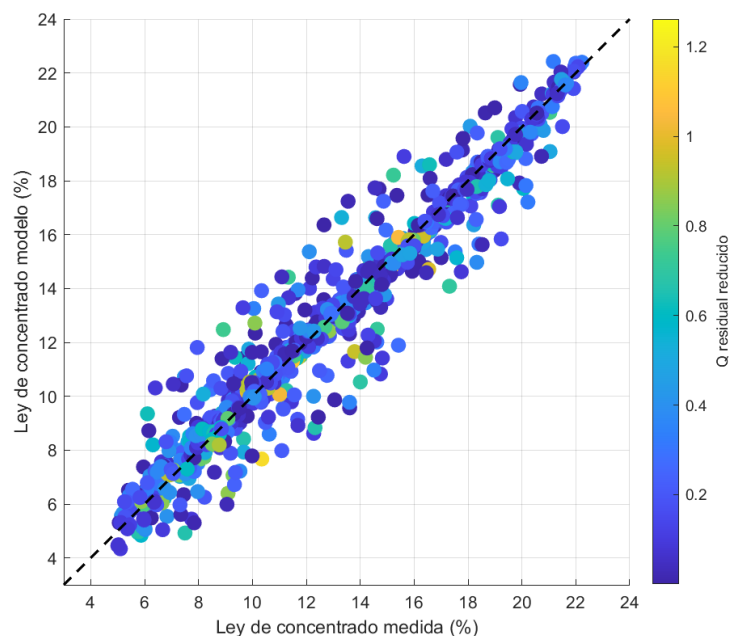


Figura 21. Simulador vs modelo ARX, Q residual reducido.

Una vez completadas las etapas anteriores y obtenido el modelo ARX definitivo, construido con seis componentes principales y utilizando un total de 2 *lags* o valores pasados, se procede a la extracción del vector de regresión. El vector de regresión completo se encuentra disponible en la Tabla 19 del Anexo 1, donde se presentan todos los coeficientes asociados, tanto a los valores pasados de las variables de entrada como a los de la variable de salida (ley de concentrado), que fueron utilizados en la construcción del modelo ARX final.

4.7.4. Residuos modelo ARX

Al igual que en los modelos previamente analizados, los residuos representan la diferencia entre la predicción de la ley de concentrado de cobre realizada por el modelo y los valores experimentales, permitiendo así evaluar de forma directa el error de estimación asociado a cada observación.

El *software* SOLO permite distinguir entre datos de calibración y validación, por lo que en la Figura 22 se presentan los residuos obtenidos para el modelo ARX, diferenciando visualmente ambos subconjuntos. Por esta razón, el análisis se enfoca principalmente en el subconjunto de validación. Para complementar el análisis, en el gráfico se incluyen las líneas correspondientes a los percentiles 2,5 % y 97,5 % para el subconjunto de validación, representadas mediante líneas segmentadas de color naranja y verde, respectivamente.

Al centrar el análisis en los residuos del conjunto de validación y apoyándose en los percentiles mencionados, se observa que la distribución de residuos es simétrica y altamente centrada en torno a

cero. La mayoría de los residuos se encuentra dentro del rango ± 1 , reflejando un comportamiento predictivo estable y consistente, especialmente en comparación con modelos evaluados previamente.

La dispersión de los residuos se extiende aproximadamente entre leyes de concentrado del 6 % al 21 %, con una leve tendencia a reducirse o mantenerse constante al avanzar hacia valores más altos de concentración. No obstante, este patrón no representa el comportamiento predominante, dado que la mayoría de las observaciones están fuertemente agrupadas alrededor del eje cero. Esto evidencia una mejora significativa en la precisión del modelo a lo largo de casi toda su zona de operación (entre 5 % y 22 %), en comparación con modelos anteriores. Este avance se refleja también en un valor de varianza de 0,72, lo que indica una considerable reducción en la dispersión y variabilidad de los residuos en el subconjunto de validación, respecto a los modelos basados en la metodología PLS-FIR. Este resultado resalta la importancia e impacto de incorporar valores pasados de la variable de salida en la estimación de su valor actual.

Además, se identifican zonas donde el modelo tiende a sobreestimar o subestimar ligeramente la ley de concentrado. En el tramo entre 7 % y 13 %, se observa una leve sobreestimación, aunque los residuos raramente superan +1. Por otro lado, entre 15 % y 19 %, se presenta una leve infraestimación, sin que los residuos inferiores a -1 sean significativos. Estas desviaciones son pequeñas y no generan un sesgo relevante, manteniéndose los residuos en un rango moderado y con buena simetría.

Las mejoras observadas en el modelo ARX, especialmente en comparación con el modelo FIR global (Figura 16), se atribuyen principalmente a la incorporación de valores pasados de la variable de salida. Esta componente autorregresiva captura de manera más efectiva la dinámica temporal del sistema, mejorando la calidad de las predicciones y disminuyendo notablemente la dispersión y varianza de los residuos. En conjunto, las diferencias en el comportamiento de los residuos fortalecen la ventaja del modelo ARX para generalizar de manera robusta bajo condiciones operativas no vistas, consolidándose como una alternativa eficiente.

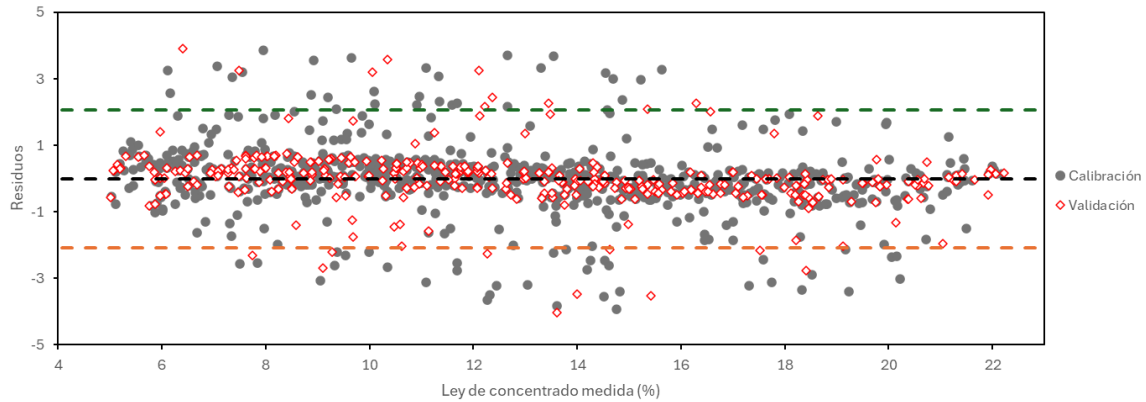


Figura 22. Residuos modelo ARX.

4.7.5. Modelo ARX no ideal

Como se mencionó previamente en la Sección 4.7.1, la principal ventaja del modelo ARX radica en la incorporación de una componente autorregresiva, es decir, la capacidad del modelo para expresar la salida actual como una combinación lineal de sus propias salidas pasadas, junto con las entradas pasadas. Esta característica distingue al modelo ARX del modelo FIR, que solo considera las variables de entrada. Sin embargo, esta ventaja también representa una limitación importante en aplicaciones prácticas, especialmente cuando se busca implementar el modelo como un sensor virtual. Para que el modelo ARX pueda realizar predicciones en su forma ideal, requiere contar con valores reales previos de la variable de salida (en este caso, la ley de concentrado de cobre) como insumo. En el modelo desarrollado, se necesitan al menos dos mediciones pasadas, con un intervalo temporal de 10,53 minutos entre cada una, tal como fue detallado en secciones anteriores.

Esta exigencia no es compatible con el propósito típico de un sensor virtual, cuya función es precisamente evitar la necesidad de realizar mediciones directas de la variable de interés. Esto puede deberse a restricciones tecnológicas, altos costos, tiempos de análisis prolongados o la imposibilidad de obtener datos de manera continua. Por lo tanto, si se quisiera implementar este modelo ARX como sensor virtual en tiempo real, no sería posible contar con las mediciones reales pasadas necesarias para alimentar el modelo y obtener predicciones actualizadas.

Frente a esta limitación, se propone una alternativa de uso del modelo ARX en modo no ideal, donde se reemplazan los valores reales pasados de la variable de salida por las predicciones previas generadas por el propio modelo. En otras palabras, tras una fase inicial de arranque en la que se ingresan las primeras dos mediciones reales para activar el modelo, cada predicción futura se alimenta

con las predicciones anteriores, generando una secuencia de estimaciones autorregresivas puramente basadas en el modelo.

Este enfoque permite simular un escenario de uso realista como sensor virtual, prescindiendo de mediciones continuas de la variable de salida. La estructura del modelo (coeficientes de regresión, auto escalado y número de componentes) se mantiene idéntica a la del modelo ARX ideal, diferenciándose únicamente en el origen de los valores pasados de salida utilizados en las predicciones.

En este contexto, la Figura 23 presenta los residuos asociados al modelo ARX no ideal, definidos como la diferencia entre la predicción de la ley de concentrado realizada por el modelo y los valores experimentales obtenidos mediante simulación. Adicionalmente, se incluye como referencia visual el intervalo de error esperado del modelo ARX ideal, representado por los percentiles 2,5% y 97,5% mediante líneas segmentadas de color naranja y verde, respectivamente. Esta comparación gráfica permite visualizar con claridad el grado de dispersión adicional que presenta el modelo ARX no ideal en comparación con su versión ideal, y evaluar de forma cualitativa la pérdida de precisión asociada a la ausencia de observaciones reales pasadas en su implementación.

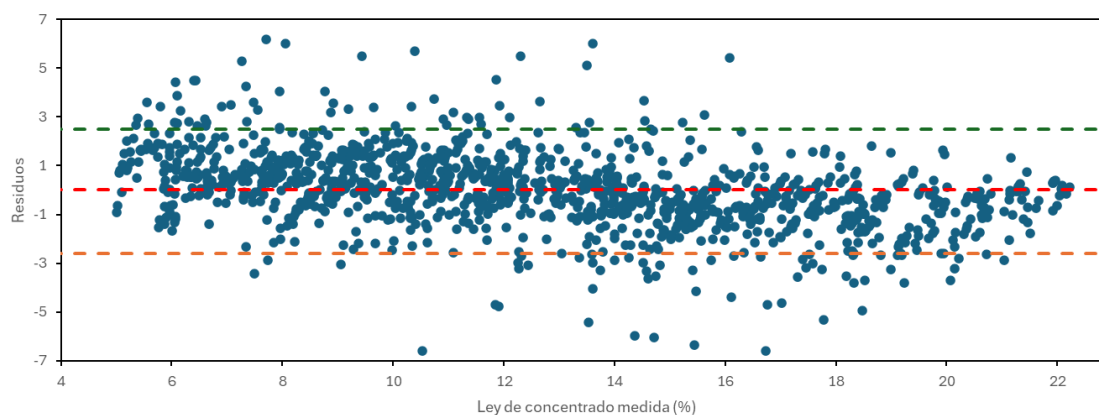


Figura 23. Residuos modelo ARX no ideal.

A partir de la Figura 22 y Figura 23, se puede concluir que el modelo ARX no ideal, aquel que estima el valor actual de la variable de salida utilizando únicamente predicciones anteriores en lugar de mediciones reales, presenta residuos que tienden a distribuirse en torno al cero y a mantener cierta simetría. No obstante, se observa una clara acentuación de comportamientos sistemáticos, manifestándose como una tendencia a la infraestimación en la región de leyes entre 15 % y 22 %, y

una leve sobreestimación en el rango de 5 % a 11 %. A lo anterior se suma un incremento notable en la dispersión de los residuos en comparación con el modelo ARX ideal.

Es importante destacar que, a pesar del carácter autorregresivo del modelo no ideal, no se evidencia una estructura que sugiera deriva sistemática o acumulación progresiva del error a lo largo del rango de operación. Esta ausencia de patrones de error persistentes indica que el modelo mantiene una relativa estabilidad temporal, incluso al alimentarse exclusivamente de sus propias predicciones. Esto refuerza la idea de que, aunque degradado, el modelo ARX no ideal conserva un nivel básico de estabilidad operativa.

Al comparar ambas figuras, resulta evidente que el modelo ARX no ideal presenta una dispersión significativamente mayor en sus residuos. Esta diferencia se hace más notoria al superponer en la Figura 23 los intervalos de error esperados del modelo ARX ideal (percentiles 2,5% y 97,5%), junto con la línea segmentada que indica el centrado respecto al cero. Las observaciones del modelo no ideal exceden con frecuencia dichos límites, lo que refleja un deterioro claro en la precisión del modelo al depender de sus propias estimaciones pasadas.

Esta observación visual se ve respaldada cuantitativamente al comparar los valores globales de RMSE obtenidos para ambos modelos. El modelo ARX ideal presenta un RMSE de 0,94, mientras que el ARX no ideal alcanza un RMSE de 1,56. Esto representa un aumento absoluto de 0,62 unidades, equivalente a un incremento relativo del 65,5 % en el error. Este resultado evidencia una pérdida significativa en el desempeño predictivo del modelo cuando se prescinde de valores reales de la variable de salida como entrada del modelo.

Esta pérdida de desempeño se explica, en parte, por la estructura del modelo ARX utilizado y por el número de retardos considerados. Al tratarse de un modelo ARX, la estimación actual de la ley de concentrado presenta una alta dependencia de sus valores pasados. Tal como se demuestra en la Sección 5.3, esta variable ejerce una influencia dominante en la predicción del valor presente, dependencia que se ve acentuada cuando se emplea un número reducido de retardos.

En consecuencia, la estimación del modelo ARX no ideal queda fuertemente condicionada por la calidad de las dos predicciones previas utilizadas como entrada. Dado que estas predicciones, a su vez, dependen de estimaciones anteriores, cualquier error significativo presente en ellas tiende a propagarse hacia la estimación actual. Este efecto se intensifica al contar con solo dos retardos, ya que no existe un número suficiente de valores históricos que permita amortiguar o diluir el impacto del error acumulado en la predicción final.

4.7.6. Determinación tiempo de reinicio del sensor

A partir de los resultados obtenidos en la sección anterior para el modelo ARX no ideal, resta establecer el momento óptimo en el cual el sensor virtual debe ser reiniciado. Si bien se demostró que los residuos de este modelo presentan una distribución centrada en torno al cero y no evidencian una correlación aparente, el principal efecto derivado del uso de predicciones pasadas en lugar de valores reales ha sido un incremento significativo en la dispersión de los residuos. Este comportamiento denota una pérdida de precisión en las estimaciones, manifestada a través de un mayor error de predicción, producto de la dependencia autorregresiva del modelo.

Frente a esta condición, surge la necesidad de definir un criterio temporal que permita identificar el momento adecuado para reiniciar el modelo. El objetivo de este procedimiento es evitar que las predicciones de la ley de concentrado se alejen excesivamente de sus valores reales o simulados, comprometiendo así la confiabilidad del sensor virtual. Con este fin, se analiza el comportamiento temporal de los residuos generados por el modelo ARX no ideal dentro de un intervalo acotado de 2.500 minutos, el cual representa una ventana temporal suficiente para evaluar el desempeño sostenido del modelo en un contexto operativo continuo.

En la Figura 24 se presenta dicho análisis, donde se grafica la evolución de los residuos a lo largo del tiempo. Al igual que en la Figura 23, se incorporan como referencias visuales los límites de error esperados para el modelo ARX ideal. Estos están representados por los percentiles 2,5% y 97,5%, correspondientes a las líneas segmentadas de color naranja (límite inferior) y verde (límite superior), respectivamente. La inclusión de estos umbrales permite contrastar el comportamiento del modelo no ideal respecto al desempeño ideal, facilitando la identificación de desviaciones que excedan los márgenes aceptables y justificando la aplicación de un reinicio preventivo del sensor.

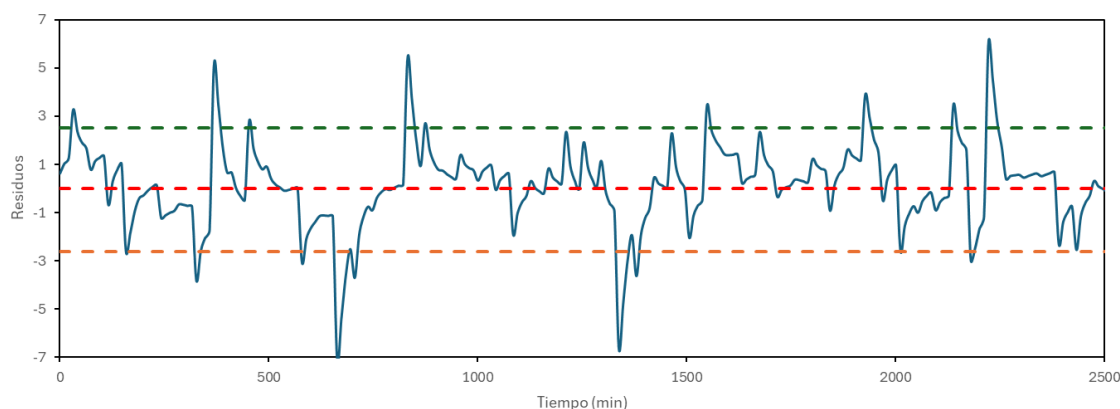


Figura 24. Residuos modelo ARX no ideal temporales.

Al observar la serie temporal de residuos del modelo ARX no ideal, se aprecia que, en su mayoría, estos se mantienen dentro del rango definido por los percentiles del error del modelo ARX ideal (líneas punteadas). Las desviaciones que exceden estos límites son eventos puntuales y breves; tras superarlos, los residuos tienden a regresar rápidamente hacia valores cercanos a cero, indicando que corresponden más a estimaciones aisladas que a tramos prolongados de mal desempeño.

Aunque se evidencia una mayor dispersión en comparación con el modelo ideal, el comportamiento temporal general de los residuos continúa siendo estadísticamente aceptable. No se observan patrones de deriva sistemática ni acumulación progresiva del error, lo cual sugiere que el modelo mantiene estabilidad operativa incluso al alimentarse exclusivamente de sus propias predicciones. Aun en presencia de desviaciones, los residuos conservan un centrado adecuado en torno al cero a lo largo del dominio temporal, lo que respalda la validez del modelo en su configuración no ideal.

Sin embargo, como se mencionó, es posible identificar ciertos momentos donde los residuos superan los límites esperados. Aunque estos no implican necesariamente un mal ajuste estructural, sí indican posibles episodios de pérdida momentánea de precisión. Este comportamiento es característico de modelos ARX no ideales, donde pequeños errores de predicción pueden amplificarse si no se detectan y gestionan oportunamente.

Por esta razón, se justifica la implementación de un análisis con ventanas deslizantes para monitorear de forma continua el desempeño del modelo. Esta técnica permite calcular métricas de error en segmentos temporales sucesivos, facilitando la identificación de tramos donde la capacidad predictiva del modelo se ve degradada. En el contexto de sensores virtuales operando en tiempo real, este enfoque resulta especialmente útil para mantener la confiabilidad del sistema frente a datos nuevos o no vistos.

En este análisis se utilizó el RMSE (*Root Mean Square Error*) como métrica principal para evaluar el desempeño del modelo ARX en el tiempo. Para ello, se aplicó una ventana deslizante, permitiendo observar la evolución local del error y detectar posibles pérdidas de precisión del modelo. El tamaño de la ventana fue definido considerando tanto el número total de observaciones disponibles como la dinámica operativa del proceso de flotación en la celda piloto. Dado que el sistema registra una observación cada 10,53 minutos ($\tau_{m\acute{a}x}$) y se dispone de un total de 1.287 observaciones, se seleccionó una ventana de 65 muestras, lo que equivale a un intervalo temporal de aproximadamente 11,4 horas. Esta elección equilibra dos aspectos clave: evita una resolución demasiado gruesa, como ocurriría con ventanas grandes, y a la vez reduce la sensibilidad excesiva al ruido puntual, como sucedería con ventanas muy pequeñas.

Bajo esta estructura, se comparó el comportamiento del modelo ARX no ideal con el del modelo ARX ideal (alimentado con salidas reales pasadas). Para detectar posibles pérdidas de precisión del modelo no ideal, se definió un umbral de error basado en el desempeño del modelo ideal. Este umbral representa un límite de confianza que refleja la variabilidad esperada cuando el modelo opera correctamente.

El umbral de error fue calculado a partir de la serie de valores del RMSE local correspondiente al modelo ARX ideal, es decir, aquel cuya predicción se basa en salidas reales pasadas. Esta serie se obtuvo aplicando una ventana deslizante de 65 observaciones, lo que permitió capturar la variabilidad natural del modelo bajo condiciones normales. A partir de dicha serie, se estimó la media y la desviación estándar del RMSE, definiéndose el umbral como la suma de la media más dos veces la desviación estándar ($\text{media} + 2\sigma$). Cuando el RMSE del modelo ARX no ideal supera este umbral, se interpreta como una señal de que el sensor virtual ha comenzado a deteriorarse. Para otorgar mayor robustez al criterio y evitar decisiones precipitadas frente a fluctuaciones puntuales del error, se establece que esta condición debe mantenerse durante un mínimo de 5 ventanas consecutivas antes de recomendar un reinicio. Esto permite distinguir entre fallos reales del modelo y variaciones esporádicas no significativas.

Este análisis se resume en la Figura 25, la cual presenta la evolución temporal del RMSE del modelo ARX no ideal, el comportamiento del modelo ARX ideal y el umbral de control previamente definido. A partir de esta información, se estimó que el modelo ARX no ideal mantiene un desempeño aceptable durante un promedio de 90 ventanas, lo que equivale a aproximadamente 949 minutos o 15,8 horas de funcionamiento continuo antes de requerir un reinicio. En consecuencia, este valor se propone como una estimación del tiempo medio recomendado para el reinicio del sensor virtual. Si bien dicho valor puede ajustarse según las condiciones específicas del proceso, proporciona una base cuantitativa y objetiva para definir intervalos de reinicio operativos que aseguren la continuidad y confiabilidad del sistema predictivo en escenarios reales de monitoreo industrial.

Cabe destacar que el análisis con ventanas deslizantes se realiza utilizando valores reales pasados de la ley de concentrado, lo cual impide su implementación directa en una planta industrial donde dicha información no está disponible en línea. Sin embargo, el resultado principal obtenido (tiempo promedio de reinicio) sí puede ser aplicado en un entorno real. Bajo condiciones operacionales similares y utilizando el mismo modelo ARX no ideal construido a partir del modelo ideal, es razonable esperar un comportamiento de degradación temporal del mismo orden de magnitud, siempre que el modelo y las condiciones operacionales se mantengan dentro de los rangos considerados en la simulación. En este contexto, el tiempo estimado de 15,8 horas constituye una

referencia práctica para definir políticas de reinicio preventivo del sensor virtual en entornos reales, por ejemplo, estableciendo reinicios cada 8 o 12 horas, con el fin de preservar su confiabilidad antes de que el deterioro de la estimación se vuelva significativo.

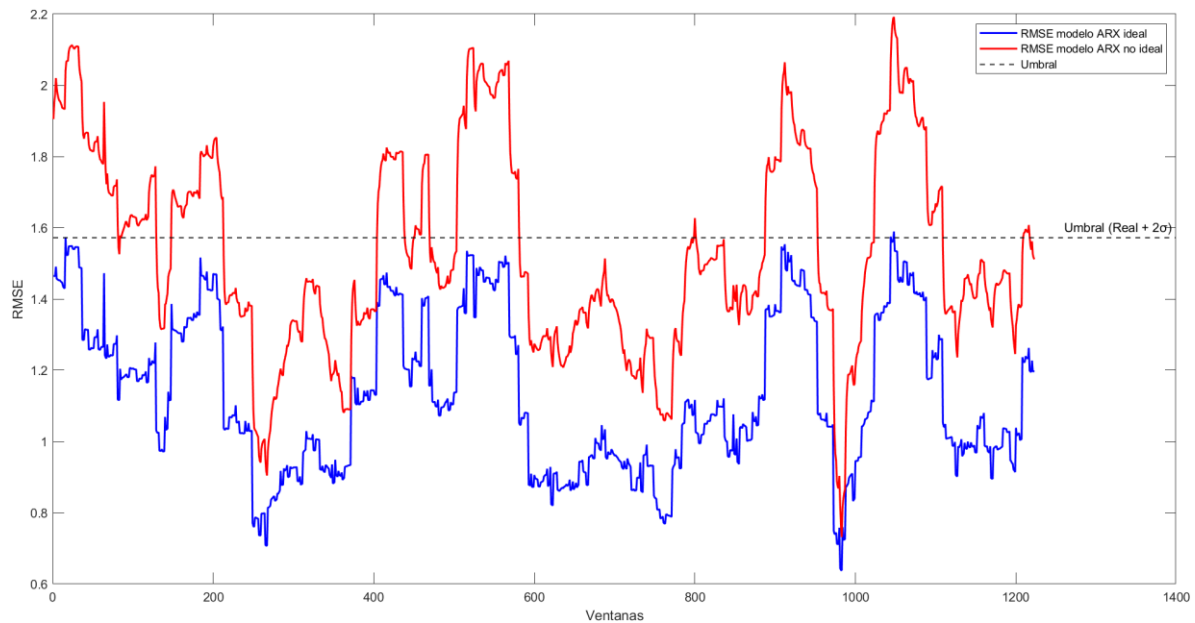


Figura 25. Comparación de RMSE (Ventana deslizante).

5. Análisis y discusión de resultados

5.1. Resumen modelos generados

La Tabla 14 presenta un resumen del desempeño de cuatro modelos predictivos evaluados en función de su capacidad explicativa y precisión estimativa en la predicción de la ley de concentrado de cobre. Los modelos considerados incluyen FIR, FIR locales, ARX ideal y ARX no ideal. Todos los modelos FIR, tanto el global como los locales, se calibraron utilizando un total de siete componentes principales. En cambio, tanto el modelo ARX ideal como el ARX no ideal emplean seis componentes en el nuevo espacio construido a partir de la metodología PLS.

Respecto a los indicadores de error, se observa que el modelo ARX ideal presenta el mejor desempeño promedio, con un RMSE de 0,94, el menor entre todos los modelos. Esta métrica sugiere que, en términos generales, sus predicciones están más próximas a los valores reales, siendo el modelo que mejor se ajusta a los datos experimentales. Además, exhibe el menor error absoluto máximo (4,04) junto con la mayor capacidad explicativa en el bloque Y (94,79 %), lo cual refleja la gran consistencia del modelo. Estos resultados son coherentes con lo esperado al emplear la metodología PLS-ARX, ya que este enfoque incorpora tanto valores pasados de las variables de entrada como de la variable de salida, permitiendo modelar de forma más precisa la dinámica del proceso. Sin embargo, pese a su excelente desempeño estadístico, el modelo ARX ideal no resulta operativamente viable como sensor virtual en línea, dado que requiere disponer de salidas reales pasadas. Esta dependencia limita su aplicación a entornos simulados o a sistemas donde la medición de la salida esté frecuentemente disponible y sea confiable, condición que no siempre se cumple en escenarios industriales reales.

En contraposición, el modelo ARX no ideal, alimentado únicamente por sus propias predicciones, representa una alternativa operativa viable como sensor virtual. Sin embargo, muestra un claro deterioro en su desempeño predictivo. A pesar de mantener la misma varianza explicada que su contraparte ideal, su RMSE aumenta a 1,56, el mayor entre todos los modelos comparados. Asimismo, presenta el segundo mayor error absoluto máximo (7,05), solo superado por los modelos FIR locales. Estos resultados evidencian una acumulación progresiva de errores, inherente a la naturaleza autorregresiva del modelo, lo cual subraya la necesidad de implementar estrategias de monitoreo. En esta investigación, se propone el uso de ventanas deslizantes para determinar momentos óptimos de reinicio del sensor virtual y evitar la pérdida de confiabilidad en las estimaciones.

Tabla 14. Resumen estadístico de los modelos obtenidos.

Modelo	Componentes principales	Varianza bloque X acumulativa	Varianza bloque Y acumulativa	Error absoluto máximo	RMSE
FIR	7	80,08	90,59	6,43	1,37
FIR locales	7 7 7	85,03 80,35 80,08	81,57 72,05 90,59	8,04	1,22
ARX	6	82,44	94,79	4,04	0,94
ARX no ideal	6	82,44	94,79	7,05	1,56

En cuanto a los modelos FIR y FIR locales, el modelo FIR global presenta un RMSE de 1,37, siendo el segundo peor valor entre los modelos evaluados, aunque registra un error absoluto máximo relativamente bajo (6,43). Por su parte, los modelos FIR locales, a pesar de presentar el mayor error absoluto máximo (8,04), logran una mejor precisión media con un RMSE de 1,22, lo que sugiere una mayor adaptabilidad a dinámicas locales. Esto evidencia que la segmentación del dominio operativo mediante modelos locales permite mejorar la precisión media, aunque incrementa el riesgo de errores puntuales más elevados.

Desde la perspectiva explicativa, todos los modelos muestran una alta varianza acumulada en Y, con excepción de los modelos FIR locales 1 y 2, que alcanzan únicamente el 81,57 % y 72,05 % respectivamente. Los modelos ARX y FIR global, en cambio, superan el 90 %, destacando en su capacidad para capturar la dinámica del proceso. En cuanto al bloque X, los modelos ARX presentan una varianza explicada ligeramente superior (82,44 %) frente al FIR global (80,08 %) o el modelo FIR local 2 (80,35 %). No obstante, todos los modelos cumplen con el criterio de selección de componentes basado en varianza acumulada explicada.

En conjunto, los resultados muestran que los modelos FIR locales ofrecen la mejor precisión media (exceptuando el modelo ARX ideal) gracias a su capacidad de adaptación a condiciones particulares. Por otro lado, los modelos ARX destacan por su robustez estructural y capacidad explicativa, pero requieren vigilancia activa en su configuración no ideal. Finalmente, el modelo FIR global, aunque superado por los otros enfoques, mantiene un desempeño razonable y sirve como línea base en la comparación.

Por ende, al considerar qué modelo implementar como sensor virtual, se deben ponderar no solo métricas como el RMSE, sino también la estabilidad frente a errores acumulativos y la viabilidad operativa en tiempo real. En este contexto, el modelo ARX no ideal representa una opción realista al

no requerir salidas reales como entradas. No obstante, su tendencia a degradarse en el tiempo obliga a definir estrategias de control del error, como umbrales y reinicios periódicos. Esta necesidad de monitoreo continuo lo vuelve menos robusto que el ARX ideal, pero más funcional en contextos reales. Por su parte, los modelos FIR locales, aunque precisos, requieren una implementación más compleja, ya que implican clasificar observaciones en función de su error porcentual y aplicar el modelo adecuado para cada subconjunto, como se expone en la Sección 4.6.1. Esta estrategia demanda mayor capacidad de procesamiento, pero puede brindar resultados significativamente más ajustados si se dispone de información confiable para segmentar correctamente.

En definitiva, la elección del modelo más adecuado dependerá del equilibrio entre precisión, robustez, adaptabilidad y factibilidad operativa que se desee priorizar en la implementación final del sensor virtual.

5.2. Distribución y normalidad de los residuos

Con el objetivo de evaluar el comportamiento de los residuos generados por los distintos modelos, se aplicaron dos herramientas gráficas: histogramas de densidad (Figura 26) y gráficos P-P (Figura 27). Los histogramas permiten visualizar la forma de la distribución de los residuos, mientras que los gráficos P-P permiten evaluar visualmente su ajuste a una distribución normal.

En primer lugar, la Figura 26 presenta los histogramas de residuos correspondientes a cada uno de los modelos evaluados. A nivel general, se observa que todos los modelos generan residuos centrados en torno al cero, lo cual constituye una propiedad deseable, ya que indica ausencia de sesgo sistemático en las estimaciones (es decir, no hay sobreestimaciones ni subestimaciones recurrentes). Sin embargo, se identifican diferencias significativas en la dispersión y la forma de las distribuciones, lo que permite caracterizar con mayor profundidad el desempeño de cada modelo.

Los modelos FIR locales destacan por presentar una distribución de residuos bastante concentrada y simétrica en torno al cero, lo que denota buena precisión general. No obstante, también muestran colas más extendidas en comparación con su versión global, lo que evidencia la presencia de errores puntuales más elevados. Este comportamiento sugiere que los modelos locales, si bien logran mejorar la precisión media al adaptarse a segmentos específicos de los datos, también conllevan un mayor riesgo de errores extremos. La Figura 26 permite visualizar claramente esta doble condición: centrado adecuado, pero con colas más largas que los otros modelos.

En contraste, el modelo FIR global presenta una distribución con mayor densidad en torno al cero, pero con una dispersión general algo más elevada que los modelos locales. Su distribución es

simétrica y más contenida en términos de amplitud (entre -6 y +6), lo que sugiere que, si bien los errores no son extremos, están más uniformemente distribuidos. Esto coincide con lo observado en secciones anteriores, donde el modelo FIR global mostraba una mayor dispersión media del error, pero sin la presencia de residuos atípicos tan marcados como los de los modelos locales.

Respecto a los modelos ARX, el modelo ARX ideal presenta la distribución de residuos más acotada y concentrada de todos los modelos analizados. Sus residuos están fuertemente centrados en torno al cero y se encuentran contenidos en un rango estrecho (aproximadamente entre -4 y +4), lo que refleja un alto nivel de precisión y exactitud en sus estimaciones. La simetría y forma de la distribución también refuerzan su carácter robusto, siendo coherente con sus indicadores de desempeño previos.

Por su parte, el modelo ARX no ideal exhibe una distribución más dispersa y levemente asimétrica, evidenciando una tendencia hacia la infraestimación. Su rango de residuos es más amplio (desde aproximadamente -8 hasta +6), lo que representa una pérdida clara de precisión en comparación con su contraparte ideal. Esta dispersión refleja el impacto negativo de la alimentación autorregresiva del modelo, donde pequeños errores se amplifican progresivamente al no contar con valores reales de la variable de salida como entradas. Estos resultados refuerzan la necesidad de aplicar estrategias de reinicio periódico al emplear este modelo como sensor virtual, como se recomendó en secciones anteriores con una frecuencia de aproximadamente 15,8 horas.

En conjunto, esta comparación entre histogramas permite concluir que, si bien todos los modelos muestran un comportamiento razonable, aquellos que operan con datos reales pasados (como el ARX ideal) o con segmentación adecuada (FIR locales) tienden a ofrecer distribuciones de error más concentradas y simétricas, características deseables en un modelo predictivo. Por ende, el ARX ideal sobresale como el modelo más preciso, mientras que el ARX no ideal, aunque operativamente viable, requiere un monitoreo constante para evitar deterioros en su desempeño.

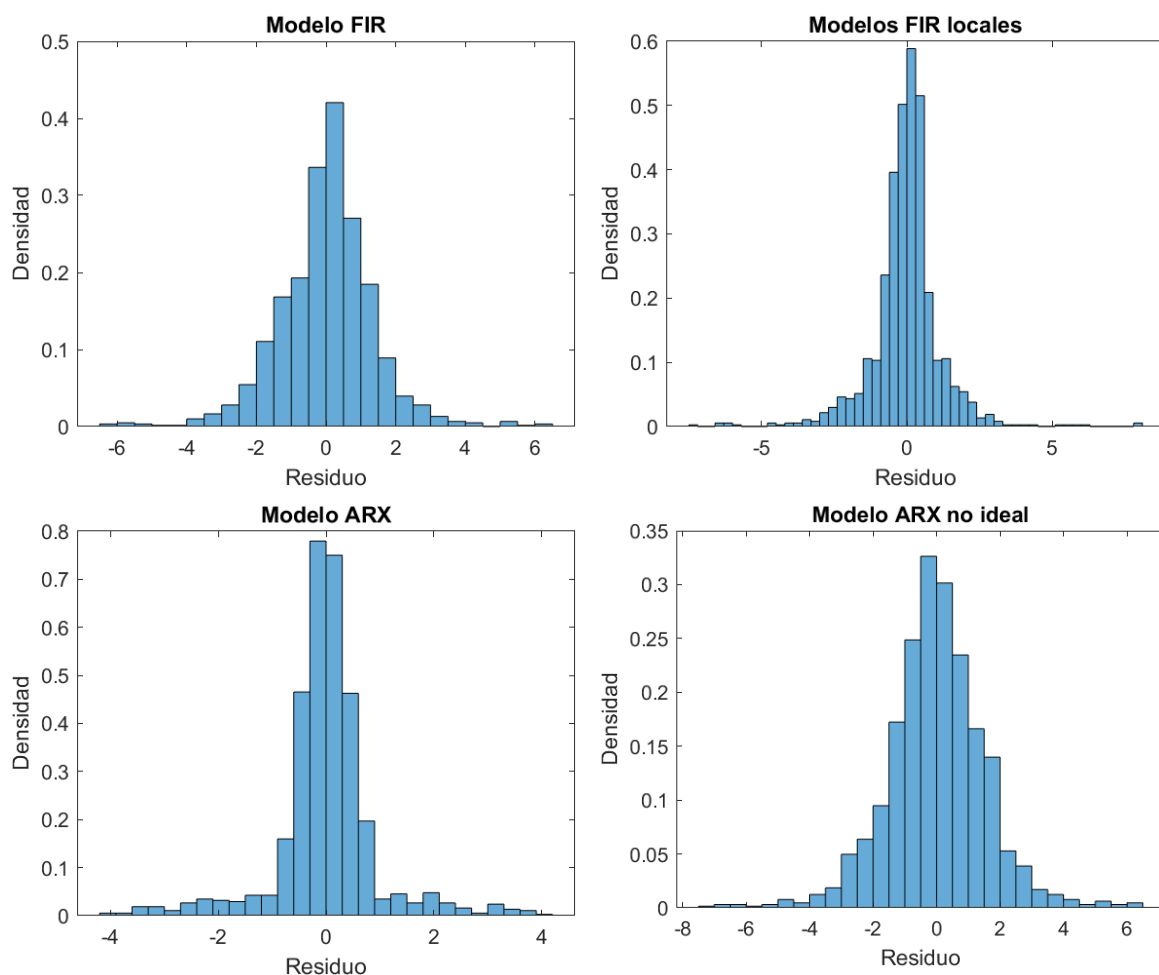


Figura 26. Distribución de errores (residuos) para cada modelo evaluado.

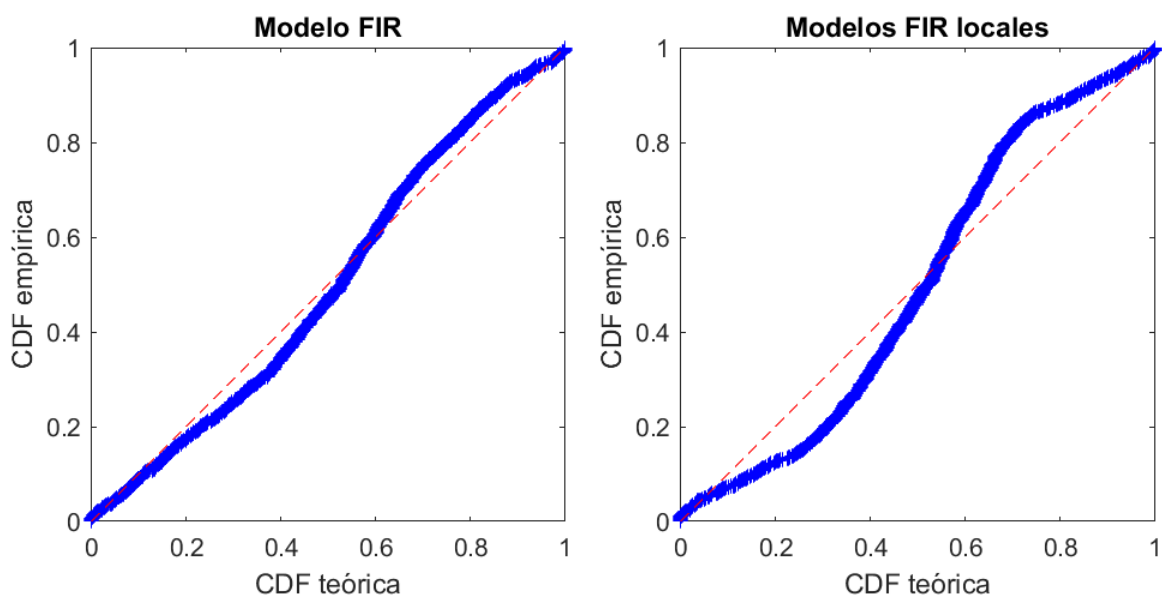
Por otro lado, la Figura 27 presenta los gráficos P-P correspondientes a los residuos de cada modelo, los cuales permiten evaluar el grado de normalidad de sus distribuciones. En este tipo de gráficos se compara la función de distribución acumulada (CDF) empírica de los residuos con la CDF teórica de una distribución normal. Un buen ajuste se manifiesta cuando los puntos se alinean sobre la diagonal de 45°, lo que indica que los residuos se distribuyen de manera aproximadamente normal.

En términos generales, el modelo ARX no ideal es el que exhibe el comportamiento más cercano a la normalidad, ya que sus puntos se alinean con mayor precisión a lo largo de la diagonal del gráfico P-P, especialmente en la zona central y con menores desviaciones en los extremos. Esta alineación indica que los residuos siguen, en buena medida, una distribución acumulada similar a la normal teórica. A pesar de ello, aún se observan desviaciones sutiles en los extremos, por lo que no puede considerarse una normalidad perfecta.

En segundo lugar, el modelo FIR global también muestra una aproximación razonable a la normalidad, aunque con desviaciones más marcadas que el modelo ARX no ideal, tanto en la zona central como en los extremos. Si bien el ajuste es aceptable, las discrepancias respecto a la línea diagonal indican una menor consistencia en los residuos con la distribución normal esperada.

En contraste, los modelos FIR locales presentan notorias desviaciones a lo largo de todo el gráfico, tanto en el centro como en los extremos, lo que sugiere una mayor divergencia respecto a la normalidad. Esto puede explicarse por la naturaleza segmentada de estos modelos, ya que, al estar calibrados sobre subconjuntos específicos del espacio de entrada, pueden capturar comportamientos particulares que no necesariamente siguen distribuciones normales. Aunque sus residuos tienden a ser simétricos y bien centrados (según su histograma), la forma de su distribución acumulada evidencia un claro alejamiento del comportamiento normal esperado.

Finalmente, el modelo ARX ideal muestra el patrón más evidente de desviación respecto a la normalidad, con curvaturas marcadas a lo largo de todo el gráfico. Esto indica que los residuos del modelo presentan una distribución acumulada significativamente distinta de la normal, posiblemente influenciada por la dinámica real del sistema y el uso de valores reales de salida como entradas. A pesar de tener residuos muy centrados y con baja dispersión (como se evidenció en su histograma), estos no siguen una distribución normal en sentido estricto.



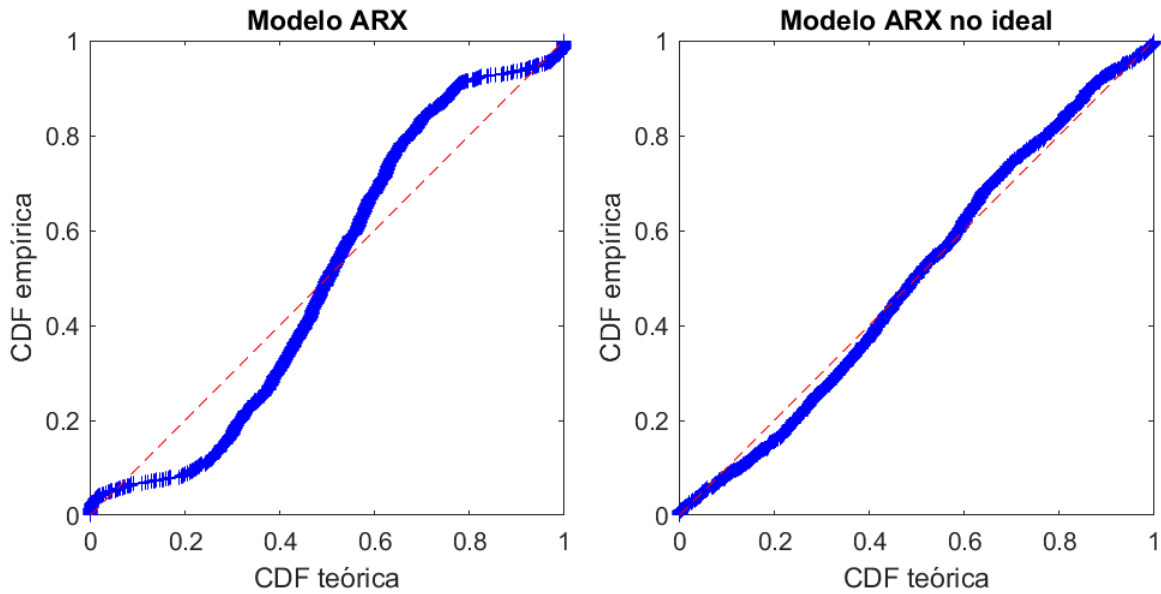


Figura 27. P-P plots de residuos: evaluación de la normalidad en los modelos.

En conjunto, los resultados de los gráficos P-P permiten evaluar la calidad estadística de los modelos desde la perspectiva de la normalidad. Se confirma que el modelo ARX no ideal presenta la mejor aproximación a una distribución normal, seguido por el FIR global. Por otro lado, tanto los FIR locales como el ARX ideal se alejan significativamente de la normalidad. Sin embargo, es importante destacar que un buen ajuste a la normalidad no siempre se traduce en un mejor desempeño predictivo. Por ejemplo, el modelo ARX ideal, a pesar de alejarse más de la normalidad según el gráfico P-P, mostró la mejor precisión y exactitud en términos de RMSE y dispersión acotada de residuos. En contraste, el modelo ARX no ideal, aunque con una forma de distribución acumulada más próxima a la normal, presenta un mayor error absoluto y una dispersión más amplia en los residuos, lo que evidencia una menor precisión en sus estimaciones. Esta observación refuerza la importancia de analizar múltiples dimensiones del desempeño del modelo para una evaluación completa, especialmente cuando se consideran aplicaciones como sensores virtuales en procesos industriales.

5.3. Análisis de propagación del error

En el contexto de los procesos de flotación, la medición en línea de variables metalúrgicas críticas como la ley de concentrado representa un desafío técnico y económico. Dada la dificultad de implementar sensores físicos que entreguen esta información en tiempo real de forma precisa y confiable, se han desarrollado modelos de tipo FIR y ARX, los cuales actúan como sensores virtuales capaces de estimar la ley de concentrado a partir de variables de proceso que sí son accesibles mediante instrumentación convencional.

Sin embargo, estos modelos dependen completamente de los valores entregados por sensores físicos, los cuales están sujetos a diversas fuentes de error, tales como ruido, deriva o fallas transitorias. Esto implica que errores en las mediciones de estas variables pueden traducirse en estimaciones incorrectas por parte del modelo, afectando la confiabilidad de la predicción entregada por el sensor virtual.

Ante esta situación, se llevó a cabo un análisis de propagación del error de medición, con el objetivo de evaluar la sensibilidad de la ley de concentrado estimada frente a errores individuales en las variables que alimentan a los modelos desarrollados (ARX y FIR). Para ello, se simulan perturbaciones controladas del 5%, 10%, 20% y 40% sobre cada variable, de forma individual, y se cuantifica el impacto generado en la salida del modelo. Este impacto se determina como la diferencia entre la predicción realizada con entradas sin error y aquella generada bajo el efecto del error de medición introducido.

Cabe destacar que este análisis se aplica sobre modelos de naturaleza lineal, como lo son los modelos FIR y ARX. En este contexto, un incremento proporcional en el error de una variable de entrada genera un impacto proporcional en la salida del modelo. Por ejemplo, al duplicar una perturbación del 5 % al 10 %, el efecto sobre la salida también se duplica. En este sentido, el impacto asociado a una perturbación del 40 % puede inferirse a partir del impacto generado por una perturbación del 5 %. Sin embargo, la inclusión explícita de perturbaciones del 10 %, 20 % y 40 % permite visualizar de forma directa el efecto de errores de mayor magnitud y, además, corroborar empíricamente el comportamiento lineal de los modelos desarrollados.

Este enfoque permite identificar cuáles variables son más críticas para la estabilidad del modelo y, por tanto, requieren mayor confiabilidad en su medición para no comprometer la calidad de la estimación entregada por el sensor virtual.

Para realizar este análisis fue necesario definir una condición de operación específica sobre la cual se aplicaron las perturbaciones, tanto para el modelo FIR como para el ARX. La Tabla 15 muestra la observación utilizada como referencia, seleccionada dentro del espacio de operación de la celda piloto.

Tabla 15. Observación utilizada en el análisis.

Q_F	S_f	X_f	G_f	J_g	H_f
3,5	30	0,65	20	0,75	2,5

Cabe destacar que, para este análisis, no se consideran todas las variables de entrada de cada modelo, sino únicamente aquellas que presentan una alta relevancia en su construcción, determinada a partir

del análisis de *VIP scores* (*Variable Importance in the Projection*) entregado por el *software* de simulación SOLO. Este índice permite identificar qué variables contribuyen significativamente a la capacidad predictiva del modelo, facilitando la selección de aquellas que son más relevantes para la ley de concentrado. Los resultados correspondientes al análisis de los modelos FIR y ARX se presentan en las Tabla 22 y Tabla 23, incluidas en el Anexo 3.

La Figura 28 presenta los resultados obtenidos al aplicar el análisis de propagación del error al modelo FIR. En este caso, las variables consideradas como entradas fueron: J_g (velocidad superficial del aire), porcentaje de gruesos, ley de alimentación y porcentaje de sólidos. Entre las variables analizadas, J_g se destaca como la más influyente, provocando un incremento de hasta +2,94 puntos porcentuales en la predicción de la ley de concentrado cuando el error de medición alcanza el 40%. El porcentaje de sólidos también muestra una contribución positiva relevante, aunque de menor magnitud que J_g .

En contraste, tanto el porcentaje de gruesos como la ley de alimentación inducen impactos negativos en la salida del modelo, con un efecto máximo de -1,87 puntos porcentuales en el caso de la ley de alimentación. Estos resultados indican que errores de medición en estas variables pueden llevar a una sobreestimación de la ley de concentrado, lo que puede comprometer la calidad del control del proceso si no se dispone de una instrumentación suficientemente precisa y confiable.

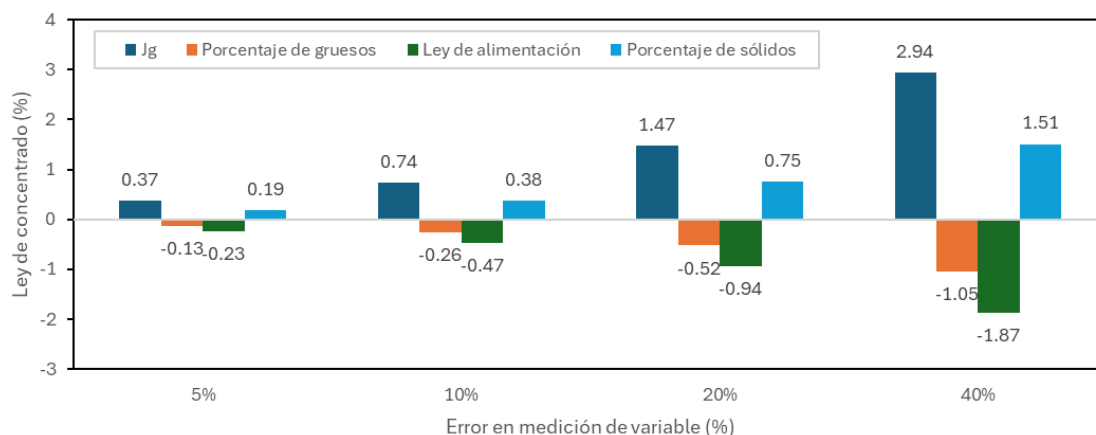


Figura 28. Propagación del error de medición en la ley de concentrado modelo FIR.

La Figura 29 presenta los resultados del análisis de propagación del error de medición aplicado al modelo ARX. A diferencia del modelo FIR, el modelo ARX incorpora como entrada adicional la ley de concentrado en el instante anterior ($t-1$), además de variables comunes como el flujo de aire superficial (J_g), la ley de alimentación y el porcentaje de sólidos. Aunque el modelo ARX puede considerar múltiples retardos de la salida (por ejemplo, hasta $t-2$), para este análisis se ha optado por evaluar exclusivamente el efecto del retardo inmediato ($t-1$). Esta decisión se fundamenta en dos

razones principales: en primer lugar, dicho retardo suele tener la mayor influencia sobre la predicción actual y, en segundo lugar, su inclusión permite representar de forma general el efecto dinámico sin complejizar innecesariamente el análisis.

Los resultados muestran que, tras incluir la variable de ley de concentrado en $(t-1)$, esta se convierte en la de mayor impacto en la estimación final, con un efecto de hasta $-3,41$ puntos porcentuales al introducir un error de medición del 40% . Este valor es el más alto entre todas las variables evaluadas, desplazando a J_g , que en el modelo FIR se identificó como la de mayor sensibilidad. Este cambio pone de manifiesto el papel fundamental de las salidas pasadas en los modelos ARX, donde la estimación actual se construye a partir de información generada previamente.

El impacto negativo observado indica que, ante errores de medición en la variable $(t-1)$, el modelo tiende a sobreestimar la ley de concentrado, lo cual puede comprometer la confiabilidad del sensor si no se gestiona adecuadamente. Por su parte, las otras tres variables (J_g , ley de alimentación y porcentaje de sólidos) mantienen el mismo orden de importancia observado en el análisis del modelo FIR, aunque con una magnitud de impacto algo menor. Por ejemplo, el impacto de J_g disminuye de $2,94\%$ a $1,99\%$, mientras que el de la ley de alimentación se reduce de $-1,87\%$ a $-1,44\%$. Esto evidencia cómo la inclusión de la variable de salida pasada no solo mejora la precisión del modelo, sino que también redistribuye la carga predictiva entre las distintas variables, reduciendo la sensibilidad del modelo a errores en las variables de entrada.

Finalmente, estos hallazgos destacan la necesidad de garantizar una buena calidad en los valores pasados de la variable de salida utilizados como entradas. Esto es especialmente relevante en la implementación del modelo ARX como sensor virtual no ideal, donde la alimentación se basa en predicciones anteriores. En este contexto, se vuelve imprescindible incorporar estrategias de monitoreo y reinicio periódico que limiten la acumulación de error, asegurando así una estimación confiable de la ley de concentrado.

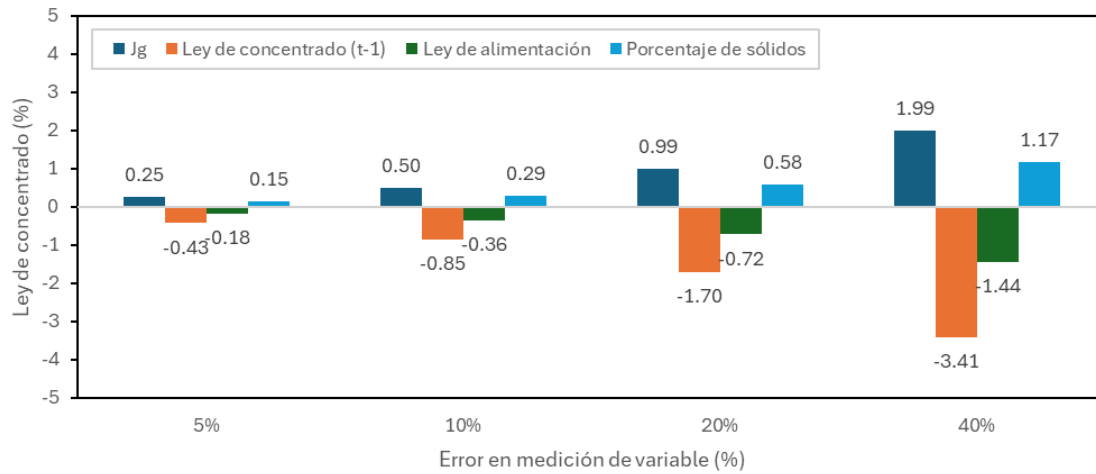


Figura 29. Propagación del error de medición en la ley de concentrado modelo ARX.

La comparación entre los modelos FIR y ARX evidencia un cambio en la influencia de las variables sobre la ley de concentrado. En el FIR, la variable más influyente es el flujo superficial de gas (J_g), mientras que, en el ARX, al incorporar la ley de concentrado en el instante anterior ($t-1$), esta se convierte en la de mayor impacto. Este ajuste otorga al ARX una mayor capacidad para estabilizar la predicción frente a perturbaciones transitorias, pero también implica el riesgo de acumulación de errores cuando se alimenta de sus propias estimaciones (implementación no ideal). Por ello, aunque el ARX presenta ventajas en robustez y estabilidad, requiere mecanismos de control y reinicio periódico para mantener su confiabilidad como sensor virtual en aplicaciones industriales continuas.

5.4. Comparación del comportamiento predictivo de los modelos

Para finalizar el desarrollo y evaluación de los sensores virtuales propuestos, la Figura 30 presenta una comparación visual entre la ley de concentrado estimada por cada uno de los modelos desarrollados (ARX, ARX no ideal, FIR y FIR locales) y los valores medidos experimentalmente. El modelo FIR global exhibe una dispersión moderada a alta a lo largo de todo el rango de operación del sensor. Se observa una tendencia marcada a la sobreestimación en leyes bajas y a la infraestimación en leyes altas. Este comportamiento sugiere una capacidad limitada de generalización ante variaciones en las condiciones del proceso, reflejando un desempeño más débil respecto al resto de los modelos.

En contraste, los modelos ARX ideal y FIR locales, especialmente el primero, presentan una concentración notable de puntos alrededor de la línea 1:1, indicando una menor dispersión y mayor precisión. El modelo ARX ideal destaca por su capacidad para reducir significativamente la variabilidad en todo el rango operativo, mientras que los FIR locales, si bien también mejoran el

ajuste, todavía presentan algunas desviaciones puntuales importantes. Esta mayor alineación con la diagonal de igualdad demuestra que ambos enfoques logran capturar de forma más efectiva la dinámica de la ley de concentrado, lo que los posiciona como alternativas confiables en condiciones controladas. El modelo ARX no ideal muestra una mayor dispersión que el FIR global, aunque con datos algo más alineados a la diagonal. En leyes bajas tiende a sobreestimar y en leyes altas a infraestimar, lo cual se explica por la propagación de errores autorregresivos al utilizar predicciones pasadas. Si bien es menos preciso, representa una configuración más realista para operación en planta, donde no se dispone de salidas reales.

En comparación, los modelos FIR locales y ARX ideal ofrecen el mejor desempeño. El ARX ideal resalta por su alta precisión, aunque su dependencia de datos reales limita su aplicabilidad. Los FIR locales, al no requerir retroalimentación, se adaptan mejor a distintas condiciones operacionales y muestran buena estabilidad. Por su parte, el ARX no ideal, aunque menos exacto, es más robusto frente a perturbaciones y errores en las entradas, lo que lo hace útil en escenarios con baja calidad de datos. En contraste, el FIR global se descarta por su alto nivel de error y sensibilidad. En definitiva, la elección práctica se resume en un compromiso: los FIR locales son recomendables con datos confiables, mientras que el ARX no ideal resulta más adecuado en contextos donde la robustez frente a perturbaciones es prioritaria.

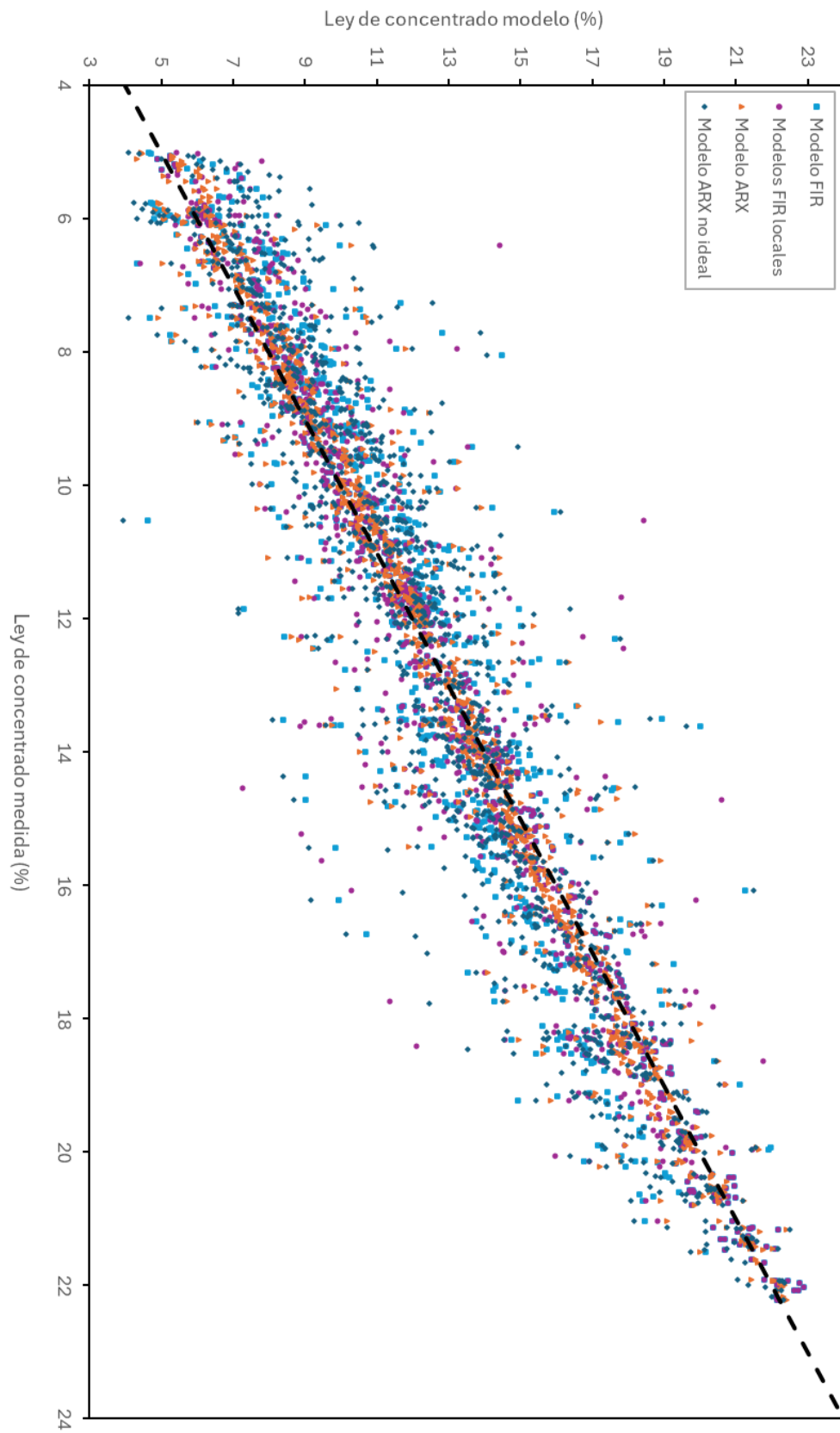


Figura 30. Comparación del comportamiento predictivo de los modelos.

6. Conclusiones

En esta memoria se desarrolló un sensor virtual dinámico para estimar la ley de cobre en una celda de flotación piloto tipo *rougher*, integrando modelos dinámicos clásicos (FIR y ARX) con técnicas multivariadas modernas (PLS). El uso de un simulador dinámico–estático acoplado, con seis variables (cuatro de alimentación y dos de operación) en tres niveles, permitió generar 1.283 observaciones para FIR y 1.287 para ARX, reproduciendo de forma controlada las dinámicas del proceso, aunque bajo supuestos que difieren de la operación real.

En cuanto al desempeño, el FIR global (seis *lags*) obtuvo un RMSE de 1,37, reduciéndose a 1,22 (–11,4 %) con tres modelos FIR locales, lo que disminuyó la dispersión de residuos. El ARX ideal (dos *lags*) logró el mejor ajuste (RMSE = 0,94) y residuos más simétricos, pero su dependencia de valores reales pasados limita su uso en línea. Para superar esta restricción se desarrolló un ARX no ideal, que usa sus propias predicciones y un reinicio cada 15,8 h, alcanzando un RMSE de 1,56, con mayor dispersión, pero viabilidad operativa.

El análisis estadístico con T^2 y Q residual destacó la necesidad de excluir observaciones atípicas, mejorando la estabilidad de los modelos. La sensibilidad y propagación de error identificaron al J_g como variable más influyente en FIR y al valor pasado de la ley de concentrado ($t-1$) en ARX, resaltando la importancia de su medición precisa.

Comparativamente, los modelos FIR locales y el ARX ideal fueron los más equilibrados en precisión y robustez. El FIR global fue descartado, mientras que el ARX no ideal, aunque menos preciso, resultó estable y aplicable en condiciones reales con una estrategia adecuada de reinicio. Ningún modelo presentó residuos perfectamente normales, pero sí simétricos y centrados, especialmente en FIR locales y ARX ideal.

En síntesis, la metodología demostró ser efectiva en el entorno de simulación considerado, evidenciando viabilidad técnica para el desarrollo de sensores virtuales dinámicos robustos y mostrando potencial para su futura implementación industrial, previa validación con datos reales de planta. La elección del modelo depende del balance entre precisión y estabilidad: FIR locales son recomendables con datos de alta calidad, y ARX no ideal es preferible cuando se requiere mayor robustez frente a perturbaciones o limitaciones de medición.

7. Recomendaciones

A partir de los resultados obtenidos en este trabajo, se identifican diversas oportunidades de mejora y proyecciones futuras que permitirían fortalecer y ampliar el alcance del sensor virtual desarrollado. Una primera recomendación consiste en validar los modelos construidos utilizando datos provenientes de una planta piloto o industrial en operación real. Si bien las metodologías empleadas resultan adecuadas para el modelamiento dinámico multivariable, su desempeño depende directamente de la calidad y representatividad de los datos utilizados. En este estudio, dichos datos fueron generados mediante un código de simulación que incorpora ciertos supuestos sobre el comportamiento dinámico de la variable de salida. La incorporación de datos experimentales reales permitiría ajustar los modelos a condiciones propias del sistema físico, considerando efectos como retardos adicionales, ruido de sensores, no linealidades del proceso y diferencias en las frecuencias de muestreo, aumentando así la robustez y aplicabilidad práctica de los modelos.

Asimismo, se recomienda establecer metodologías claras para la detección y descarte de observaciones atípicas, dado que estas pueden afectar significativamente la capacidad predictiva. El uso de estadígrafos como Hotelling's T^2 y Q residual, tal como se aplicó en este trabajo, debería integrarse de forma sistemática en el proceso de modelado para mejorar la confiabilidad del sensor virtual.

Otra recomendación es implementar los modelos FIR, FIR locales, ARX y ARX no ideal en estudios experimentales de laboratorio, bajo condiciones similares a la celda piloto simulada. Esto permitiría evaluar su desempeño frente a datos reales, verificando la validez de los supuestos de simulación y su comportamiento en escenarios más representativos. En particular, para los modelos ARX se sugiere evaluar el uso de un mayor número de retardos (*lags*), con el fin de reducir la dependencia de valores pasados de la salida estimada $y(t - l)$. Como se observó en este trabajo, cuando estas salidas no corresponden a valores reales sino a estimaciones del propio modelo, incorporan errores que se propagan en el tiempo.

También se sugiere ampliar el conjunto de variables de entrada del modelado, incorporando parámetros como pH, densidad de pulpa, granulometría o tipo de colector. Esto no solo podría mejorar la precisión de los modelos, sino también identificar variables clave que actualmente no son monitoreadas o controladas.

De igual forma, se recomienda profundizar en la aplicación de metodologías de modelos locales, considerando criterios de importancia de variables distintos al utilizado en este trabajo. En particular, podría evaluarse la relevancia de las variables a partir de su impacto sobre parámetros clave del

sistema, como el efecto del flujo de alimentación de pulpa (Q_F) en el tiempo de residencia (τ) o la influencia individual de cada variable (S_f, X_f, J_g, H_f , entre otras) sobre la ganancia (K_i o g_i), ya sea positiva o negativa. Asimismo, resulta de interés extender esta metodología de modelos locales a otras variables de salida y a modelos que siguen la metodología ARX, con el objetivo de comparar la calidad de ajuste obtenida frente a distintas configuraciones de modelado.

Otra proyección relevante corresponde a la implementación de análisis con ventanas deslizantes para el monitoreo continuo del desempeño del modelo ARX no ideal. Si bien en este trabajo dicho análisis se realizó en un entorno simulado, su aplicación en escenarios reales resulta especialmente atractiva en situaciones donde se disponga de sensores de ley de concentrado que presenten fallas intermitentes. En estos casos, el análisis permitiría evaluar el desempeño del modelo incluso frente a datos incompletos, posibilitando que el sensor virtual continúe operando y entregando estimaciones durante los periodos de indisponibilidad del sensor físico, hasta que este recupere su funcionamiento normal. Cabe destacar que este enfoque solo resulta inviable en plantas donde no exista ningún tipo de medición de la ley de concentrado, dado que el análisis requiere información real de la variable de salida para su correcta aplicación.

Finalmente, se plantea extender la aplicación del sensor virtual a otras variables metalúrgicas relevantes, como la recuperación o la ley de cobre en colas. El desarrollo de sensores virtuales multivariantes permitiría optimizar de forma simultánea distintos objetivos operacionales, potenciando el enfoque hacia sistemas de control más robustos y completos.

8. Referencias

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 433-459.
- Bravo Rain, G. L. (2017). *Sistema en línea de diagnóstico de estado de operación y detección de fallas en una planta piloto de flotación rougher*. Valparaíso: Tesis Magister Ingeniería Civil Química, UTFSM.
- Diaz, F., & Yianatos, J. (2010). Residence time distribution in large industrial flotation cells. *Atoms for Peace an International Journal*, 2-10.
- Durán, J. P. (2014). *Desarrollo de una estrategia de control para un circuito de flotación rougher piloto*. Valparaíso: Memoria Ingeniería Civil Química, UTFSM.
- Eigenvector Research Inc. (s.f.). *Advanced Machine Learning software, consulting & Training - Eigenvector*. Obtenido de Eigenvector: <https://eigenvector.com>
- Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikström, C., & Wold, S. (2006). *Multi- and Megavariate Data Analysis Part II: Advanced Applications and Method Extensions*. Umetrics Academy.
- Finch, J. A. (1995). Column Flotation: A selected review Part IV: Novel flotation devices. *Minerals Engineering*, 587-602.
- Finch, J. A., & Dobby, G. S. (1991). Column flotation: a selected review, part II. *Minerals Engineering*, 911-923.
- Frank, I. E., & Friedman, J. H. (1993). A Statistical View of Some Chemometrics Regression Tools. *Technometrics*, 109-135.
- Fuerstenau, M. C., Jameson, G., & Yoon, R. H. (2007). *Froth Flotation: A Century of Innovation*. SME.
- Gaudin, A. M. (1922). *Flotation*. New York: McGraw-Hill book company.
- Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, 1-17.

- Gharai, M., & Venugopal, R. (2016). Modeling of Flotation Process—An Overview of Different Approaches. *Mineral Processing and Extractive Metallurgy Review*, 120-133.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- Herrmann, C., & Zappettini, E. O. (2014). *Recursos minerales, minería y medio ambiente*. Buenos Aires: Instituto de geología y recursos minerales, SEGEMAR.
- Hodouin, D., Jämsä-Jounela, S. -L., Carvalho, M. T., & Bergh, L. (2001). State of the art and challenges in mineral processing control. *Control Engineering Practice*, 995-1005.
- Jackson, J. E. (2003). *A user's guide to principal components*. John Wiley & Sons.
- Jolliffe, I. T. (2002). *Principal Component Analysis (2da edición)*. Springer.
- Jong, S. (1993). SIMPLS: An efficient algorithm for partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 251-263.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 141-151.
- Kaspar, M. H., & Ray, W. H. (1992). Chemometric methods for process monitoring and high-performance controller design. *AIChE Journal*, 1593-1608.
- Kaspar, M. H., & Ray, W. H. (1993). Dynamic PLS modelling for process control. *Chemical Engineering Science*, 3447-3461.
- Kelly, E. G., & Spottiswood, D. J. (1982). *Introduction to Mineral Processing*. John Wiley & Sons.
- Ljung, L. (1999). *System Identification: Theory for the User (2da ed.)*. Prentice Hall.
- Lynch, A. J., Johnson, N. W., Manlapig, E. V., & Thorne, B. J. (1981). *Mineral and Coal Flotation Circuits: Their Simulation and Control*. Elsevier Scientific Publishing Company.
- MacGregor, J. F., & Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, 403-414.
- Martens, H., & Næs, T. (1992). *Multivariate Calibration*. John Wiley & Sons.

- Miettinen, T., Ralston, J., & Fornasiero, D. (2010). The limits of fine particle flotation. *Minerals Engineering*, 420-437.
- Ministerio de Minería. (2023). *Anuario de estadísticas del cobre y otros minerales*. Comisión Chilena del Cobre (COCHILCO).
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis (5ta ed.)*. John Wiley & Sons.
- Napier-Munn, T. J., Morrell, S., Morrison, R. D., & Kojovic, T. (1996). *Mineral Comminution Circuits: Their Operation and Optimisation*. Julius Kruttschnitt Mineral Research Centre, University of Queensland.
- Ndoro, T., & Witika, L. (2017). A Review of the Flotation of Copper Minerals. *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, 145-165.
- Ormeño Puentes, F. A. (2024). *Balance hídrico de procesos metalúrgicos molienda y flotación en planta concentradora, minera Los Pelambres ANTOFAGASTA MINERALS*. Concepción: Universidad de Concepción.
- Rojas Santibañez, R. R. (2023). *Desarrollo de un sensor virtual dinámico de ley de concentrado en una columna de flotación usando metodología PLS*. Valparaíso: Memoria Ingeniería Civil Química, UTFSM.
- Ronen, D., Sanders, C. F., Tan, H. S., Mort, P. R., & Doyle, F. J. (2009). Predictive Modeling of Key Process Variables in Granulation Processes based on Dynamic Partial Least Squares. *IFAC Proceedings Volumes*, 411-416.
- Schwarz, S., & Alexander, D. (2006). Gas dispersion measurements in industrial flotation cells. *Minerals Engineering*, 554-560.
- Villela, D., Kutscher, C., Castillo, E., & Cantallopis, J. (2017). *Sulfuros primarios: desafíos y oportunidades*. Comisión Chilena del Cobre (COCHILCO).
- Wills, B. A., & Finch, J. (2015). *Wills' Mineral Processing Technology An Introduction to the Practical Aspects of Ore Treatment and Mineral Recovery*. Butterworth-Heinemann.
- Wills, B. A., & Napier-Munn, T. (2006). *Wills' Mineral Processing Technology (7th ed.)*. Elsevier Science & Technology Books.

- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, 37-52.
- Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 109-130.
- Yianatos Bernardino, J. (2005). *Flotación de minerales*. Valparaíso: UTFSM.
- Yianatos, J., & Vinnett, L. (2015). *Flotación de minerales: Fundamentos, tecnología y aplicación*. Valparaíso: UTFSM.
- Yianatos, J., Bergh, L., & Aguilera, J. (2003). Flotation scale up: use of separability curves. *Minerals Engineering*, 347-352.
- Zambrano Ríos, B. A. (2024). *Control experto-difuso de atributos en espumas de flotación medidas por procesamiento de imágenes en un circuito rougher piloto*. Valparaíso: Tesis Magíster Ingeniería Civil Química, UTFSM.
- Zheng, X., Franzidis, J.-P., & Manlapig, E. (2004). Modelling of froth transportation in industrial flotation cells Part I. Development of froth transportation models for attached particles. *Minerals Engineering*, 981–988.

9. Anexos

9.1. Anexo 1: Coeficientes de regresión de los modelos

Tabla 16. Matriz de coeficientes de regresión del modelo FIR global.

Variables	Coeficiente asociado a cada variable $X_{FIR}(t - l)$						
	$(t - 6)$	$(t - 5)$	$(t - 4)$	$(t - 3)$	$(t - 2)$	$(t - 1)$	(t)
Q_F	0,0127	-0,0267	0,0048	-0,0088	0,0028	-0,0092	0
S_f	0,0039	-0,0219	-0,0067	-0,0094	-0,1091	-0,3562	0
X_f	-0,0660	-0,0261	0,0226	0,0676	0,1960	0,4637	0
G_f	-0,0545	-0,0464	0,0061	0,0185	0,0581	0,1606	0
J_g	0,1117	-0,0213	-0,0583	-0,0660	-0,1519	-0,4448	0
H_f	-0,0193	-0,0048	-0,0085	0,0049	0,0438	0,1298	0

Tabla 17. Matriz de coeficientes de regresión del modelo FIR local 1.

Variables	Coeficiente asociado a cada variable $X_{FIR}(t - l)$						
	$(t - 6)$	$(t - 5)$	$(t - 4)$	$(t - 3)$	$(t - 2)$	$(t - 1)$	(t)
Q_F	0,0041	-0,0281	-0,0082	-0,0139	0,0365	-0,0503	0
S_f	-0,0576	0,0318	0,0305	0,0491	-0,0936	-0,5194	0
X_f	-0,0784	-0,0080	0,0239	0,0330	0,1211	0,4991	0
G_f	-0,0308	-0,0078	-0,0302	0,0278	0,1181	0,2053	0
J_g	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0
H_f	-0,1359	0,0272	0,0369	0,0521	0,0842	0,2092	0

Tabla 18. Matriz de coeficientes de regresión del modelo FIR local 2.

Variables	Coeficiente asociado a cada variable $X_{FIR}(t - l)$						
	$(t - 6)$	$(t - 5)$	$(t - 4)$	$(t - 3)$	$(t - 2)$	$(t - 1)$	(t)
Q_F	0,0158	0,0408	-0,0055	-0,0272	-0,0289	-0,0641	0
S_f	-0,0255	0,0626	0,0261	-0,0077	-0,1409	-0,4103	0
X_f	0,0348	-0,0401	-0,0218	-0,0044	0,1897	0,6544	0
G_f	-0,0082	-0,0262	-0,0557	-0,0055	0,0409	0,1800	0
J_g	0,1478	0,0177	-0,0449	-0,1206	-0,1610	-0,2853	0
H_f	0,0519	0,0037	-0,0386	-0,0083	-0,0082	0,0524	0

Tabla 19. Matriz de coeficientes de regresión del modelo ARX.

Variables	Coeficiente asociado a cada variable $X_{ARX}(t - l)$		
	$(t - 2)$	$(t - 1)$	(t)
Q_F	-0,0036	-0,0167	0
S_f	0,0524	-0,2761	0
X_f	-0,0751	0,3557	0
G_f	-0,0218	0,0990	0
J_g	-0,0207	-0,2995	0
H_f	-0,0136	0,0866	0
X_c	-0,0383	0,5870	-

9.2. Anexo 2: Modelos locales

Tabla 20. Resumen estadístico del modelo FIR local 1.

Componente principal	Varianza bloque X	Varianza bloque X acumulativa	Varianza bloque Y	Varianza bloque Y acumulativa	RMSECV Ley de concentrado
1	21,15	21,15	56,80	56,80	2,5562
2	14,13	35,28	10,95	67,75	2,2461
3	8,96	44,24	7,24	75,00	2,005
4	11,41	55,65	3,56	78,55	1,8832
5	10,33	65,98	1,10	79,65	1,8645
6	4,74	70,72	1,66	81,30	1,7926
7	14,31	85,03	0,27	81,57	1,7899
8	3,11	88,14	0,50	82,07	1,7861
9	1,18	89,32	0,46	82,53	1,8038
10	1,47	90,79	0,13	82,67	1,8178
11	1,04	91,83	0,08	82,75	1,8357

Tabla 21. Resumen estadístico del modelo FIR local 2.

Componente principal	Varianza bloque X	Varianza bloque X acumulativa	Varianza bloque Y	Varianza bloque Y acumulativa	RMSECV Ley de concentrado
1	16,97	16,97	35,73	35,73	2,4089
2	12,20	29,17	18,74	54,47	2,0708
3	12,94	42,11	7,36	61,83	1,9244
4	7,66	49,78	5,66	67,50	1,8087
5	14,65	64,43	1,52	69,02	1,781
6	6,35	70,78	2,14	71,17	1,7456
7	9,57	80,35	0,89	72,05	1,7341
8	3,43	83,78	0,83	72,88	1,7259
9	2,16	85,94	0,66	73,54	1,7209
10	2,21	88,16	0,27	73,81	1,728
11	1,12	89,27	0,23	74,04	1,7439

9.3. Anexo 3: Relevancia de variables en modelos estimados según VIP scores

Tabla 22. Relevancia de variables según VIP scores en el modelo FIR.

Variables relevantes	Variables irrelevantes
Velocidad superficial del aire (J_g)	Flujo de alimentación de pulpa (Q_F)
Ley de cobre en alimentación (X_f)	Altura de espuma (H_f)
Porcentaje de solidos (S_f)	
Porcentaje de gruesos en alimentación (G_f)	

Tabla 23. Relevancia de variables según VIP scores en el modelo ARX.

Variables relevantes	Variables irrelevantes
Ley de concentrado pasada ($X_c(t-l)$)	Porcentaje de gruesos en alimentación (G_f)
Velocidad superficial del aire (J_g)	Flujo de alimentación de pulpa (Q_F)
Ley de cobre en alimentación (X_f)	Altura de espuma (H_f)
Porcentaje de solidos (S_f)	

9.4. Anexo 4: Gráficos de calibración y rangos operativos de variables clave

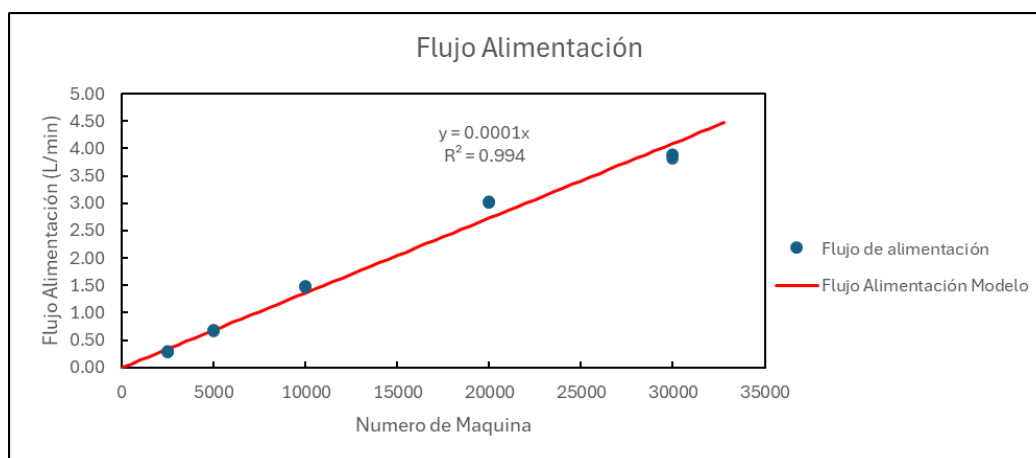


Figura 31. Calibración bomba de alimentación.

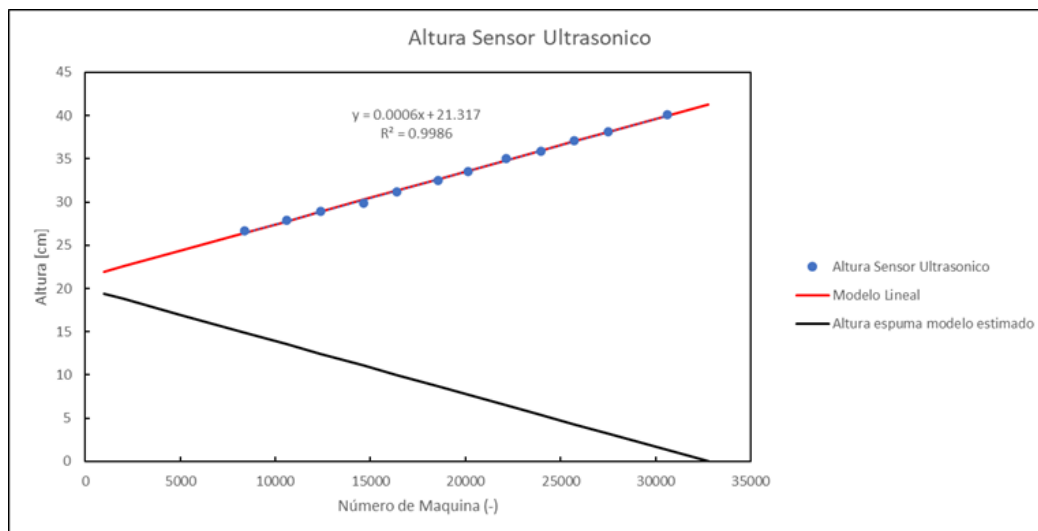


Figura 32. Calibración de nivel sensor ultrasónico

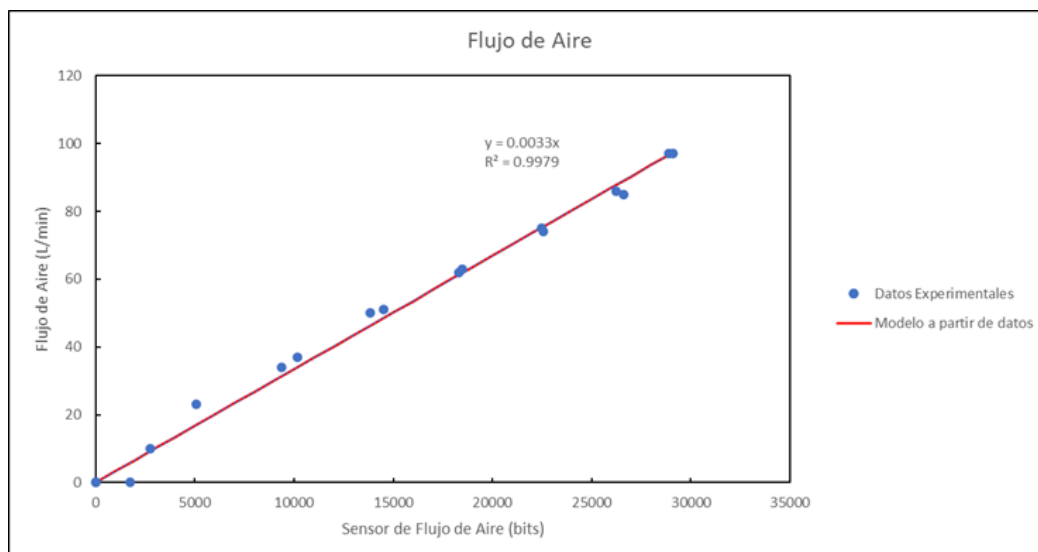


Figura 33. Calibración sensor flujo de aire.

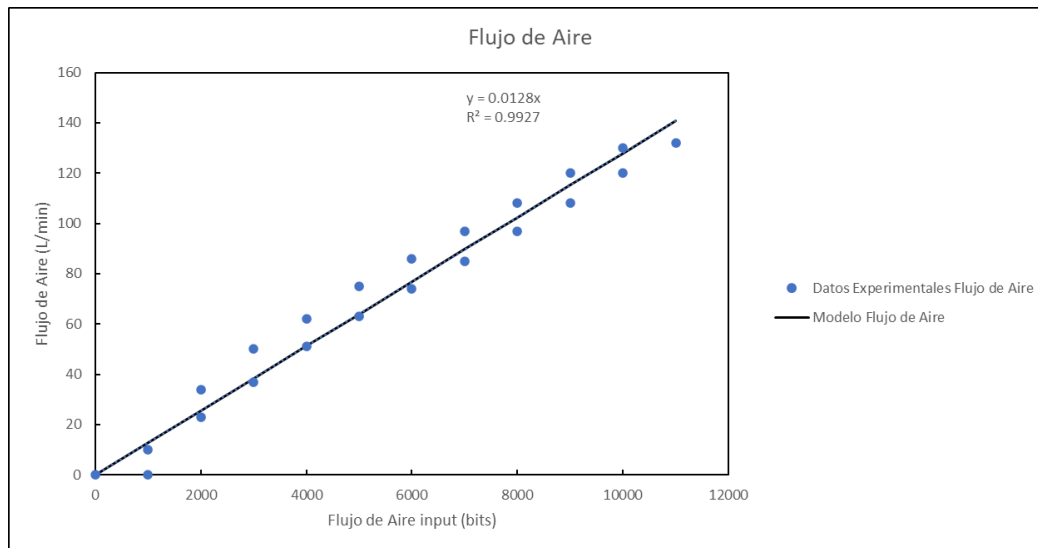


Figura 34. Relación flujo de aire con porcentaje de apertura de válvula.

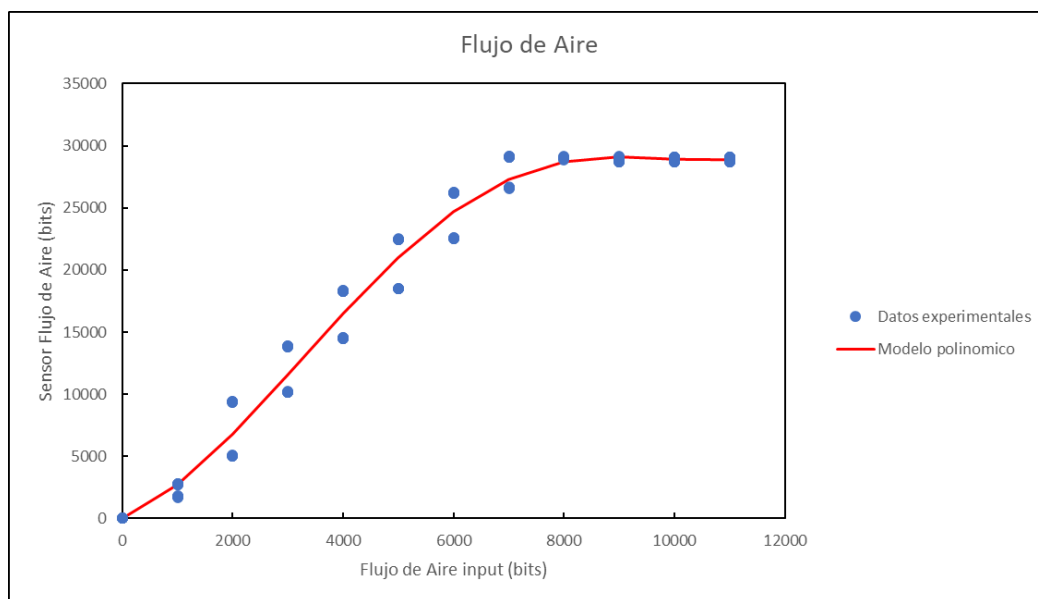


Figura 35. Calibración de relación entre sensor y válvula de aire.

9.5. Anexo digital 1

Los datos experimentales sin procesar utilizados en este trabajo se encuentran disponibles en formato Excel en el siguiente enlace:

https://docs.google.com/spreadsheets/d/1si99YxJFPxmxOwmb4upa6xIIP0W3IQen/edit?usp=drive_link&oid=107779268971931397782&rtpof=true&sd=true