

UNIVERSIDAD DE CONCEPCIÓN - CHILE  
FACULTAD DE INGENIERÍA  
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

Estimación cuantitativa de la composición mineralógica de  
concentrados de cobre mediante fusión de datos hiperespectrales  
VNIR–SWIR y regresión de vectores de soporte

*por*

Lorenzo Andres Muñoz Molina

*Profesor guía*

Dr. Daniel Gerónimo Sbarbaro Hofer.

*Profesores comisión*

Dr. Pablo Coelho Caro.

Dr. Sebastián Godoy.

Dra. Rosario del Pilar Castillo Felices.

Concepción, Junio de 2026

Tesis presentada a la

ESCUELA DE GRADUADOS  
DE LA UNIVERSIDAD DE CONCEPCIÓN



*para optar al grado de*

MAGÍSTER EN INGENIERÍA ELÉCTRICA

Estimación cuantitativa de la composición mineralógica de  
concentrados de cobre mediante fusión de datos hiperespectrales  
VNIR–SWIR y regresión de vectores de soporte

Lorenzo Andres Muñoz Molina

Una Tesis del  
Departamento de Ingeniería Eléctrica

Presentada en Cumplimiento Parcial de los Requerimientos del Grado de  
Magíster en ciencias con Mención en Ingeniería Eléctrica de la Escuela de  
Graduados de la Universidad de Concepción, Chile

Enero 2026

© Lorenzo Andres Muñoz Molina, 2026

# Resumen

Estimación cuantitativa de la composición mineralógica de concentrados de cobre mediante fusión de datos hiperespectrales VNIR–SWIR y regresión de vectores de soporte

Lorenzo Andres Muñoz Molina  
Universidad de Concepción, 2026

La composición mineralógica de concentrados de cobre influye directamente en el desempeño de los procesos pirometalúrgicos, particularmente en la fusión flash utilizada en fundiciones modernas. Variaciones en la proporción de fases sulfuradas pueden afectar la estabilidad térmica del horno y la eficiencia del proceso. Sin embargo, la caracterización del concentrado se realiza principalmente mediante técnicas de laboratorio fuera de línea, como análisis químicos o microscopía automatizada, que presentan altos costos y tiempos de respuesta prolongados, limitando su uso para apoyar la operación en tiempo real. En este contexto, resulta relevante disponer de técnicas rápidas y no destructivas que permitan estimar la composición del concentrado sin interferir en el material analizado. La espectroscopía de reflectancia difusa mediante sensores hiperespectrales permite obtener información composicional a partir de la interacción entre radiación electromagnética y propiedades mineralógicas del material, constituyendo una alternativa prometedora para el monitoreo continuo de concentrados. El presente trabajo propone una metodología para estimar cuantitativamente %Cu, %Calcopirita ( $\text{CuFeS}_2$ ) y %Pirita ( $\text{FeS}_2$ ) en concentrados de cobre utilizando datos hiperespectrales adquiridos en las regiones VNIR (400–1000 nm) y SWIR (900–1700 nm). Ambos dominios espectrales fueron analizados de forma independiente y posteriormente integrados mediante una estrategia de fusión de datos a nivel medio (Mid-Level Data Fusion, MLDF). Sobre esta información se desarrollaron modelos de regresión multivariada utilizando PLS y un enfoque híbrido no lineal PLS-SVR. Los espectros fueron preprocesados mediante filtrado Savitzky–Golay de primera derivada, corrección de dispersión multiplicativa (MSC), autoescalado y selección de longitudes de onda mediante el criterio VIP. Los resultados muestran que los modelos PLS-SVR presentan un desempeño superior al modelo lineal PLS, especialmente cuando se emplea la estrategia de fusión MLDF. En la etapa de predicción independiente, el modelo PLS-SVR MLDF alcanzó errores relativos de predicción (REP%) de aproximadamente 4.5% para %Cu, 5.3% para %Calcopirita ( $\text{CuFeS}_2$ ) y 13% para %Pirita ( $\text{FeS}_2$ ), con coeficientes de determinación ( $R^2$ ) de 0.87, 0.90 y 0.80, respectivamente. Estos resultados evidencian el potencial de la metodología propuesta para la caracterización rápida de concentrados de cobre mediante sensores hiperespectrales en aplicaciones industriales.

*A mi familia, novia y amigos más cercanos.*

# Agradecimientos

En primer lugar, agradezco a Dios por haberme dado la oportunidad de continuar estudiando y explorando nuevas oportunidades. Este ha sido un camino largo, incluso más de lo esperado, pero que ha valido completamente la pena, ya que me permitió aprender y superar muchos desafíos tanto en lo técnico como en la vida cotidiana.

Agradezco profundamente a mi familia por su apoyo incondicional en la decisión de seguir estudiando y acompañarme durante todo este proceso. También quiero agradecer a Sofía, mi novia, quien gracias a Dios logró titularse y hoy ejerce como doctora. Su apoyo fue fundamental, ya que muchas veces me ayudó a sobrellevar los momentos de ensayo y error propios del aprendizaje, además de brindarnos siempre apoyo mutuo.

Quiero agradecer también al fútbol del Interfacultades de la Universidad de Concepción, al equipo de Biomédica del Departamento de Ingeniería Eléctrica y el equipo Transmisión, quienes me brindaron un espacio recreativo muy importante durante este proceso y me permitieron disfrutar del deporte mientras avanzaba en mis estudios.

Asimismo, agradezco a la profesora Eugenia Araneda y al profesor Víctor Parra por su apoyo en mi trabajo en el área metalúrgica, y a Rodrigo Fuentes por enseñarme paso a paso conceptos fundamentales de óptica y fotónica.

De manera muy especial, agradezco a mi profesor guía de tesis, el Dr. Daniel Sbarbaro, por su apoyo, paciencia y constante disposición a escuchar y orientar durante este proceso. Me considero muy afortunado de haber tenido la oportunidad de aprender con él, no solo conocimientos académicos, sino también enseñanzas valiosas para la vida. Sus enseñanzas siempre permanecerán conmigo, y le deseo lo mejor a él y a su familia.

Finalmente, quiero agradecer también a mi yo del pasado, que a pesar de los momentos difíciles nunca se rindió y decidió seguir adelante.

# Tabla de Contenidos

|                                                                                       |             |
|---------------------------------------------------------------------------------------|-------------|
| <b>LISTA DE TABLAS .....</b>                                                          | <b>IX</b>   |
| <b>LISTA DE FIGURAS .....</b>                                                         | <b>X</b>    |
| <b>NOMENCLATURA.....</b>                                                              | <b>XIII</b> |
| <b>ABREVIACIONES .....</b>                                                            | <b>XV</b>   |
| <b>1. INTRODUCCIÓN .....</b>                                                          | <b>1</b>    |
| 1.1. INTRODUCCIÓN GENERAL .....                                                       | 1           |
| 1.2. PROBLEMÁTICA .....                                                               | 6           |
| 1.3. RELEVANCIA DE LAS COMPOSICIONES Y EL USO DE LAS CÁMARAS HIPERESPECTRALES.....    | 8           |
| 1.4. PREGUNTA DE INVESTIGACIÓN INICIAL .....                                          | 11          |
| <b>2. ESTADO DEL ARTE.....</b>                                                        | <b>12</b>   |
| 2.1. MÉTODOS DE CUANTIFICACIÓN:.....                                                  | 12          |
| 2.1.1 Regresión lineal múltiple por Mínimos Cuadrados Ordinarios (MLR-OLS) .....      | 12          |
| 2.1.2 Regresión de Componentes Principales (PCR) .....                                | 14          |
| 2.1.3 Regresión por Mínimos Cuadrados Parciales (PLS) .....                           | 15          |
| 2.1.4 Redes neuronales (ANN).....                                                     | 16          |
| 2.1.5 Random Forest Regression (RFR).....                                             | 20          |
| 2.1.6 Extreme gradient boosting (XGB). .....                                          | 22          |
| 2.1.7 Error-in-variables (EIV) en las Regresiones. ....                               | 23          |
| 2.1.8 Support Vector Machine Regresion (SVR). ....                                    | 25          |
| 2.2. CONSIDERACIONES ADICIONALES EN LAS REGIONES DE INTERÉS (ROI).....                | 30          |
| 2.3. CONSIDERACIONES EN LA REDUCCIÓN DE DIMENSIONALIDAD. ....                         | 31          |
| 2.4. ¿QUÉ TIPO DE TÉCNICA DE REGRESIÓN ES LA MÁS ADECUADA AL CONTEXTO? .....          | 32          |
| 2.5. FUSION DE DATOS. ....                                                            | 34          |
| 2.6. PREPROCESAMIENTOS.....                                                           | 36          |
| 2.6.1 Centrado en la media (Mean centering).....                                      | 37          |
| 2.6.2 Autoescalado (Autoscaling).....                                                 | 38          |
| 2.6.3 Suavizado y derivadas de Savitzky–Golay (SG) .....                              | 38          |
| 2.6.4 Corrección de dispersión multiplicativa (MSC) .....                             | 40          |
| 2.6.5 Normalización Standard Normal Variate (SNV) .....                               | 41          |
| 2.6.6 Selección de bandas espectrales relevantes .....                                | 42          |
| 2.7. MÉTRICAS DE EVALUACIÓN DE LAS REGRESIONES.....                                   | 44          |
| 2.7.1 Coeficiente de determinación ( $R^2$ ) .....                                    | 45          |
| 2.7.2 Root Mean Square Error (RMSE) .....                                             | 45          |
| 2.7.3 Normalized Root Mean Square Error (NRMSE) por rango. ....                       | 45          |
| 2.7.4 Ratio of Performance to Deviation (RPD).....                                    | 46          |
| 2.7.5 Relative Error of Prediction (REP%) .....                                       | 47          |
| 2.7.6 Range Error Ratio (RER) .....                                                   | 47          |
| 2.8. CONCLUSIONES ESTADO DEL ARTE. ....                                               | 48          |
| <b>3. PROPUESTA DE TESIS PLANTEADA.....</b>                                           | <b>50</b>   |
| 3.1. PLANTEAMIENTO DE LA HIPÓTESIS DE TRABAJO. ....                                   | 50          |
| 3.2. JUSTIFICACIÓN DE LA HIPÓTESIS DE TRABAJO. ....                                   | 51          |
| 3.3. OBJETIVO GENERAL.....                                                            | 53          |
| 3.4. OBJETIVOS ESPECÍFICOS. ....                                                      | 53          |
| 3.4.1 Evaluación de modelos en dominios espectrales individuales.....                 | 54          |
| 3.4.2 Implementación de estrategias de fusión de datos espectrales.....               | 54          |
| 3.4.3 Comparación del desempeño predictivo con y sin integración de información ..... | 54          |
| 3.4.4 Análisis de la influencia del tipo de modelo de regresión .....                 | 54          |
| 3.4.5 Interpretación de resultados en función de la información espectral .....       | 54          |

|           |                                                                                               |            |
|-----------|-----------------------------------------------------------------------------------------------|------------|
| <b>4.</b> | <b>ALCANCES Y LIMITACIONES .....</b>                                                          | <b>55</b>  |
| 4.1.      | SENSORES A UTILIZAR.....                                                                      | 55         |
| 4.1.1     | Camara VNIR .....                                                                             | 56         |
| 4.1.2     | Desarrollo del sistema de adquisición de imagenes VNIR. ....                                  | 59         |
| 4.1.3     | Camara SWIR.....                                                                              | 62         |
| 4.2.      | CANTIDAD Y TIPOS DE MUESTRAS A MEDIR.....                                                     | 63         |
| 4.3.      | TÉCNICA DE REFERENCIA PARA LAS COMPOSICIONES MINERALÓGICAS Y ELEMENTALES.....                 | 64         |
| 4.4.      | USO DE SOFTWARE.....                                                                          | 66         |
| <b>5.</b> | <b>METODOLOGÍA EXPERIMENTAL Y RESULTADOS.....</b>                                             | <b>68</b>  |
| 5.1.      | PREPARACIÓN DE LAS MUESTRAS: .....                                                            | 68         |
| 5.1.1     | Fabricación de briquetas y obtención de los datos de referencia .....                         | 69         |
| 5.1.2     | Distribución de las muestras preparadas. ....                                                 | 71         |
| 5.2.      | ADQUISICIÓN DE LOS ESPECTROS:.....                                                            | 73         |
| 5.2.1     | Espectros VNIR.....                                                                           | 74         |
| 5.2.2     | Espectros SWIR .....                                                                          | 76         |
| 5.3.      | ANÁLISIS EXPLORATORIO PCA.....                                                                | 77         |
| 5.3.1     | PCA VNIR.....                                                                                 | 78         |
| 5.3.2     | PCA SWIR.....                                                                                 | 81         |
| 5.3.3     | Conclusiones análisis PCA.....                                                                | 84         |
| 5.4.      | PREPROCESAMIENTO Y CONSTRUCCIÓN DE LA FUNCIÓN DE PERDIDA: .....                               | 85         |
| 5.4.1     | Primer preprocesamiento: Savgol.....                                                          | 85         |
| 5.4.2     | Segundo preprocesamiento MSC: .....                                                           | 86         |
| 5.4.3     | Tercer preprocesamiento autoescalado: .....                                                   | 87         |
| 5.4.4     | Cuarto preprocesamiento VIP-kPLS:.....                                                        | 88         |
| 5.5.      | ESTRATEGIA DE FUSIÓN DE DATOS.....                                                            | 89         |
| 5.6.      | CONSTRUCCIÓN DE LOS MODELOS DE REGRESIÓN.....                                                 | 91         |
| 5.6.1     | Construcción PLS:.....                                                                        | 92         |
| 5.6.2     | Construcción PLS-SVR:.....                                                                    | 93         |
| 5.7.      | MÉTRICAS DE DESEMPEÑO:.....                                                                   | 95         |
| 5.8.      | ENTRENAMIENTO Y VALIDACIÓN DE LOS MODELOS .....                                               | 97         |
| 5.9.      | OPTIMIZACIÓN DE LA FUNCIÓN OBJETIVO. ....                                                     | 99         |
| 5.10.     | DEFINICIÓN DE LOS HIPERPARÁMETROS:.....                                                       | 101        |
| 5.10.1    | Hiperparametros Modelos PLS. ....                                                             | 101        |
| 5.10.2    | Hiperparametros Modelos PLS-SVR.....                                                          | 101        |
| 5.10.3    | Modelos MLDF.....                                                                             | 102        |
| 5.11.     | ESTIMACIÓN DE LA INCERTIDUMBRE. ....                                                          | 103        |
| 5.12.     | RESULTADOS .....                                                                              | 105        |
| 5.13.     | ANÁLISIS DE DESEMPEÑO PARA LA ESTIMACIÓN DE %CU .....                                         | 109        |
| 5.13.1    | Desempeño %Cu con fusión de datos (MLDF) .....                                                | 109        |
| 5.13.2    | Desempeño %Cu con sensores individuales (VNIR y SWIR).....                                    | 110        |
| 5.14.     | ANÁLISIS DE DESEMPEÑO PARA LA ESTIMACIÓN DE %CALCOPIRITA .....                                | 111        |
| 5.14.1    | Desempeño %Calcopirita con fusión de datos (MLDF).....                                        | 111        |
| 5.14.2    | Desempeño %Calcopirita con sensores individuales (VNIR y SWIR).....                           | 111        |
| 5.15.     | ANÁLISIS DE DESEMPEÑO PARA LA ESTIMACIÓN DE %PIRITA .....                                     | 112        |
| 5.15.1    | Desempeño %Pirita con fusión de datos (MLDF).....                                             | 112        |
| 5.15.2    | Desempeño %Pirita con sensores individuales (VNIR y SWIR).....                                | 113        |
| 5.16.     | VISUALIZACIÓN DEL DESEMPEÑO PREDICTIVO: MEJORES MODELOS PARA %CU, %CALCOPIRITA Y %PIRITA..... | 113        |
| 5.17.     | COMPARACION GENERAL DE TODOS LOS MODELOS.....                                                 | 116        |
| 5.18.     | ANÁLISIS TRANSVERSAL DEL DESEMPEÑO Y CAPACIDAD DE GENERALIZACIÓN .....                        | 117        |
| <b>6.</b> | <b>CONCLUSIONES .....</b>                                                                     | <b>119</b> |
| 6.1.      | SUMARIO .....                                                                                 | 119        |
| 6.2.      | CONCLUSIONES .....                                                                            | 120        |
| 6.3.      | TRABAJO FUTURO.....                                                                           | 122        |

|                                                  |            |
|--------------------------------------------------|------------|
| <b>REFERENCIAS .....</b>                         | <b>124</b> |
| <b>A. ANEXO TABLAS Y GRÁFICOS ASOCIADOS.....</b> | <b>133</b> |
| A.1. TABLAS DE HIPERPARAMETROS ASOCIADOS .....   | 133        |
| A.2. GRAFICAS DE LOS MODELOS .....               | 135        |
| A.3. DATASHEET DE LOS SENSORES VNIR Y SWIR ..... | 143        |

# Lista de Tablas

|                                                                                                                                                                   |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Tabla 1 Interpretación de las métricas de desempeño predictivo en concentrados minerales.....                                                                     | 97  |
| Tabla 2 Hiperparámetros optimizados según la configuración de modelado.....                                                                                       | 102 |
| Tabla 3 Resultados de las Métricas de Desempeño (RMSE, $R^2$ , REP%, RPD, RER) para la Estimación de %Cu según Modelo, Sensor y Etapa de Evaluación.....          | 106 |
| Tabla 4 Resultados de las Métricas de Desempeño (RMSE, $R^2$ , REP%, RPD, RER) para la Estimación de %Calcopirita según Modelo, Sensor y Etapa de Evaluación..... | 107 |
| Tabla 5 Resultados de las Métricas de Desempeño (RMSE, $R^2$ , REP%, RPD, RER) para la Estimación de %Pirita según Modelo, Sensor y Etapa de Evaluación.....      | 108 |
| Tabla A.1 Hiperparámetros Óptimos para la Estimación de %Cu según Modelo y Configuración Espectral.<br>.....                                                      | 133 |
| Tabla A.2 Hiperparámetros Óptimos para la Estimación de %Calcopirita según Modelo y Configuración Espectral. ....                                                 | 134 |
| Tabla A.3 Hiperparámetros Óptimos para la Estimación de %Pirita según Modelo y Configuración Espectral.<br>.....                                                  | 134 |

# Lista de Figuras

|                                                                                                                                                                                                                                                                                                                                                                                                                |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 1.1 Esquema conceptual simplificado de un horno de fusión flash para concentrados de cobre.. Basado en[4,5] .....                                                                                                                                                                                                                                                                                       | 2  |
| Figura 1.2 Representación industrial de una instalación de fusión flash para concentrados de cobre. El sistema incluye la zona de reacción donde se alimenta el concentrado junto con aire u oxígeno enriquecido, la cámara de sedimentación para la separación de mata y escoria, y los sistemas asociados de manejo de gases. Fuente: Metso (adaptado de material técnico corporativo). Basado en [78] ..... | 4  |
| Figura 2.1 esquema de una neurona artificial, elaboración propia basado en [27]. .....                                                                                                                                                                                                                                                                                                                         | 17 |
| Figura 2.2. Estructura de un MLP [28].....                                                                                                                                                                                                                                                                                                                                                                     | 19 |
| Figura 2.3. Esquema visual de cómo funciona el algoritmo random forest modificado de: <a href="https://levelup.gitconnected.com/random-forest-regression-209c0f354c84">[https://levelup.gitconnected.com/random-forest-regression-209c0f354c84.]</a> .....                                                                                                                                                     | 21 |
| Figura 2.4. Representación esquemática del algoritmo XGBoost [48,49]. .....                                                                                                                                                                                                                                                                                                                                    | 22 |
| Figura 2.5. diferencia entre las minimizaciones OLS y TLS, Elaboración propia basada en [55].....                                                                                                                                                                                                                                                                                                              | 24 |
| Figura 2.6. Comparación conceptual entre la regresión OLS (izquierda) y la regresión SVR (derecha). Basado en [37]. .....                                                                                                                                                                                                                                                                                      | 26 |
| Figura 2.7 Metodología clásica para representar una muestra en modelos de regresión utilizando imágenes hiperespectrales [22].....                                                                                                                                                                                                                                                                             | 30 |
| Figura 2.8 Esquema conceptual de los distintos niveles de fusión de datos: fusión a bajo nivel (LLDF), fusión a nivel medio (MLDF) y fusión a alto nivel (HLDF). En LLDF se combinan directamente los datos originales, en MLDF se integran características extraídas de cada fuente y en HLDF se fusionan las salidas de modelos independientes. ....                                                         | 35 |
| Figura 4.1 muestra la cámara hiperespectral VNIR HSI V4 utilizada en este trabajo junto con sus principales parámetros ópticos y geométricos. ....                                                                                                                                                                                                                                                             | 57 |
| Figura 4.2 ilustra esquemáticamente el procedimiento de calibración espectral, la identificación de las líneas de emisión de la lámpara CFL y el ajuste de la regresión utilizada para la asignación longitud de onda-píxel. ....                                                                                                                                                                              | 59 |
| Figura 4.3 Sistema experimental de adquisición hiperespectral VNIR: (A) mesa óptica; (B) plataforma lineal de escaneo Zaber Technologies X-ASR100B120B-SE03D12; (C) blanco estándar de dióxido de titanio (TiO <sub>2</sub> ); (D) cámara hiperespectral VNIR; (E) focos halógenos de tungsteno (2700 K); (F) fuente de iluminación LED en el rango 400–450 nm. ....                                           | 61 |
| Figura 4.4 Sistema experimental de adquisición hiperespectral SWIR: (A) cámara hiperespectral Pika IR+ (900–1700 nm); (B) fuente de iluminación tipo Fiber Line; (C) plataforma de traslación lineal. El sistema completo es operado mediante el software Spectronon Pro. ....                                                                                                                                 | 63 |
| Figura 4.5 Flujo de trabajo del sistema QEMSCAN para la cuantificación mineralógica y elemental. Basada en [71]. ....                                                                                                                                                                                                                                                                                          | 65 |

|                                                                                                                                                                                                                    |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figura 5.1 Proceso de preparación de muestras y fabricación de briquetas para análisis QEMSCAN. (Elaboración propia).....                                                                                          | 70  |
| Figura 5.2 Briqueta de concentrado de cobre preparada para análisis QEMSCAN. (muestra tomada con microscopio).....                                                                                                 | 70  |
| Figura 5.3 Grafica de la distribución porcentual para %Cu (azul), %Calcopirita (naranja) y %pirita (amarillo). .....                                                                                               | 72  |
| Figura 5.4 Reconstrucción recortada (área roja) para una muestra concentrada. ....                                                                                                                                 | 73  |
| Figura 5.5 Espectros VNIR de las muestras de concentrado de cobre. ....                                                                                                                                            | 74  |
| Figura 5.6 Eficiencia cuántica del sensor UI-3360CP-NIR-GL (Rev. 2) [72]. ....                                                                                                                                     | 75  |
| Figura 5.7 Espectros SWIR de las muestras de concentrado de cobre. ....                                                                                                                                            | 76  |
| Figura 5.8 Varianza explicada del PCA aplicado a los espectros VNIR. ....                                                                                                                                          | 78  |
| Figura 5.9 Scores PC1–PC2 del PCA aplicado a los espectros VNIR. ....                                                                                                                                              | 79  |
| Figura 5.10 Loadings PC1 y PC2 del PCA aplicado a los espectros VNIR. ....                                                                                                                                         | 80  |
| Figura 5.11 Varianza explicada del PCA aplicado a los espectros VNIR. ....                                                                                                                                         | 81  |
| Figura 5.12 Loadings PC1–PC2 del PCA aplicado a los espectros SWIR. ....                                                                                                                                           | 82  |
| Figura 5.13 Scores PC1–PC2 del PCA aplicado a los espectros SWIR. ....                                                                                                                                             | 83  |
| Figura 5.14 Estrategia planteada para la fusión MLDF.[Elaboración propia] .....                                                                                                                                    | 91  |
| Figura 5.15 Esquema del proceso de entrenamiento, validación y predicción de los modelos.[Elaboración propia] .....                                                                                                | 99  |
| Figura 5.16. Desempeño del modelo PLS-SVR (MLDF) para %Cu en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+. ....          | 114 |
| Figura 5.17. Desempeño del modelo PLS-SVR (MLDF) para %Calcopirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+. .... | 114 |
| Figura 5.18. Desempeño del modelo PLS-SVR (MLDF) para %Pirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+. ....      | 115 |
| Figura A.1 Desempeño del modelo PLS-SVR (VNIR) para %Cu en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.....             | 135 |
| Figura A.2 Desempeño del modelo PLS-SVR (SWIR) para %Cu en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.....             | 136 |
| Figura A.3 Desempeño del modelo PLS (MLDF) para %Cu en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.....                 | 136 |

|                                                                                                                                                                                                                                                 |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figura A.4 Desempeño del modelo PLS (VNIR) para %Cu en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.....                                              | 137 |
| Figura A.5 Desempeño del modelo PLS (SWIR) para %Cu en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.....                                              | 137 |
| Figura A.6 Desempeño del modelo PLS-SVR (VNIR) para %Calcopirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+. ....                                | 138 |
| Figura A.7 Desempeño del modelo PLS-SVR (SWIR) para %Calcopirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+. ....                                | 138 |
| Figura A.8 Desempeño del modelo PLS (MLDF) para %Calcopirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+. ....                                    | 139 |
| Figura A.9 Desempeño del modelo PLS (VNIR) para %Calcopirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+. ....                                    | 139 |
| Figura A.10 Desempeño del modelo PLS (SWIR) para %Calcopirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+. ....                                   | 140 |
| Figura A.11 Desempeño del modelo PLS-SVR (VNIR) para %Pirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+. ....                                    | 140 |
| Figura A.12 Desempeño del modelo PLS-SVR (SWIR) para %Pirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+. ....                                    | 141 |
| Figura A.13 Desempeño del modelo PLS (MLDF) para %Pirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.....                                         | 141 |
| Figura A.14 Desempeño del modelo PLS (VNIR) para %Pirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.....                                         | 142 |
| Figura A.15 Desempeño del modelo PLS (SWIR) para %Pirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.....                                         | 142 |
| Figura A.16 Extracto del datasheet del sensor CMOS monocromático utilizado en la cámara hiperespectral VNIR, incluyendo sus principales especificaciones técnicas y la respuesta espectral cuántica en el rango visible–infrarrojo cercano..... | 143 |
| Figura A.17 Extracto del datasheet de la cámara hiperespectral Pika IR+, mostrando el rango espectral SWIR (900–1700 nm), sus principales características técnicas y ejemplos de respuesta espectral. ....                                      | 144 |

# Nomenclatura

## Matrices

|                    |                                                                                           |
|--------------------|-------------------------------------------------------------------------------------------|
| $X$                | : matriz de datos espectrales originales, de dimension $n \cdot m$ .                      |
| $X_{VNIR}$         | : matriz espectral correspondiente al rango VNIR, de dimensión $n \cdot m_{VNIR}$ .       |
| $X_{SWIR}$         | : matriz espectral correspondiente al rango SWIR, de dimensión $n \cdot m_{SWIR}$ .       |
| $\tilde{X}$        | : matriz de datos espectrales preprocesados, de dimensión $n \cdot k$ .                   |
| $\tilde{X}_{MLDF}$ | : matriz de la fusión de datos a nivel medio (MLDF), de $n \cdot (A_{VNIR} + A_{SWIR})$ . |
| $W$                | : matriz de pesos del modelo PLS.                                                         |
| $P$                | : matriz de cargas (loadings) del modelo PLS.                                             |

## Vectores

|                       |                                                                                                                        |
|-----------------------|------------------------------------------------------------------------------------------------------------------------|
| $x_i$                 | : vector espectral de la $i$ -ésima muestra, de dimensión, $1 \cdot m$ , $x_i = [x_{i,1} \ x_{i,2} \ \dots \ x_{i,n}]$ |
| $\bar{x}$             | : Vector espectral promedio del conjunto de calibración.                                                               |
| $Y$                   | : vector de valores reales de la variable respuesta, de dimensión $n \cdot 1$ .                                        |
| $\hat{Y}$             | : vector de valores predichos por el modelo, de dimensión $n \cdot 1$                                                  |
| $t_k$                 | : <i>scores</i> correspondientes al $k$ -ésimo componente latente de PLS, de dimensión $n \cdot 1$                     |
| $\tilde{\beta}_{OLS}$ | : vector de regresion lineal multiple (MLR).                                                                           |
| $\tilde{\beta}_{PCR}$ | : vector de regresion por componentes principales (PCR).                                                               |
| $\hat{Y}_{(-i)}$      | : Vector de predicciones obtenido al excluir la muestra $i$ del entrenamiento (LOO).                                   |

## Escalares

|             |                                                                                    |
|-------------|------------------------------------------------------------------------------------|
| $n$         | : Número de muestras.                                                              |
| $m$         | : Número de variables espectrales.                                                 |
| $A$         | : Número de componentes principales o variables latentes retenidas.                |
| $a$         | : Índice de la componente latente.                                                 |
| $x_{ij}$    | : Valor de la variable espectral $j$ en la muestra $i$ .                           |
| $\bar{x}_j$ | : Media de la variable espectral $j$ , calculada sobre el conjunto de calibración. |
| $\bar{x}_i$ | : Media del espectro $x_i$ .                                                       |
| $s_j$       | : Desviación estándar de la variable espectral $j$ .                               |
| $s_i$       | : Desviación estándar del espectro $x_i$ .                                         |
| $a_i$       | : Coeficiente multiplicativo de dispersión para la muestra $i$ (MSC).              |
| $b_i$       | : Término aditivo de dispersión para la muestra $i$ (MSC).                         |
| $p$         | : Grado del polinomio local.                                                       |
| $d$         | : Orden de la derivada ( $d = 0$ corresponde a suavizado).                         |
| $w$         | : Tamaño de la ventana del filtro.                                                 |
| $c_i^{(d)}$ | : Coeficiente del filtro Savitzky–Golay para la derivada de orden $d$ .            |
| $VIP_j$     | : Valor de importancia de la variable espectral $j$ .                              |
| $SSY_a$     | : Varianza de la variable respuesta explicada por la componente latente $a$ .      |
| $w_{j,a}$   | : Peso de la variable $j$ en la componente latente $a$ .                           |

|              |                                                                                 |
|--------------|---------------------------------------------------------------------------------|
| $y_i$        | : Valor de referencia correspondiente a la muestra $i$ .                        |
| $\hat{y}_i$  | : Valor predicho por el modelo para la muestra $i$ .                            |
| $\bar{y}$    | : Valor medio de los datos de referencia.                                       |
| $y_{max}$    | : Valor máximo de la variable de referencia.                                    |
| $y_{min}$    | : Valor mínimo de la variable de referencia.                                    |
| $s_y$        | : Desviación estándar de los valores de referencia.                             |
| $R^2_c$      | : Coeficiente de determinación en Calibración.                                  |
| $R^2_{cv}$   | : Coeficiente de determinación en Validación cruzada.                           |
| $R^2_v$      | : Coeficiente de determinación en Validación externa.                           |
| $R_p$        | : Coeficiente de determinación en Predicción.                                   |
| $RMSE_c$     | : <i>Root Mean Square Error</i> en Calibración.                                 |
| $RMSE_{cv}$  | : <i>Root Mean Square Error</i> en Validación cruzada.                          |
| $RMSE_v$     | : <i>Root Mean Square Error</i> en Validación externa.                          |
| $RMSE_p$     | : <i>Root Mean Square Error</i> en Predicción.                                  |
| $RPD_c$      | : <i>Ratio of Performance to Deviation</i> en Calibración.                      |
| $RPD_{cv}$   | : <i>Ratio of Performance to Deviation</i> en Validación cruzada.               |
| $RPD_v$      | : <i>Ratio of Performance to Deviation</i> en Validación externa.               |
| $RPD_p$      | : <i>Ratio of Performance to Deviation</i> en Predicción.                       |
| $REP\%_c$    | : <i>Relative Error of Prediction</i> en Calibración.                           |
| $REP\%_{cv}$ | : <i>Relative Error of Prediction</i> en Validación cruzada.                    |
| $REP\%_v$    | : <i>Relative Error of Prediction</i> en Validación externa.                    |
| $REP\%_p$    | : <i>Relative Error of Prediction</i> en Predicción.                            |
| $RER_c$      | : <i>Range Error Ratio</i> en Calibración.                                      |
| $RER_{cv}$   | : <i>Range Error Ratio</i> en Validación cruzada.                               |
| $RER_v$      | : <i>Range Error Ratio</i> en Validación externa.                               |
| $RER_p$      | : <i>Range Error Ratio</i> en Predicción.                                       |
| $OBJ$        | : Valor de la función objetivo utilizada para la selección del modelo.          |
| $savgol$     | : Longitud de ventana del filtro Savitzky–Golay.                                |
| $k$          | : Número de variables espectrales seleccionadas mediante <i>VIP</i> .           |
| $nComp$      | : Número de variables latentes del modelo PLS.                                  |
| $C$          | : ( <i>BoxConstraint</i> ) Parámetro de penalización de errores del modelo SVR. |
| $\epsilon$   | : Margen de tolerancia insensible al error en SVR.                              |
| $\sigma$     | : ( <i>KernelScale</i> ) Parámetro de escala del kernel <i>RBF</i> .            |
| $\gamma$     | : Parámetro que controla la contribución relativa en la fusión MLDF.            |

# Abreviaciones

## Mayúsculas

|         |                                      |
|---------|--------------------------------------|
| PLS     | : Partial Least Squares.             |
| PCR     | : Principal Component Regression.    |
| PCA     | : Principal Component Analysis.      |
| SVR     | : Support Vector Regression.         |
| SVM     | : Support Vector Machine.            |
| TLS     | : Total Least Squares.               |
| RFR     | : Random Forest Regression.          |
| XGBoost | : Extreme Gradient Boosting.         |
| ANN     | : Artificial Neural Network.         |
| VNIR    | : Visible and Near Infrared.         |
| SWIR    | : Short-Wave Infrared.               |
| LLDF    | : Low-Level Data Fusion.             |
| MLDF    | : Mid-Level Data Fusion.             |
| HLDF    | : High-Level Data Fusion.            |
|         |                                      |
| SG      | : Savitzky–Golay.                    |
| MSC     | : Multiplicative Scatter Correction. |
| SNV     | : Standard Normal Variate.           |
| VIP     | : Variable Importance in Projection. |
|         |                                      |
| OLS     | : Ordinary Least Squares.            |
| LOO     | : Leave-One-Out.                     |
| BO      | : Bayesian Optimization.             |
| JK+     | : Jackknife+.                        |

## **Minúsculas**

|      |                          |
|------|--------------------------|
| rbf  | : radial basis function. |
| sg   | : Savitzky–Golay filter. |
| cv   | : cross-validation.      |
| loo  | : leave-one-out.         |
| cal  | : calibration.           |
| val  | : validation.            |
| pred | : prediction.            |
| prep | : preprocessing.         |

# 1. Introducción

---

## 1.1. Introducción General

El cobre es un recurso estratégico para la economía chilena. Su relevancia no solo se explica por los volúmenes producidos, sino también por la forma en que estos deben ser procesados. En los próximos años, la eficiencia en el tratamiento de minerales sulfurados será un factor determinante.

De acuerdo con la Proyección de la Producción de Cobre en Chile para el período 2024–2034, elaborada por la Comisión Chilena del Cobre (COCHILCO), la producción nacional alcanzaría cerca de 5,54 millones de toneladas de cobre fino hacia el año 2034 [1]. Este nivel permitiría mantener a Chile entre los principales productores a escala mundial. Sin embargo, el mismo informe revela un cambio estructural relevante en la matriz productiva. La producción hidrometalúrgica tiende a disminuir, mientras que el concentrado de cobre se consolida como el producto dominante. En la próxima década, este concentrado representaría más del 65 % de la producción total [1].

Este escenario no se desarrolla en condiciones favorables para la industria de fundiciones. A nivel internacional, el mercado de concentrados enfrenta un déficit estructural que se arrastra desde 2019. Según antecedentes de COCHILCO, esta situación responde a un crecimiento desalineado entre la oferta minera y la capacidad de fundición instalada [2]. El impacto ha sido directo. Los cargos de tratamiento y refinación (TC/RC) han mostrado una caída sostenida, reduciendo la rentabilidad del sector y aumentando su exposición operacional.

En Chile, el problema adquiere una dimensión adicional. Altos costos operacionales, márgenes de caja inferiores al promedio mundial y una utilización incompleta de la capacidad instalada afectan la competitividad de las fundiciones nacionales. Esto ocurre pese a que el país continúa siendo uno de los principales productores de concentrado de cobre a nivel global [2].

Frente a este contexto, la fusión flash se ha consolidado como la tecnología predominante para la fusión primaria de concentrados de cobre. Hoy, la mayoría de las fundiciones modernas opera con este tipo de hornos. En ellos, el concentrado seco es alimentado junto con aire u oxígeno enriquecido. Las reacciones de oxidación parcial de los sulfuros generan el calor necesario para sostener una fusión

de carácter autógeno. No se requieren aportes energéticos externos significativos. El proceso produce dos fases líquidas principales mata rica en cobre y escoria además de una corriente gaseosa con alto contenido de  $\text{SO}_2$  [3,4].

El horno flash opera de manera autógena. Este rasgo es clave. Su desempeño depende de forma directa de la composición del concentrado alimentado. Incluso variaciones relativamente pequeñas en los contenidos de cobre, hierro o azufre pueden generar cambios apreciables en la estabilidad térmica, en la formación de escoria y en la eficiencia global del proceso.

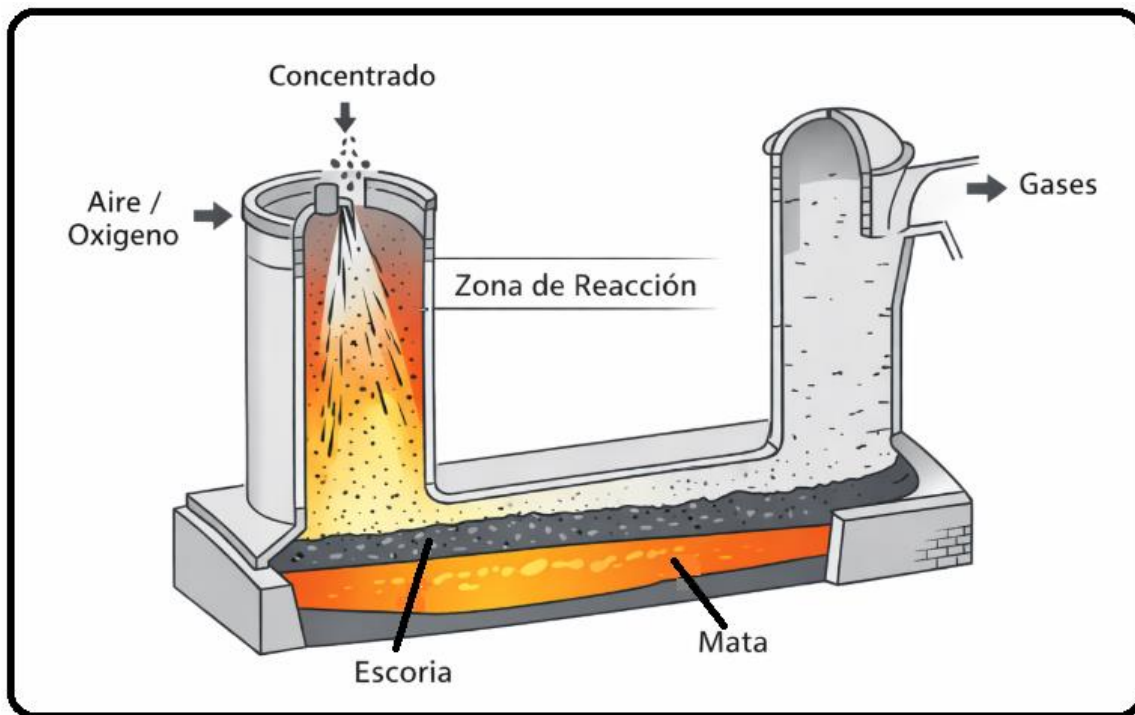


**Figura 1.1 Esquema conceptual simplificado de un horno de fusión flash para concentrados de cobre..**  
**Basado en[4,5]**

La Figura 1.1 muestra un esquema conceptual simplificado de un horno de fusión flash. En él se representan las etapas principales del proceso: la inyección del concentrado y del agente oxidante, la reacción en suspensión y la posterior sedimentación gravitacional de las fases líquidas. En esta última zona, la escoria de menor densidad se dispone sobre la mata de mayor densidad, permitiendo la evacuación separada de ambas corrientes. Este comportamiento ha sido ampliamente documentado en la literatura metalúrgica especializada y en material docente universitario [3,4]. No es un detalle menor. Refleja la fuerte dependencia del proceso respecto de las características del concentrado de entrada.

La mata obtenida en la fusión primaria es tratada posteriormente en reactores de conversión. Entre ellos, el Convertidor Teniente ocupa un lugar destacado dentro de la pirometalurgia del cobre. En este reactor, la mata líquida y, en algunos esquemas operacionales, fracciones controladas de concentrado seco es sometida al soplado de aire u oxígeno enriquecido. El objetivo es eliminar hierro y azufre adicionales hasta alcanzar metal blanco, con contenidos típicos cercanos al 75 % de cobre. El proceso genera, además, escoria y gases [5].

Esta etapa es fundamental. Sin embargo, su desempeño no es independiente. Depende directamente de la calidad de la mata generada en el horno flash. Cuando la fusión primaria presenta inestabilidades, estas tienden a propagarse hacia la conversión.

Desde una perspectiva operacional, la variabilidad en la composición elemental y mineralógica del concentrado no puede considerarse secundaria. Cambios en la proporción de las principales fases sulfuradas afectan con mayor intensidad la etapa de fusión primaria, donde el sistema es más sensible. Las perturbaciones generadas en esta etapa suelen transmitirse aguas abajo, impactando el rendimiento del convertidor y la calidad del producto final. Este efecto es real. No es marginal.

Desde una perspectiva mineralógica, la composición del concentrado resulta especialmente relevante debido a que las principales especies sulfuradas participan de manera distinta en las reacciones de oxidación que sostienen la fusión autógena. La calcopirita ( $\text{CuFeS}_2$ ), principal portadora de cobre en muchos concentrados, y la pirita ( $\text{FeS}_2$ ), frecuentemente presente como ganga sulfurada o mineral asociado, aportan cantidades diferentes de hierro y azufre al sistema. **Variaciones en su proporción relativa modifican el balance térmico del horno, la cantidad de escoria generada y la composición de la mata producida. Como consecuencia, cambios en la mineralogía de alimentación pueden afectar la estabilidad operacional y la eficiencia metalúrgica del proceso de fundición.** En este contexto, la estimación temprana de la abundancia de estas fases minerales adquiere relevancia para apoyar el control y la optimización de la operación.

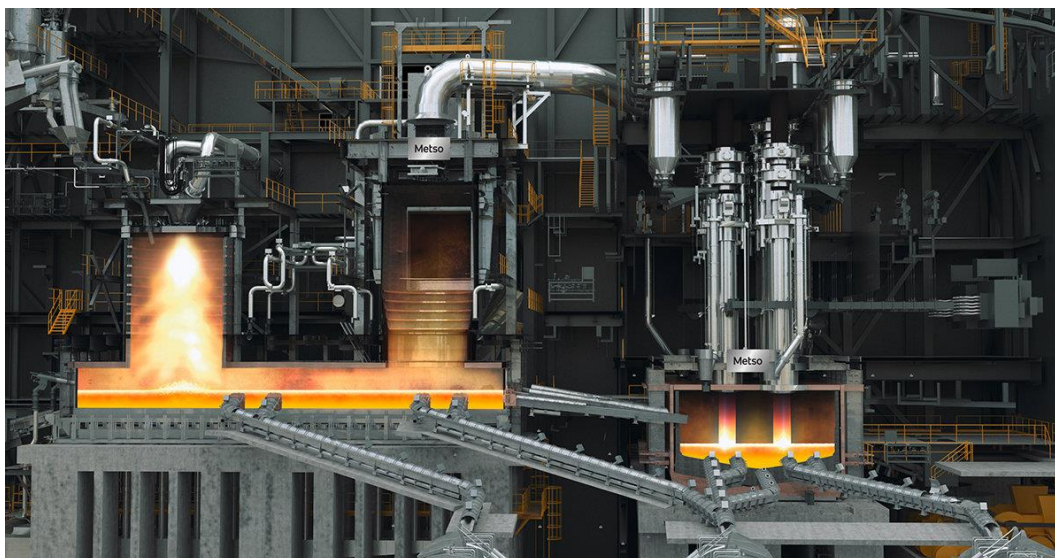


**Figura 1.2 Representación industrial de una instalación de fusión flash para concentrados de cobre. El sistema incluye la zona de reacción donde se alimenta el concentrado junto con aire u oxígeno enriquecido, la cámara de sedimentación para la separación de mata y escoria, y los sistemas asociados de manejo de gases. Fuente: Metso (adaptado de material técnico corporativo). Basado en[78]**

La Figura 1.2 muestra una representación industrial de una instalación de fusión flash. A diferencia del esquema conceptual presentado anteriormente, esta imagen permite visualizar la complejidad de los equipos y sistemas auxiliares involucrados en la operación de una fundición moderna. En este contexto, la composición del concentrado alimentado adquiere una importancia fundamental, ya que influye directamente en el comportamiento térmico del horno, la formación de mata y escoria, y la estabilidad general del proceso.

Por esta razón, la caracterización temprana del concentrado de entrada adquiere un rol central en la optimización global del proceso pirometalúrgico. Disponer de información previa permite anticipar desviaciones operacionales antes de que estas se manifiesten en productos intermedios o finales.

En los últimos años, distintas técnicas espectroscópicas han comenzado a utilizarse para la caracterización de minerales y concentrados. El interés por estas técnicas no surge solo desde la investigación académica. También responde a limitaciones prácticas de los métodos clásicos utilizados en mineralogía de procesos.

La caracterización mineralógica tradicional se basa principalmente en análisis fuera de línea. Entre ellos se incluyen la difracción de rayos X (XRD), los análisis químicos y la microscopía automatizada, como SEM/QEMSCAN [6–9]. Estas técnicas son ampliamente reconocidas por su alta precisión analítica y por su capacidad para identificar y cuantificar fases minerales y composición elemental en minerales y concentrados de cobre.

Sin embargo, su aplicación presenta restricciones claras. Los tiempos de respuesta suelen ser elevados. Los costos operacionales también son significativos. Además, la preparación de muestras requiere procedimientos específicos que dificultan su uso en esquemas de monitoreo continuo o en tiempo real.

Frente a estas limitaciones, la imagen hiperespectral (HSI) ha comenzado a explorarse como una alternativa complementaria. En particular, los rangos visible–infrarrojo cercano (VNIR) e infrarrojo de onda corta (SWIR) permiten capturar información asociada a propiedades mineralógicas del material. La medición es no destructiva y puede realizarse sin interferir directamente con el proceso [10-11].

En el caso de materiales particulados, como los concentrados de cobre, la espectroscopía de reflectancia difusa resulta especialmente adecuada. La medición se realiza directamente sobre la superficie del material, sin necesidad de preparación adicional. Esto la hace compatible con aplicaciones en línea. En conjunto, la variabilidad composicional y mineralógica del concentrado influye de manera directa en la estabilidad y eficiencia de los procesos pirometalúrgicos, particularmente en la fusión primaria. **Contar con herramientas que permitan caracterizar el concentrado antes de su ingreso al horno representa una oportunidad para mejorar la operación del proceso y reducir la dependencia de mediciones tardías.** Bajo este enfoque, el desarrollo de estrategias de sensado avanzadas adquiere interés desde un punto de vista operativo. **La posibilidad de obtener información del concentrado de forma rápida, no destructiva y con potencial aplicación en monitoreo en línea abre nuevas oportunidades para apoyar la toma de decisiones en fundiciones modernas.**

## 1.2. Problemática

En la minería chilena, la caracterización mineralógica de minerales y concentrados de cobre se apoya, en su mayoría, en técnicas de laboratorio. Este tipo de análisis se realiza sobre muestras discretas y de manera fuera de línea. Aunque estos métodos permiten obtener resultados con alta precisión analítica, no facilitan una caracterización continua del material en proceso y suelen involucrar tiempos de respuesta prolongados. Como consecuencia, en la operación diaria no se dispone de información en tiempo real sobre corrientes de proceso consideradas críticas.

En el contexto nacional, no se cuenta con una implementación industrial consolidada que permita el sensado directo del concentrado de cobre a la entrada del horno flash con fines de estimación mineralógica continua. La falta de este tipo de medición impide anticipar variaciones en la alimentación y limita el uso de la información composicional como herramienta efectiva de apoyo a la operación.

La importancia de esta problemática se vuelve aún más evidente al considerar la sensibilidad del proceso de fusión flash frente a la composición del concentrado alimentado. Cambios en las proporciones de las fases sulfuradas, así como variaciones en los contenidos de cobre, hierro y azufre, pueden afectar la estabilidad térmica del horno, los mecanismos de formación de escoria y el rendimiento global del proceso pirometalúrgico. Diversos estudios e informes técnicos señalan que el horno flash responde de manera sensible a fluctuaciones en la calidad químico-mineralógica del concentrado, generando condiciones operacionales inestables cuando estas variaciones no son identificadas oportunamente [12]. No obstante, en la práctica, tales diferencias suelen detectarse cuando el material ya ha sido procesado, reduciendo de forma significativa la capacidad de reacción operativa.

A esta situación se suma la naturaleza física del concentrado de cobre. Se trata de un material particulado fino y heterogéneo, lo que restringe el uso de técnicas de sensado convencionales. Métodos que requieren contacto directo con el material o preparación previa de muestras y superficies homogéneas resultan poco compatibles con una operación en tiempo real bajo condiciones industriales. En este contexto, adquieren relevancia aportes provenientes de otras áreas de aplicación donde se enfrenta el desafío de caracterizar directamente materiales particulados. La literatura metodológica sobre espectroscopía de reflectancia difusa, desarrollada en ámbitos distintos a la

minería, reporta que este tipo de técnicas permite caracterizar polvos de forma no invasiva, sin preparación de muestra, de manera no destructiva y con potencial de operación en línea [13]. Si bien estas aplicaciones no se enfocan en concentrados minerales, los principios físicos asociados a la interacción entre la radiación y materiales particulados son directamente aplicables al caso del concentrado de cobre.

En consecuencia, la caracterización en línea de un material en polvo requiere metodologías no invasivas, capaces de operar en ambientes industriales y de proporcionar información representativa del material alimentado sin interferir con el proceso. Desde esta perspectiva, la espectroscopía de reflectancia difusa en las regiones visible e infrarroja cercana se presenta como una alternativa técnicamente viable, ya que se adapta de forma natural a superficies irregulares y muestra sensibilidad frente a variaciones en la composición química y mineralógica del material.

Sin embargo, la existencia de una relación entre la respuesta espectral y la mineralogía del concentrado no garantiza, por sí sola, que dicha relación pueda ser utilizada de manera cuantitativa y confiable en un entorno industrial. Uno de los principales desafíos radica en el desarrollo de modelos predictivos robustos. El desempeño de estos modelos depende de diversos factores, entre ellos la forma en que se representen los datos espectrales, la selección de la región de interés que mejor caracterice el material alimentado y las estrategias de preprocesamiento aplicadas para mitigar efectos de dispersión, variaciones de iluminación y ruido instrumental. A ello se suma la necesidad de emplear enfoques de modelado multivariado capaces de describir adecuadamente la relación entre la información espectral y la composición mineralógica del concentrado.

Como resultado, se identifica una brecha técnica y metodológica asociada a la ausencia de estrategias de sensado en línea para la caracterización mineralógica del concentrado de cobre a la entrada del horno flash en Chile. **Si bien se reconoce la existencia de una relación entre los espectros de reflectancia difusa y la composición del material, esta ha sido abordada de manera limitada desde un punto de vista cuantitativo.** En un escenario donde la tecnología de fusión flash se encuentra ampliamente implementada y la variabilidad del concentrado representa un desafío operacional relevante, surge la necesidad de optimizar la relación espectro–composición. **Esto implica estudiar de forma sistemática qué combinaciones de representación de datos, preprocesamiento y modelos de cuantificación permiten alcanzar mayores niveles de precisión**

Adicionalmente, la información espectral utilizada para la caracterización del material no es necesariamente única, ya que puede provenir de distintas regiones del espectro electromagnético. Estas regiones presentan sensibilidades diferenciadas frente a propiedades físicas y químicas del material, lo que introduce un grado adicional de complejidad en la forma en que los datos son representados y utilizados en el modelado. En este sentido, no solo resulta relevante seleccionar modelos de cuantificación adecuados, sino también definir cómo se estructura la información espectral de entrada, considerando que distintas fuentes pueden aportar información complementaria sobre la composición del concentrado (Fusión de datos).

### **1.3. Relevancia de las composiciones y el uso de las cámaras hiperespectrales.**

La composición de un concentrado de cobre es un aspecto relevante tanto desde el punto de vista económico como desde su comportamiento metalúrgico. Esta importancia no se explica únicamente por el contenido elemental del material, sino también por la forma en que dichos elementos se encuentran distribuidos entre las distintas fases minerales presentes. En la práctica, ambos enfoques —mineralógico y elemental— influyen de manera conjunta en la respuesta del concentrado durante su procesamiento, en particular en procesos pirometalúrgicos como la fusión.

Un concentrado de cobre está compuesto por un número limitado de fases minerales sulfuradas. Entre las más habituales se encuentran la calcopirita ( $\text{CuFeS}_2$ ), la bornita ( $\text{Cu}_5\text{FeS}_4$ ), la pirita ( $\text{FeS}_2$ ), la pirrotita ( $\text{Fe}_{1-x}\text{S}$ ) y otros sulfuros de cobre, además de minerales de ganga. Estas fases no se comportan de la misma forma desde el punto de vista metalúrgico. Algunas concentran el valor metálico y favorecen la recuperación de cobre, mientras que otras no aportan directamente al contenido de Cu y solo incrementan el volumen de material a procesar, afectando la eficiencia global del sistema [14].

En los concentrados de cobre procesados en Chile, la calcopirita suele ser la fase sulfurada dominante y, en muchos casos, el principal portador de cobre. La pirita aparece con frecuencia como fase acompañante, aportando hierro y azufre, pero sin contribuir al contenido de Cu del concentrado. Como resultado, la mineralogía de estos materiales suele estar dominada por pocas fases principales, principalmente calcopirita y pirita. La proporción relativa entre ambas tiene un impacto directo tanto en la ley de cobre como en el comportamiento del concentrado durante la fundición y otras etapas metalúrgicas [14-15].

Cuando la calcopirita es la fase mayoritaria que contiene cobre, la ley de %Cu del concentrado queda estrechamente relacionada con su proporción mineralógica, dado que la composición química de esta fase define un contenido de cobre relativamente fijo. Sin embargo, esta relación no es completamente directa. La presencia de pirita, la coexistencia de distintas fases sulfuradas, la mezcla de partículas y la heterogeneidad propia del material introducen variabilidad que no puede ser explicada únicamente a partir del análisis elemental. En este sentido, medir solo los contenidos globales de Cu, Fe o S no permite identificar cómo dichos elementos se distribuyen entre las fases minerales presentes, lo que limita la interpretación metalúrgica del concentrado [15].

Esta situación hace que la caracterización del material sea más compleja. No resulta suficiente observar una fase de manera aislada ni analizar únicamente la composición elemental promedio. Es necesario contar con información que permita describir cómo las fases coexisten y cómo se distribuyen dentro del concentrado, especialmente considerando que se trata de un material fino y heterogéneo.

En este contexto, las cámaras hiperspectrales aparecen como una alternativa de interés para la caracterización de concentrados minerales. En particular, aquellas que operan en el rango visible–infrarrojo cercano (VNIR, 400–1000 nm) permiten registrar firmas espectrales asociadas a propiedades electrónicas y estructurales de los minerales sulfurados. Este tipo de medición es no destructiva y, además, entrega información espacial, lo que permite analizar la variabilidad interna del material observado. Estudios aplicados han mostrado que los datos hiperspectrales VNIR pueden emplearse para modelar y discriminar fases mineralógicas presentes en concentrados de cobre, incluyendo la diferenciación entre calcopirita y pirita, aun cuando ambas comparten elementos comunes como hierro y azufre [16].

Por otra parte, la composición elemental continúa siendo una variable fundamental para la valorización del concentrado y para el diseño de los procesos metalúrgicos. Elementos como cobre, hierro y azufre condicionan tanto el valor económico como la operación de las etapas posteriores. Sin embargo, el análisis puramente elemental no entrega información directa sobre la forma en que estos elementos se encuentran asociados a las distintas fases minerales. En este sentido, el uso de

información hiperespectral ofrece la posibilidad de integrar ambos enfoques y de analizar de manera conjunta la composición mineralógica y elemental del concentrado.

Si bien el desarrollo de modelos basados en datos hiperespectrales presenta desafíos metodológicos, su correcta implementación abre una vía concreta para mejorar la caracterización de concentrados de cobre. La posibilidad de capturar información espectral, mineralógica y espacial de forma simultánea constituye una herramienta con potencial para apoyar la optimización de los procesos metalúrgicos asociados, especialmente en contextos donde la variabilidad del concentrado representa un desafío operacional relevante.

En este contexto, es importante destacar que distintas regiones del espectro hiperespectral capturan fenómenos físicos complementarios. Mientras la región visible–infrarrojo cercano (VNIR) está asociada principalmente a transiciones electrónicas y propiedades estructurales de los minerales, la región infrarroja de onda corta (SWIR) se relaciona con vibraciones moleculares y enlaces químicos. Esta diferencia en los mecanismos de interacción sugiere que ambas regiones pueden aportar información no redundante sobre la composición del concentrado.

En consecuencia, el uso conjunto de información proveniente de múltiples regiones espectrales representa una oportunidad para mejorar la caracterización del material. En particular, estrategias de **fusión de datos** permiten integrar estas fuentes de información con el objetivo de construir representaciones más completas del sistema analizado. No obstante, la forma en que dicha información es combinada e incorporada dentro de modelos de cuantificación constituye un desafío metodológico relevante, ya que puede influir directamente en el desempeño predictivo alcanzado.

#### **1.4. Pregunta de investigación inicial**

Existe evidencia de una relación entre la respuesta espectral de reflectancia difusa y la composición mineralógica y elemental de minerales y concentrados de cobre. No obstante, la forma en que esta relación puede ser modelada de manera precisa y robusta no es única, ya que depende de diversas decisiones metodológicas, tales como la representación de los datos espectrales, las estrategias de preprocesamiento, los esquemas de calibración y validación, así como los criterios utilizados para ajustar y comparar modelos de cuantificación.

Adicionalmente, considerando que distintas regiones del espectro hiperespectral pueden aportar información complementaria sobre el material, surge la necesidad de evaluar no solo el tipo de modelo utilizado, sino también la forma en que estas fuentes de información son empleadas e integradas dentro del proceso de modelado. En este sentido, tanto la selección del enfoque de cuantificación como la estrategia de utilización de los datos espectrales pueden influir de manera significativa en el desempeño predictivo alcanzado.

En este contexto, surge la siguiente pregunta de investigación que guía el desarrollo del presente trabajo:

**¿Cómo influyen las estrategias de modelado lineal y no lineal, junto con el uso e integración de información espectral proveniente de distintas regiones (VNIR–SWIR), en la capacidad de**

**estimar de manera precisa y robusta la composición mineralógica y elemental de concentrados de cobre mediante espectroscopía de reflectancia difusa?**

## **2. Estado del arte**

---

Para abordar esta pregunta de investigación inicial, se ha realizado una revisión bibliográfica que abarca la teoría y exploración de los modelos de cuantificación más recientes utilizados hasta la fecha. Además, se ha considerado la posibilidad de emplear un modelo que se adapte según las propiedades específicas y el contexto de lo que se pretende cuantificar, lo cual podría conducir a resultados más precisos y pertinentes. **Se abordó la revisión del estado del arte desde lo más fundamental hasta lo más específico y actual.**

### **2.1. Métodos de cuantificación:**

Para estudiar los modelos de cuantificación basados en datos hiperespectrales es necesario revisar los distintos tipos de ajuste y estrategias de modelado utilizados en la literatura. La relación entre la respuesta espectral y la composición mineralógica o elemental puede representarse mediante enfoques de diferente complejidad, los cuales varían en su forma de tratar la alta dimensionalidad de los datos y en los criterios empleados para evaluar su desempeño. En este sentido, la revisión de los métodos de cuantificación permite establecer un marco general para comprender las diferencias entre los enfoques existentes y sustentar las decisiones metodológicas adoptadas posteriormente.

#### **2.1.1 Regresión lineal múltiple por Mínimos Cuadrados Ordinarios (MLR-OLS)**

La regresión por mínimos cuadrados ordinarios (OLS) es una técnica clásica que busca minimizar la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por el modelo. Este enfoque asume que las variables independientes se encuentran libres de error y, aunque es conceptualmente simple, puede resultar limitado en el análisis de datos hiperespectrales debido a la alta dimensionalidad y a la fuerte multicolinealidad presente entre las variables espectrales [17].

Sea  $\mathbf{X}$  como una matriz de dimensiones  $n \cdot m$  que representan  $n$  analitos y  $m$  variables independientes en el caso de espectros serían  $n$  espectros y  $m$  bandas, además, sea  $\mathbf{Y}$  un vector de dimensiones  $p \cdot 1$  que representa los valores de referencia, cabe decir que  $n=p$  en este contexto, se busca encontrar los coeficientes que mejor ajusten la relación entre las variables independientes y los valores de referencia, minimizando la suma de los cuadrados de las diferencias entre las predicciones del modelo y los valores reales observados.

El modelo de regresión lineal en el contexto de datos hiperespectrales puede expresarse como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (1)$$

Donde:

$\boldsymbol{\beta}$  es el vector de coeficientes del modelo que queremos estimar y  $\boldsymbol{\epsilon}$  es el vector de errores asociados.

El objetivo de la regresión OLS es encontrar los valores óptimos para  $\boldsymbol{\beta}$  que minimizan la suma de los cuadrados de los errores, lo que conduce a la siguiente solución analítica:

$$\check{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y} \quad (2)$$

De esta manera, se obtiene el vector de regresión óptimo que ajusta los datos de acuerdo con el criterio de mínimos cuadrados ordinarios. La regresión lineal múltiple (MLR) ha sido ampliamente utilizada en espectroscopia de reflectancia basada en cámaras hiperespectrales, con aplicaciones relevantes en áreas como el censo remoto [18] y la detección de plagas en cultivos [19].

No obstante, este enfoque presenta desventajas importantes cuando la matriz  $\mathbf{X}$  posee una alta dimensionalidad, ya que la presencia de un gran número de bandas espectrales suele inducir problemas de colinealidad entre las variables. Esta situación es común en sistemas hiperespectrales, donde el elevado valor de  $m$  puede afectar la estabilidad numérica del modelo y la interpretabilidad de los coeficientes estimados. Para mitigar este problema, algunos autores optan por identificar las bandas

espectrales más relevantes mediante análisis exploratorios o por recortar regiones específicas del espectro, reduciendo así la colinealidad.

Frente a estas limitaciones, surgen técnicas alternativas como la regresión por componentes principales (PCR) y la regresión por mínimos cuadrados parciales (PLS), las cuales permiten abordar de manera más robusta la alta dimensionalidad y la colinealidad de los datos, y se describen en las siguientes subsecciones.

### 2.1.2 Regresión de Componentes Principales (PCR)

La regresión por componentes principales (PCR) es una técnica que reduce la dimensionalidad de los datos espectrales mediante la aplicación del análisis de componentes principales (PCA) sobre la matriz  $X$ , utilizando los scores como nuevas variables independientes [20]. Este enfoque surge como una alternativa para mitigar los problemas de multicolinealidad propios de datos hiperespectrales, permitiendo comprimir la información contenida en un gran número de bandas espectrales en un conjunto reducido de variables no correlacionadas.

Sea  $X$  la misma matriz de data  $n \cdot m$ . Mediante PCA, esta matriz puede descomponerse en términos de scores y loadings según:

$$X = TP^T + E \quad (3)$$

Donde  $T$  representa la matriz de scores de dimensión  $n \cdot n$  con  $n$  la cantidad máxima de componentes principales,  $P$  representa la matriz de loadings de dimensión  $n \cdot m$  y  $E$  representa el error.

Se escoge una matriz truncada  $T_A$  de dimensión  $n \cdot a$  y  $P_A$  de dimensión  $m \cdot a$  con  $a$  la cantidad de componentes principales que en las primeras componentes se obtenga la máxima covarianza.

De esta manera aplicando el mismo criterio OLS de la regresión MLR podemos expresar que la minimización para obtener el vector de regresión es de la forma con dimensiones  $a \cdot 1$ :

$$\check{\beta}_{OLS-PCR} = (T_A^T T_A)^{-1} T_A^T Y \quad (4)$$

Por ende el coeficiente en el plano de la matriz  $\mathbf{X}$  se obtiene multiplicando la matriz de loadings  $\mathbf{P}_A$  así obteniéndose el vector de regresión.

$$\check{\beta}_{PCR} = \mathbf{P}_A \check{\beta}_{OLS-PCR} \quad (5)$$

La regresión PCR presenta como principal ventaja la eliminación de la colinealidad, ya que los scores  $\mathbf{T}_A$  contienen únicamente información no correlacionada derivada de  $\mathbf{X}$ . Sin embargo, esta misma característica constituye su principal limitación, puesto que el PCA considera exclusivamente la varianza de  $\mathbf{X}$  sin incorporar información sobre la covarianza entre  $\mathbf{X}$  e  $\mathbf{Y}$ . Como consecuencia, las componentes seleccionadas no necesariamente son las más relevantes para la predicción de la variable de interés, lo que motiva el uso de métodos alternativos como la regresión por mínimos cuadrados parciales (PLS), descrita en el siguiente apartado.

### 2.1.3 Regresión por Mínimos Cuadrados Parciales (PLS)

La regresión por mínimos cuadrados parciales (PLS) constituye un marco general de modelado que combina la proyección de las variables independientes con la información contenida en la variable de respuesta, con el objetivo de extraer variables latentes que maximicen la covarianza entre  $\mathbf{X}$  e  $\mathbf{Y}$ . Este enfoque ha sido ampliamente adoptado en quimiometría y análisis espectral debido a su capacidad para tratar datos altamente correlacionados y de gran dimensionalidad [21], siendo particularmente adecuado para el análisis de imágenes hiperespectrales [22]. A diferencia de la regresión por componentes principales (PCR), en PLS las proyecciones no se determinan únicamente a partir de la varianza de  $\mathbf{X}$ , sino que incorporan explícitamente la información de  $\mathbf{Y}$ , permitiendo obtener variables latentes más relevantes desde el punto de vista predictivo.

En su formulación general, la regresión PLS se basa en la descomposición simultánea de las matrices de datos  $\mathbf{X}$  e  $\mathbf{Y}$  de la siguiente forma [21,23]:

$$\mathbf{X} = \mathbf{T} \mathbf{P}^T + \mathbf{E} \quad (6)$$

$$\mathbf{Y} = \mathbf{U} \mathbf{Q}^T + \mathbf{F} \quad (7)$$

$$\mathbf{U} = \mathbf{B} \mathbf{T} \quad (8)$$

donde  $\mathbf{T}$  y  $\mathbf{U}$  representan las matrices de *scores* asociadas a  $\mathbf{X}$  e  $\mathbf{Y}$ , respectivamente,  $\mathbf{P}$  y  $\mathbf{Q}$  corresponden a las matrices de *loadings*, y  $\mathbf{E}$  y  $\mathbf{F}$  son las matrices de error. La relación fundamental entre ambas proyecciones puede expresarse como:

donde  $\mathbf{B}$  es una matriz diagonal que captura la relación lineal entre los *scores* de  $\mathbf{X}$  y  $\mathbf{Y}$ . A partir de esta relación se obtiene el modelo de regresión, el cual permite estimar la variable de respuesta a partir de los datos espectrales [21]. Si bien existen diversas implementaciones algorítmicas de la regresión PLS como NIPALS, PLS1, PLS2 o SIMPLS todas ellas comparten esta estructura conceptual y difieren principalmente en el procedimiento utilizado para construir la matriz de pesos  $\mathbf{W}$ , la cual define las direcciones latentes en el espacio de  $\mathbf{X}$  [21,23]. En este sentido, la elección del algoritmo no modifica la formulación general del modelo, sino la forma en que se estiman las variables latentes. En particular, el algoritmo SIMPLS permite obtener directamente los pesos latentes maximizando la covarianza entre  $\mathbf{X}$  e  $\mathbf{Y}$ , sin requerir una deflación explícita de la matriz  $\mathbf{X}$  [25]. En este contexto, la construcción de las variables latentes puede describirse mediante:

$$\mathbf{w}_i = \frac{\mathbf{X}^T \mathbf{y}_{i-1}}{\|\mathbf{X}^T \mathbf{y}_{i-1}\|}, \quad \mathbf{W}_a = [\mathbf{w}_1, \dots, \mathbf{w}_a], \quad \mathbf{T}_a = \mathbf{X} \mathbf{W}_a \quad (9)$$

$$\mathbf{p}_i = \frac{\mathbf{X}^T \mathbf{t}_i}{\mathbf{t}_i^T \mathbf{t}_i}, \quad \mathbf{P}_a = [\mathbf{p}_1, \dots, \mathbf{p}_a], \quad \mathbf{X}_a \approx \mathbf{T}_a \mathbf{P}_a^T \quad (10)$$

donde  $\mathbf{W}_a$  es la matriz de pesos latentes que define las direcciones en el espacio espectral  $\mathbf{X}$  que maximizan la covarianza con la variable de respuesta  $\mathbf{Y}$ ,  $\mathbf{T}_a$  corresponde a la matriz de scores asociada a un número  $a$  de variables latentes, y  $\mathbf{P}_a$  representa la matriz de loadings del modelo. La matriz  $\mathbf{X}_a$  corresponde a la reconstrucción aproximada de  $\mathbf{X}$  utilizando dichas variables latentes.

De este modo, la regresión PLS puede interpretarse como un método que busca simultáneamente reducir la dimensionalidad de los datos espectrales y maximizar su capacidad predictiva, siendo la construcción de la matriz de pesos  $\mathbf{W}$  el elemento central que distingue las distintas implementaciones del método [21,22]. Finalmente, el desempeño del modelo PLS depende de una correcta selección del número de variables latentes y de la aplicación de estrategias de preprocesamiento adecuadas.

#### 2.1.4 Redes neuronales (ANN)

Las redes neuronales artificiales (Artificial Neural Networks, ANN) son modelos computacionales inspirados en el funcionamiento del cerebro humano y se utilizan para describir relaciones complejas

entre variables. En el análisis de datos hiperespectrales, las ANN han sido empleadas con frecuencia debido a su capacidad para capturar relaciones no lineales entre los espectros y las propiedades de interés. Sin embargo, su desempeño depende de varios factores, como la arquitectura de la red, el preprocesamiento aplicado a los datos y la cantidad y calidad de la información disponible para el entrenamiento [26].

Desde un punto de vista matemático, una red neuronal busca aproximar una función a partir de un conjunto de datos, especialmente cuando dicha relación no puede describirse adecuadamente mediante un modelo lineal. Esto se logra mediante la combinación de capas de neuronas que aplican transformaciones no lineales a las señales de entrada, permitiendo representar relaciones complejas entre variables de entrada y salida [27].

En problemas de regresión, las ANN permiten modelar la relación entre espectros hiperespectrales y propiedades químicas o físicas de interés. Para ello, la red se entrena utilizando pares de datos formados por espectros y valores de referencia conocidos. Durante este proceso, los pesos internos se ajustan de manera iterativa para reducir el error entre las predicciones del modelo y los valores reales. Una vez entrenada, la red puede utilizarse para estimar propiedades desconocidas a partir de nuevos espectros, lo que resulta especialmente útil cuando se trabaja con grandes volúmenes de datos y relaciones no lineales difíciles de modelar de forma explícita. El principio de funcionamiento de una red neuronal se basa en la neurona artificial, cuyo esquema se muestra en la Figura 2.2.

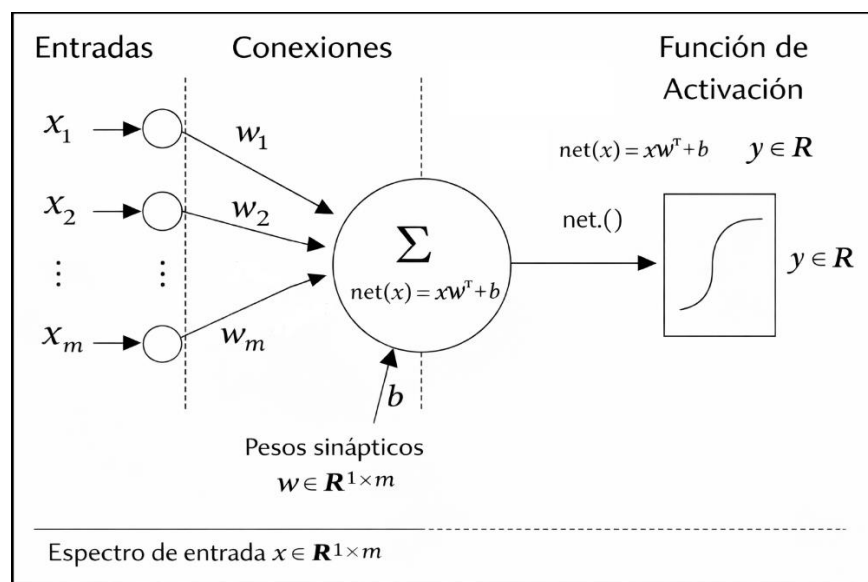


**Figura 2.1** esquema de una neurona artificial, elaboración propia basado en [27].

Una neurona realiza dos operaciones principales: la combinación lineal de las entradas y la aplicación de una función de activación. La neurona recibe como entrada un vector  $\mathbf{x} \in \mathbb{R}^{1 \times m}$ , donde cada componente  $x_i$  representa una característica de entrada (por ejemplo, una banda espectral). A cada entrada se le asocia un peso  $w_i$ , y la neurona incluye además un término de sesgo  $b$ , que actúa como intercepto. La combinación lineal de las entradas se calcula como:

$$\mathbf{z} = \sum_{i=1}^m w_i x_i + b \quad (11)$$

donde  $\mathbf{z}$  corresponde a la suma ponderada de las entradas.

Posteriormente, la neurona aplica una función de activación  $\sigma(\cdot)$  con el objetivo de introducir no linealidad en el modelo. La salida de la neurona, denotada como  $\mathbf{a}$ , se expresa como:

$$\mathbf{a} = \sigma(\mathbf{z}) \quad (12)$$

La elección de la función de activación es un aspecto clave en el diseño de redes neuronales. A continuación, se describen algunas de las funciones de activación más utilizadas.

#### **Función sigmoide:**

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (13)$$

Esta función transforma su entrada a un valor en el rango  $[0,1]$  y es comúnmente utilizada cuando la salida del modelo puede interpretarse como una probabilidad.

#### **Función tangente hiperbólica (tanh):**

$$\sigma(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (14)$$

La función **tanh** produce salidas en el rango  $(-1,1)$ , lo que permite obtener activaciones centradas en cero y puede favorecer una convergencia más rápida durante el entrenamiento en comparación con la función sigmoide.

#### **Función Rectified Linear Unit (ReLU):**

$$\sigma(z) = \max(0, z) \quad (15)$$

La función **ReLU** anula los valores negativos y mantiene los valores positivos sin modificación. Es ampliamente utilizada en redes neuronales profundas debido a su eficiencia computacional y a su capacidad para mitigar el problema de la saturación del gradiente.

Cada función de activación presenta ventajas y limitaciones, y su selección depende del tipo de problema y de la arquitectura de la red. En general, la función sigmoide se emplea en salidas probabilísticas, la función tanh cuando se requiere una salida centrada en cero, y la función ReLU en redes profundas debido a su eficiencia y estabilidad numérica.

Existen diversas arquitecturas de redes neuronales, siendo una de las más conocidas y utilizadas el perceptrón multicapa (Multilayer Perceptron, MLP). La estructura general de un MLP se presenta en la Figura 2.3.

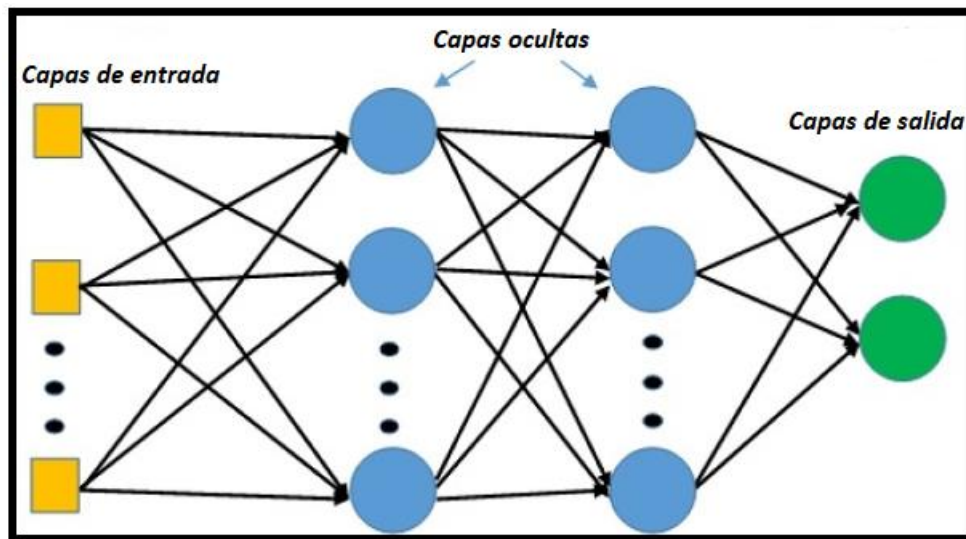


**Figura 2.2. Estructura de un MLP [28].**

Un perceptrón multicapa (MLP) se estructura en tres tipos de capas. La capa de entrada recibe los datos y cada una de sus neuronas representa una variable del conjunto de entrada. Esta capa no realiza procesamiento, sino que transfiere la información a la siguiente etapa. Las capas ocultas son las encargadas de procesar los datos, permitiendo que la red aprenda relaciones complejas y no lineales.

En estas capas, cada neurona combina las entradas mediante pesos, aplica una función de activación y transmite el resultado, incorporando además un término de sesgo. Finalmente, la capa de salida entrega la predicción del modelo. Su configuración depende del problema abordado: en regresión suele emplearse una función de activación lineal, mientras que en clasificación se utilizan funciones no lineales adecuadas.

Diversos estudios han utilizado regresión basada en redes neuronales artificiales (ANN) para la caracterización de variables mineralógicas que no presentan una relación lineal evidente, como la estimación de abundancias minerales a partir de datos hiperespectrales en el rango VNIR–SWIR. En este contexto, se ha reportado el uso de modelos de regresión no lineales, incluyendo ANN, para estimar abundancias minerales en núcleos de perforación a partir de datos hiperespectrales, empleando información mineralógica de alta resolución como referencia [28].

#### **2.1.5 Random Forest Regression (RFR).**

Random Forest Regression (RFR) es un algoritmo de aprendizaje supervisado basado en un conjunto de árboles de decisión. Su principal ventaja radica en que combina las predicciones de múltiples árboles, lo que permite mejorar la precisión del modelo y reducir el riesgo de sobreajuste. Debido a estas características, RFR ha sido ampliamente utilizado en el análisis de datos hiperespectrales, donde suele enfrentarse a problemas de alta dimensionalidad y multicolinealidad entre las variables espectrales [29,30]. El funcionamiento general del algoritmo se ilustra en la Figura 2.4.

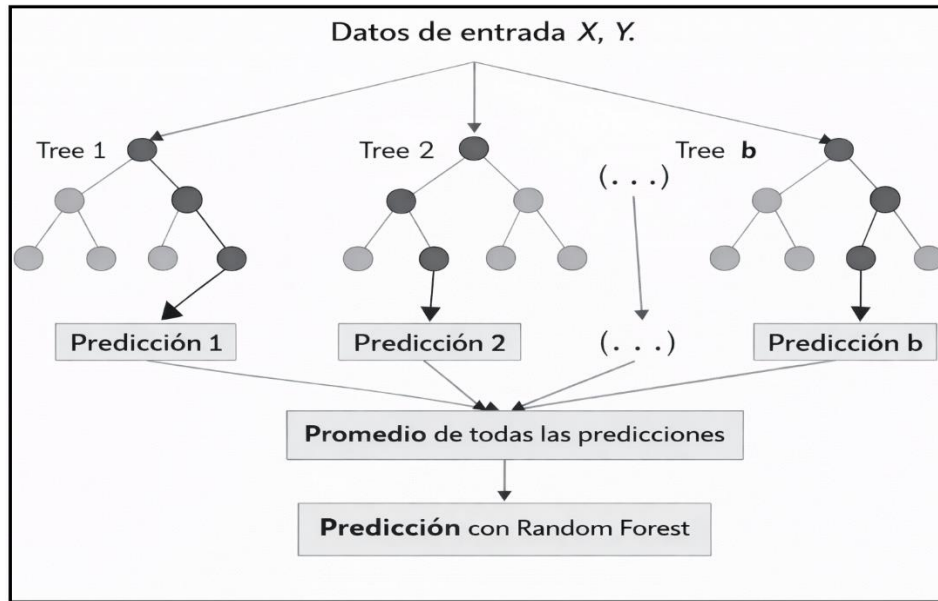


**Figura 2.3. Esquema visual de cómo funciona el algoritmo random forest modificado de:**  
[\[https://levelup.gitconnected.com/random-forest-regression-209c0f354c84.\].](https://levelup.gitconnected.com/random-forest-regression-209c0f354c84)

El principio fundamental de RFR consiste en la generación de múltiples subconjuntos de datos a partir de las matrices  $X$  e  $Y$  mediante la técnica de muestreo *bootstrap*. Este procedimiento selecciona, de forma aleatoria y con reemplazo, muestras del conjunto de datos original, dando lugar a distintos conjuntos de entrenamiento. Si  $X \in \mathbb{R}^{n \times m}$  representa una matriz de  $n$  espectros con  $m$  bandas y  $Y \in \mathbb{R}^{n \times 1}$  el vector de valores de referencia, se generan  $b$  subconjuntos  $(X^{(b)}, Y^{(b)})$ , con  $b = 1, \dots, B$ , los cuales se utilizan para entrenar de manera independiente cada árbol de decisión.

Durante la construcción de cada árbol, en cada nodo se selecciona aleatoriamente un subconjunto de variables predictoras, lo que introduce variabilidad entre los árboles y contribuye a disminuir la correlación entre ellos. La división de los nodos se realiza minimizando la varianza de la variable de salida dentro de los nodos resultantes, favoreciendo particiones más homogéneas respecto a  $Y$ .

Una vez entrenados todos los árboles, la predicción final del modelo se obtiene promediando las predicciones individuales de cada árbol. Este proceso de agregación permite obtener una estimación más robusta y estable que la entregada por un único árbol de decisión.

En aplicaciones relacionadas al contexto, Random Forest Regression ha sido utilizada para la

cuantificación mineralógica y la estimación de parámetros geoquímicos a partir de espectros reflectantes, mostrando un desempeño robusto frente a la complejidad de los datos [31,32]. No obstante, su uso implica un mayor costo computacional y una interpretación menos directa del modelo, por lo que la elección de hiperparámetros y el preprocesamiento de los datos resultan aspectos relevantes para su aplicación práctica.

### 2.1.6 Extreme gradient boosting (XGB).

XGBoost (Extreme Gradient Boosting) es una técnica de aprendizaje automático basada en árboles de decisión, similar en concepto a Random Forest, pero con una diferencia fundamental en su estrategia de entrenamiento. Mientras que Random Forest construye múltiples árboles de forma independiente y combina sus predicciones mediante bagging, XGBoost utiliza un enfoque de boosting, en el cual los árboles se entrenan de manera secuencial [33].

En este esquema, cada nuevo árbol se ajusta con el objetivo de corregir los errores cometidos por los modelos anteriores. El principio del boosting consiste en combinar varios modelos débiles para construir un modelo más robusto. En el caso de XGBoost, este proceso se basa en el gradient boosting, donde la optimización se realiza minimizando una función de pérdida mediante el uso de su gradiente. En cada iteración, el modelo calcula los residuos asociados a la predicción actual y entrena un nuevo árbol de decisión para aproximar dichos errores, mejorando progresivamente el desempeño global del modelo. La Figura 2.5. muestra un esquema conceptual del funcionamiento de XGBoost.

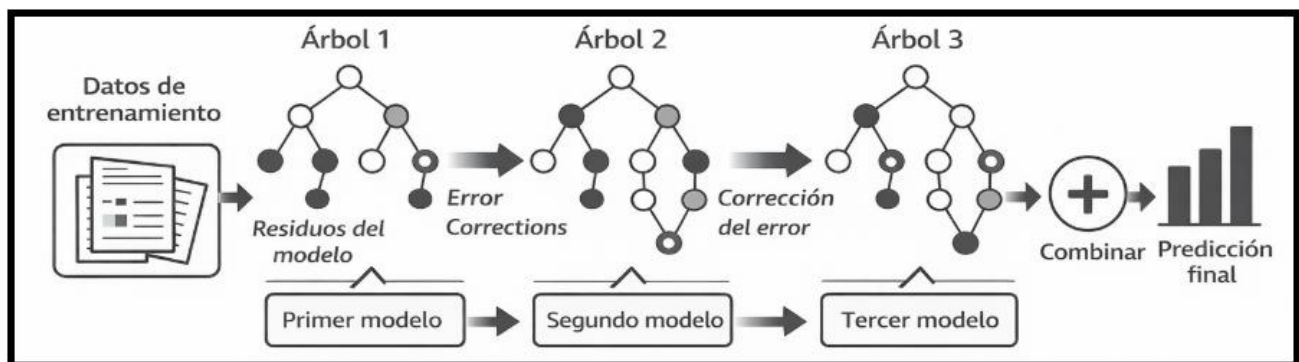


Figura 2.4. Representación esquemática del algoritmo XGBoost [48,49].

A partir de los datos de entrenamiento, representados por la matriz de espectros  $X$  y el vector de valores de referencia  $Y$ , se entrena un primer árbol que entrega una estimación inicial. A continuación, se calculan los residuos del modelo y se entrena un segundo árbol que introduce una corrección sobre

el error previo. Este proceso se repite de forma iterativa, dando lugar a una secuencia de árboles que van ajustando progresivamente la predicción. Finalmente, las salidas de todos los árboles se combinan para obtener la predicción final del modelo.

Gracias a este enfoque secuencial, XGBoost es capaz de modelar relaciones complejas y no lineales, mostrando un buen desempeño en problemas con alta dimensionalidad y gran volumen de datos. Además, el algoritmo incorpora mecanismos de regularización que ayudan a controlar el sobreajuste y mejorar la capacidad de generalización del modelo. Por estas razones, XGBoost ha sido utilizado en diversas aplicaciones, incluyendo problemas de quimiometría como la estimación de la calidad de alimentos y otras tareas de predicción basadas en datos espectrales [34].

### **2.1.7 Error-in-variables (EIV) en las Regresiones.**

La regresión por mínimos cuadrados ordinarios (OLS) busca minimizar las discrepancias entre los valores observados y los predichos, centrándose en los errores residuales en  $\mathbf{Y}$ . Sin embargo, en el caso de cámaras hiperespectrales, donde cada píxel puede capturar espectros con mediciones variables, los errores en  $\mathbf{X}$  pueden ser significativos debido a la variabilidad entre los espectros capturados y la sensibilidad de la cámara a diferentes bandas. Este error implícito no se aborda directamente con OLS.

Por otro lado, técnicas como PCA o PLS pueden mitigar una parte de este error implícitamente al comprimir la información en un conjunto reducido de variables latentes, lo que permite despreciar o

minimizar los efectos del error de los sensores. Dado que el error de los sensores tiende a concentrarse en las últimas componentes o variables latentes en lugar de las primeras.

Un enfoque que aborda de manera directa el problema del error en ambas variables es la regresión por mínimos cuadrados totales (Total Least Squares, TLS) [55]. En este tipo de regresión, tanto

$X$  como  $Y$  se consideran variables aleatorias afectadas por error, lo que permite estimar la relación subyacente entre ellas sin asumir que una de las variables es exacta. A diferencia de OLS, donde la minimización se realiza únicamente en la dirección vertical, TLS busca minimizar la distancia ortogonal entre los datos y el modelo, incorporando así los errores presentes en ambas dimensiones. Este enfoque ha sido ampliamente discutido en el contexto de modelos error-in-variables y resulta especialmente adecuado cuando las mediciones experimentales presentan incertidumbre en todas las variables involucradas [55].

La **Figura 2.5** ilustra de forma conceptual la diferencia entre los criterios de minimización empleados en OLS y TLS. Mientras que OLS considera únicamente los residuos en la dirección de  $Y$ , TLS tiene en cuenta las desviaciones tanto en  $X$  como en  $Y$ , reflejando de manera más realista la naturaleza de los errores de medición en sistemas experimentales.

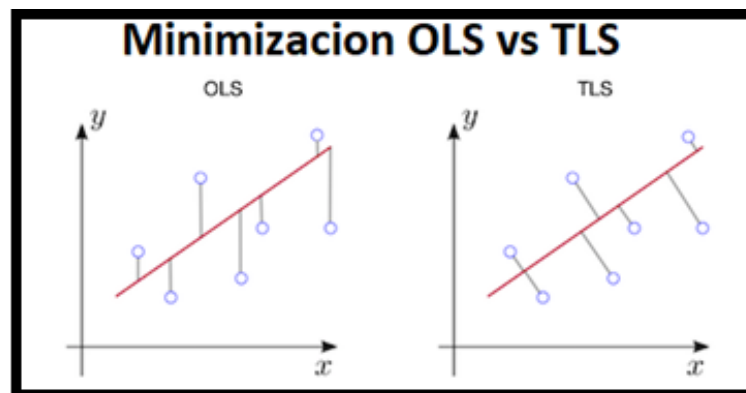


**Figura 2.5.** diferencia entre las minimizaciones OLS y TLS, Elaboración propia basada en [55].

Sin embargo, la regresión TLS enfrenta un desafío significativo al aplicarse a datos hiperespectrales debido a la fuerte multicolinealidad existente entre las bandas espectrales. En estos casos, la alta correlación entre las variables predictoras puede dar lugar a relaciones casi lineales durante la

calibración del modelo, lo que afecta su estabilidad numérica y compromete la capacidad de validación. Una posible estrategia para mitigar este problema consiste en reducir el número de bandas espectrales y conservar únicamente aquellas más informativas; no obstante, esta reducción puede implicar la pérdida de información relevante. Por esta razón, la regresión TLS suele resultar más adecuada en sistemas multiespectrales o en configuraciones con reducción dimensional previa, mientras que en el ámbito hiperespectral su uso se ha explorado principalmente a través de formulaciones específicas, como variantes estructuradas de mínimos cuadrados totales aplicadas en problemas de unmixing espectral y contextos geo-mineralógicos [36].

### **2.1.8 Support Vector Machine Regresion (SVR).**

La Regresión de Vectores de Soporte (Support Vector Regression, SVR) es un algoritmo de aprendizaje supervisado derivado de las máquinas de soporte vectorial (Support Vector Machines, SVM), orientado a problemas de regresión. En el contexto del análisis de datos hiperespectrales, SVM puede utilizarse tanto para clasificación como para regresión; en este último caso, el objetivo es encontrar una función que describa adecuadamente la relación entre las variables de entrada y la variable de salida, controlando simultáneamente el error de ajuste y la complejidad del modelo [37]. A diferencia de la regresión lineal clásica basada en mínimos cuadrados ordinarios (OLS), que busca minimizar directamente la distancia entre los datos observados y la recta de regresión, SVR adopta un enfoque alternativo. Como se ilustra en la Figura 2.6, SVR define una banda de tolerancia de ancho  $2\epsilon$  alrededor de la función de regresión, dentro de la cual los errores no son penalizados.

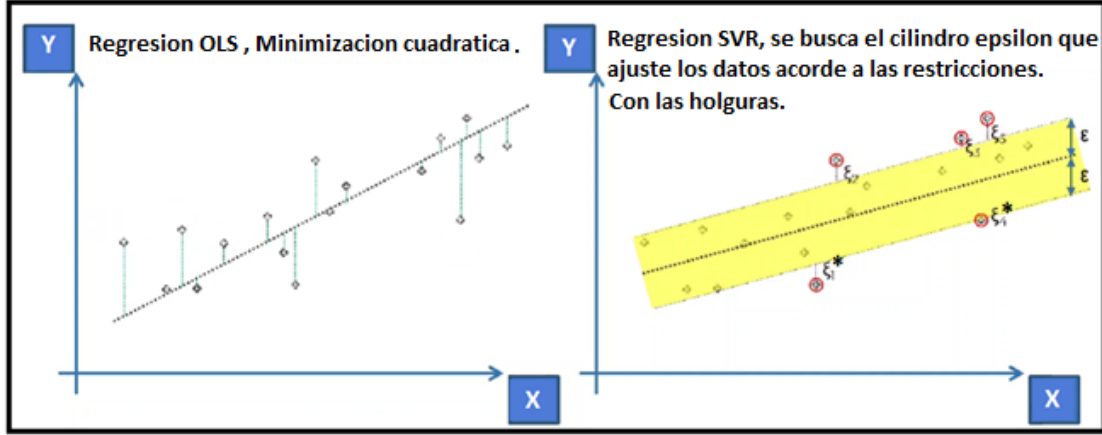


**Figura 2.6. Comparación conceptual entre la regresión OLS (izquierda) y la regresión SVR (derecha). Basado en [37].**

El parámetro  $\epsilon$  representa el margen de tolerancia permitido en las predicciones. Mientras las estimaciones del modelo se mantengan dentro de este margen, no se incurre en penalización. De esta forma, SVR introduce flexibilidad en el ajuste, permitiendo cierto grado de error admisible sin afectar la función objetivo.

No obstante, los datos que se encuentran fuera del margen  $\epsilon$  son penalizados mediante variables de holgura, denotadas como  $\xi_i$  y  $\xi_i^*$ , las cuales representan las desviaciones por encima y por debajo del margen, respectivamente. Estos puntos suelen asociarse a observaciones atípicas o *outliers*.

**Formulación del problema SVR.** Para el caso de un kernel lineal, la función de regresión puede expresarse como:

$$f(x) = \mathbf{w}^T \mathbf{x} + b_0 \quad (16)$$

Donde  $\mathbf{x} \in \mathbb{R}^m$  representa un vector de entrada correspondiente a un espectro individual, mientras que  $\mathbf{X} \in \mathbb{R}^{n \times m}$  denota la matriz completa de datos formada por  $n$  espectros con  $m$  bandas espectrales. El vector  $\mathbf{w}$  corresponde a los pesos asociados a los vectores de soporte y  $b_0$  es el término de sesgo del modelo. El vector  $\mathbf{Y} \in \mathbb{R}^{n \times 1}$  contiene los valores de referencia asociados a cada espectro.

El objetivo inicial consiste en minimizar la norma del vector de pesos, lo que conduce a la siguiente función objetivo:

$$J(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} = \frac{1}{2} \|\mathbf{w}\|^2 \quad (17)$$

sujeta a la restricción:

$$\forall n: |y_n - (w^T x_n + b_0)| \leq \varepsilon \quad (18)$$

En la práctica, es frecuente que esta restricción no pueda cumplirse estrictamente para todos los datos.

Para abordar este problema, SVR introduce una función de pérdida  $\varepsilon$ -insensible, definida como:

$$L(y, f(x)) = \begin{cases} 0, & |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon, & \text{en caso contrario} \end{cases} \quad (19)$$

Esta formulación permite ignorar errores pequeños dentro del margen y penalizar únicamente las desviaciones significativas.

Para relajar las restricciones y permitir soluciones factibles, el problema de optimización se reformula incorporando las variables de holgura  $\xi_i$  y  $\xi_i^*$ , junto con el parámetro de regularización  $C > 0$ , que controla la penalización asociada a los errores fuera del margen:

$$\min_{w, b_0, \xi, \xi^*} \left( \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right) \quad (20)$$

sujeto a:

$$y_i - w^T x_i - b_0 \leq \varepsilon + \xi_i \quad (21)$$

$$w^T x_i + b_0 - y_i \leq \varepsilon + \xi_i^* \quad (22)$$

$$\xi_i \geq 0, \xi_i^* \geq 0 \quad (23)$$

El problema de SVR puede resolverse de forma eficiente mediante su formulación dual, basada en multiplicadores de Lagrange. Esta formulación permite expresar el problema como:

$$\max_{\alpha, \alpha^*} \left( -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle + \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) \right) \quad (24)$$

sujeto a:

$$\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \quad (25)$$

$$0 \leq \alpha_i, \alpha_i^* \leq C \quad (26)$$

A partir de la solución dual, el vector de pesos se obtiene como:

$$w = \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i \quad (27)$$

y el sesgo  $b_0$  se calcula utilizando las condiciones de Karush–Kuhn–Tucker (KKT). La predicción para un nuevo vector  $x$  se expresa como:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b_0 \quad (28)$$

Una de las principales fortalezas de SVR es la posibilidad de modelar relaciones no lineales mediante el uso de funciones kernel:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x, x_i) + b_0 \quad (29)$$

Donde  $K(x, x_i)$  es una función kernel que define la similitud entre los vectores. Entre los kernels más utilizados se encuentran:

El kernel lineal es el caso más simple y asume una relación aproximadamente lineal entre los espectros y la variable de salida. Su uso resulta adecuado cuando los datos han sido previamente transformados o cuando se espera que la relación espectro–propiedad pueda describirse mediante una combinación lineal de las bandas. Además, este kernel ofrece una mayor interpretabilidad del modelo y un menor costo computacional.

$$K(x, x_i) = x^T x_i \quad (30)$$

El kernel polinomial permite modelar relaciones no lineales de orden superior, siendo útil cuando se sospecha que la relación entre las variables sigue una estructura polinómica. Sin embargo, el uso de grados elevados puede aumentar el riesgo de sobreajuste, especialmente en conjuntos de datos con ruido espectral o número limitado de muestras.

$$K(x, x_i) = (x^T x_i + c)^d \quad (31)$$

El kernel radial de base gaussiana (RBF) es uno de los más utilizados en aplicaciones hiperespectrales debido a su alta flexibilidad. Este kernel es capaz de capturar relaciones complejas y no lineales sin requerir una suposición previa sobre la forma de la relación entre  $\mathbf{X}$  e  $\mathbf{Y}$ . No obstante, su desempeño depende fuertemente de la correcta selección del parámetro  $\gamma$ , el cual controla el alcance de la influencia de cada vector de soporte. Valores inadecuados de este parámetro pueden conducir a sobreajuste o subajuste del modelo.

$$K(x, x_i) = \exp(-\gamma \|x - x_i\|^2) \quad (32)$$

El kernel sigmoideal introduce una relación no lineal de tipo sigmoide, similar a la función de activación utilizada en redes neuronales. Aunque es menos común en regresión hiperespectral, puede resultar útil en casos específicos donde la relación entre los datos espectrales y la variable de interés presenta un comportamiento saturante. Su aplicación requiere una cuidadosa selección de parámetros para asegurar la estabilidad del modelo.

$$K(x, x_i) = \tanh(\alpha x^T x_i + c) \quad (33)$$

La elección del kernel y de sus parámetros asociados es un aspecto clave del desempeño del modelo, ya que determina la capacidad de SVR para capturar relaciones complejas entre las variables.

En el ámbito de la mineralogía y la estimación cuantitativa a partir de datos espectrales, la regresión de vectores de soporte (SVR) ha demostrado ser una herramienta eficaz para modelar relaciones no lineales entre los espectros y las propiedades de interés. En particular, SVR ha sido empleada como modelo principal para la estimación de contenidos mineralógicos a partir de datos hiperespectrales de campo, mostrando una alta capacidad predictiva y una adecuada generalización del modelo [38]. Asimismo, este enfoque ha sido aplicado para relacionar parámetros espectrales extraídos de imágenes hiperespectrales en el rango SWIR con indicadores de mineralización, permitiendo vincular directamente características espectrales con procesos hidrotermales y zonación mineralógica [39]. Estos antecedentes respaldan el uso de SVR como una alternativa robusta para la cuantificación mineralógica y elemental en contextos geológicos donde las relaciones entre las variables no son estrictamente lineales.

## 2.2. Consideraciones adicionales en las regiones de interés (ROI).

En imágenes hiperespectrales, la región de interés (ROI) corresponde al conjunto de píxeles que representa la muestra a analizar, al cual se le asocia un valor de referencia  $Y$ . De manera habitual, la representación del ROI en modelos de regresión se realiza a partir del cálculo de un espectro promedio de los píxeles seleccionados, reduciendo así el cubo hiperespectral a un único vector espectral por muestra, como se muestra en la Figura 2.7.

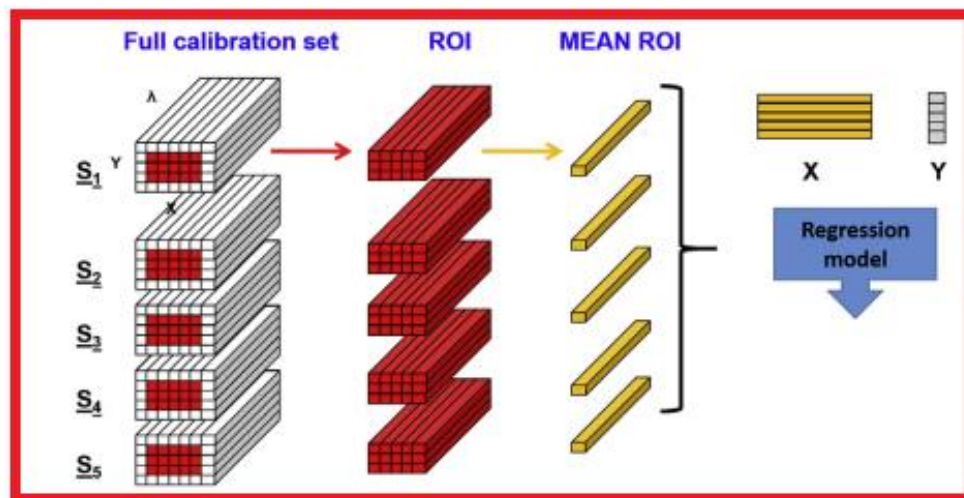


**Figura 2.7 Metodología clásica para representar una muestra en modelos de regresión utilizando imágenes hiperespectrales [22].**

Si bien este enfoque es simple y ampliamente utilizado, puede presentar limitaciones cuando existe una alta variabilidad espectral dentro del ROI. En muestras heterogéneas, como materiales particulados o mezclas minerales, los espectros de los píxeles pueden diferir entre sí debido a factores como la textura superficial, la dispersión del material, la presencia de micro-sombras, variaciones en la granulometría o efectos de mezcla a escala subpíxel. En estas situaciones, el espectro promedio puede no ser completamente representativo del comportamiento espectral real de la muestra, lo que puede afectar la robustez del modelo de calibración.

Para abordar esta problemática de manera práctica, en este trabajo se optó por incorporar la variabilidad mediante replicación experimental. Dado que las muestras se midieron en formato de polvo sobre placas Petri cuadradas, cada muestra fue medida en tres ocasiones independientes, modificando deliberadamente la disposición del material en cada adquisición. De esta forma, se capturaron variaciones reales asociadas a la dispersión del polvo y a la preparación de la muestra, considerando cada medición como una observación adicional dentro del conjunto de calibración.

Este enfoque permite preservar la variabilidad física del sistema de medición y resulta especialmente adecuado para aplicaciones de cuantificación mineralógica y elemental, ya que mejora la representatividad del conjunto de entrenamiento y aumenta la robustez del modelo frente a variaciones inevitables en la preparación y medición de muestras particuladas.

### **2.3. Consideraciones en la reducción de dimensionalidad.**

La reducción de dimensionalidad es un paso importante en el análisis de datos hiperespectrales, ya que éstos poseen un número muy grande de bandas espectrales que están altamente correlacionadas entre sí. En estos datos, técnicas lineales como PCA pueden proyectar los datos originales en un espacio de menor dimensión de forma determinista, preservando la estructura global de los datos y facilitando su interpretación en las fases posteriores de modelado [40].

Igualmente, la Regresión por Mínimos Cuadrados Parciales (PLS) integra reducción de dimensionalidad y análisis multivariante, siendo ampliamente utilizada en problemas de calibración y regresión espectral. Este tipo de aproximación es especialmente apropiada en situaciones de cuantificación, ya que proporciona modelos estables, reproducibles y con una interpretación física, incluso en datos de alta dimensión [40].

Por otro lado, métodos no lineales como t-SNE y UMAP se aplican principalmente para exploración y visualización. Estos enfoques buscan mantener vecindades locales entre las muestras y están sujetos a algoritmos de optimización no deterministas, lo que puede dar lugar a resultados sensibles a hiperparámetros y entre ejecuciones [41]. Por ende, su uso directo en modelos de regresión cuantitativa no es el más apropiado.

Por estas razones, en aplicaciones de cuantificación mineralógica y elemental basadas en imágenes hiperespectrales se opta por técnicas lineales y determinísticas como PCA y PLS para la reducción de dimensionalidad, mientras que los métodos no lineales se reservan para análisis exploratorios complementarios [40,41].

#### **2.4. ¿Qué tipo de técnica de regresión es la más adecuada al contexto?**

Al analizar los distintos modelos disponibles para la cuantificación mediante cámaras hiperespectrales, surge naturalmente la necesidad de evaluar qué técnica de regresión resulta más adecuada para capturar la relación entre la información espectral y las variables de interés. En este contexto, esta elección implica un compromiso entre capacidad de ajuste, robustez y capacidad de generalización del modelo, especialmente cuando se trabaja con datos de alta dimensionalidad y un número limitado de muestras.

Desde el enfoque clásico de la quimiometría, técnicas como la regresión lineal múltiple (MLR), la regresión de componentes principales (PCR) y la regresión por mínimos cuadrados parciales (PLS) han sido ampliamente utilizadas. En particular, PLS destaca por su capacidad de construir variables latentes que maximizan la covarianza entre las variables espectrales y la variable de respuesta, lo que la convierte en una herramienta especialmente adecuada en presencia de multicolinealidad y alta dimensionalidad, condiciones típicas de los datos hiperespectrales.

Por otro lado, el aprendizaje automático ofrece alternativas adicionales, como la regresión de vectores de soporte (SVR), las redes neuronales artificiales basadas en perceptrones multicapa (MLP) y los

métodos basados en árboles de decisión, como Random Forest (RFR) y XGBoost (XGB). Estos enfoques permiten modelar relaciones no lineales complejas; sin embargo, su aplicación directa puede conducir a sobreajuste cuando el número de muestras es reducido, una situación frecuente en estudios experimentales con datos espectrales.

En este contexto, SVR y MLP representan alternativas intermedias entre modelos estrictamente lineales y métodos altamente no lineales. Desde un punto de vista teórico, los perceptrones multicapa constituyen aproximadores universales de funciones, lo que implica que, bajo condiciones adecuadas, pueden modelar cualquier relación no lineal entre las variables de entrada y salida. No obstante, en la práctica, el desempeño de MLP depende fuertemente del tamaño del conjunto de entrenamiento, del diseño de la arquitectura y del proceso de optimización, lo que puede favorecer el sobreajuste y una mayor variabilidad del modelo cuando el número de muestras es limitado.

Estudios comparativos han mostrado que, en escenarios con tamaños de muestra reducidos, los modelos basados en SVR tienden a presentar un desempeño más estable y robusto que las redes neuronales, manteniendo errores de predicción más bajos y menor sensibilidad a la variabilidad del entrenamiento [42].

SVR permite controlar explícitamente la complejidad del modelo mediante el parámetro de regularización y la selección del kernel, lo que resulta particularmente ventajoso en datos espectrales de alta dimensión. Sin embargo, una limitación práctica de SVR aplicada directamente sobre espectros completos es su sensibilidad a la multicolinealidad presente en las variables de entrada, lo que puede afectar la estabilidad del modelo.

Para abordar esta limitación, se han propuesto enfoques híbridos que combinan técnicas de reducción de dimensionalidad con modelos SVR. En particular, la combinación de PLS con SVR, conocida como PLS-SVR, ha sido ampliamente estudiada en el ámbito del análisis espectroscópico, mostrando mejoras consistentes en precisión y robustez frente al uso individual de PLS o SVR [43]. En este enfoque, las variables latentes obtenidas mediante PLS se utilizan como entrada del modelo SVR, lo que permite reducir la dimensionalidad, mitigar la multicolinealidad y preservar la información relevante asociada a la variable de respuesta.

Resultados similares han sido reportados en estudios de minería de datos y análisis multivariante, donde la integración de PLS con máquinas de soporte vectorial ha demostrado un mejor desempeño en conjuntos de datos complejos y altamente correlacionados [44]. Asimismo, la efectividad de estos enfoques híbridos ha sido validada en aplicaciones industriales de soft sensing, como la estimación de variables de proceso en fermentación de penicilina, donde la combinación PLS-SVR presenta mejoras en precisión y estabilidad respecto a modelos individuales [45]. De manera análoga, esquemas basados en componentes extraídos mediante PCA y PLS, utilizados como entrada para modelos de regresión posteriores, han mostrado un desempeño robusto en problemas de estimación con conjuntos de datos limitados, como la predicción de edad cerebral a partir de características altamente correlacionadas [46].

PLS-SVR se presenta como una alternativa llamativa para la cuantificación mineralógica y elemental a partir de imágenes hiperespectrales. Este enfoque integra la estructura latente y la reducción de dimensionalidad propias de PLS con la flexibilidad y el control de complejidad de SVR, permitiendo capturar relaciones no estrictamente lineales sin sacrificar robustez ni estabilidad. Además, esta formulación resulta naturalmente extensible a esquemas de fusión de datos en distintos niveles.

## **2.5. Fusión de datos.**

La fusión de datos corresponde a un conjunto de estrategias orientadas a integrar información proveniente de distintas fuentes con el objetivo de mejorar la calidad, robustez y capacidad predictiva de los modelos. En el contexto del análisis hiperespectral y multiesensorial, este enfoque permite aprovechar información complementaria capturada por diferentes sensores, lo que resulta especialmente relevante cuando se busca cuantificar propiedades complejas, como la composición mineralógica o elemental de minerales y concentrados.

Dependiendo del nivel en el que se realice la integración de la información, la fusión de datos suele clasificarse en tres enfoques principales: fusión a bajo nivel (Low-Level Data Fusion, LLDF), fusión a nivel medio (Mid-Level Data Fusion, MLDF) y fusión a alto nivel (High-Level Data Fusion, HLDF), tal como se esquematiza en la Figura 2.8

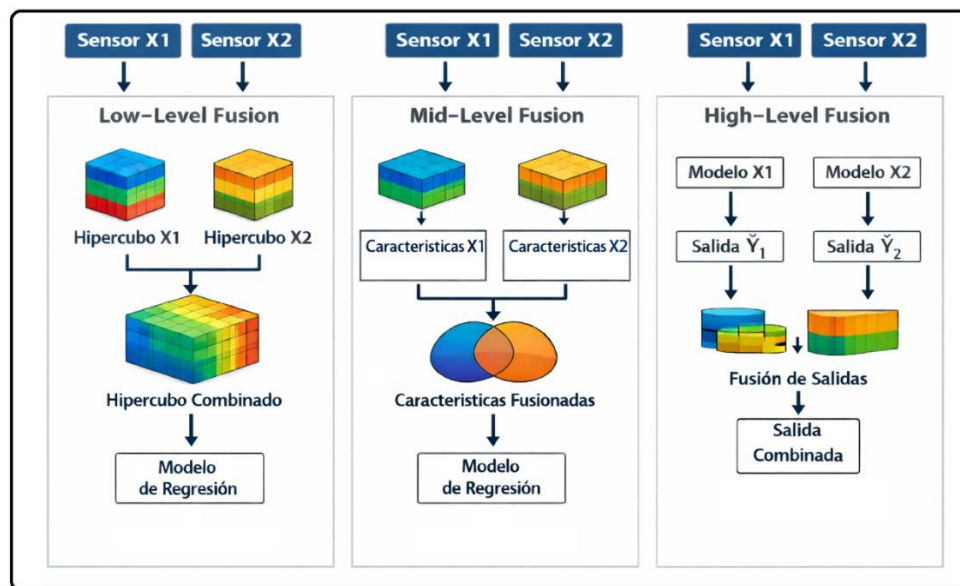


**Figura 2.8** Esquema conceptual de los distintos niveles de fusión de datos: fusión a bajo nivel (LLDF), fusión a nivel medio (MLDF) y fusión a alto nivel (HLDF). En LLDF se combinan los datos originales, en MLDF se integran características extraídas de cada fuente y en HLDF se fusionan las salidas de modelos independientes.

La fusión a bajo nivel se basa en la combinación directa de los datos originales o preprocesados provenientes de distintas fuentes, siendo la concatenación de la información la estrategia más comúnmente utilizada. En este enfoque, los espectros o hipercubos adquiridos por diferentes sensores, por ejemplo en distintos rangos espectrales como VIS-NIR y SWIR, o mediante técnicas de distinta naturaleza como reflectancia difusa y LIBS, se integran en una única matriz de entrada. Este tipo de fusión ha mostrado resultados favorables en el análisis de concentrados de cobre, donde la concatenación directa de señales espectrales ha permitido mejorar la caracterización de especies minerales [47], así como en la clasificación conjunta a partir de imágenes hiperespectrales multibanda [49]. No obstante, al preservar toda la información original, LLDF suele generar conjuntos de datos de muy alta dimensionalidad, lo que incrementa la multicolinealidad y el riesgo de sobreajuste, especialmente cuando el número de muestras disponibles es limitado.

En este contexto, la fusión a nivel medio se presenta como una alternativa más adecuada cuando se trabaja con múltiples sensores y un elevado número de variables. En MLDF, los datos provenientes de cada sensor se preprocesan de forma independiente y se someten a técnicas de extracción o selección de características, con el objetivo de retener únicamente la información más relevante de cada fuente. Posteriormente, estas características reducidas se concatenan en un espacio común para el entrenamiento del modelo. Este enfoque permite disminuir la redundancia espectral, controlar la

dimensionalidad y mejorar la estabilidad de los modelos predictivos. Diversos estudios han demostrado que la fusión a nivel medio ofrece un equilibrio favorable entre complejidad y desempeño, mostrando mejoras consistentes en precisión y robustez tanto en la identificación de minerales mediante la combinación de espectroscopía Raman y LIBS [48], como en el análisis mineralógico de concentrados de cobre a partir de datos hiperespectrales y LIBS [50]. Por estas razones, MLDF ha sido ampliamente adoptada en aplicaciones espectroscópicas y multiesensoriales.

Por su parte, la fusión a alto nivel se realiza a partir de las salidas de modelos independientes entrenados para cada fuente de datos. En este esquema, las predicciones individuales se combinan mediante reglas de decisión, promedios ponderados u otros mecanismos de agregación. Si bien este enfoque presenta una mayor simplicidad conceptual, su aplicación en problemas de cuantificación espectral es menos frecuente, ya que requiere la combinación de múltiples modelos y no explota de forma directa las relaciones internas entre las variables originales.

En general, HLDF suele considerarse más apropiada en etapas finales de implementación o en contextos industriales más avanzados, donde existen modelos individuales bien validados y el objetivo principal es integrar decisiones a nivel de sistema.

En todos los esquemas de fusión de datos, las técnicas de compresión de información y selección de características juegan un rol fundamental, ya que permiten reducir la redundancia espectral, mejorar la capacidad de generalización de los modelos y disminuir los costos computacionales asociados al entrenamiento. En particular, la evidencia reportada en la literatura sugiere que la fusión a nivel medio constituye una estrategia especialmente efectiva para aplicaciones multiesensoriales en minerales y concentrados, reforzando el potencial de la fusión de datos como una herramienta clave en sistemas avanzados de cuantificación y monitoreo industrial [47–50].

## **2.6. Preprocesamientos.**

En el análisis quimiométrico de datos espectrales, el preprocesamiento constituye una etapa fundamental previa al modelado, independientemente de si los datos provienen de **un único sensor** o de **múltiples fuentes de información** combinadas mediante esquemas de fusión de datos. Su objetivo

principal es reducir la influencia de variaciones no relacionadas directamente con la propiedad de interés, tales como ruido instrumental, efectos de dispersión de la radiación, desplazamientos de línea base o diferencias de escala entre variables. Desde un punto de vista formal, los datos espectrales se organizan en una matriz  $\mathbf{X} \in \mathbb{R}^{n \times m}$ , donde  $n$  representa el número de muestras y  $m$  el número de variables espectrales (longitudes de onda, bandas o características extraídas). La variable respuesta, denotada como  $\mathbf{Y}$ , corresponde a la propiedad química, mineralógica o composicional que se desea modelar mediante técnicas de calibración multivariante. En este contexto, el preprocesamiento no busca modificar la información química contenida en  $\mathbf{X}$ , sino **mejorar la relación entre la variabilidad relevante de los datos y la respuesta**  $y$ , facilitando la extracción de patrones significativos por parte de los modelos quimiométricos, tal como se discute en el marco general de la calibración multivariante [51].

### 2.6.1 Centrado en la media (Mean centering)

El **centrado en la media** es un preprocesamiento básico del análisis multivariado que consiste en restar, a cada variable espectral, su valor medio calculado sobre el conjunto de calibración. En el análisis espectral, los datos se organizan en una matriz  $\mathbf{X}$ , donde cada fila  $\mathbf{x}_i$  representa el espectro correspondiente a la muestra  $i$ , y cada columna corresponde a una variable espectral.

Este preprocesamiento no modifica la forma relativa de los espectros individuales  $\mathbf{x}_i$ , pero resulta especialmente relevante para el correcto funcionamiento de **modelos de regresión basados en variables latentes**, como la **Regresión por Componentes Principales (PCR)** y la **Regresión por Mínimos Cuadrados Parciales (PLS)**. El centrado en la media permite que las variables latentes extraídas por estos modelos describan variaciones alrededor del comportamiento promedio de los datos, evitando que las regresiones estén dominadas por desplazamientos absolutos de intensidad entre variables.

Formalmente, para una matriz de datos  $\mathbf{X}$ , el centrado en la media para un espectro  $\mathbf{x}_i$  se expresa como:

$$x_{ij}^{(c)} = x_{ij} - \bar{x}_j \quad (34)$$

Donde  $x_{ij}$  corresponde al valor de la variable espectral  $j$  en el espectro  $\mathbf{x}_i$ , y  $\bar{x}_j$  representa la media de dicha variable calculada sobre todas las muestras del conjunto de calibración.

En la práctica, el centrado en la media es **uno de los preprocesamientos más utilizados** en análisis quimiométrico, ya que facilita la interpretación de las variables latentes y contribuye a la estabilidad numérica de los modelos de regresión multivariante, aunque su conveniencia final depende del tipo de datos y del objetivo del análisis [51].

### 2.6.2 Autoescalado (Autoscaling)

El **autoescalado**, también conocido como normalización tipo **z-score**, es un preprocesamiento estadístico que se aplica **variable a variable**, restando la media y dividiendo por la desviación estándar de cada variable espectral [52]. Su objetivo principal es **homogeneizar la varianza de las variables**, evitando que aquellas con mayor amplitud dominen el proceso de modelado.

En el análisis espectral, el autoescalado no corrige efectos físicos de dispersión —a diferencia de técnicas como MSC o SNV—, sino que actúa a nivel estadístico, equilibrando la contribución relativa de las variables al modelo. Este preprocesamiento es particularmente útil cuando las variables presentan diferentes escalas o niveles de variabilidad, como ocurre en problemas de **calibración multivariante** y en escenarios de **fusión de datos multifuente** [52].

Para una matriz de datos  $\mathbf{X}$ , para un espectro  $\mathbf{x}_i$  el autoescalado se define como:

$$\mathbf{x}_{ij}^{(auto)} = \frac{\mathbf{x}_{ij} - \bar{\mathbf{x}}_j}{s_j} \quad (35)$$

Donde  $\mathbf{x}_{ij}$  corresponde al valor de la variable espectral  $j$  en el espectro  $\mathbf{x}_i$ , y  $\bar{\mathbf{x}}_j$  y  $s_j$  representan la media y la desviación estándar de la variable  $j$ , respectivamente.

El uso del autoescalado contribuye a mejorar la estabilidad numérica y la interpretación de los modelos de regresión multivariante, aunque su aplicación debe evaluarse en función del tipo de datos y del objetivo del análisis [52].

### 2.6.3 Suavizado y derivadas de Savitzky–Golay (SG)

El método de **Savitzky–Golay (SG)** es una técnica de preprocesamiento espectral basada en el ajuste local de un polinomio de bajo orden dentro de una ventana móvil de puntos espectrales. A partir de

este ajuste, es posible obtener tanto una versión suavizada del espectro como sus derivadas analíticas, preservando la forma y posición local de las bandas espectrales [53].

Desde un punto de vista conceptual, el filtro de Savitzky–Golay reemplaza cada valor espectral por la evaluación de un modelo polinomial local ajustado en torno a dicho punto. Para cada longitud de onda  $k$ , se selecciona una ventana simétrica de  $w = 2m_0 + 1$  puntos vecinos del espectro  $x_i$ , y se aplican coeficientes de ponderación fijos  $c_l^{(d)}$  a los valores circundantes. Estos coeficientes se derivan analíticamente de manera que la suma ponderada sea matemáticamente equivalente al ajuste por mínimos cuadrados de un polinomio de grado  $p$  y a la evaluación de su derivada de orden  $d$  en el centro de la ventana [53,54]. Formalmente, la estimación del valor suavizado o derivado de orden  $d$  para el espectro  $x_i$  en la posición  $k$  se expresa como:

$$x_{i,k}^{(d)} = \sum_{l=-m_0}^{m_0} c_l^{(d)} x_{i,k+l} \quad (36)$$

En el caso particular del **suavizado** ( $d = 0$ ), la expresión puede escribirse como:

$$x_i^{(SG)} = \sum_{k=-m}^m c_k x_{i+k} \quad (37)$$

Donde  $2m + 1$  corresponde al tamaño de la ventana y  $c_k$  son coeficientes determinados por el ajuste polinomial. El suavizado mediante Savitzky–Golay se emplea principalmente para reducir el ruido aleatorio de alta frecuencia, preservando simultáneamente la estructura local de los espectros. A partir del mismo esquema, es posible calcular **derivadas espectrales**, las cuales cumplen distintos propósitos: la **derivada de primer orden** permite corregir desplazamientos constantes de línea base y resaltar variaciones locales en la señal, mientras que la **derivada de segundo orden** contribuye a eliminar curvaturas lentas de fondo y mejorar la resolución de bandas espectrales solapadas [53].

El desempeño del filtro depende de manera crítica de la selección de sus parámetros, en particular del **grado del polinomio  $p$**  y del **tamaño de la ventana  $w$** . En aplicaciones espectroscópicas, se utilizan habitualmente polinomios de **bajo orden**, típicamente de **segundo o tercer grado**, ya que permiten describir adecuadamente la curvatura local de las bandas sin introducir oscilaciones artificiales

asociadas a un ajuste excesivo [53,54]. De forma complementaria, el tamaño de la ventana suele seleccionarse de manera moderada, con longitudes del orden de **5 a 21 puntos**, buscando un compromiso entre la reducción de ruido y la preservación de características espectrales relevantes [54]. Dado que el uso de derivadas incrementa la sensibilidad al ruido, la elección conjunta del grado del polinomio ***p***, del tamaño de la ventana ***w*** y del orden de la derivada ***d*** debe realizarse cuidadosamente y evaluarse mediante procedimientos de validación cruzada o análisis comparativo del desempeño de los modelos quimiométricos, tal como se recomienda en la literatura clásica de calibración multivariante y preprocesamiento espectral.

#### 2.6.4 Corrección de dispersión multiplicativa (MSC)

La **Multiplicative Scatter Correction (MSC)** es una técnica de preprocesamiento diseñada para compensar efectos de dispersión de la radiación que suelen aparecer en espectros de reflectancia difusa, especialmente cuando se analizan muestras sólidas o pulverulentas. Estos efectos físicos, no relacionados directamente con la composición química del material, pueden generar variaciones **aditivas** (desplazamientos de línea base) y **multiplicativas** (cambios de escala) en la señal espectral, asociadas, por ejemplo, a diferencias en tamaño de partícula, rugosidad superficial, densidad de empaquetamiento o trayectorias ópticas variables.

Desde el punto de vista conceptual, MSC parte de la hipótesis de que cada espectro medido puede interpretarse como una versión linealmente distorsionada de un espectro de referencia. Dicho espectro de referencia se define, por lo general, como el espectro promedio obtenido a partir del conjunto de calibración. Bajo esta hipótesis, para cada muestra ***i***, se ajusta un modelo lineal entre el espectro individual ***x<sub>i</sub>*** y el espectro de referencia ***̄x***, permitiendo estimar un coeficiente multiplicativo ***a<sub>i</sub>*** y un término aditivo ***b<sub>i</sub>***, los cuales representan los efectos de dispersión presentes en la medición.

Formalmente, el modelo de MSC se expresa como:

$$\mathbf{x}_i = \mathbf{a}_i \bar{\mathbf{x}} + \mathbf{b}_i + \boldsymbol{\varepsilon}_i \quad (38)$$

Donde ***ε<sub>i</sub>*** representa el término de error residual.

El espectro de referencia ***̄x*** se calcula como el promedio de todos los espectros del conjunto de datos:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (39)$$

Los coeficientes  $a_i$  y  $b_i$  se estiman mediante mínimos cuadrados. Una vez obtenidos estos parámetros, el espectro corregido mediante MSC se calcula como:

$$x_i^{(\text{MSC})} = \frac{x_i - b_i}{a_i} \quad (40)$$

Desde un punto de vista práctico, la aplicación de MSC permite reducir de forma efectiva la variabilidad espectral asociada a efectos físicos, contribuyendo a que los espectros individuales se alineen mejor con el comportamiento medio del conjunto de datos. Este proceso se realiza sin modificar de manera significativa la información química contenida en las bandas espectrales. En la literatura se ha reportado que el uso de MSC puede mejorar la relación de tipo lineal entre la señal espectral y la propiedad de interés, además de aumentar la robustez de los modelos de calibración multivariante, particularmente en estudios de espectroscopía de reflectancia difusa en el infrarrojo cercano [55].

Cabe señalar que MSC actúa sobre cada espectro de forma individual, utilizando un referente global común. Por esta razón, su efecto depende en gran medida de la calidad del conjunto de calibración y de cuán representativo sea el espectro promedio empleado como referencia. En consecuencia, aunque su aplicación puede traducirse en mejoras apreciables en la estabilidad y el desempeño de modelos de regresión multivariante, su conveniencia final debe evaluarse considerando el tipo de material analizado, el sistema de medición utilizado y el objetivo analítico del estudio [52,55].

### 2.6.5 Normalización Standard Normal Variate (SNV)

La **Standard Normal Variate (SNV)** es una técnica de normalización que se aplica **espectro a espectro**, cuyo objetivo es corregir efectos físicos de dispersión presentes en datos espectrales, especialmente en mediciones de reflectancia difusa. Desde un punto de vista quimiométrico, SNV busca eliminar variaciones aditivas y multiplicativas que no están directamente relacionadas con la composición del material, mejorando así la relación entre la señal espectral y la variable de interés [51].

A diferencia de otros preprocesamientos estadísticos, SNV actúa de manera independiente sobre cada muestra, sin utilizar información del resto del conjunto de datos ni requerir la definición de un espectro de referencia. En este sentido, su funcionamiento es conceptualmente distinto al del autoescalado, que se aplica variable a variable, y similar al de MSC en cuanto a su orientación hacia la corrección de efectos físicos, pero con una formulación más simple [52].

Formalmente, para un espectro individual  $\mathbf{x}_i$ , la transformación SNV se expresa como:

$$\mathbf{x}_{ij}^{(SNV)} = \frac{\mathbf{x}_{ij} - \bar{\mathbf{x}}_i}{s_i} \quad (41)$$

donde  $\bar{\mathbf{x}}_i$  y  $s_i$  corresponden a la media y a la desviación estándar de todas las variables espectrales del espectro  $\mathbf{x}_i$ , respectivamente.

Mediante esta transformación, cada espectro queda centrado y normalizado en función de su propia distribución de intensidades, reduciendo la influencia de variaciones globales de offset y escala. En la práctica, SNV resulta especialmente útil en el análisis de espectros de reflectancia difusa y de imágenes hiperespectrales de materiales sólidos, donde las heterogeneidades físicas pueden dominar la variabilidad espectral no asociada a la composición química. Su aplicación contribuye a mejorar la estabilidad y el desempeño de modelos de regresión multivariante, aunque su conveniencia final depende del tipo de material analizado y del objetivo del estudio [51,52].

### 2.6.6 Selección de bandas espectrales relevantes

La selección de bandas espectrales se utiliza principalmente para reducir la dimensionalidad de los datos espectrales, eliminando variables redundantes o con escaso contenido informativo. Con ello, se busca no solo simplificar el problema, sino también mejorar la interpretabilidad y la estabilidad de los modelos multivariados, así como su capacidad de generalización. En conjuntos de datos espectrales, donde existe una alta colinealidad entre longitudes de onda adyacentes, esta etapa adquiere especial importancia, ya que permite mitigar el riesgo de sobreajuste tanto en análisis basados en un único sensor como en esquemas de fusión de datos [51].

Desde la perspectiva de la quimiometría clásica, la selección de variables ha sido abordada a través de distintos criterios que se complementan entre sí. Entre los enfoques más tempranos se encuentra el análisis de los coeficientes de regresión obtenidos a partir de modelos lineales, como la regresión por

componentes principales (PCR) o la regresión por mínimos cuadrados parciales (PLS). Estos coeficientes permiten identificar longitudes de onda con una mayor influencia sobre la variable respuesta y ofrecen una primera aproximación interpretativa de la relevancia espectral. Sin embargo, su interpretación puede verse limitada cuando existe una fuerte colinealidad entre variables, situación habitual en datos hiperespectrales.

Con el objetivo de abordar esta limitación, se desarrollaron métodos de selección por intervalos espectrales, los cuales dividen el espectro en regiones contiguas y evalúan su capacidad predictiva mediante modelos multivariados.

Dentro de esta familia de enfoques, el método Interval Partial Least Squares (iPLS) propone fragmentar el rango espectral en intervalos y construir modelos PLS independientes para cada uno de ellos, de modo que sea posible comparar su desempeño y seleccionar aquellas regiones que concentran mayor información relevante para la predicción [56].

Siguiendo el mismo principio, el método Synergy Interval Partial Least Squares (siPLS) amplía esta idea al evaluar combinaciones de intervalos espectrales. De esta forma, se busca identificar regiones que actúan de manera conjunta y cuya contribución informativa podría no ser evidente cuando los intervalos se analizan de forma aislada [57].

A partir del uso práctico de modelos PLS, la literatura ha propuesto criterios derivados del propio modelo para evaluar la importancia relativa de las variables espectrales. Uno de los más difundidos es el **Variable Importance in Projection (VIP)**, que cuantifica la contribución de cada variable espectral a la explicación conjunta de la variabilidad de la matriz de predictores  $\mathbf{X}$  y de su relación con la variable respuesta  $y$ . Este tipo de criterio resulta especialmente útil en datos hiperespectrales de alta dimensionalidad, donde solo una fracción de las bandas contiene información químicamente relevante, permitiendo una reducción eficiente del número de variables consideradas.

Además de los enfoques anteriores, se han desarrollado métodos de selección de variables basados en estrategias de búsqueda y optimización, tales como los **algoritmos genéticos**, técnicas de **regularización** y métodos adaptativos como el **Competitive Adaptive Reweighted Sampling (CARS)**. Estos enfoques permiten explorar subconjuntos de variables de manera iterativa, priorizando aquellas bandas que contribuyen de forma más significativa al desempeño del modelo. En aplicaciones

a concentrados de cobre, se ha reportado que el uso de CARS en combinación con esquemas de fusión de datos contribuye a reducir la redundancia espectral y a mejorar la cuantificación de especies minerales [47].

Desde un punto de vista formal, el criterio VIP para la variable espectral  $j$  puede expresarse como:

$$VIP_j = \sqrt{m \frac{\sum_{a=1}^A SSY_a w_{j,a}^2}{\sum_{a=1}^A SSY_a}} \quad (42)$$

Donde  $m$  corresponde al número total de variables espectrales,  $a$  es el número de componentes latentes del modelo PLS,  $SSY_a$  representa la varianza de la variable respuesta explicada por la componente  $a$ , y  $w_{j,a}$  es el peso de la variable  $j$  en dicha componente. Esta formulación permite obtener una medida adimensional de importancia relativa, en la cual valores de VIP mayores a uno suelen asociarse a variables con relevancia predictiva superior al promedio, mientras que valores inferiores a uno se asocian generalmente a regiones espectrales redundantes o dominadas por ruido. En conjunto, la aplicación de distintos criterios de selección de bandas espectrales permite reducir el número de variables consideradas, disminuir el costo computacional y mejorar la robustez y generalización de los modelos multivariados.

La elección del método más adecuado depende del tipo de datos, del modelo de calibración empleado y del objetivo analítico del estudio, siendo habitual combinar distintos enfoques para lograr un compromiso adecuado entre simplicidad, interpretabilidad y desempeño predictivo [51].

## 2.7. Métricas de evaluación de las regresiones.

El desempeño de los modelos de regresión ha sido tradicionalmente evaluado mediante un conjunto de métricas quimiométricas que permiten caracterizar distintos aspectos de su comportamiento predictivo. En el contexto de la espectroscopía y la calibración multivariante, estas métricas buscan cuantificar tanto la capacidad explicativa del modelo como la magnitud y relevancia del error de predicción, así como su relación con la variabilidad intrínseca del conjunto de datos [41,42].

Entre las métricas más comúnmente reportadas en la literatura se encuentran el **coeficiente de determinación ( $R^2$ )**, el **Root Mean Square Error (RMSE)**, el **Ratio of Performance to Deviation (RPD)**, el **Relative Error of Prediction (REP%)** y el **Range Error Ratio (RER)**. Cada una de ellas

aporta una perspectiva diferente sobre el desempeño del modelo, por lo que su interpretación debe realizarse de manera conjunta y considerando el objetivo del análisis.

### 2.7.1 Coeficiente de determinación ( $R^2$ )

El **coeficiente de determinación**  $R^2$  cuantifica la fracción de la varianza de los valores de referencia que es explicada por las predicciones del modelo y se define como:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (43)$$

Donde  $y_i$  corresponde al valor de referencia de la muestra  $i$ ,  $\hat{y}_i$  es el valor predicho por el modelo y  $\bar{y}$  representa la media de los valores de referencia. Valores elevados de  $R^2$  indican una buena concordancia global entre valores predichos y medidos. Sin embargo, diversos autores han señalado que esta métrica es esencialmente descriptiva y no proporciona información directa sobre la magnitud del error de predicción ni sobre el desempeño fuera de la muestra, por lo que su interpretación debe complementarse con métricas basadas en el error [58,60].

### 2.7.2 Root Mean Square Error (RMSE)

El **Root Mean Square Error (RMSE)** representa la magnitud promedio del error de predicción, expresada en las mismas unidades que la variable de interés:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (44)$$

El **RMSE** constituye una medida absoluta de la precisión del modelo y permite evaluar directamente cuánto se desvían, en promedio, las predicciones respecto de los valores reales. Debido a su interpretación directa y a su independencia del rango del conjunto de datos, esta métrica es ampliamente considerada una de las más informativas para evaluar el desempeño predictivo de modelos de regresión, especialmente en aplicaciones quimiométricas con orientación práctica o industrial [58,60].

### 2.7.3 Normalized Root Mean Square Error (NRMSE) por rango.

El **Normalized Root Mean Square Error (NRMSE)** corresponde a una forma normalizada del error

cuadrático medio (RMSE), cuyo objetivo es expresar la magnitud del error de predicción en términos relativos a la escala de los datos analizados. A diferencia del RMSE, que se expresa en las mismas unidades de la variable de interés, el NRMSE permite comparar el desempeño de modelos en contextos donde la variabilidad de los datos puede diferir significativamente.

Existen distintas formas de normalizar el RMSE, tales como su relación con la media o la desviación estándar de los datos. No obstante, en este trabajo se adopta una normalización basada en el rango de variación de la variable objetivo, definida como:

$$NRMSE(\%) = \frac{RMSE}{y_{max} - y_{min}} \times 100 \quad (45)$$

Donde  $y_{max}$  y  $y_{min}$  corresponden a los valores máximo y mínimo de la variable analizada.

La elección de esta forma de normalización resulta particularmente adecuada en el contexto de concentrados de cobre, donde variables como el %Cu y las fracciones mineralógicas (calcopirita y pirita) presentan amplios rangos de variación asociados a cambios en la proporción relativa de fases minerales. En este escenario, el uso del rango permite contextualizar el error en función de la variabilidad total del sistema, facilitando su interpretación y comparación entre modelos.

En conjunto, el *NRMSE* basado en el rango constituye una métrica complementaria al RMSE y al REP%, permitiendo evaluar el desempeño predictivo tanto en términos absolutos como relativos a la variabilidad composicional del sistema.

#### 2.7.4 Ratio of Performance to Deviation (RPD)

El **Ratio of Performance to Deviation (RPD)** relaciona el error de predicción con la variabilidad intrínseca del conjunto de datos y se define como:

$$RPD = \frac{s_y}{RMSE} \quad (46)$$

Donde  $s_y$  corresponde a la desviación estándar de los valores de referencia. Esta métrica permite contextualizar el error de predicción en función de la dispersión natural de la propiedad analizada. En la literatura quimiométrica, se ha propuesto que valores de **RPD** inferiores a 1.5 se asocian a modelos con bajo desempeño predictivo, valores entre 1.5 y 2.0 corresponden a aplicaciones exploratorias, y valores superiores a 2.0 indican modelos con capacidad predictiva aceptable a buena [60].

No obstante, distintos autores han advertido que el **RPD** no debe interpretarse de manera aislada, ya

que depende tanto del **RMSE** como de la distribución del conjunto de datos. En particular, el uso acrítico de umbrales de **RPD** puede conducir a interpretaciones simplificadas del desempeño del modelo, lo que ha dado lugar al denominado “*RPD myth*” en la práctica quimiométrica [59]. Aun así, cuando se utiliza de forma complementaria, el **RPD** sigue siendo una métrica ampliamente reportada para contextualizar el error de predicción.

### 2.7.5 Relative Error of Prediction (REP%)

El **Relative Error of Prediction (REP%)** corresponde a una forma normalizada del RMSE respecto a la media de la variable de interés, y se expresa como:

$$REP\% = \frac{RMSE}{\bar{y}} \times 100 \quad (47)$$

Esta métrica proporciona una medida relativa del error de predicción en términos porcentuales, lo que facilita la comparación entre modelos aplicados a analitos con distintos rangos de concentración. En aplicaciones de cuantificación mineralógica o elemental, valores de REP% por debajo de aproximadamente 10–15% suelen considerarse indicativos de un desempeño aceptable, dependiendo del objetivo analítico y de la variabilidad del material [58,60]. Sin embargo, su interpretación puede verse afectada cuando la media del conjunto de datos es reducida.

### 2.7.6 Range Error Ratio (RER)

El **Range Error Ratio (RER)** se define como la razón entre el rango de la variable dependiente y el error de predicción:

$$RER = \frac{y_{max} - y_{min}}{RMSE} \quad (48)$$

Donde  $y_{max}$  y  $y_{min}$  corresponden a los valores máximo y mínimo de la variable de referencia, respectivamente. Conceptualmente, el RER busca relacionar la amplitud total del rango de la propiedad de interés con la magnitud del error de predicción del modelo, y ha sido reportado en algunos estudios como un indicador complementario de la calidad del modelo [60].

No obstante, revisiones críticas recientes han señalado que el RER presenta limitaciones importantes cuando el objetivo es evaluar la **capacidad predictiva real** de modelos quimiométricos. En particular,

su fuerte dependencia del rango experimental y, por ende, de valores extremos o del diseño del conjunto de datos puede conducir a interpretaciones poco robustas al comparar modelos desarrollados sobre diferentes datasets [58]. En este contexto, se observa en la literatura una preferencia sistemática por métricas basadas en el error absoluto y en la variabilidad estadística de los datos, como ***RMSE***, ***R<sup>2</sup>***, ***RPD*** y ***REP%***, las cuales ofrecen una evaluación más homogénea y comparable del desempeño predictivo entre distintas aplicaciones [58,59,60].

## 2.8. Conclusiones estado del arte.

Como se puede constatar en el estado del arte, la cuantificación de propiedades mineralógicas y elementales a partir de datos espectrales es un problema multivariado en esencia, de alta dimensionalidad, colineal y muy susceptible a efectos físicos no directamente asociados a la composición química. Estas propiedades han impulsado el desarrollo y la aplicación de métodos quimiométricos y de aprendizaje automático para extraer información de datos espectrales complejos y variables.

Dentro de los modelos de cuantificación, en la literatura se encuentran muy extendidas las metodologías basadas en regresión multivariante (PCR, PLS), así como modelos no lineales (SVR, redes neuronales, modelos basados en árboles). Los modelos lineales se siguen utilizando por su interpretabilidad y estabilidad, y los modelos no lineales han dado mejores resultados cuando la relación entre la señal espectral y la propiedad de interés tiene un comportamiento más complejo. En la práctica, la elección del modelo está guiada por el balance entre capacidad predictiva, robustez y complejidad computacional.

Debido a la alta dimensionalidad de los datos espectrales e hiperespectrales, la reducción y compresión de datos se convierte en una etapa esencial. Técnicas como PCA, PLS o métodos de selección de variables reducen la dimensionalidad efectiva del problema, eliminan la colinealidad y mejoran la estabilidad numérica de los modelos. En este sentido, la elección de bandas espectrales (basándose en coeficientes de regresión, métodos por intervalos o criterios modelados) no solo mejora la capacidad predictiva, sino que también permite la interpretabilidad de los resultados.

La revisión también muestra el creciente interés por enfoques de fusión de datos, especialmente en situaciones donde se tiene información de múltiples sensores o métodos de análisis. Entre los niveles

de fusión que se han llegado a informar, la fusión media es un buen punto intermedio entre la riqueza de información y la restricción de dimensionalidad, al ser capaz de combinar antes del modelado las características más pertinentes extraídas de cada fuente. Sin embargo, la literatura está de acuerdo en que la fusión es tan efectiva como lo haya sido su etapa de preprocesamiento y extracción de características.

En cuanto a los preprocesamientos espectrales, no hay un método que sea siempre el mejor. Técnicas como el centrado a la media, autoescalado, SNV, MSC y Savitzky–Golay se aplican en conjunto para corregir efectos físicos, igualar la varianza de las variables y mejorar la relación señal–ruido. La selección y encadenamiento de estos preprocesamientos es altamente dependiente del tipo de material, el sistema de medición y la meta analítica, y es común medir su efecto en términos de validación cruzada o conjuntos de prueba independientes.

Finalmente, la validación de los modelos se basa en métricas quimiométricas que miden diferentes aspectos de la capacidad predictiva. La literatura actual hace énfasis en métricas basadas en el error absoluto y la variabilidad estadística de los datos (RMSE,  $R^2$ , RPD, REP%) para una evaluación justa y comparable del rendimiento. Pero, a la vez, se sabe que métricas basadas en rangos, como el RER, pueden ser menos indicativas de la verdadera capacidad predictiva, por depender mucho del conjunto de datos y los valores extremos.

El estado del arte muestra que la cuantificación espectral necesita un tratamiento integral en el que la elección del modelo, la reducción de dimensionalidad, la fusión de datos, el preprocesamiento y la métrica de evaluación cobran relevancia si se quiere obtener un modelo con el desempeño deseado para este contexto. Esta perspectiva integral es el fundamento de la metodología propuesta en este trabajo.

### 3. Propuesta de tesis planteada.

---

#### 3.1. Planteamiento de la Hipótesis de Trabajo.

El modelado de la composición mineralógica y elemental de minerales y concentrados de cobre a partir de datos hiperspectrales implica enfrentar una serie de desafíos asociados a la alta dimensionalidad de los datos, la colinealidad entre variables espectrales y la posible presencia de relaciones no lineales entre la señal medida y las propiedades de interés. En este contexto, tanto la elección del modelo de regresión como las estrategias de compresión y representación de los datos espectrales resultan determinantes para alcanzar modelos predictivos robustos y con capacidad de generalización. La literatura revisada muestra que la evaluación del desempeño de modelos de cuantificación debe considerar métricas complementarias que permitan caracterizar su comportamiento en distintas etapas del proceso de validación, incluyendo calibración, validación cruzada y validación externa. Métricas como *RMSE*, *RPD*,  $R^2$ , *REP* y *RER* han sido ampliamente utilizadas para describir distintos aspectos del desempeño predictivo, tales como la magnitud del error, la capacidad explicativa y la relación entre el error y la variabilidad del conjunto de datos.

Dentro de este marco, se ha reportado que los modelos basados en regresión de vectores de soporte (SVR) constituyen una alternativa adecuada para el análisis de datos espectrales complejos, especialmente cuando se combinan con técnicas de compresión o reducción de dimensionalidad, como PLS o PCR, que permiten mitigar los efectos de la multicolinealidad y mejorar la estabilidad del proceso de modelado. La capacidad de SVR para ajustar relaciones no lineales mediante el uso de funciones kernel resulta particularmente relevante en sistemas minerales, donde la variabilidad espacial y composicional puede inducir comportamientos espectrales complejos.

No obstante, el desempeño de los modelos de cuantificación no depende únicamente del tipo de algoritmo de regresión utilizado, sino también de la naturaleza y calidad de la información espectral empleada como entrada. En el caso de datos hiperespectrales, distintas regiones del espectro pueden capturar fenómenos físicos complementarios asociados a la composición del material, lo que sugiere que la representación de los datos constituye un factor tan relevante como la elección del modelo en el proceso de cuantificación.

En particular, las regiones visible–infrarrojo cercano (VNIR) e infrarrojo de onda corta (SWIR) presentan sensibilidades diferenciadas frente a propiedades electrónicas y vibracionales de los minerales, respectivamente. Esta complementariedad abre la posibilidad de integrar ambas fuentes de información mediante estrategias de fusión de datos, con el objetivo de construir representaciones más completas del sistema analizado. Sin embargo, la forma en que esta integración es realizada, así como su interacción con el tipo de modelo de cuantificación utilizado, constituye un aspecto que requiere ser evaluado de manera sistemática.

**Se plantea como hipótesis que el desempeño predictivo en la cuantificación de la composición mineralógica y elemental de concentrados de cobre a partir de datos hiperespectrales depende tanto del tipo de modelo de regresión empleado como de la información espectral utilizada. En particular, se espera que los modelos no lineales, como PLS-SVR, presenten un mejor desempeño que los modelos lineales tradicionales, y que la integración de información espectral proveniente de las regiones VNIR y SWIR, mediante estrategias de fusión de datos, permita mejorar la capacidad predictiva en comparación con el uso de cada región por separado, debido a la complementariedad de la información física capturada en ambos rangos espectrales.**

### **3.2. Justificación de la hipótesis de trabajo.**

La cuantificación de la composición mineralógica y elemental de concentrados de cobre a partir de datos hiperespectrales presenta desafíos asociados a la alta dimensionalidad de los datos, la fuerte colinealidad entre bandas espectrales y la presencia de relaciones no lineales entre la señal espectral y las propiedades de interés. Estas características hacen necesario el uso de modelos capaces de capturar dicha complejidad sin incurrir en un aumento excesivo de la complejidad computacional ni en problemas de sobreajuste.

En este contexto, la regresión por mínimos cuadrados parciales (PLS) ha sido ampliamente utilizada debido a su robustez y a su capacidad para manejar datos colineales mediante la proyección en un espacio de variables latentes. Sin embargo, su naturaleza esencialmente lineal puede limitar su desempeño cuando la relación entre los espectros y la variable respuesta presenta comportamientos no lineales, como ocurre en sistemas minerales complejos y heterogéneos.

Por esta razón, el uso de modelos no lineales como Support Vector Regression (SVR), en combinación con técnicas de compresión de datos como PLS, se presenta como una alternativa adecuada para capturar relaciones más complejas entre la información espectral y la composición del material. La elección de SVR frente a otros modelos de aprendizaje automático más complejos se justifica por su capacidad para modelar relaciones no lineales mediante el uso de un número relativamente reducido de hiperparámetros, lo que facilita su control y disminuye el riesgo de sobreajuste, especialmente en conjuntos de datos de tamaño limitado.

En particular, el uso del kernel de base radial (RBF) permite aproximar una amplia variedad de funciones no lineales, lo que resulta especialmente adecuado para el modelado de respuestas espectrales complejas, donde la relación entre la señal y la composición mineralógica no puede asumirse estrictamente lineal. De este modo, el enfoque PLS-SVR permite mantener un equilibrio entre capacidad predictiva y complejidad del modelo.

Por otra parte, la información espectral disponible para la caracterización del material no es única, ya que distintas regiones del espectro hiperespectral, como VNIR y SWIR, capturan fenómenos físicos complementarios asociados a propiedades electrónicas y vibracionales de los minerales. En este sentido, la utilización de una única fuente espectral puede limitar la capacidad del modelo para representar adecuadamente la complejidad del sistema analizado.

En este contexto, las estrategias de fusión de datos permiten integrar información proveniente de múltiples fuentes espectrales con el objetivo de construir representaciones más completas del material. En particular, los enfoques de fusión a nivel medio (Mid-Level Data Fusion, MLDF) permiten combinar información relevante de distintas fuentes en un espacio de características reducido, favoreciendo la incorporación de información complementaria sin un crecimiento excesivo de la dimensionalidad.

Adicionalmente, el desempeño del modelo se evalúa utilizando un conjunto coherente de preprocesamientos espectrales, que incluyen filtrado de Savitzky–Golay para la mejora de la relación señal-ruido, corrección de dispersión mediante Multiplicative Scatter Correction (MSC), centrado y autoescalado de las variables, y selección de variables basada en el criterio Variable Importance in Projection (VIP). Estos preprocesamientos se aplican de manera consistente tanto en modelos lineales como no lineales, permitiendo realizar comparaciones en condiciones equivalentes.

En conjunto, la combinación de modelos de regresión lineales y no lineales, junto con el uso de estrategias de integración de información espectral, se plantea como una alternativa capaz de mejorar la precisión predictiva en la cuantificación de concentrados de cobre. Bajo este marco, resulta razonable esperar mejoras en métricas de evaluación como RMSE, RPD,  $R^2$ , REP% y RER en comparación con enfoques que utilizan una única fuente espectral o modelos exclusivamente lineales, lo que proporciona una base sólida para la hipótesis de trabajo planteada.

### **3.3. Objetivo General.**

Con el fin de abordar de manera integral la problemática y el objetivo general planteado en esta investigación, se definen los siguientes objetivos específicos, orientados a evaluar y analizar el desempeño de distintos enfoques de modelado en la cuantificación de la composición mineralógica y elemental de concentrados de cobre a partir de datos hiperespectrales. En particular, estos objetivos buscan estudiar tanto la influencia del tipo de modelo de regresión como el impacto de la integración de información espectral proveniente de distintas regiones, considerando su efecto en la capacidad predictiva de los modelos.

### **3.4. Objetivos Específicos.**

Con el fin de abordar de manera integral la problemática y el objetivo general planteado en esta investigación, se definió un conjunto de objetivos específicos que reflejan las distintas etapas necesarias para el desarrollo del trabajo. Dada la complejidad del proceso de cuantificación mineralógica y elemental a partir de datos hiperespectrales que involucra la preparación y caracterización de muestras, la adquisición controlada de datos, el análisis exploratorio, el tratamiento espectral, el modelado quimiométrico y la evaluación del desempeño predictivo, resulta necesario descomponer el objetivo general en objetivos específicos.

#### **3.4.1 Evaluación de modelos en dominios espectrales individuales**

Evaluar el desempeño de modelos de regresión lineales (PLS) y no lineales (PLS-SVR) en la estimación de la composición mineralógica y elemental de concentrados de cobre, utilizando información espectral proveniente de las regiones VNIR y SWIR de manera independiente, con el fin de establecer una línea base de comparación para cada dominio espectral.

#### **3.4.2 Implementación de estrategias de fusión de datos espectrales**

Implementar estrategias de fusión de datos que permitan integrar información espectral proveniente de las regiones VNIR y SWIR, generando representaciones conjuntas del material analizado que incorporen la complementariedad de ambas fuentes.

#### **3.4.3 Comparación del desempeño predictivo con y sin integración de información**

Comparar el desempeño predictivo de los modelos de cuantificación al utilizar información espectral individual (VNIR o SWIR) frente a información integrada mediante fusión de datos, evaluando su impacto mediante métricas quimiométricas y esquemas de validación consistentes.

#### **3.4.4 Análisis de la influencia del tipo de modelo de regresión**

Analizar la influencia del tipo de modelo de regresión (lineal y no lineal) en la capacidad de capturar relaciones entre la información espectral y la composición del material, considerando tanto datos individuales como integrados, con el objetivo de identificar diferencias en su comportamiento predictivo.

#### **3.4.5 Interpretación de resultados en función de la información espectral**

Interpretar los resultados obtenidos en función de la naturaleza de la información espectral y de los mecanismos físicos asociados a las regiones VNIR y SWIR, identificando las principales ventajas, limitaciones y alcances de cada enfoque de modelado en el contexto de la caracterización de concentrados de cobre.

## 4. Alcances y Limitaciones

---

Aquí se indican claramente las limitaciones asumidas en el desarrollo de su trabajo. Por ejemplo, que usa solamente simulación o un set-up a escala para verificar su hipótesis y cumplir su objetivo. También podría indicar que su solución es sólo válida en Chile, etc.. Respecto de este documento, los alcances son: los detalles no cubiertos por esta guía quedan libres para ser definidos por los memoristas. Por otro lado, dado que hay varios procesadores de texto disponibles, este documento es solamente referencial para los formatos indicados; sin embargo, se debe considerar en forma rigurosa respecto de los contenidos.

### 4.1. Sensores a utilizar

En el desarrollo de esta investigación se emplearon dos sistemas de adquisición hiperespectral que cubren rangos espectrales complementarios, con el propósito de capturar información relevante asociada a la composición mineralógica y elemental de concentrados de cobre. La selección de estos sensores responde a criterios de disponibilidad experimental, relevancia espectral y aplicabilidad en contextos de laboratorio, permitiendo evaluar tanto enfoques de modelado con un solo sensor como estrategias complementarias de fusión de datos.

Por una parte, se utilizó un sistema hiperespectral en el rango visible–infrarrojo cercano (VNIR) de desarrollo propio, diseñado e implementado específicamente para este estudio. Este sistema fue concebido como una plataforma flexible de laboratorio, orientada a la adquisición de datos espectrales de muestras pulverulentas bajo condiciones controladas. Si bien este prototipo permite explorar el potencial del rango VNIR para la cuantificación de concentrados de cobre, su desempeño óptico y radiométrico corresponde al de un sistema experimental, por lo que no pretende replicar directamente las prestaciones de cámaras comerciales de mayor costo.

Complementariamente, se empleó una cámara hiperespectral comercial en el rango de infrarrojo de onda corta (SWIR), modelo PIKA IR, de la empresa Resonon. Este sensor fue seleccionado debido a

su uso extendido en aplicaciones industriales y de investigación, así como por su capacidad para capturar información espectral asociada a enlaces químicos y vibraciones moleculares relevantes en minerales sulfurados. El uso de un sistema comercial en el rango SWIR permite, además, establecer un punto de referencia más cercano a escenarios de aplicación real.

La utilización conjunta de ambos sensores posibilita analizar de manera comparativa el aporte espectral de los rangos VNIR y SWIR, así como explorar esquemas de integración de información mediante fusión de datos a nivel medio (MLDF) como un complemento metodológico. No obstante, el alcance experimental del estudio se restringe a estos dos sistemas de adquisición y a las configuraciones definidas durante su implementación, sin considerar otros rangos espectrales ni técnicas espectroscópicas adicionales.

#### 4.1.1 Cámara VNIR

Para la adquisición de datos en el rango visible–infrarrojo cercano (VNIR), se utilizó una cámara hiperespectral de barrido lineal (*pushbroom*) denominada **HSI V4**, basada en un diseño modular tipo *do-it-yourself* descrito por Sigernes et al. [61]. El sistema fue configurado para mediciones de laboratorio, priorizando estabilidad geométrica y repetibilidad espectral.

La cámara integra un cabezal **UI-3360CP-NIR-GL R2**, con un tamaño de imagen de **1088 × 2048 píxeles**, donde una dimensión corresponde al eje espacial y la otra al eje espectral. El rango espectral nominal del sistema abarca aproximadamente **297.5–1005.5 nm**; no obstante, en este trabajo se empleó una **fuentes de iluminación continua con emisión efectiva entre 400 y 1000 nm**, asegurando condiciones de iluminación adecuadas dentro del rango de interés espectral.

El sistema óptico incorpora una **rejilla de transmisión de 600 líneas/mm**, acoplada a una rendija de entrada con **3 mm de altura y 25 µm de ancho**, parámetros que determinan la resolución espectral y la eficiencia luminosa del sistema. Aunque el detector dispone de **1088 píxeles en el eje espacial**, aproximadamente **600 píxeles fueron efectivos** para la adquisición de información de la muestra, correspondiendo a la región iluminada y enfocada del objeto de interés. La magnificación efectiva de la rendija en el eje espacial es de **1.28**.

De acuerdo con las especificaciones del diseño óptico, el valor de **FWHM máximo teórico** del sistema es del orden de **1.4 nm**. Sin embargo, bajo las condiciones reales de operación empleadas en este estudio, con una distancia de trabajo de aproximadamente **40 cm entre la cámara y la muestra**, la resolución espectral efectiva medida fue del orden de **2.5 nm (FWHM)**.

Este valor representa de forma más realista el desempeño espectral del sistema durante las mediciones experimentales y considera efectos geométricos, ópticos y de iluminación. La **Figura 4.1** muestra la cámara hiperespectral VNIR HSI V4 utilizada en este trabajo junto con sus principales parámetros ópticos y geométricos.

| Part                      | HSI V4                                 |
|---------------------------|----------------------------------------|
| Camera head               | UI-3360CP-NIR-GL R2                    |
| Image size                | (1088, 2048)                           |
| Spectral range            | 297.5 - 1005.5 nm                      |
| FWHM                      | 1.4 nm                                 |
| Grating                   | 600 grooves/mm<br>transmission grating |
| Slit height               | 3 mm                                   |
| Slit width                | 25 $\mu\text{m}$                       |
| Slit height magnification | 1.28                                   |



**HSI V4**

**Figura 4.1 muestra la cámara hiperespectral VNIR HSI V4 utilizada en este trabajo junto con sus principales parámetros ópticos y geométricos.**

Previo a la adquisición de datos hiperespectrales, la cámara VNIR HSI V4 fue sometida a un proceso de calibración espacial, espectral y radiométrica, con el objetivo de asegurar la calidad geométrica, espectral y radiométrica de las mediciones.

La **calibración espacial** se realizó mediante el ajuste manual del enfoque y de la geometría de adquisición, buscando maximizar el contraste y la nitidez de la imagen sobre la región de interés. Este procedimiento permitió definir una distancia de trabajo estable y reproducible, garantizando una adecuada proyección del objeto sobre el eje espacial del detector y una correcta utilización de los píxeles efectivos.

La **calibración espectral** se llevó a cabo utilizando una **lámpara fluorescente compacta (CFL)** como fuente de referencia, aprovechando sus **líneas de emisión discretas y bien definidas** en el rango VNIR, comúnmente empleadas en espectroscopía óptica para la calibración de instrumentos [62–65]. La lámpara se utilizó para iluminar difusamente un **estándar blanco**, ubicado a una distancia

aproximada de **40 cm** de la cámara. A partir del espectro registrado, se identificaron las principales líneas de emisión características de la CFL y se determinó, para cada una de ellas, el **píxel central asociado al máximo de intensidad**. A estos píxeles se les asignaron las **longitudes de onda nominales** reportadas en la literatura [63–65].

La relación entre la posición del píxel ( $p$ ) y la longitud de onda ( $\lambda$ ) no es estrictamente lineal debido a la geometría del sistema dispersivo basado en una rejilla de difracción y a efectos ópticos residuales. Por esta razón, la correspondencia píxel–longitud de onda se modeló mediante una **regresión polinómica de segundo orden**, un enfoque ampliamente documentado en manuales y textos de espectroscopía e instrumentación óptica [63–65]. El modelo adoptado se expresa como:

$$\lambda(p) = a p^2 + b p + c \quad (49)$$

Donde  **$a$** ,  **$b$**  y  **$c$**  son los coeficientes obtenidos mediante un ajuste por mínimos cuadrados a partir de los pares ( $p_i, \lambda_i$ ) correspondientes a los picos de emisión identificados. Este ajuste permite capturar la **leve no linealidad** presente en la dispersión espectral del sistema, la cual se manifiesta como una curvatura suave en la relación píxel–longitud de onda.

Adicionalmente, el análisis de la **forma espectral de los picos de emisión** permitió estimar la **resolución espectral efectiva del sistema**. De acuerdo con la literatura clásica de espectroscopía, la resolución espectral de un instrumento puede determinarse a partir del **ancho completo a media altura (Full Width at Half Maximum, FWHM)** de una línea de emisión estrecha, el cual representa el *bandpass* instrumental [62–64]. En este trabajo, el FWHM se estimó a partir de picos aislados del espectro de la CFL, inicialmente en unidades de píxel y posteriormente convertido a nanómetros utilizando la relación de calibración espectral obtenida. Este procedimiento permitió estimar una **resolución espectral efectiva del orden de 2.5 nm** bajo las condiciones reales de adquisición empleadas.

La **Figura 4.2** ilustra esquemáticamente el procedimiento de calibración espectral, incluyendo la identificación de las líneas de emisión de la lámpara CFL y el ajuste de la regresión utilizada para la asignación longitud de onda–píxel.

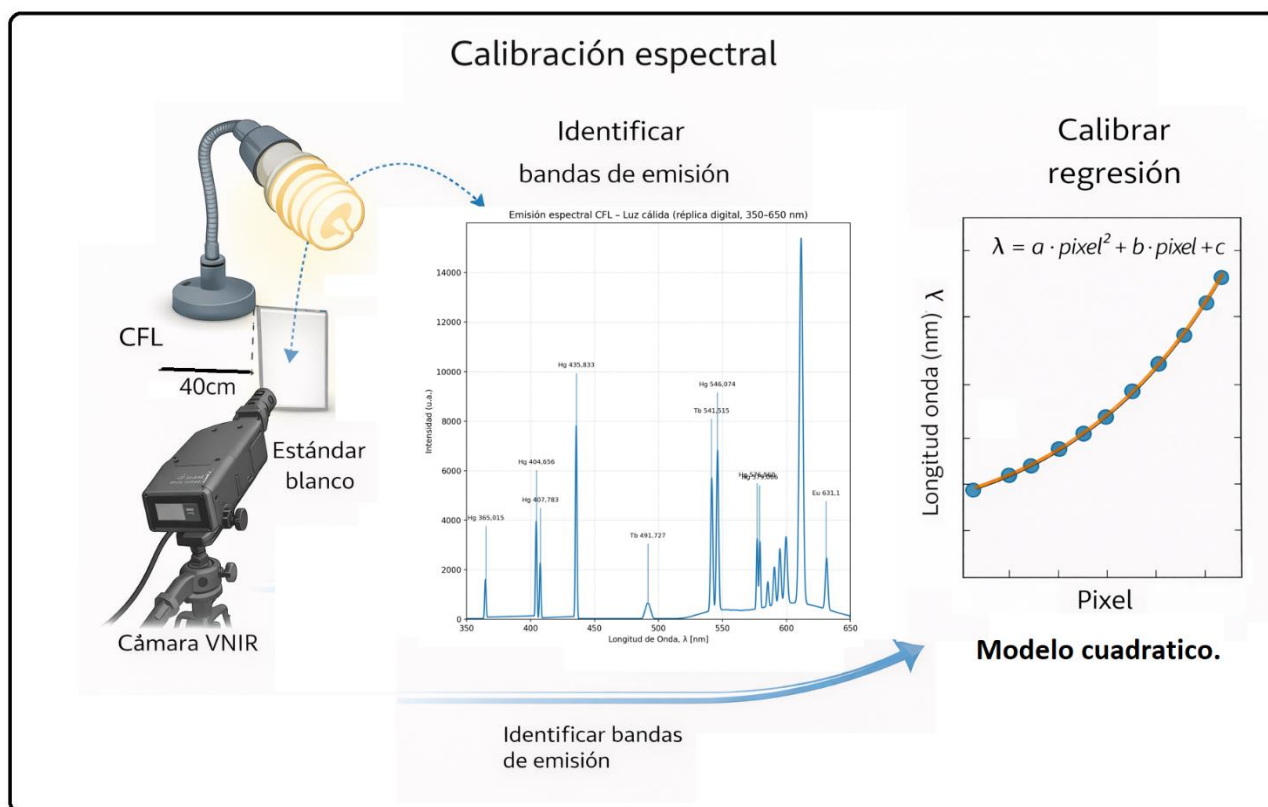


**Figura 4.2** ilustra esquemáticamente el procedimiento de calibración espectral, la identificación de las líneas de emisión de la lámpara CFL y el ajuste de la regresión utilizada para la asignación longitud de onda-píxel.

Finalmente, la **calibración radiométrica** se realizó mediante la corrección estándar de **blanco y negro**. Para ello, se adquirieron imágenes de referencia utilizando un **patrón blanco de dióxido de titanio (TiO<sub>2</sub>)**, empleado como aproximación a una reflectancia difusa cercana al 100%, y una **referencia oscura (dark reference)** obtenida mediante la tapa del lente de la cámara en ausencia de iluminación. Estas referencias se utilizaron para corregir las variaciones de respuesta del detector y de la iluminación, permitiendo expresar los datos hiperespectrales en términos de **reflectancia relativa**.

#### 4.1.2 Desarrollo del sistema de adquisición de imágenes VNIR.

En este trabajo se decidió utilizar la cámara hiperespectral VNIR bajo un esquema de adquisición **tipo line-scan (pushbroom)**, considerando tanto las características del sensor como los requerimientos de precisión del estudio. En este tipo de configuración, cada captura corresponde a una única línea espacial de la muestra, para la cual se registra la información espectral completa.

En consecuencia, la reconstrucción de una imagen bidimensional con información espectral requiere introducir un movimiento relativo controlado entre la muestra y la cámara.

La elección del modo pushbroom se fundamentó principalmente en la necesidad de asegurar una adquisición estable y repetible. En particular, la cámara no dispone de una tasa de fotogramas de video suficientemente alta como para realizar adquisiciones confiables en modo área. Bajo estas condiciones, intentar sincronizar la captura de imágenes completas con un desplazamiento continuo podría introducir distorsiones geométricas o irregularidades en el muestreo espacial. En cambio, el esquema pushbroom permitió una sincronización más precisa entre la captura de cada línea espectral y el movimiento mecánico del sistema, mejorando el control del proceso de adquisición.

Para implementar esta configuración, se utilizó una plataforma lineal de escaneo de alta precisión de **Zaber Technologies**, modelo **X-ASR100B120B-SE03D12** (B), montada sobre una mesa óptica (A). Esta configuración permitió asegurar una adecuada estabilidad mecánica y minimizar vibraciones durante la adquisición. La cámara hiperespectral VNIR (D) se mantuvo fija durante todo el proceso, mientras que la muestra fue desplazada de forma uniforme bajo su campo de visión, permitiendo la adquisición secuencial de líneas espectrales necesarias para la reconstrucción del cubo hiperespectral.

La operación conjunta de la cámara y la plataforma se realizó mediante una interfaz gráfica desarrollada en **LabVIEW**, la cual permitió controlar de manera integrada los principales parámetros de adquisición, tales como la velocidad de desplazamiento, el tiempo de exposición y el número de líneas adquiridas. Este enfoque facilitó la repetibilidad de las mediciones y el control del proceso experimental.

El sistema incorporó además un blanco estándar de referencia basado en dióxido de titanio ( $\text{TiO}_2$ ) (C), utilizado para los procedimientos de calibración y normalización de la respuesta espectral del sistema.

En cuanto al sistema de iluminación, se empleó una combinación de fuentes de luz con el objetivo de cubrir de manera adecuada todo el rango espectral VNIR. En primer lugar, se utilizaron focos halógenos de tungsteno con una temperatura de color de 2700 K (E), los cuales proporcionaron una iluminación estable y continua en gran parte del espectro visible. No obstante, debido a la naturaleza cálida de este tipo de fuentes, su emisión en la región violeta-azul (400–450 nm) es limitada. Para compensar esta deficiencia, se incorporó una fuente de iluminación LED en el rango de 400–450 nm

(F), basada en LEDs de alta potencia con difusores de aproximadamente 45°, lo que permitió reforzar dicha región espectral y obtener una iluminación más homogénea sobre la superficie de la muestra.

La combinación de ambas fuentes de iluminación resultó ser una solución eficiente y de bajo costo, permitiendo obtener una iluminación uniforme y estable sin recurrir a sistemas de iluminación experimentales convencionales, los cuales suelen implicar costos significativamente mayores.

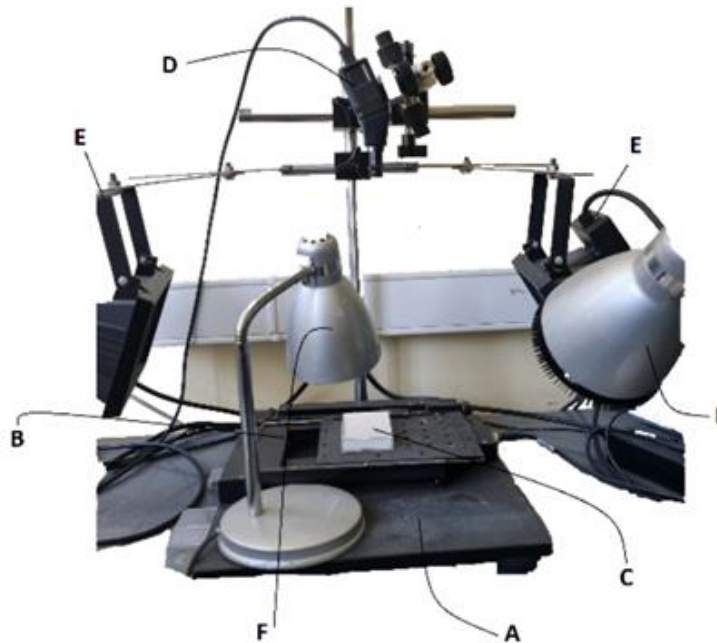


**Figura 4.3 Sistema experimental de adquisición hiperespectral VNIR: (A) mesa óptica; (B) plataforma lineal de escaneo Zaber Technologies X-ASR100B120B-SE03D12; (C) blanco estándar de dióxido de titanio ( $\text{TiO}_2$ ); (D) cámara hiperespectral VNIR; (E) focos halógenos de tungsteno (2700 K); (F) fuente de iluminación LED en el rango 400–450 nm.**

Finalmente, a partir de la adquisición secuencial de líneas espectrales, se reconstruyeron cubos hiperespectrales con dos dimensiones espaciales y una dimensión espectral ( $x$ ,  $y$ ,  $\lambda$ ). Los cubos generados fueron almacenados en formato **ENVI**, lo que facilitó su posterior procesamiento, análisis quimiométrico y modelado predictivo mediante software especializado.

#### 4.1.3 Camara SWIR

Además del sistema VNIR, en este estudio se utilizó una cámara hiperespectral en el rango SWIR, correspondiente al modelo Pika IR+, perteneciente al Centro de Biotecnología de la Universidad de Concepción. Esta cámara opera en el rango del infrarrojo de onda corta, cubriendo aproximadamente longitudes de onda entre 900 y 1700 nm, y fue utilizada con el objetivo de complementar la información espectral obtenida en la región VNIR.

El sistema presenta una resolución espectral de **5.6 nm**, con un número aproximado de **336 bandas espectrales**, lo que permite capturar con alto nivel de detalle las características espectrales del material. A diferencia del sistema VNIR, la cámara Pika IR+ corresponde a un sistema integrado que incluye su propia fuente de iluminación y una plataforma de desplazamiento lineal, diseñadas específicamente para la adquisición hiperespectral tipo line-scan (pushbroom). En este esquema, la cámara se mantiene fija mientras la muestra se desplaza de manera controlada bajo su campo de visión, permitiendo la adquisición secuencial de líneas espectrales y la posterior reconstrucción del cubo hiperespectral.

La iluminación del sistema SWIR se realizó mediante una fuente de luz tipo Fiber Line (B), integrada al sistema de la cámara, la cual proporciona una iluminación uniforme y estable en todo el rango espectral de operación. Esta configuración permitió realizar adquisiciones consistentes sin necesidad de incorporar fuentes de iluminación externas adicionales.

El desplazamiento de la muestra se efectuó mediante una plataforma de traslación lineal integrada al sistema (C), lo que permitió controlar de forma precisa la **velocidad de escaneo**, así como el número de líneas adquiridas durante cada medición.

La velocidad de adquisición y el **framerate del sistema** fueron ajustados de manera que se asegurara una adecuada resolución espacial y una correcta reconstrucción del cubo hiperespectral, evitando distorsiones geométricas en la imagen. Tanto la cámara hiperespectral (A), como la iluminación y la plataforma de desplazamiento fueron operadas de manera conjunta mediante el software Spectronon Pro, el cual permitió configurar los parámetros de adquisición, visualizar los datos en tiempo real y almacenar los cubos hiperespectrales generados.

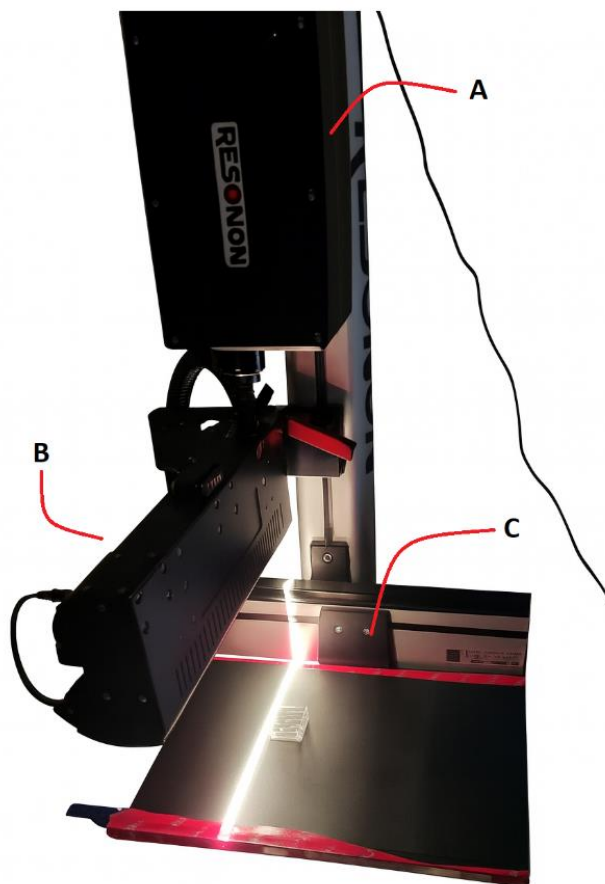


**Figura 4.4** Sistema experimental de adquisición hiperspectral SWIR: (A) cámara hiperspectral Pika IR+ (900–1700 nm); (B) fuente de iluminación tipo Fiber Line; (C) plataforma de traslación lineal. El sistema completo es operado mediante el software Spectronon Pro.

Este enfoque integrado facilitó la adquisición de cubos hiperspectrales SWIR de forma estable y reproducible, los cuales fueron posteriormente utilizados junto con los datos VNIR en las etapas de análisis y modelado desarrolladas en este trabajo.

#### **4.2. Cantidad y tipos de muestras a medir.**

En este estudio se analizaron un total de **95 muestras de concentrado de cobre**, compuestas predominantemente por **calcopirita y pirita**. Las muestras provinieron de distintas fuentes industriales y experimentales, con el fin de contar con un conjunto de datos representativo y variado. Parte de las muestras correspondió a concentrados comerciales obtenidos de **Minera Centinela**, los cuales se encontraban distribuidos en tres lotes comerciales suministrados por **Pan Pacific Copper, Aurubis y Trafigura**.

Adicionalmente, el conjunto de datos incluyó muestras propias, preparadas específicamente para este estudio.

Todas las muestras se encontraban en **formato de polvo**, presentando distintos **tamaños de partícula** y características texturales, representativas de concentrados de cobre de origen industrial. Algunas de las muestras propias fueron obtenidas a partir de material a granel que fue posteriormente pulverizado, con el objetivo de adecuar el material a los requerimientos de la adquisición hiperspectral y aumentar la diversidad del conjunto de muestras.

Con el fin de ampliar aún más la variabilidad mineralógica y composicional, algunas muestras fueron **mezcladas en proporciones controladas**, generando muestras compuestas adicionales. Esta estrategia permitió construir un conjunto de datos más diverso, abarcando un rango más amplio de composiciones y distribuciones de tamaño de partícula, lo cual resulta especialmente relevante para el desarrollo y validación de modelos predictivos robustos.

Cada muestra fue recibida o preparada en bolsas selladas, debidamente identificadas mediante un código alfanumérico, y almacenada en un desecador hasta su análisis, con el fin de minimizar la absorción de humedad. Para la adquisición hiperspectral, una pequeña fracción de cada muestra en polvo fue distribuida de manera uniforme sobre una placa Petri cuadrada y transparente (4,5 cm × 4,5 cm), formando una superficie homogénea adecuada para la adquisición de imágenes y el posterior recorte de regiones de interés.

#### **4.3. Técnica de referencia para las composiciones mineralógicas y elementales.**

a cuantificación mineralógica de referencia se realizó utilizando un sistema **FEI QEMSCAN Quanta 650 FEG SEM** (Thermo Fisher Scientific, Hillsboro, OR, USA), siguiendo flujos de trabajo consolidados de **mineralogía automatizada** para la caracterización de concentrados de cobre. La elección de **QEMSCAN** (del inglés *Quantitative Evaluation of Minerals by Scanning Electron Microscopy*) como técnica de referencia responde a que esta plataforma es ampliamente reconocida como el **estándar oro** en el análisis mineralógico cuantitativo dentro de la industria minera y metalúrgica [66].

El sistema QEMSCAN permite una **caracterización integral de las muestras**, ya que entrega de forma simultánea información **mineralógica, elemental y textural**. Esta capacidad se basa en la integración de la microscopía electrónica de barrido (SEM) con el análisis por espectroscopía de rayos X de energía dispersiva (EDS), lo que permite identificar automáticamente las fases minerales en función de su composición química.

Este enfoque resulta especialmente adecuado para el análisis de materiales complejos, como los concentrados de cobre, y permite obtener resultados consistentes y reproducibles [67].

Debido a estas características, los resultados obtenidos mediante QEMSCAN son comúnmente utilizados como valores de referencia confiables (*ground truth*) en estudios de caracterización mineralógica. En particular, esta técnica se emplea de manera frecuente para la validación y comparación de métodos analíticos alternativos o emergentes, incluyendo técnicas espectroscópicas y métodos basados en imágenes, dado el alto nivel de detalle y la completitud de la información que proporciona [68].

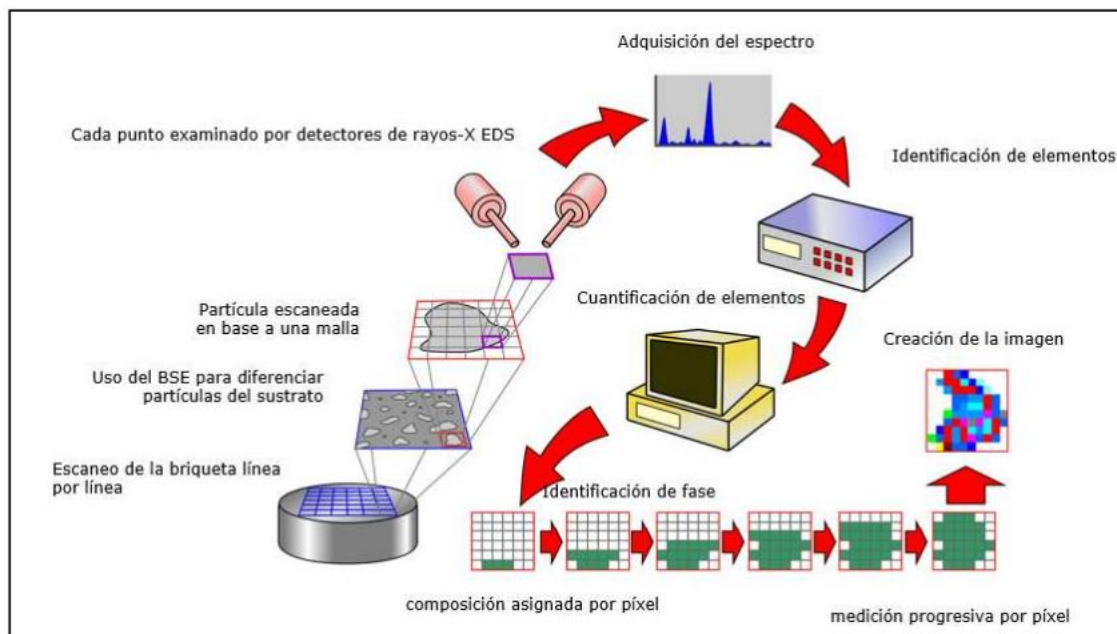


**Figura 4.5 Flujo de trabajo del sistema QEMSCAN para la cuantificación mineralógica y elemental.**

**Basada en [71].**

La **Figura 4.5** muestra de manera esquemática el principio de funcionamiento del sistema QEMSCAN y el flujo de procesamiento utilizado durante el análisis mineralógico automatizado. El procedimiento comienza con el escaneo sistemático de la muestra, la cual es preparada generalmente en forma de briqueta pulida. Esta briqueta es analizada línea por línea mediante SEM, lo que permite un muestreo controlado y repetible de toda la superficie analizada.

Durante el escaneo, la muestra es examinada sobre la base de una **mallla regular de píxeles**, donde cada punto corresponde a una posición específica de análisis. En una primera etapa, la señal de **electrones retrodispersados (BSE)** se utiliza para diferenciar las partículas minerales del sustrato y delimitar bordes de grano y estructuras texturales relevantes. Posteriormente, en cada píxel de la mallla se adquiere un **espectro de rayos X mediante EDS**, el cual contiene la información elemental característica del material analizado.

A partir de estos espectros, se realiza la **identificación y cuantificación de los elementos químicos** presentes en cada punto. Esta información elemental es comparada con una base de datos mineralógica previamente definida, lo que permite asignar una **fase mineral específica a cada píxel** en función de su composición química. Este proceso se repite de manera progresiva para todos los píxeles de la mallla, generando una medición completamente **pixel a pixel** sobre la superficie de la muestra.

Como resultado de este procedimiento, se obtiene una imagen mineralógica segmentada, en la cual cada píxel queda asociado a una fase mineral determinada. A partir de esta información es posible calcular directamente la **composición modal de los minerales**, así como analizar su distribución espacial, tamaño de partícula, asociaciones mineralógicas y grado de liberación. Estas características son especialmente relevantes en el análisis de concentrados de cobre y en la interpretación metalúrgica de los resultados [69].

El uso extendido de QEMSCAN en la industria minera, tanto en exploración como en evaluación metalúrgica y control de procesos, ha demostrado que los resultados entregados por esta técnica presentan una **alta validez y exactitud**, consolidándola como una herramienta de referencia en estudios de mineralogía cuantitativa. Por esta razón, los datos mineralógicos y elementales obtenidos mediante QEMSCAN fueron considerados como la **técnica de referencia** para la validación de los modelos desarrollados en este trabajo [70].

#### **4.4. Uso de software.**

Para el desarrollo del sistema de adquisición, procesamiento y análisis de los datos hiperespectrales se utilizaron distintas herramientas de software, las cuales cumplieron funciones complementarias dentro del flujo de trabajo experimental. En particular, se emplearon los programas **Spectronon Pro**, **MATLAB**, **Python** y **LabVIEW**, seleccionados en función de sus capacidades específicas y de la necesidad de integrar adquisición, control y análisis de datos.

El software **Spectronon Pro** se utilizó principalmente para la adquisición y el análisis inicial de los datos obtenidos con la cámara hiperspectral SWIR. Este programa permitió capturar directamente los cubos hiperspectrales generados por la cámara, así como visualizar y explorar de forma preliminar los espectros adquiridos. Además, Spectronon Pro facilitó la exportación de los cubos y de la información espectral a formatos compatibles con MATLAB, lo que permitió incorporar estos datos de manera directa en las etapas posteriores de procesamiento y análisis.

La mayor parte del procesamiento de los datos y del análisis quimiométrico se llevó a cabo en **MATLAB**, el cual se utilizó como el entorno principal de trabajo. En este software se realizaron tareas como el manejo de los cubos hiperspectrales, el preprocesamiento espectral, la extracción de información relevante y el desarrollo de los modelos predictivos. MATLAB resultó especialmente adecuado para estas tareas debido a su facilidad para trabajar con grandes matrices de datos, sus herramientas de análisis multivariado y la flexibilidad para implementar algoritmos personalizados.

Por su parte, **Python** se utilizó en conjunto con LabVIEW principalmente para integrar librerías asociadas al control de la plataforma móvil de desplazamiento y a algunos controladores de la cámara VNIR. En este contexto, Python permitió acceder a bibliotecas y drivers que no se encontraban disponibles de forma nativa en LabVIEW, lo que facilitó la comunicación con el hardware y la automatización de ciertas tareas de control y sincronización.

**LabVIEW** se empleó para el desarrollo de la **interfaz gráfica de usuario**, la cual permitió centralizar el control del sistema experimental. A través de esta interfaz se integraron rutinas programadas tanto en Python como en MATLAB, facilitando la operación conjunta de la cámara, la plataforma de desplazamiento y la gestión de los datos adquiridos. De este modo, LabVIEW actuó como un entorno de integración que permitió coordinar de manera práctica los distintos lenguajes de programación utilizados en el desarrollo del sistema.

## 5. Metodología Experimental y Resultados.

---

### 5.1. Preparación de las muestras:

Se utilizaron **95 muestras de concentrado de cobre**, provenientes de distintas compañías mineras. Todas las muestras se encontraban en **formato de polvo**, presentando variabilidad tanto en el tamaño de partícula como en su composición mineralógica y elemental, lo que permitió conformar un conjunto de datos representativo de concentrados industriales reales.

Para cada muestra se obtuvo un **valor de referencia mineralógico y elemental** mediante análisis QEMSCAN, el cual fue empleado posteriormente como referencia para la calibración y validación de los modelos de cuantificación desarrollados en este trabajo. Dicho análisis se realizó a partir de una **porción representativa del polvo de cada concentrado**, la cual fue sometida a un proceso de **fabricación de briquetas** previo al análisis.

Con el objetivo de evaluar la **repetibilidad y la variabilidad del procedimiento**, algunas muestras fueron preparadas y analizadas en más de una ocasión, comparándose los resultados obtenidos entre repeticiones. No se observaron diferencias significativas, lo que confirma la **consistencia y reproducibilidad** tanto del proceso de preparación como del análisis QEMSCAN.

Cabe destacar que la **etapa más demandante en términos de tiempo** correspondió a la preparación de las briquetas y su posterior análisis, debido a que el equipo QEMSCAN disponible en el laboratorio GEA de la Universidad de Concepción opera por **lotes de hasta 14 muestras**, lo que requirió una planificación secuencial del trabajo experimental y generó retrasos en el desarrollo de la investigación.

### 5.1.1 Fabricación de briquetas y obtención de los datos de referencia

Para cada muestra de concentrado se fabricó una briketa siguiendo el procedimiento que se describe a continuación:

1. **Pesaje y mezcla inicial:** se pesaron **0,64 g de muestra de concentrado** y **1,96 g de carbono**, los cuales fueron mezclados en presencia de alcohol con el fin de facilitar la dispersión inicial del material.
2. **Homogeneización:** la suspensión obtenida fue sometida a un proceso de **ultrasonido**, con el objetivo de garantizar una mezcla homogénea. Posteriormente, el alcohol fue dejado evaporar de forma natural durante un período aproximado de **tres días**.
3. **Incorporación en resina:** una vez seco el material, el polvo fue colocado en un molde junto con **resina epóxica y su respectivo endurecedor**, mezclándose durante aproximadamente **30 minutos** hasta obtener una pasta uniforme.
4. **Curado:** la mezcla fue dejada en reposo sobre una superficie plana durante un mínimo de **24 horas**, permitiendo la solidificación completa de la briketa.
5. **Pulido:** tras el desmolde, la briketa sólida fue sometida a un proceso de **pulido progresivo**, finalizando con pulido con **alúmina**, con el objetivo de obtener una superficie lisa, plana y adecuada para la correcta identificación de las fases minerales.
6. **Carbonización:** finalmente, la superficie pulida fue recubierta con una **capa conductora de carbono**, requisito necesario para su posterior análisis mediante microscopía electrónica y QEMSCAN.
- 7.

Como resultado de este procedimiento, se obtuvo una **briketa de geometría circular**, en la cual es posible observar claramente la distribución de las fases minerales presentes en cada concentrado.

El proceso completo de preparación de las muestras y fabricación de las briquetas para el análisis QEMSCAN se resume esquemáticamente en la **Figura 5.1**. A modo de complemento, la **Figura 5.2** muestra una **briketa real obtenida tras el proceso de preparación**, donde se observa la superficie pulida utilizada para el análisis QEMSCAN, evidenciando la adecuada compactación del material y la presencia de las fases minerales embebidas en la matriz.



**Figura 5.1** Proceso de preparación de muestras y fabricación de briquetas para análisis QEMSCAN. (Elaboración propia)

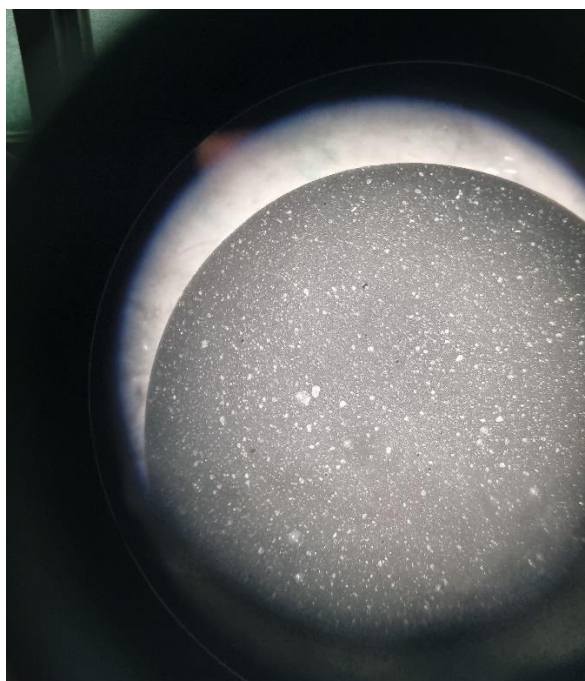


**Figura 5.2** Briqueta de concentrado de cobre preparada para análisis QEMSCAN. (muestra tomada con microscopio)

### 5.1.2 Distribución de las muestras preparadas.

Con el fin de garantizar la capacidad de generalización de los modelos de cuantificación desarrollados, se conformó un conjunto de muestras lo más diverso y representativo posible de concentrados de cobre industriales. Para ello, se analizó la distribución mineralógica y elemental de todas las muestras preparadas, a partir de los valores de referencia obtenidos mediante QEMSCAN.

Las distribuciones de las concentraciones de los principales constituyentes se representaron mediante gráficos tipo box plot, los cuales permiten visualizar de manera conjunta la dispersión de los datos, los valores centrales y la posible presencia de valores atípicos. Este análisis exploratorio permitió evaluar la amplitud de los rangos composicionales y detectar muestras con concentraciones significativamente alejadas del comportamiento general.

Si bien inicialmente se contaba con un número mayor de muestras, se observó que algunas presentaban concentraciones extremadamente alejadas del rango habitual, lo que podía introducir outliers no representativos y afectar negativamente tanto la calibración de los modelos como su capacidad de generalización. En consecuencia, se procedió a definir un número óptimo de muestras, buscando un compromiso entre representatividad composicional y estabilidad estadística.

En este proceso se equilibraron dos aspectos fundamentales: por una parte, disponer de suficiente variabilidad para capturar distintos escenarios mineralógicos presentes en concentrados reales, y por otra, evitar cambios demasiado abruptos en las concentraciones, que dificultaran el ajuste de los modelos predictivos. De este modo, el conjunto final de 95 muestras constituye una base experimental robusta y adecuada tanto para el entrenamiento como para la validación, asegurando un balance apropiado entre variabilidad, homogeneidad relativa y capacidad de generalización.

La Figura 5.3 presenta la distribución de las tres variables de interés empleadas en el entrenamiento de los modelos de cuantificación, correspondientes al contenido de cobre (%Cu), calcopirita (%Calcopirita) y pirita (%Pirita).

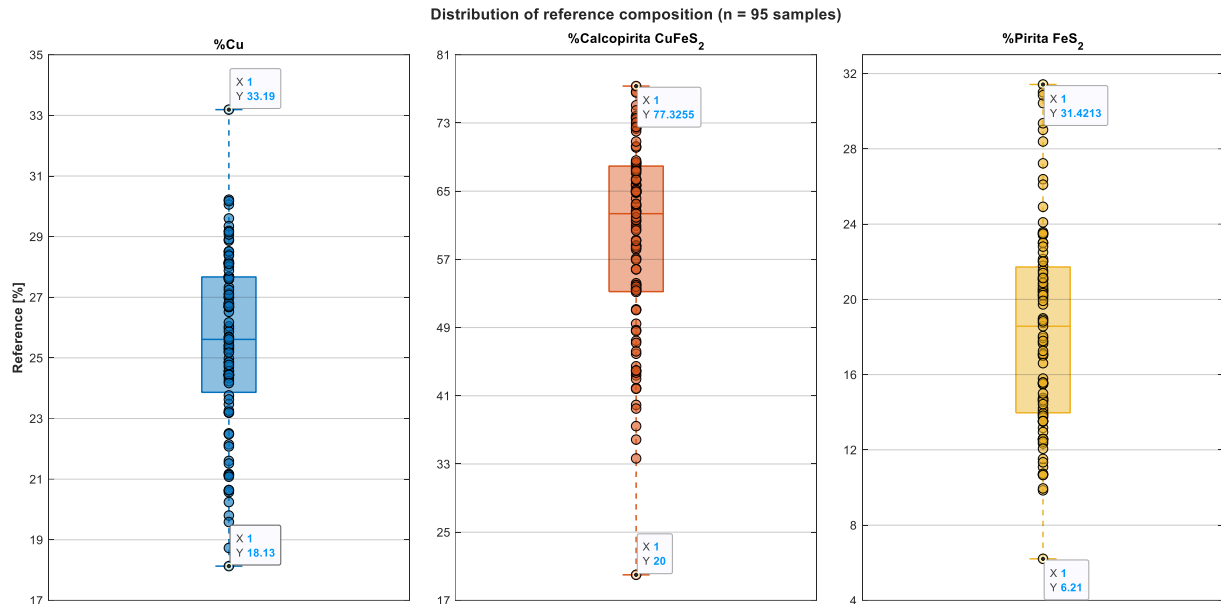


**Figura 5.3** Grafica de la distribución porcentual para %Cu (azul), %Calcopirita (naranja) y %pirita (amarillo).

Los rangos composicionales observados en el conjunto de muestras corresponden a **concentrados de cobre con leyes medias a altas**, coherentes con productos finales de flotación industrial. En particular, el contenido de cobre (%Cu) varía entre **18,13% y 33,19%**, valores que reflejan concentrados con una fracción significativa de minerales portadores de cobre. Los contenidos más elevados dentro de este rango están asociados a una mayor proporción relativa de fases cupríferas y a una menor dilución por minerales no portadores de cobre, lo que es consistente con concentrados de buena calidad metalúrgica.

Desde el punto de vista mineralógico, la **calcopirita** se presenta como la fase dominante, con concentraciones que alcanzan hasta **77,32%**, lo que explica directamente los niveles elevados de cobre observados en parte de las muestras. La **pirita**, en cambio, muestra un rango entre **6,21% y 31,42%**, actuando principalmente como mineral acompañante. Si bien su presencia es habitual en concentrados sulfurados, variaciones en su contenido reflejan diferencias en la selectividad del proceso de flotación y en las asociaciones mineralógicas del mineral de alimentación. En conjunto, estos rangos representan escenarios realistas de concentrados industriales de alta ley, proporcionando una base adecuada para el desarrollo y evaluación de modelos de cuantificación robustos.

## 5.2. Adquisición de los espectros:

La adquisición de los espectros hiperespectrales se realizó de manera controlada con el objetivo de obtener mediciones representativas y reproducibles de cada muestra de concentrado. Para cada una de las 95 muestras, se adquirieron tres espectros independientes, todos obtenidos a partir del mismo lote de material, pero mediante reacomodamientos sucesivos del polvo, con el fin de capturar la variabilidad asociada a la disposición superficial de las partículas y minimizar efectos locales no representativos.

Cada muestra fue dispuesta en una placa Petri de geometría cuadrada, cubriendo completamente el área de medición. A partir de cada preparación, se seleccionó una región cuadrada de interés, desde la cual se extrajo el espectro hiperespectral representativo. Este proceso fue repetido tres veces por muestra, obteniéndose un total de 285 espectros VNIR, los cuales constituyen la base de datos espectral utilizada en este trabajo. El procedimiento seguido para la selección de las regiones de interés durante la adquisición se ilustra en la Figura 5.4.

**De manera análoga, la adquisición de los espectros en la región SWIR se realizó siguiendo el mismo protocolo experimental, manteniendo las condiciones de preparación de muestra, selección de regiones de interés y número de repeticiones por muestra. Esto permitió asegurar la comparabilidad entre ambas fuentes espectrales y la coherencia en el análisis posterior, particularmente en el contexto de las estrategias de fusión de datos implementadas.**

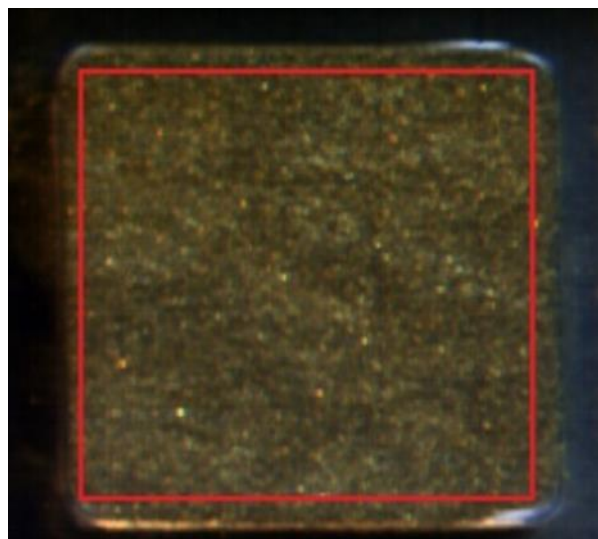


**Figura 5.4 Reconstrucción recortada (área roja) para una muestra concentrada.**

### 5.2.1 Espectros VNIR.

Los espectros VNIR obtenidos a partir de las muestras de concentrado de cobre se presentan en la **Figura 5.5**, donde se muestran los espectros correspondientes al conjunto completo de mediciones realizadas. Debido al carácter ruidoso de los espectros crudos, asociado principalmente a la naturaleza granular del material en polvo y a variaciones locales de iluminación y dispersión superficial, los espectros mostrados fueron sometidos a un **suavizado ligero**, aplicado **exclusivamente con fines de visualización**. Este suavizado permite atenuar el ruido de alta frecuencia sin modificar de manera significativa la forma global de los espectros ni desplazar las características espectrales relevantes. Cabe destacar que este procedimiento **no corresponde al preprocesamiento empleado posteriormente en el desarrollo de los modelos de cuantificación**, el cual se describe en detalle en la Sección 5.3.

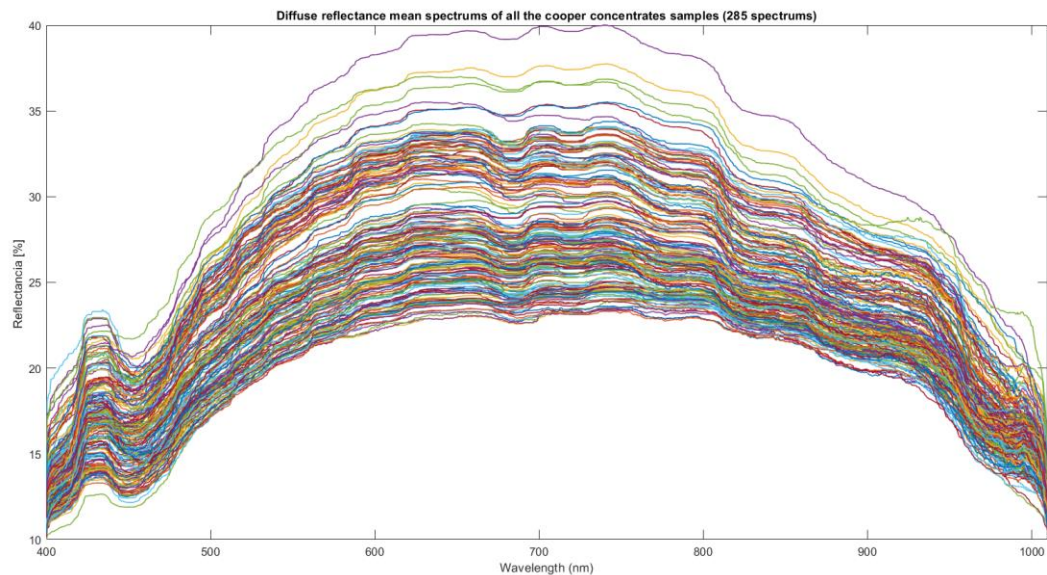


**Figura 5.5 Espectros VNIR de las muestras de concentrado de cobre.**

Los espectros VNIR obtenidos a partir de las muestras de concentrado de cobre se presentan en la **Figura 5.5**, donde se muestran los espectros correspondientes al conjunto completo de mediciones realizadas. Debido al carácter ruidoso de los espectros crudos, asociado principalmente a la naturaleza granular del material en polvo y a variaciones locales de iluminación y dispersión superficial, los espectros mostrados fueron sometidos a un **suavizado ligero**, aplicado **exclusivamente con fines de visualización**.

Este suavizado permite atenuar el ruido de alta frecuencia sin modificar de manera significativa la forma global de los espectros ni desplazar las características espectrales relevantes. Cabe destacar que este procedimiento **no corresponde al preprocesamiento empleado posteriormente en el desarrollo de los modelos de cuantificación**, el cual se describe en detalle en la Sección 5.3.

La figura permite observar una respuesta espectral consistente entre las distintas muestras, con variaciones principalmente en el nivel de reflectancia y en la pendiente espectral, atribuibles a diferencias en la composición mineralógica y en la granulometría del material. No obstante, la forma general de los espectros medidos está influenciada no solo por las propiedades ópticas de las muestras, sino también por la **respuesta espectral propia del sistema sensor**.

En este contexto, la Figura 5.6 muestra la respuesta espectral del sensor VNIR utilizada en la adquisición de datos, obtenida a partir del datasheet del fabricante. En particular, la curva morada corresponde a la eficiencia cuántica del sensor UI-3360CP-NIR-GL (Rev. 2) en función de la longitud de onda. Se observa que la respuesta no sigue una tendencia continua y monótona, sino que presenta un comportamiento ondulado, con sucesivos aumentos y disminuciones de sensibilidad a lo largo del rango VNIR, además de una reducción hacia los extremos del rango de medición. Este comportamiento es propio del sistema sensor y su óptica asociada, y contribuye a explicar la forma general observada en los espectros VNIR adquiridos [72].

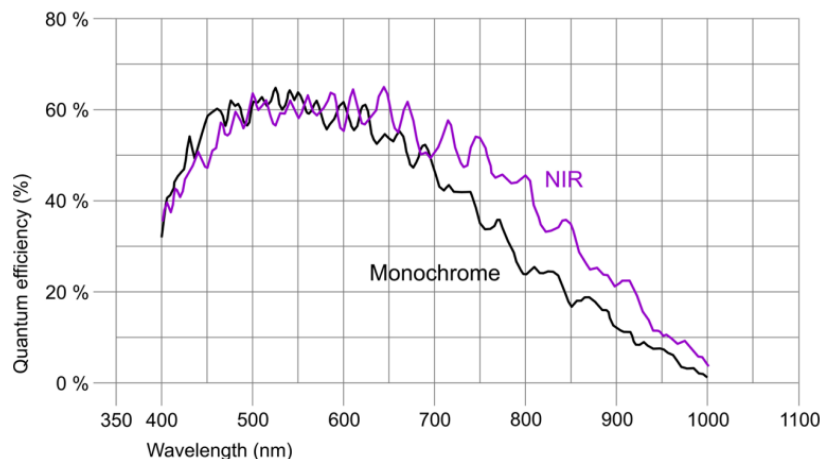


**Figura 5.6** Eficiencia cuántica del sensor UI-3360CP-NIR-GL (Rev. 2) [72].

Estas características instrumentales refuerzan la necesidad de considerar cuidadosamente el tratamiento posterior de los datos espectrales. En particular, las variaciones de sensibilidad a lo largo del rango VNIR y la disminución de señal hacia los extremos justifican la aplicación de estrategias de preprocesamiento orientadas a reducir el ruido y compensar efectos instrumentales, las cuales se abordan en detalle en la Sección 5.3. De este modo, se asegura que la información espectral utilizada por los modelos de cuantificación refleje principalmente las propiedades del material y no limitaciones propias del sistema de adquisición.

### 5.2.2 Espectros SWIR

Los espectros obtenidos en el rango SWIR (900–1700 nm) se muestran en la Figura 5.7, donde se observa una respuesta espectral consistente entre las distintas muestras, con variaciones principalmente asociadas al nivel de reflectancia. A diferencia del rango VNIR, los espectros adquiridos con la cámara Pika IR+ presentan una mayor estabilidad y menor nivel de ruido, por lo que no fue necesario aplicar ningún tipo de suavizado para su visualización.

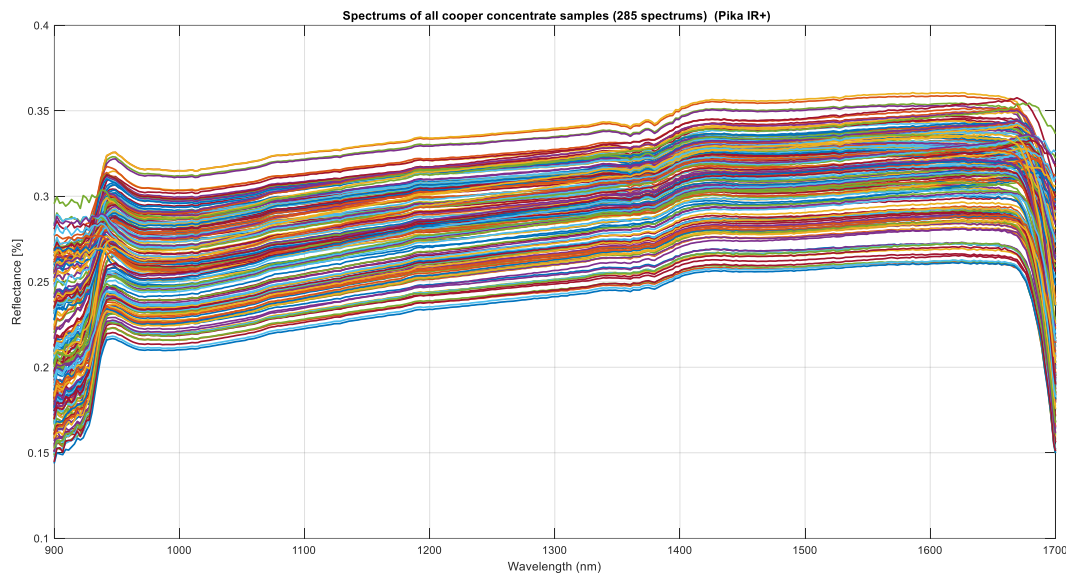


**Figura 5.7 Espectros SWIR de las muestras de concentrado de cobre.**

En la figura se aprecia un comportamiento más plano y uniforme de la reflectancia a lo largo del rango SWIR, especialmente en la región central del espectro, lo que refleja una mejor relación señal-ruido del sistema de adquisición en este rango espectral. La disminución de la señal hacia los extremos del rango de medición es atribuible principalmente a la respuesta instrumental del sensor y no a irregularidades propias de las muestras. Las diferencias observadas entre espectros reflejan la variabilidad composicional de los concentrados y efectos asociados a la dispersión superficial del material en polvo.

### **5.3. Análisis exploratorio PCA.**

Con el propósito de comprender mejor el comportamiento de los datos espectrales y analizar el origen de la variabilidad observada en las mediciones, se realizó un análisis exploratorio inicial mediante Análisis de Componentes Principales (PCA) sobre los espectros VNIR y SWIR. Esta técnica multivariante no supervisada permite simplificar el conjunto de datos original, representando la información contenida en un gran número de bandas espectrales mediante un número reducido de componentes que concentran la mayor parte de la variabilidad del sistema.

**Este análisis se llevó a cabo sobre los datos espectrales en su estado original (sin preprocesamiento), con el objetivo de identificar patrones globales, posibles outliers y fuentes de variabilidad asociadas tanto al material como a condiciones de adquisición. Posteriormente, en las etapas siguientes del trabajo, se aplicaron distintos métodos de preprocesamiento espectral con el fin de mejorar la calidad de los datos para el modelado.**

En esta sección, el PCA se aplica por separado a los espectros correspondientes a los rangos VNIR y SWIR, con el objetivo de analizar las características propias de cada uno sin introducir aún efectos asociados a la fusión de datos. A partir del análisis de los scores, es posible identificar patrones generales, tendencias y posibles agrupamientos entre las muestras, mientras que el estudio de los loadings permite reconocer qué regiones del espectro son las que más contribuyen a dichas variaciones. Este análisis exploratorio sirve como base para interpretar el comportamiento espectral del conjunto de datos y apoya las decisiones metodológicas adoptadas en las etapas posteriores de preprocesamiento y modelamiento.

### 5.3.1 PCA VNIR

La **Figura 5.8** presenta el gráfico de varianza explicada del análisis PCA aplicado al conjunto de espectros VNIR. Se observa que la **primera componente principal (PC1)** concentra aproximadamente el **83% de la varianza total**, mientras que la **segunda componente (PC2)** explica cerca de un **6,6%**, lo que indica que la mayor parte de la variabilidad espectral está dominada por un efecto global común a las muestras. Como se aprecia en la figura, el aporte de las componentes posteriores es considerablemente menor, lo que justifica centrar el análisis exploratorio en las dos primeras componentes principales.

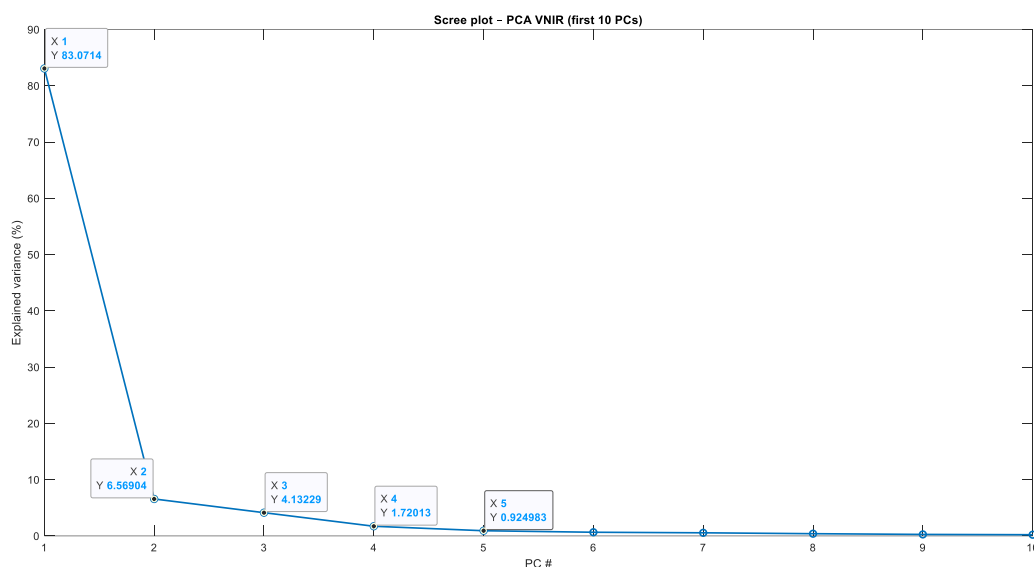


**Figura 5.8 Varianza explicada del PCA aplicado a los espectros VNIR.**

Tal como se muestra en la **Figura 5.9**, correspondiente al gráfico de **scores PC1–PC2**, las muestras se distribuyen de manera continua en el espacio reducido, sin separaciones abruptas ni agrupamientos completamente definidos. En la figura se distinguen **cinco tipos de muestras**, asociadas a distintos orígenes industriales y de laboratorio: concentrados provenientes de **Aurubis (A)**, **Trafigura (T)**, **Pan Pacific Copper (P)**, **Chuquicamata (Δ)** y **mezclas (MM)** preparadas en laboratorio a partir de concentrados de diversa procedencia. A pesar de estas diferencias de origen, se observa una **superposición significativa entre los grupos**, evidenciada por la intersección de las elipses de dispersión, lo que sugiere que los concentrados comparten un dominio espectral común y presentan variaciones composicionales graduales más que clases claramente separables.

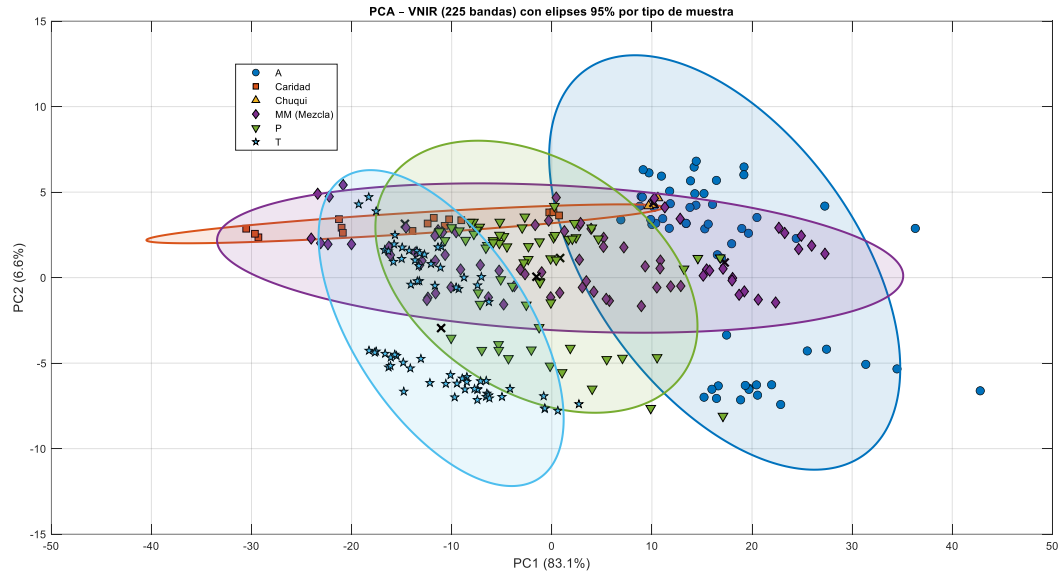


**Figura 5.9 Scores PC1–PC2 del PCA aplicado a los espectros VNIR.**

La **intersección entre la mayoría de las elipses** es un resultado esperado y positivo, ya que indica que las muestras, aun provenientes de distintas compañías o preparaciones, pertenecen a un mismo espacio espectral y no conforman poblaciones artificialmente separadas. Asimismo, los **centroides** de los distintos grupos se encuentran relativamente próximos entre sí, lo que refuerza la idea de un comportamiento espectral globalmente consistente entre los concentrados analizados.

Adicionalmente, los **tres puntos asociados a cada muestra individual**, correspondientes a adquisiciones espectrales independientes, se agrupan espacialmente muy cerca unos de otros en el espacio PCA. Esta cercanía entre réplicas confirma una **alta reproducibilidad en la adquisición espectral**, evidenciando que, pese a tratarse de mediciones distintas, los espectros representan de manera consistente una misma muestra.

El análisis de los **loadings**, presentado en la **Figura 5.10**, permite comprender qué zonas del espectro son las que más influyen en la variabilidad capturada por las primeras componentes principales. En este sentido, **PC1** muestra una contribución positiva y bastante uniforme a lo largo de gran parte del rango VNIR, lo que sugiere que esta componente está asociada principalmente a **cambios globales en el nivel de reflectancia** de las muestras. Este tipo de variación es habitual en materiales en polvo y suele estar relacionado con efectos generales como la dispersión superficial, la granulometría y la forma en que el material se dispone durante la medición.

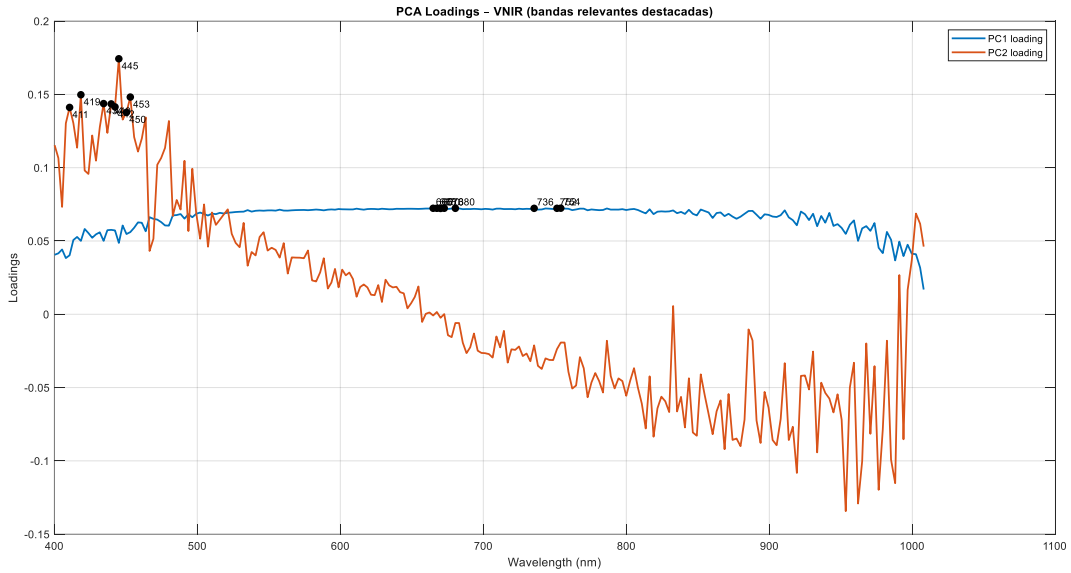


**Figura 5.10 Loadings PC1 y PC2 del PCA aplicado a los espectros VNIR.**

En particular, en el rango aproximado entre **650 y 750 nm**, los loadings de PC1 se mantienen **estables y sin picos marcados**, lo que indica que en esta zona no existe una banda espectral dominante. Más bien, la información se encuentra **repartida de manera continua** a lo largo del espectro, y son las distintas bandas, en conjunto, las que contribuyen a explicar la variabilidad observada. Este comportamiento es característico de espectros de concentrados minerales, donde los cambios responden a propiedades globales del material más que a rasgos espectrales aislados.

Por otro lado, **PC2** presenta un comportamiento más dependiente de regiones específicas del espectro, mostrando un cambio de signo entre distintas zonas del VNIR. Se observa una contribución más intensa en el **visible corto**, aproximadamente entre **400 y 450–500 nm**, región que coincide con una mayor intensidad relativa del sistema de iluminación utilizado durante la adquisición. En este trabajo, dicha iluminación fue implementada mediante una **fuentes LED de fabricación propia**, diseñada para reforzar la señal en esa zona del espectro, lo que se refleja directamente en los loadings y explica parte de la variabilidad recogida por esta componente.

Hacia los extremos del rango VNIR se observan mayores oscilaciones en los loadings, lo que es consistente con una menor estabilidad de la señal en dichas regiones y con la influencia de la respuesta instrumental del sistema de adquisición. En contraste, la región central del espectro presenta un comportamiento más suave y estable, reflejando una contribución espectral más uniforme en ese rango.

El análisis de loadings muestra que la variabilidad capturada por las primeras componentes principales combina efectos globales del espectro con variaciones más localizadas en determinadas regiones, proporcionando una visión exploratoria del comportamiento espectral del conjunto de datos.

### 5.3.2 PCA SWIR

La **Figura 5.11** presenta la varianza explicada correspondiente al análisis PCA aplicado al conjunto de espectros SWIR. Se observa que la primera componente principal (PC1) concentra aproximadamente el **81,9 % de la varianza total**, mientras que la segunda componente principal (PC2) explica cerca del **14,2 %**. A partir de la tercera componente, la contribución de varianza es marcadamente menor, indicando que su aporte resulta marginal desde una perspectiva exploratoria. Este comportamiento sugiere que la estructura espectral en el rango SWIR puede describirse adecuadamente mediante un espacio de baja dimensionalidad, donde las dos primeras componentes capturan la mayor parte de la variabilidad relevante. En contraste con escenarios dominados por una única componente, la contribución significativa de PC2 indica la presencia de variaciones espectrales adicionales no estrictamente globales.

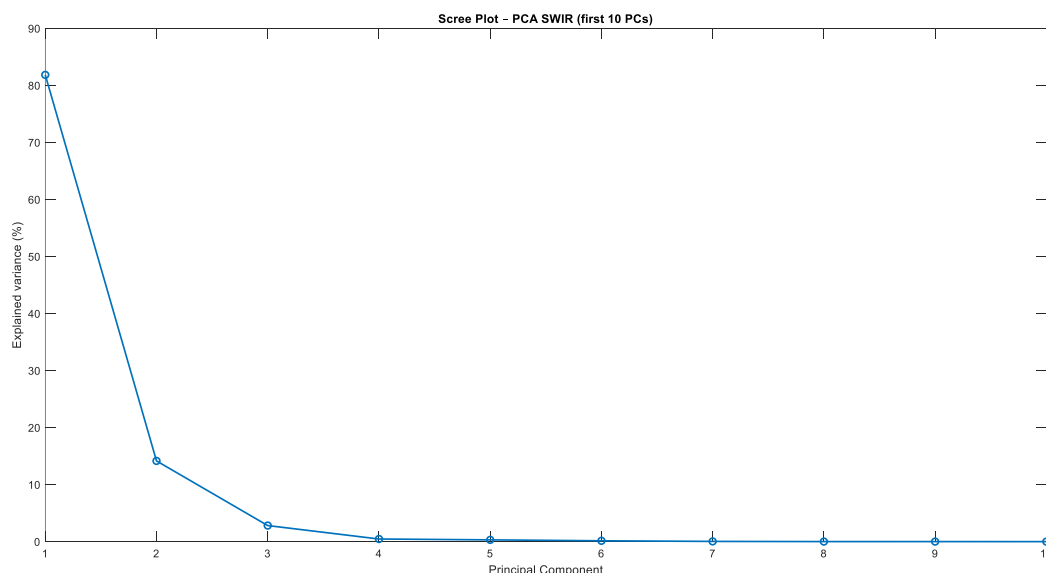


**Figura 5.11** Varianza explicada del PCA aplicado a los espectros VNIR.

Tal como se observa en la **Figura 5.12** as muestras exhiben una organización estructurada en el espacio PCA definido por PC1 y PC2. Los concentrados industriales correspondientes a Aurubis (A), Pan Pacific Copper (P) y Trafigura (T) tienden a agruparse en una región común del espacio multivariado, mostrando una superposición parcial de sus dominios de dispersión. Este patrón indica que dichos concentrados comparten características espectrales globales similares dentro del rango SWIR. Por otra parte, las muestras asociadas a **Caridad (C)** presentan una distribución más localizada y parcialmente diferenciada, sugiriendo un comportamiento espectral global distinto respecto de los concentrados industriales predominantes. De manera análoga, las muestras **MM (mezclas)** muestran una dispersión extendida a lo largo de PC1, lo cual es coherente con la naturaleza composicionalmente variable de este conjunto. La distribución observada refleja la capacidad del PCA para capturar diferencias sistemáticas entre grupos de muestras, donde PC1 describe principalmente variaciones globales de reflectancia y PC2 introduce un gradiente adicional asociado a cambios espectrales más específicos.

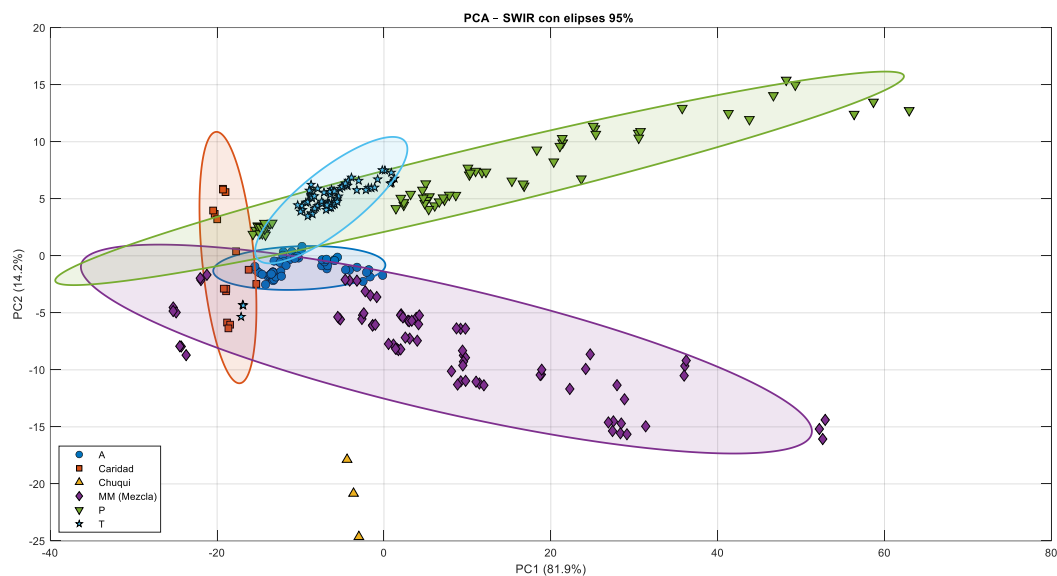


**Figura 5.12 Loadings PC1–PC2 del PCA aplicado a los espectros SWIR.**

El análisis de los **loadings**, presentado en la **Figura 5.13**, permite interpretar la naturaleza espectral de las componentes principales.

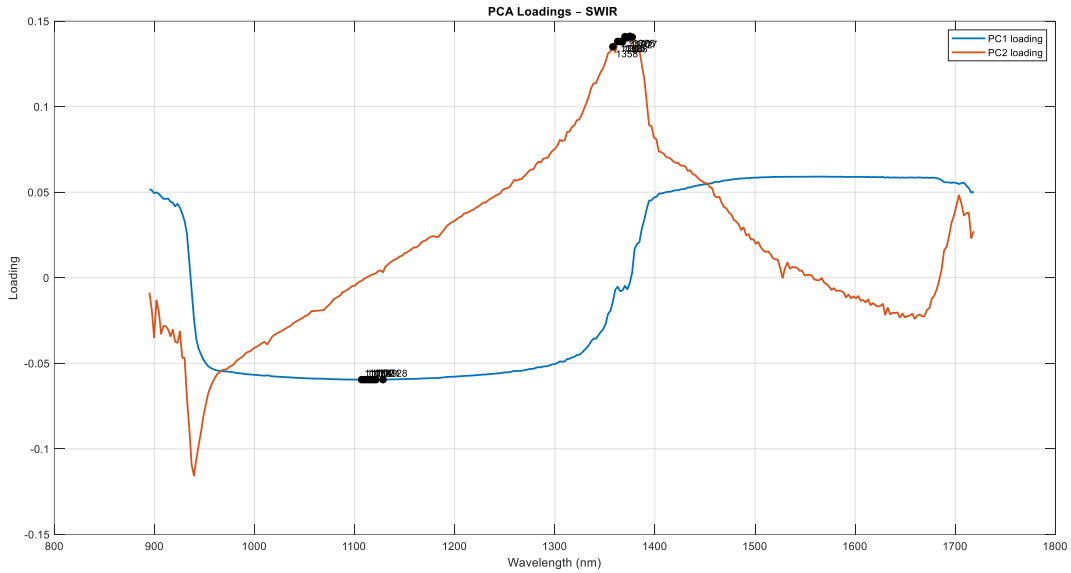


**Figura 5.13 Scores PC1–PC2 del PCA aplicado a los espectros SWIR.**

La componente PC1 exhibe un comportamiento relativamente suave y de baja variabilidad a lo largo del rango SWIR, lo que sugiere su asociación con efectos globales del nivel de reflectancia. En contraste, PC2 presenta variaciones más pronunciadas, destacándose regiones espectrales específicas donde la contribución alcanza valores extremos. Particularmente, se observan cambios significativos en la zona de **~1350–1400 nm**, indicando que esta componente captura variaciones más localizadas del espectro. El análisis PCA aplicado a los espectros SWIR pone de manifiesto un comportamiento espectral dominado por efectos globales, con una estructura más simple y estable en comparación con el rango VNIR.

### **5.3.3 Conclusiones análisis PCA**

El análisis mediante PCA aplicado a los espectros VNIR y SWIR se realizó con la intención de tener una primera aproximación al comportamiento de los datos, más que de llevar a cabo un estudio exhaustivo. En esta etapa, el PCA se utilizó como una herramienta exploratoria sencilla, orientada principalmente a observar cómo se distribuye la variabilidad espectral y a familiarizarse con la estructura general del conjunto de muestras antes de avanzar hacia análisis más detallados.

Desde este punto de vista, el rango VNIR se consideró el más relevante, ya que concentra una mayor diversidad de información espectral y muestra una estructura más rica, con variaciones distribuidas en más de una componente principal. En contraste, el rango SWIR presenta un comportamiento más estable y dominado por efectos globales, por lo que su aporte se entiende principalmente como complementario al VNIR, ayudando a reforzar y contextualizar la información obtenida en este último.

Este análisis permitió comprobar que los datos se comportan de manera consistente, que las réplicas espectrales de una misma muestra se agrupan de forma cercana y que no se observan comportamientos claramente anómalos. De esta forma, el PCA cumplió su rol como una etapa previa de exploración, entregando una visión general del conjunto de datos y sirviendo como base para abordar con mayor claridad las etapas posteriores de preprocesamiento, fusión de datos y modelamiento.

#### 5.4. Preprocesamiento y construcción de la función de pérdida:

Previo al desarrollo de los modelos de cuantificación, los espectros adquiridos fueron sometidos a distintas etapas de preprocesamiento, con el objetivo de reducir efectos no deseados asociados a la adquisición y resaltar la información espectral relevante para el modelamiento. En datos hiperespectrales, este paso es fundamental, ya que los espectros crudos pueden verse influenciados por factores como ruido instrumental, variaciones de iluminación, dispersión superficial del material y diferencias en la disposición de las muestras.

##### 5.4.1 Primer preprocesamiento: Savgol.

En este trabajo se utilizó el filtro de Savitzky–Golay **únicamente en modo de suavizado** ( $d = 0$ ), sin aplicar la primera derivada, debido a consideraciones instrumentales y metodológicas asociadas a la naturaleza de los datos espectrales. Los espectros adquiridos, en particular en el rango VNIR, presentan un nivel apreciable de ruido de alta frecuencia, relacionado con la iluminación, el sensor y el carácter difuso del concentrado en polvo. En este contexto, si bien la primera derivada puede resaltar variaciones locales, también tiende a amplificar el ruido, lo que puede afectar la estabilidad del modelamiento.

Adicionalmente, dado que el objetivo principal fue la **cuantificación multivariada** de propiedades mineralógicas y elementales, y no la detección puntual de bandas estrechas, se priorizó preservar las **tendencias globales del espectro**. Por esta razón, el uso del suavizado permitió reducir el ruido manteniendo la forma general de las curvas espectrales y una relación más directa entre los datos preprocesados y las variables de interés.

Desde un punto de vista matemático, el filtro de Savitzky–Golay puede expresarse como un promedio ponderado local. Para cada muestra  $i$  y longitud de onda  $k$ , el valor suavizado se calcula como:

$$x_{i,k}^{(0)} = \sum_{l=-m_0}^{m_0} c_l^{(0)} x_{i,k+l} \quad (50)$$

donde  $w = 2m_0 + 1$  corresponde al ancho de la ventana y  $c_l^{(0)}$  son los coeficientes del filtro asociados a un polinomio de grado  $p$ .

Definiendo el espectro de la muestra  $i$  como el vector fila  $\mathbf{x}_i \in \mathbb{R}^{1 \times m}$ , el espectro suavizado queda representado por el vector  $\mathbf{x}_i^{(SG)}$ . Agrupando los espectros suavizados de todas las muestras, la transformación Savitzky–Golay aplicada al conjunto completo se define como:

$$\text{Savgol}(\mathbf{X}) = \mathbf{X}^{(SG)} = \begin{bmatrix} \mathbf{x}_1^{(SG)} \\ \mathbf{x}_2^{(SG)} \\ \vdots \\ \mathbf{x}_n^{(SG)} \end{bmatrix} \in \mathbb{R}^{n \times m} \quad (51)$$

#### 5.4.2 Segundo preprocesamiento MSC:

La corrección de dispersión multiplicativa (**MSC**) se aplicó con el objetivo de **reducir efectos no químicos** presentes en los espectros, asociados principalmente a la naturaleza granular del concentrado de cobre. En mediciones de reflectancia sobre materiales en polvo, factores como el tamaño de partícula, la rugosidad superficial, la compactación del material y la distribución espacial de las partículas pueden introducir variaciones aditivas (desplazamientos de línea base) y multiplicativas (cambios de amplitud) que no están directamente relacionadas con la composición mineralógica.

El método MSC asume que cada espectro medido puede describirse como una versión linealmente distorsionada de un **espectro de referencia**, el cual se define como el espectro medio del conjunto de muestras. Al estimar y corregir estas distorsiones de forma individual para cada muestra, MSC permite **homogeneizar la escala y el nivel basal de los espectros**, preservando al mismo tiempo la información espectral relevante para el modelamiento multivariado. Esta etapa resulta especialmente útil previo al ajuste de modelos como PLS o SVR, ya que contribuye a mejorar la estabilidad numérica y la comparabilidad entre muestras.

Desde un punto de vista matemático, se define primero el espectro de referencia como el promedio de los espectros suavizados:

$$\bar{\mathbf{x}}^{(SG)} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^{(SG)} \quad (52)$$

Para cada muestra  $i$ , se asume el siguiente modelo lineal:

$$x_i^{(SG)} = a_i \tilde{x}^{(SG)} + b_i \quad (53)$$

donde  $a_i$  representa el efecto multiplicativo,  $b_i$  el efecto aditivo. Los coeficientes  $a_i$  y  $b_i$  se estiman mediante mínimos cuadrados y la corrección MSC se obtiene invirtiendo la distorsión estimada:

$$x_i^{(MSC)} = \frac{x_i^{(SG)} - b_i}{a_i} \quad (54)$$

Finalmente, agrupando los espectros corregidos de todas las muestras, la matriz de predictores corregida mediante MSC se define como:

$$MSC(X^{(SG)}) = X^{(MSC)} = \begin{bmatrix} x_1^{(MSC)} \\ x_2^{(MSC)} \\ \vdots \\ x_n^{(MSC)} \end{bmatrix} \in \mathbb{R}^{n \times m} \quad (55)$$

### 5.4.3 Tercer preprocesamiento autoescalado:

El autoscaling se aplicó para asegurar que todas las longitudes de onda aportaran de manera comparable al modelamiento. En lugar de utilizar únicamente un centrado a la media, se prefirió escalar las variables a varianza unitaria, ya que esto permitió trabajar con todas las variables en una misma escala. **Esta elección facilitó la optimización de hiperparámetros, especialmente en modelos como SVR, al evitar que diferencias de magnitud entre variables influyeran artificialmente en el ajuste y en la estabilidad del proceso de búsqueda.**

Desde un punto de vista matemático, el autoscaling se define como una transformación aplicada de forma independiente a cada longitud de onda. Sea  $X^{(MSC)}$  la matriz corregida, se define el vector de medias y desviaciones estándar como:

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i^{(MSC)}, \sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i^{(MSC)} - \mu)^2} \quad (56)$$

El espectro autoscalado de la muestra  $i$  queda definido como:

$$x_i^{(auto)} = \frac{x_i^{(MSC)} - \mu}{\sigma} \quad (57)$$

Finalmente, agrupando los espectros autoscalados de todas las muestras, se obtiene la matriz de autoscalada:

$$Auto(X^{(MSC)}) = X^{(auto)} = \begin{bmatrix} x_1^{(auto)} \\ x_2^{(auto)} \\ \vdots \\ x_n^{(auto)} \end{bmatrix} \in \mathbb{R}^{n \times m} \quad (58)$$

#### 5.4.4 Cuarto preprocesamiento VIP-kPLS:

El cuarto preprocesamiento aplicado corresponde a la **selección de variables espectrales mediante el criterio de Variable Importance in Projection (VIP)**, implementado sobre un modelo PLS preliminar. En datos hiperespectrales, los espectros suelen estar compuestos por cientos de longitudes de onda altamente correlacionadas, de las cuales solo una fracción aporta información relevante para la cuantificación de las variables de interés. El uso directo de todas las variables puede introducir redundancia, aumentar la complejidad del modelo y afectar su estabilidad numérica.

En este contexto, la selección de variables basada en VIP permite **identificar y retener únicamente las longitudes de onda más informativas**, manteniendo aquellas que contribuyen de manera significativa a la explicación de la variable de respuesta.

Desde un punto de vista matemático, el procedimiento se basa en el ajuste de un modelo PLS sobre la matriz autoscalada  $X^{(auto)}$  y la variable de respuesta, extrayendo  $A$  componentes latentes. A partir de este modelo, la importancia de cada longitud de onda  $j$  se cuantifica mediante el índice  $VIP$ , definido como:

$$VIP_j = \sqrt{\frac{m \sum_{a=1}^A SSY_a w_{j,a}^2}{\sum_{a=1}^A SSY_a}} \quad (59)$$

donde  $m$  es el número total de variables espectrales,  $A$  el número de componentes latentes del modelo PLS,  $SSY_a$  representa la varianza de la variable de respuesta explicada por la componente  $a$ , y  $w_{j,a}$  corresponde al peso de la variable  $j$  en dicha componente. Este índice entrega una medida adimensional de la relevancia de cada longitud de onda en el modelo.

Una vez calculados los valores VIP, se selecciona el conjunto de índices correspondiente a las  $k$  variables con mayor importancia  $\mathcal{S}_k = \{j_1, j_2, \dots, j_k\}$ , a partir de este conjunto, se construye la matriz reducida de espectros preprocesados  $\tilde{\mathbf{X}}$  como:

$$\tilde{\mathbf{X}} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_n \end{bmatrix} = \mathbf{X}_{[:, \mathcal{S}_k]}^{(auto)} \in \mathbb{R}^{n \times k} \quad (60)$$

Donde cada fila  $\tilde{x}_i$  corresponde al espectro de la muestra  $i$  restringido a las  $k$  longitudes de onda seleccionadas. Con esta última etapa, el flujo completo de preprocesamiento queda definido como:

$$\mathbf{X} \xrightarrow{SG} \mathbf{X}^{(SG)} \xrightarrow{MSC} \mathbf{X}^{(MSC)} \xrightarrow{Auto} \mathbf{X}^{(auto)} \xrightarrow{VIP-kPLS} \tilde{\mathbf{X}}$$

La secuencia de preprocesamientos aplicada se definió como una forma ordenada de tratar las distintas fuentes de variabilidad presentes en los datos espectrales, considerando tanto efectos propios del sistema de adquisición como características físicas del material y aspectos estadísticos del modelamiento. El suavizado mediante Savitzky–Golay permitió disminuir el ruido de alta frecuencia sin alterar la forma general de los espectros, la corrección MSC ayudó a compensar diferencias de escala y desplazamientos de línea base asociados a la naturaleza granular de las muestras, el autoscaling aseguró que todas las longitudes de onda aportaran de manera comparable al ajuste de los modelos, y la selección de variables basada en VIP permitió enfocar el análisis en las regiones espectrales con mayor contenido informativo.

### 5.5. Estrategia de fusión de datos

Una vez completada la cadena de preprocesamientos aplicada de forma independiente a cada bloque espectral incluyendo el filtrado Savitzky–Golay, la corrección por dispersión mediante MSC, el autoescalado y la selección de variables basada en VIP–kPLS, el siguiente paso consistió en definir la estrategia mediante la cual la información proveniente de los sensores VNIR y SWIR sería integrada en un único modelo predictivo.

Previo a la etapa de fusión, los datos hiperespectrales fueron representados a nivel de muestra mediante la extracción de un espectro promedio a partir de una región de interés (ROI) definida en cada adquisición, tal como se describe en la Sección 5.2. Esta estrategia permitió reducir la variabilidad espacial a nivel de píxel y asegurar que tanto los datos VNIR como SWIR fueran comparables bajo una misma unidad de análisis, a pesar de las diferencias en resolución espacial y características instrumentales entre ambos sistemas.

En este trabajo se adoptó una estrategia de fusión de datos a nivel intermedio (*mid-level data fusion*, **MLDF**), en la cual la integración se realiza en el espacio latente, en lugar de combinar directamente los espectros originales. Tal como se ilustra en la **Figura 5.14**, este enfoque permite extraer previamente la información más relevante de cada sensor y fusionarla de manera controlada. Esta decisión resulta especialmente adecuada en el contexto del análisis hiperespectral, donde ambos bloques presentan un elevado número de longitudes de onda altamente correlacionadas. Una concatenación directa de las variables originales conduciría a un espacio predictor de muy alta dimensionalidad, aumentando la multicolinealidad y la sensibilidad al ruido.

Para evitar estas limitaciones, cada bloque espectral preprocesado,  $\tilde{X}_{VNIR}$  y  $\tilde{X}_{SWIR}$ , fue modelado de manera independiente mediante regresión por mínimos cuadrados parciales (PLS). A partir de estos modelos se obtuvieron las matrices de *scores* latentes  $T_{VNIR}$  y  $T_{SWIR}$ , las cuales condensan la variabilidad espectral más relevante desde el punto de vista químico en un número reducido de componentes. De este modo, la información esencial de cada sensor queda representada de forma compacta, robusta y menos redundante.

La representación **MLDF** final se construyó mediante la concatenación de dichas matrices de *scores*, incorporando un factor de ponderación  $\gamma$  que permite ajustar la contribución relativa de cada bloque espectral en el modelo conjunto, de acuerdo con la expresión:

$$\tilde{X}_{MLDF} = [\gamma T_{VNIR} \quad (1 - \gamma) T_{SWIR}] \in \mathbb{R}^{n \times (A_{VNIR} + A_{SWIR})} \quad (61)$$

Esto permite preservar la información complementaria aportada por los sensores VNIR y SWIR, manteniendo al mismo tiempo una dimensionalidad reducida y una estructura más estable para el modelado posterior.

Sobre esta representación latente fusionada se desarrollaron los modelos PLS–SVR, permitiendo un balance controlado entre ambos rangos espectrales y la capacidad predictiva del sistema. **Cabe destacar que esta estrategia de fusión se implementa a nivel de características (feature-level), y no a nivel de píxel ni de imagen completa, lo que resulta consistente con el enfoque de modelado adoptado en este trabajo, orientado a la cuantificación global de cada muestra.**

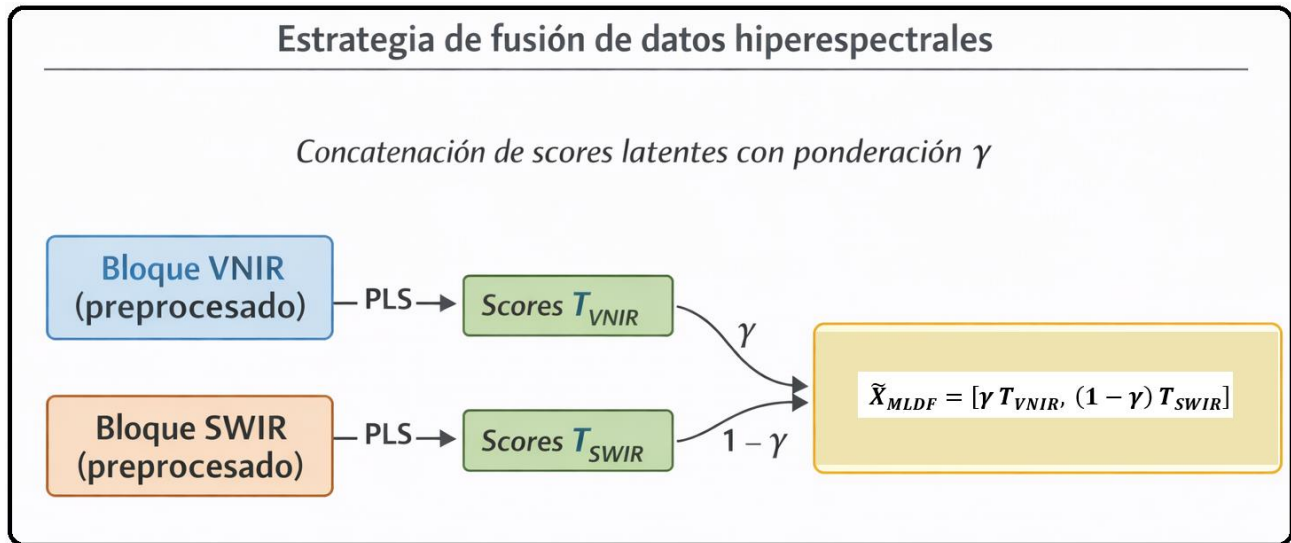


**Figura 5.14 Estrategia planteada para la fusión MLDF.[Elaboración propia]**

## 5.6. Construcción de los modelos de regresión

El preprocesamiento de los datos hiperespectrales, la construcción de las matrices espectrales VNIR y SWIR, y la estrategia de fusión de datos a nivel intermedio (*mid-level data fusion*, MLDF) fueron descritos en las secciones anteriores. En esta sección se presentan los procedimientos adoptados para el entrenamiento de los modelos de regresión, las métricas empleadas para la evaluación cuantitativa de su desempeño y el marco general utilizado para la optimización de los hiperparámetros.

Se consideraron dos enfoques de regresión: (i) la regresión lineal por Mínimos Cuadrados Parciales (PLS) y (ii) un modelo híbrido de Mínimos Cuadrados Parciales–Regresión por Vectores de Soporte (PLS–SVR). Ambos enfoques fueron entrenados y evaluados de manera independiente utilizando tres configuraciones de entrada distintas: datos VNIR únicamente, datos SWIR únicamente y la representación fusionada MLDF obtenida a partir de las variables latentes VNIR y SWIR.

### 5.6.1 Construcción PLS:

Sea  $\tilde{X} \in \mathbb{R}^{n \times k}$  la matriz espectral resultante, donde  $n$  corresponde al número de muestras y  $k$  al número de variables espectrales seleccionadas. A partir de esta matriz, se construyó el modelo de regresión por Mínimos Cuadrados Parciales (PLS), cuyo objetivo es modelar la relación entre  $\tilde{X}$  y la variable respuesta  $Y$  mediante un conjunto reducido de variables latentes que maximizan su covarianza mutua.

La regresión PLS fue implementada utilizando el algoritmo SIMPLS. En cada iteración  $i$ , se calcula un vector de pesos  $w_i$ , el cual define una dirección en el espacio de  $\tilde{X}$  que maximiza la covarianza con la variable respuesta residual  $y_{i-1}$ :

$$w_i = \frac{\tilde{X}^\top y_{i-1}}{\|\tilde{X}^\top y_{i-1}\|} \quad (62)$$

Los vectores de pesos asociados a las  $a$  variables latentes se agrupan en la matriz  $W_a = [w_1, \dots, w_a]$ . A partir de esta matriz se obtienen los *scores* latentes, que representan la proyección de las muestras originales en el espacio latente:

$$T_a = \tilde{X} W_a \quad (63)$$

Posteriormente, se calculan los *loadings*  $p_i$ , que describen la contribución de cada variable original en la reconstrucción de la matriz espectral:

$$p_i = \frac{\tilde{X}^\top t_i}{t_i^\top t_i}, P_a = [p_1, \dots, p_a] \quad (64)$$

De este modo, la matriz espectral  $\tilde{X}$  puede aproximarse mediante un modelo de rango reducido como:

$$\tilde{X}_a \approx T_a P_a^\top \quad (65)$$

Una vez definidas las matrices latentes, el vector de regresión del modelo PLS se expresa como:

$$B_a = W_a (P_a^\top W_a)^{-1} (T_a^\top T_a)^{-1} T_a^\top Y \quad (66)$$

Finalmente, la predicción para una nueva muestra espectral  $\tilde{X}_{new}$  se obtiene mediante:

$$\hat{y} = \tilde{X}_{new} B_a \quad (67)$$

Donde  $\hat{\mathbf{y}}$  corresponde al valor estimado por el modelo PLS. Esta formulación permite establecer una relación lineal estable entre  $\tilde{\mathbf{X}}$  y  $\mathbf{Y}$ , incluso en presencia de alta correlación entre las variables espectrales.

### 5.6.2 Construcción PLS-SVR:

El modelo híbrido PLS–SVR se construyó utilizando los *scores* latentes obtenidos a partir del modelo de Mínimos Cuadrados Parciales (PLS) como representación de baja dimensión de la matriz espectral original  $\tilde{\mathbf{X}}$ . En este enfoque, el PLS actúa como una etapa de extracción de características y reducción dimensional, mientras que la regresión final se realiza mediante Regresión por Vectores de Soporte (SVR), permitiendo capturar relaciones no lineales entre la información espectral y la variable de interés.

Sea  $T_a \in \mathbb{R}^{n \times a}$  la matriz de *scores* latentes obtenida a partir de  $\tilde{\mathbf{X}}$ , donde  $a$  corresponde al número de variables latentes retenidas. Estos *scores* constituyen las variables de entrada del modelo SVR. El objetivo del SVR es estimar una función de predicción  $f(\mathbf{t})$  que relacione los *scores* latentes con la variable respuesta  $\mathbf{Y}$ , de acuerdo con la expresión:

$$f(\mathbf{t}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{t}_i, \mathbf{t}) + b \quad (68)$$

donde  $\alpha_i$  y  $\alpha_i^*$  son los multiplicadores de Lagrange,  $\mathbf{t}_i$  corresponde al vector de *scores* de la muestra  $i$ ,  $K(\cdot, \cdot)$  es la función kernel y  $b$  es el término independiente del modelo.

Los coeficientes  $\alpha_i$  y  $\alpha_i^*$  se obtienen resolviendo el problema dual de optimización del SVR el cual se había descrito anteriormente en el estado del arte:

$$\min_{\alpha_i, \alpha_i^*} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(\mathbf{t}_i, \mathbf{t}_j) + \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) - \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \quad (69)$$

sujeto a las siguientes restricciones:

$$\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, 0 \leq \alpha_i, \alpha_i^* \leq C \quad (70)$$

donde  $C > 0$  es el parámetro de regularización que controla el compromiso entre la complejidad del modelo y el error de entrenamiento, mientras que  $\varepsilon$  define el ancho de la región insensible al error,

dentro de la cual las desviaciones no son penalizadas. En este trabajo, la función kernel utilizada corresponde exclusivamente al **kernel de base radial (Radial Basis Function, RBF)**, definido como:

$$K(\mathbf{t}_i, \mathbf{t}_{new}) = \exp\left(-\frac{\|\mathbf{t}_i - \mathbf{t}_{new}\|^2}{2\sigma^2}\right) \quad (71)$$

Donde  $\|\mathbf{t}_i - \mathbf{t}_{new}\|^2$  es la distancia euclidiana cuadrática entre los vectores de *scores* latentes,  $\sigma > 0$  es el parámetro de dispersión que controla el ancho de la función gaussiana. Este parámetro regula la influencia de cada vector de soporte sobre la predicción: valores pequeños de  $\sigma$  producen modelos altamente locales y sensibles a variaciones finas en el espacio latente, mientras que valores mayores generan superficies de regresión más suaves y globales.

Una vez resuelto el problema dual, el término de sesgo del modelo se calcula como:

$$\mathbf{b} = \mathbf{y}_k - \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{t}_i, \mathbf{t}_k) \quad (72)$$

donde  $\mathbf{y}_k$  corresponde al valor de referencia de la muestra  $\mathbf{k}$  que actúa como vector de soporte.

Para una nueva muestra espectral  $\tilde{\mathbf{X}}_{new}$ , los *scores* latentes se obtienen mediante la proyección sobre la matriz de pesos del PLS:

$$\mathbf{t}_{new} = \tilde{\mathbf{X}}_{new} \mathbf{W}_a \quad (73)$$

Finalmente, la predicción de la variable de interés se calcula como:

$$\hat{\mathbf{y}} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{t}_i, \mathbf{t}_{new}) + \mathbf{b} \quad (74)$$

De este modo, el modelo PLS–SVR implementado combina la reducción dimensional y estabilidad del PLS con la flexibilidad no lineal del kernel RBF, permitiendo modelar de forma robusta relaciones complejas en el espacio latente y resultando especialmente adecuado para el análisis de datos hiperespectrales y representaciones fusionadas MLDF.

### 5.7. Métricas de desempeño:

Para evaluar el desempeño de los modelos de regresión desarrollados en este estudio se emplearon las siguientes métricas: coeficiente de determinación ( $R^2$ ), error cuadrático medio ( $RMSE$ ), su versión normalizada ( $NRMSE$ ), razón de desempeño a desviación ( $RPD$ ), error relativo porcentual ( $REP\%$ ) y rango de error relativo ( $RER$ ). Estas métricas fueron seleccionadas debido a que permiten analizar el comportamiento predictivo de los modelos desde perspectivas complementarias, considerando tanto el ajuste estadístico como la magnitud y relevancia práctica del error.

El **coeficiente de determinación ( $R^2$ )** se utilizó para cuantificar la proporción de la variabilidad de los datos de referencia explicada por el modelo. Esta métrica proporciona una visión global del grado de ajuste entre los valores estimados y los valores reales, facilitando la comparación entre diferentes modelos y configuraciones espectrales. En concentrados minerales, donde la heterogeneidad mineralógica, los efectos de mezcla de fases y las incertidumbres analíticas son significativos, valores de  $R^2$  en el rango **0.80–0.90** se consideran **muy buenos**, ya que reflejan una relación sólida bajo condiciones realistas de operación.

El **error cuadrático medio ( $RMSE$ )** se empleó como una medida directa de la precisión de las predicciones, al expresar la magnitud promedio del error en las mismas unidades de la variable analizada. Su uso permite evaluar de forma clara las diferencias de desempeño entre las distintas etapas de calibración, validación cruzada y predicción, así como detectar posibles pérdidas de capacidad predictiva. Dado que el  $RMSE$  depende del rango y de la dispersión de los datos, su interpretación se realizó de forma comparativa entre modelos y configuraciones espectrales, priorizando aquellos con errores bajos y consistentes.

En particular, el  **$NRMSE$**  se incorpora como una métrica que permite interpretar el error de predicción en relación con el rango de variación de la variable de interés. A diferencia del  **$RMSE$** , que entrega una medida absoluta del error en las mismas unidades de la variable, el  **$NRMSE$**  expresa dicho error como un porcentaje del rango observado, lo que facilita la comparación entre variables con distintas escalas y dispersiones, como ocurre en la composición de cobre, calcopirita y pirita.

Esta propiedad resulta especialmente relevante en concentrados minerales, donde la amplitud composicional puede variar significativamente entre fases, permitiendo evaluar de manera más consistente el desempeño relativo de los modelos y el impacto de estrategias como la fusión de datos en la reducción del error predictivo. En términos de interpretación, valores de ***NRMSE*** inferiores al 15% indican un error bajo en relación con el rango composicional del sistema, siendo adecuados para aplicaciones cuantitativas; valores entre 15% y 25% reflejan un desempeño aceptable, mientras que valores superiores al 25% evidencian un error elevado que puede limitar la aplicabilidad práctica del modelo.

La **razón de desempeño a desviación (*RPD*)** se incorporó para evaluar el desempeño del modelo en relación con la dispersión de los datos de referencia. Esta métrica permite contextualizar el error de predicción respecto de la variabilidad natural del conjunto de datos, siendo ampliamente utilizada para estimar la utilidad práctica de los modelos en aplicaciones quimiométricas. En el contexto de concentrados minerales, valores de ***RPD* entre 2.0 y 3.0** se consideran **muy buenos** y adecuados para predicción cuantitativa confiable, mientras que valores cercanos o superiores a 3.0 indican un desempeño excelente.

El **error relativo porcentual (*REP%*)** se utilizó para expresar el error de predicción de forma normalizada y adimensional, facilitando una interpretación más intuitiva del desempeño del modelo y la comparación entre variables con diferentes rangos de concentración o magnitud. En aplicaciones metalúrgicas, valores de ***REP%* inferiores al 5%** se consideran excelentes, valores en el rango **5–10%** muy buenos y plenamente aceptables, mientras que valores entre **10% y 15%** pueden seguir siendo adecuados cuando se trabaja con variables mineralógicas complejas o concentrados con alta variabilidad interna.

Finalmente, el **rango de error relativo (*RER*)** se empleó como un indicador adicional de desempeño normalizado, relacionando el rango total de los datos de referencia con el error de predicción. Esta métrica resulta particularmente útil para evaluar la capacidad de los modelos para discriminar adecuadamente variaciones dentro del rango completo de la variable analizada. En concentrados minerales, el rango composicional de las especies está **naturalmente acotado** por límites físicos y metalúrgicos, definidos por la mineralogía dominante y las condiciones del proceso. Como consecuencia, el valor máximo alcanzable del *RER* se ve restringido.

Bajo este contexto, valores de **RER cercanos a 8–10** se consideran **buenos**, mientras que valores superiores a 10 indican **muy buen desempeño**, aun cuando no se alcancen los umbrales típicamente reportados en aplicaciones con rangos composicionales más amplios.

La Tabla 1 resume los rangos típicos observados para cada métrica y su correspondiente interpretación cualitativa.

**Tabla 1 Interpretación de las métricas de desempeño predictivo en concentrados minerales**

| <b>Métrica</b>            | <b>Rango típico observado</b> | <b>Interpretación en concentrados minerales</b>                                 |
|---------------------------|-------------------------------|---------------------------------------------------------------------------------|
| <b>(<math>R^2</math>)</b> | <b>0.80 – 0.90</b>            | Muy buen ajuste considerando heterogeneidad mineralógica y ruido analítico      |
| <b><i>RMSE</i></b>        | <b>Bajo y estable</b>         | Buena precisión predictiva cuando es consistente entre calibración y predicción |
| <b><i>NRMSE</i></b>       | <b>10 – 15%</b>               | Desempeño bueno, útil para estimaciones cuantitativas                           |
| <b><i>RPD</i></b>         | <b>2.0 – 3.0</b>              | Muy buen desempeño, adecuado para predicción cuantitativa                       |
| <b><i>REP</i>%</b>        | <b>5 – 15%</b>                | Muy buen desempeño, aceptable para aplicaciones metalúrgicas                    |
| <b><i>RER</i></b>         | <b>8 – 10</b>                 | Buen desempeño en rangos composicionales acotados                               |

## **5.8. Entrenamiento y validación de los modelos**

El entrenamiento y la evaluación de los modelos se realizaron mediante un esquema de validación escalonado basado en una partición fija del conjunto de datos. El conjunto completo de muestras fue dividido inicialmente en tres subconjuntos: el **60%** de las muestras se utilizó para la etapa de **calibración**, el **20%** se reservó para **validación externa**, y el **20% restante** se destinó al conjunto de **predicción** o prueba final. Esta partición permitió evaluar de forma progresiva el desempeño de los modelos y su capacidad de generalización.

El desarrollo de los modelos se llevó a cabo exclusivamente utilizando el conjunto de calibración. Sobre este subconjunto se aplicó validación cruzada Leave-One-Out (LOO), con el objetivo de analizar la estabilidad de los modelos y detectar posibles problemas de sobreajuste en etapas tempranas del entrenamiento. La validación externa se utilizó posteriormente como criterio independiente para guiar el ajuste de hiperparámetros, asegurando que las decisiones de optimización no dependieran ni de los datos de calibración ni de los resultados de la validación cruzada interna.

El conjunto de predicción fue mantenido completamente fuera del proceso de construcción del modelo y se empleó únicamente para evaluar el desempeño predictivo sobre muestras no vistas previamente. De este modo, se obtuvo una estimación imparcial de la capacidad de generalización de los modelos entrenados, condición fundamental para evaluar su aplicabilidad en escenarios reales.

Este esquema de validación, que combina validación cruzada interna con conjuntos independientes de validación externa y predicción, sigue prácticas ampliamente aceptadas en modelado quimiométrico y resulta particularmente adecuado para aplicaciones espectroscópicas de alta dimensionalidad.

Las métricas de desempeño ( $R^2$ , ***RMSE***, ***RPD***, ***REP%*** y ***RER***) se calcularon de manera consistente en cada una de las etapas, lo que permitió realizar una comparación fiable entre los modelos PLS lineales y los modelos no lineales PLS–SVR, así como entre las configuraciones VNIR, SWIR y la representación fusionada MLDF.

la **Figura 5.15** presenta un resumen gráfico del proceso de partición del conjunto de datos y del flujo seguido durante la construcción de los modelos. En dicha figura se ilustra el rol específico de cada subconjunto calibración, validación externa y predicción dentro del proceso de entrenamiento, optimización de hiperparámetros y evaluación final del desempeño predictivo. Esta representación permite visualizar de forma clara la separación estricta entre las etapas de desarrollo del modelo y la evaluación sobre muestras no vistas, garantizando una estimación imparcial de la capacidad de generalización.

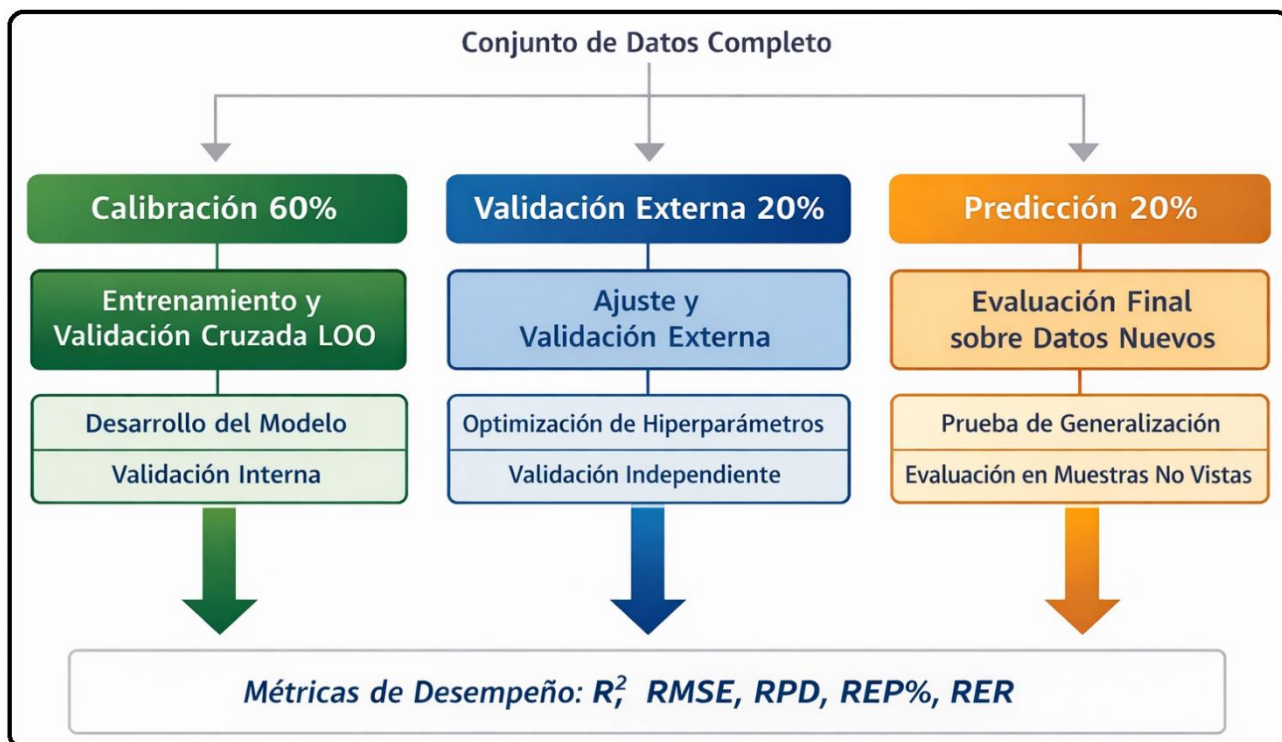


Figura 5.15 Esquema del proceso de entrenamiento, validación y predicción de los modelos.[Elaboración propia]

## 5.9. Optimización de la función objetivo.

La optimización de hiperparámetros se realizó con el objetivo de obtener modelos que presentaran un buen desempeño no solo sobre los datos de calibración, sino también cuando fueran aplicados a muestras no vistas. Tanto en los modelos PLS como en los modelos híbridos PLS–SVR considerando las configuraciones VNIR, SWIR y la representación fusionada MLDF la calidad predictiva resulta altamente sensible a la elección de los parámetros de preprocesamiento, al número de variables latentes y, en el caso de los modelos no lineales, a los parámetros de regularización asociados al kernel SVR. Por esta razón, fue necesario adoptar una estrategia de optimización estructurada pero flexible. Como estrategia de búsqueda se empleó **Optimización Bayesiana (Bayesian Optimization, BO)**.

A diferencia de enfoques exhaustivos como la búsqueda en grilla o la búsqueda aleatoria, la optimización bayesiana se basa en la construcción de un modelo probabilístico sustituto de la función objetivo, el cual es actualizado de manera iterativa a medida que se evalúan nuevas combinaciones de hiperparámetros.

A partir de este modelo sustituto, se utilizan funciones de adquisición para seleccionar los puntos más informativos del espacio de búsqueda, permitiendo balancear de forma eficiente la exploración de regiones con alta incertidumbre y la explotación de posibles soluciones [73,74].

Este enfoque permite identificar configuraciones de hiperparámetros de alto desempeño con un número reducido de evaluaciones del modelo, lo que resulta especialmente ventajoso en problemas de alta dimensionalidad y elevado costo computacional, como los asociados al análisis hiperespectral. En particular, aplicaciones recientes han demostrado el valor de este enfoque en contextos espectroscópicos reales.

Por ejemplo, en [75] se aplicó la optimización bayesiana para ajustar modelos PLSR sobre datos FT-NIR, mostrando mejoras sustanciales en precisión, robustez y eficiencia frente a métodos de búsqueda convencionales, lo que refuerza su aplicabilidad en tareas de modelado quimiométrico.

La selección del modelo se guio mediante la construcción de una función objetivo compuesta. En particular, la función objetivo combina el **RMSE** obtenido mediante validación cruzada Leave-One-Out (LOO) sobre el conjunto de calibración con el error calculado sobre el conjunto de validación externa, asignando igual peso a ambos términos:

$$\mathbf{OBJ} = 0.5 \cdot \mathbf{RMSE}_{CV} + 0.5 \cdot \mathbf{RMSE}_V \quad (75)$$

donde  $\mathbf{RMSE}_{CV}$  corresponde al error de validación cruzada LOO y  $\mathbf{RMSE}_V$  al error obtenido sobre el subconjunto de validación externa (20%). Al minimizar este criterio, se favorecen modelos que logran un equilibrio entre un buen ajuste en calibración y un desempeño estable frente a datos independientes, reduciendo el riesgo de sobreajuste.

## 5.10. Definición de los Hiperparámetros:

El número total de hiperparámetros optimizados depende de la complejidad del modelo y de la estrategia de modelado adoptada.

### 5.10.1 Hiperparametros Modelos PLS.

En la configuración PLS de bloque único (VNIR o SWIR), se optimizaron tres hiperparámetros, cada uno asociado a una etapa clave del modelado. El parámetro *savgol* (*savgolVNIR* o *savgolSWIR*) define la longitud de la ventana del filtro Savitzky–Golay y controla el nivel de suavizado espectral. El parámetro *k* (*kVNIR* o *kSWIR*) determina el número de longitudes de onda retenidas tras la selección de variables mediante VIP–PLS, regulando el grado de compresión espectral y la preservación de información químicamente relevante. Finalmente, el parámetro *nComp* (*nCompVNIR* o *nCompSWIR*) especifica el número de variables latentes extraídas en la descomposición PLS, resumiendo la estructura de covarianza entre las variables espectrales seleccionadas y la variable de interés.

### 5.10.2 Hiperparametros Modelos PLS-SVR

En los modelos PLS–SVR se introdujo regresión no lineal mediante tres hiperparámetros adicionales asociados al SVR. El parámetro *BoxConstraint* controla la penalización aplicada a los errores de predicción, donde valores mayores imponen ajustes más estrictos y valores menores favorecen soluciones más suaves y generalizables. El parámetro *Epsilon* define el ancho de la región insensible al error, dentro de la cual las desviaciones pequeñas no son penalizadas, regulando el compromiso entre tolerancia al ruido y precisión. Por su parte, el parámetro *KernelScale* controla la dispersión del kernel RBF, determinando el grado de flexibilidad del modelo: valores pequeños generan funciones de decisión altamente locales, mientras que valores mayores conducen a relaciones más suaves y globales.

### 5.10.3 Modelos MLDF

En los modelos basados en fusión de datos a nivel intermedio (MLDF), se optimizó además un hiperparámetro adicional, ***gamma*  $\gamma$** , correspondiente al peso de fusión. Este parámetro controla la contribución relativa de los espacios latentes VNIR y SWIR dentro de la matriz predictora fusionada, influyendo directamente en la forma en que se integra la información espectral complementaria de ambos sensores.

Lo siguiente puede ser resumido en la siguiente tabla 2:

**Tabla 2 Hiperparámetros optimizados según la configuración de modelado**

| Modelo                | Hiperparámetro                                                                                                                                                                                               | Descripción                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PLS (VNIR o SWIR)     | <b>savgol</b><br><b>k</b><br><b>nComp</b>                                                                                                                                                                    | Longitud de ventana del filtro Savitzky–Golay (suavizado espectral)<br>Nº de longitudes de onda seleccionadas tras VIP–PLS<br>Nº de variables latentes extraídas en la descomposición PLS                                                                                                                                                                                                                                                                                                                                                                                            |
| PLS–SVR (VNIR o SWIR) | <b>savgol</b><br><b>k</b><br><b>nComp</b><br><b>BoxConstraint</b><br><b>Epsilon</b><br><b>KernelScale</b>                                                                                                    | Igual que en PLS<br>Igual que en PLS<br>Igual que en PLS<br>Penalización aplicada a errores (control de complejidad del SVR)<br>Margen de tolerancia insensible al error en SVR<br>Ancho del kernel RBF (controla la flexibilidad del modelo no lineal)                                                                                                                                                                                                                                                                                                                              |
| MLDF–PLS              | <b>savgolVNIR</b><br><b>kVNIR</b><br><b>nCompVNIR</b><br><b>savgolSWIR</b><br><b>kSWIR</b><br><b>nCompSWIR</b><br><b><math>\gamma</math></b>                                                                 | Suavizado espectral aplicado al bloque VNIR<br>Nº de bandas seleccionadas en VNIR tras VIP–PLS<br>Nº de componentes latentes extraídas del bloque VNIR<br>Suavizado espectral aplicado al bloque SWIR<br>Nº de bandas seleccionadas en SWIR tras VIP–PLS<br>Nº de componentes latentes extraídas del bloque SWIR<br>Peso de fusión que regula la contribución relativa de VNIR y SWIR                                                                                                                                                                                                |
| MLDF–PLS–SVR          | <b>savgolVNIR</b><br><b>kVNIR</b><br><b>nCompVNIR</b><br><b>savgolSWIR</b><br><b>kSWIR</b><br><b>nCompSWIR</b><br><b><math>\gamma</math></b><br><b>BoxConstraint</b><br><b>Epsilon</b><br><b>KernelScale</b> | Suavizado espectral aplicado al bloque VNIR<br>Nº de bandas seleccionadas en VNIR tras VIP–PLS<br>Nº de componentes latentes extraídas del bloque VNIR<br>Suavizado espectral aplicado al bloque SWIR<br>Nº de bandas seleccionadas en SWIR tras VIP–PLS<br>Nº de componentes latentes extraídas del bloque SWIR<br>Peso de fusión que regula la contribución relativa de VNIR y SWIR<br>Penalización aplicada a errores (control de complejidad del SVR)<br>Margen de tolerancia insensible al error en SVR<br>Ancho del kernel RBF (controla la flexibilidad del modelo no lineal) |

### 5.11. Estimación de la incertidumbre.

En aplicaciones reales como el análisis de minerales o concentrados metálicos, no basta con generar una predicción puntual del contenido de un elemento químico, por ejemplo, decir que una muestra contiene "24.3% de cobre". Es igual de importante saber **qué tan confiable es esa predicción**. ¿El modelo está "seguro" de ese valor? ¿O podría haber una variabilidad significativa? Esta confianza o duda sobre una estimación se conoce como **incertidumbre**. Tener en cuenta la incertidumbre, permite a los usuarios del modelo ya sea en laboratorio, planta o control de calidad tomar decisiones más informadas, sabiendo si los resultados son precisos o si podrían estar sujetos a error o ambigüedad [76]. La estimación de incertidumbre, por tanto, busca no solo decir *cuánto* vale una variable objetivo, sino también proporcionar un **intervalo de predicción**, es decir, un rango dentro del cual es razonable esperar que se encuentre el valor real. Esto es especialmente importante en datos espectroscópicos, donde factores como el ruido instrumental, la variabilidad mineralógica, y la heterogeneidad física de las muestras pueden introducir errores difíciles de predecir [77]. Frente a estas limitaciones, métodos como **Jackknife+** permiten construir intervalos estadísticamente válidos que indican el grado de confianza que puede asignarse a cada predicción individual, fortaleciendo la interpretabilidad y aplicabilidad del modelo en escenarios prácticos.

Se optó por **Jackknife+** debido a sus propiedades particularmente útiles en contextos de regresión no lineal y de alta dimensionalidad, como es el caso de la fusión de datos espectrales. A diferencia de métodos tradicionales como los intervalos basados en supuestos gaussianos o en la varianza del error residual, **Jackknife+** no depende de la distribución de los errores ni de la linealidad del modelo. Además, garantiza cobertura marginal válida en muestras finitas, incluso cuando el modelo está sobreajustado o presenta complejidad estructural. Su implementación es conceptualmente sencilla y escalable, y permite una evaluación directa de la incertidumbre en modelos PLS y PLS–SVR sin requerir reentrenamientos costosos ni muestreo aleatorio adicional.

Con el objetivo de cuantificar la incertidumbre asociada a las predicciones puntuales y construir intervalos de predicción estadísticamente válidos para cada configuración de modelo, se implementó el método **Jackknife+**. Este enfoque de inferencia predictiva no paramétrico proporciona intervalos marginalmente válidos en muestras finitas bajo supuestos mínimos, lo que lo hace especialmente adecuado para modelos no lineales como PLS–SVR y para conjuntos de datos hiperespectrales de alta dimensionalidad [76, 77].

En este estudio, **Jackknife+** se aplicó utilizando el mismo esquema Leave-One-Out (LOO) empleado durante el entrenamiento de los modelos. Para cada muestra de calibración  $i$ , se reentrenó el modelo excluyéndola, generando una predicción  $\hat{y}_{(-i)}$ . Luego, para cualquier muestra nueva o reservada  $j$ , el intervalo de predicción **Jackknife+** se obtiene al combinar todas las predicciones LOO junto con sus residuos asociados. La expresión formal del intervalo es:

$$PI_j^{JK+} = [\min_i (\hat{y}_{(-i)} - (y_i - \hat{y}_{(-i)})), \max_i (\hat{y}_{(-i)} - (y_i - \hat{y}_{(-i)}))] \quad (76)$$

Donde  $y_i$  representa la concentración de referencia (por ejemplo, %Cu, %Fe, %S) de la muestra  $i$ , y  $\hat{y}_{(-i)}$  es la predicción generada sin incluir dicha muestra en el entrenamiento. Los términos  $(y_i - \hat{y}_{(-i)})$  capturan la variabilidad local del modelo, y su agregación permite construir intervalos con garantía de cobertura marginal, incluso para modelos flexibles o sobreajustados [76]. **Jackknife+** se aplicó de forma consistente a todas las configuraciones de modelado (PLS y PLS-SVR; datos VNIR, SWIR y MLDF). Dado que cada muestra física estaba representada por tres espectros promediados, que se mantuvieron agrupados durante toda la partición de datos, la estimación de incertidumbre se realizó estrictamente a nivel de muestra. Esto asegura que los intervalos reflejen la variabilidad química real, en lugar del ruido espectral intra-muestra. Este aspecto resulta especialmente relevante en el análisis de concentrados de cobre, donde la heterogeneidad textural, la distribución de tamaño de partícula y los efectos de compactación introducen incertidumbre sistemática no capturada por estimaciones puntuales. Finalmente, los intervalos generados por **Jackknife+** se integraron directamente en todas las gráficas de evaluación del modelo, calibración, validación cruzada (LOO), validación externa y predicción, facilitando una visualización clara de la incertidumbre en cada etapa. Estos intervalos complementan las métricas clásicas (**RMSE**, **R<sup>2</sup>**, **RPD**, **REP%**) al ofrecer una visión cuantitativa y gráfica de la confiabilidad de las predicciones.

## 5.12. Resultados

Para evaluar el desempeño de cada estrategia de modelado, se entrenaron todas las configuraciones MLDF, SWIR y VNIR utilizando sus mejores combinaciones de hiperparámetros, detalladas en el Apéndice. La evaluación se realizó considerando cuatro etapas clave: calibración, validación cruzada Leave-One-Out (LOO-CV), validación externa y predicción final, lo que permitió analizar tanto el ajuste del modelo como su capacidad de generalización.

En cada etapa se calcularon seis métricas: ***RMSE***, ***NRMSE***, ***R<sup>2</sup>***, ***REP%***, ***RPD*** y ***RER***. Estas permiten evaluar precisión absoluta, grado de ajuste, error relativo, relación con la dispersión del conjunto y capacidad para cubrir el rango completo de concentraciones. Las Tablas 3, 4 y 5 presentan los resultados obtenidos para la estimación de %Cu, %Calcopirita y %Pirita, respectivamente, comparando regresión lineal (PLS) y no lineal (PLS–SVR). En cada tabla se resumen los resultados para todas las combinaciones de preprocesamiento, reducción de variables y estrategia de modelado, incluyendo sensores individuales y fusión MLDF. Estas tablas sirven como punto de partida para el análisis comparativo que sigue, donde se destacan diferencias según el tipo de mineral, el aporte de cada sensor y las ventajas del modelado no lineal en relaciones espectro-mineralógicas complejas.

**Tabla 3 Resultados de las Métricas de Desempeño (RMSE, NRMSE, R<sup>2</sup>, REP%, RPD, RER) para la Estimación de %Cu según Modelo, Sensor y Etapa de Evaluación.**

| %Cu                |              |             | Metricas         |            |                |            |            |
|--------------------|--------------|-------------|------------------|------------|----------------|------------|------------|
| Etapas             | Modelo       | $RMSE_c$    | $NRMSE_c$ (%)    | $R^2_c$    | $REP_c$ (%)    | $RPD_c$    | $RER_c$    |
| Calibración        | PLS-SVR MLDF | 0.8364      | 5.5525           | 0.9151     | 3.3            | 3.44       | 18.01      |
|                    | PLS MLDF     | 0.836       | 5.5525           | 0.9152     | 3.29           | 3.44       | 18.01      |
|                    | PLS-SVR VNIR | 1.0379      | 6.8918           | 0.8693     | 4.09           | 2.77       | 14.51      |
|                    | PLS VNIR     | 1.2493      | 8.2988           | 0.8107     | 4.92           | 2.3        | 12.05      |
|                    | PLS-SVR SWIR | 1.287       | 8.5470           | 0.7991     | 5.07           | 2.24       | 11.7       |
|                    | PLS SWIR     | 1.6575      | 11.0011          | 0.6667     | 6.53           | 1.74       | 9.09       |
| Etapas             | Modelo       | $RMSE_{cv}$ | $NRMSE_{cv}$ (%) | $R^2_{cv}$ | $REP_{cv}$ (%) | $RPD_{cv}$ | $RER_{cv}$ |
| CV                 | PLS-SVR MLDF | 0.9362      | 6.2150           | 0.8937     | 3.69           | 3.08       | 16.09      |
|                    | PLS MLDF     | 0.938       | 6.2267           | 0.8933     | 3.7            | 3.07       | 16.06      |
|                    | PLS-SVR VNIR | 1.1644      | 7.7340           | 0.8355     | 4.59           | 2.47       | 12.93      |
|                    | PLS VNIR     | 1.4583      | 9.6805           | 0.742      | 5.75           | 1.97       | 10.33      |
|                    | PLS-SVR SWIR | 1.3712      | 9.1075           | 0.7719     | 5.4            | 2.1        | 10.98      |
|                    | PLS SWIR     | 1.9042      | 12.6422          | 0.5601     | 7.5            | 1.51       | 7.91       |
| Etapas             | Modelo       | $RMSE_v$    | $NRMSE_v$ (%)    | $R^2_v$    | $REP_v$ (%)    | $RPD_v$    | $RER_v$    |
| Validación externa | PLS-SVR MLDF | 0.9634      | 8.8652           | 0.8742     | 3.79           | 2.84       | 11.28      |
|                    | PLS MLDF     | 1.1789      | 10.8460          | 0.8116     | 4.63           | 2.32       | 9.22       |
|                    | PLS-SVR VNIR | 1.192       | 10.9649          | 0.8074     | 4.68           | 2.3        | 9.12       |
|                    | PLS VNIR     | 1.3558      | 12.4688          | 0.7508     | 5.33           | 2.02       | 8.02       |
|                    | PLS-SVR SWIR | 1.5016      | 13.8122          | 0.6943     | 5.9            | 1.82       | 7.24       |
|                    | PLS SWIR     | 2.2694      | 20.8768          | 0.3018     | 8.92           | 1.21       | 4.79       |
| Etapas             | Modelo       | $RMSE_P$    | $NRMSE_P$ (%)    | $R^2_P$    | $REP_P$ (%)    | $RPD_P$    | $RER_P$    |
| Predicción         | PLS-SVR MLDF | 1.1555      | 10.8696          | 0.8711     | 4.49           | 2.81       | 9.2        |
|                    | PLS MLDF     | 1.4838      | 13.9470          | 0.7874     | 5.76           | 2.19       | 7.17       |
|                    | PLS-SVR VNIR | 1.5873      | 14.9254          | 0.7567     | 6.16           | 2.05       | 6.7        |
|                    | PLS VNIR     | 1.6619      | 15.6250          | 0.7333     | 6.45           | 1.95       | 6.4        |
|                    | PLS-SVR SWIR | 1.9197      | 18.0505          | 0.6442     | 7.45           | 1.69       | 5.54       |
|                    | PLS SWIR     | 2.8694      | 26.9542          | 0.2051     | 11.14          | 1.13       | 3.71       |

**Tabla 4 Resultados de las Métricas de Desempeño (RMSE, NRMSE, R<sup>2</sup>, REP%, RPD, RER) para la Estimación de %Calcopirita según Modelo, Sensor y Etapa de Evaluación.**

| %Calcopirita       |              |             | Metricas         |            |                |            |            |
|--------------------|--------------|-------------|------------------|------------|----------------|------------|------------|
| Etapas             | Modelo       | $RMSE_c$    | $NRMSE_c$ (%)    | $R^2_c$    | $REP_c$ (%)    | $RPD_c$    | $RER_c$    |
| Calibración        | PLS-SVR MLDF | 1.9395      | 3.4282           | 0.9749     | 3.3            | 6.33       | 29.17      |
|                    | PLS MLDF     | 4.3525      | 7.6923           | 0.8737     | 7.4            | 2.82       | 13         |
|                    | PLS-SVR VNIR | 3.5012      | 9.3110           | 0.8982     | 5.9            | 3.14       | 10.74      |
|                    | PLS VNIR     | 5.2726      | 9.3197           | 0.8147     | 8.96           | 2.33       | 10.73      |
|                    | PLS-SVR SWIR | 5.2918      | 9.3545           | 0.8133     | 8.99           | 2.32       | 10.69      |
|                    | PLS SWIR     | 7.4589      | 13.1926          | 0.6292     | 12.68          | 1.65       | 7.58       |
| Etapas             | Modelo       | $RMSE_{cv}$ | $NRMSE_{cv}$ (%) | $R^2_{cv}$ | $REP_{cv}$ (%) | $RPD_{cv}$ | $RER_{cv}$ |
| CV                 | PLS-SVR MLDF | 2.8563      | 5.0480           | 0.9456     | 4.85           | 4.3        | 19.81      |
|                    | PLS MLDF     | 4.7927      | 8.4746           | 0.8469     | 8.15           | 2.56       | 11.8       |
|                    | PLS-SVR VNIR | 3.6957      | 9.8328           | 0.8866     | 6.22           | 2.98       | 10.17      |
|                    | PLS VNIR     | 5.9378      | 10.4932          | 0.765      | 10.09          | 2.07       | 9.53       |
|                    | PLS-SVR SWIR | 6.0084      | 10.6157          | 0.7594     | 10.21          | 2.04       | 9.42       |
|                    | PLS SWIR     | 7.8571      | 13.8889          | 0.5885     | 13.35          | 1.56       | 7.2        |
| Etapas             | Modelo       | $RMSE_v$    | $NRMSE_v$ (%)    | $R^2_v$    | $REP_v$ (%)    | $RPD_v$    | $RER_v$    |
| Validación externa | PLS-SVR MLDF | 3.1446      | 8.6505           | 0.8998     | 5.27           | 3.19       | 11.56      |
|                    | PLS MLDF     | 4.782       | 13.1579          | 0.7682     | 8.02           | 2.1        | 7.6        |
|                    | PLS-SVR VNIR | 6.1876      | 12.1065          | 0.7817     | 10.23          | 2.16       | 8.26       |
|                    | PLS VNIR     | 5.4715      | 15.0602          | 0.6965     | 9.17           | 1.83       | 6.64       |
|                    | PLS-SVR SWIR | 6.117       | 16.8350          | 0.6207     | 10.26          | 1.64       | 5.94       |
|                    | PLS SWIR     | 5.5712      | 15.3374          | 0.6854     | 9.34           | 1.8        | 6.52       |
| Etapas             | Modelo       | $RMSE_p$    | $NRMSE_p$ (%)    | $R^2_p$    | $REP_p$ (%)    | $RPD_p$    | $RER_p$    |
| Predicción         | PLS-SVR MLDF | 3.2722      | 9.5238           | 0.8998     | 5.29           | 3.19       | 10.5       |
|                    | PLS MLDF     | 5.4796      | 15.9490          | 0.719      | 8.86           | 1.9        | 6.27       |
|                    | PLS-SVR VNIR | 6.4028      | 11.1732          | 0.7546     | 10.89          | 2.04       | 8.95       |
|                    | PLS VNIR     | 5.8053      | 16.8919          | 0.6846     | 9.39           | 1.8        | 5.92       |
|                    | PLS-SVR SWIR | 6.8914      | 20.0401          | 0.5555     | 11.15          | 1.51       | 4.99       |
|                    | PLS SWIR     | 7.7704      | 22.6244          | 0.4349     | 12.57          | 1.34       | 4.42       |

**Tabla 5 Resultados de las Métricas de Desempeño (RMSE, NRMSE, R<sup>2</sup>, REP%, RPD, RER) para la Estimación de %Pirita según Modelo, Sensor y Etapa de Evaluación.**

| %Pirita            |              |                    | Metricas                |                              |                       |                   |                   |
|--------------------|--------------|--------------------|-------------------------|------------------------------|-----------------------|-------------------|-------------------|
| Etapas             | Modelo       | RMSE <sub>c</sub>  | NRMSE <sub>c</sub> (%)  | R <sup>2</sup> <sub>c</sub>  | REP <sub>c</sub> (%)  | RPD <sub>c</sub>  | RER <sub>c</sub>  |
| Calibración        | PLS-SVR MLDF | 2.0096             | 9.3197                  | 0.8612                       | 10.79                 | 2.69              | 10.73             |
|                    | PLS MLDF     | 2.6677             | 12.3609                 | 0.7554                       | 14.32                 | 2.03              | 8.09              |
|                    | PLS-SVR VNIR | 2.2389             | 10.3842                 | 0.8277                       | 12.02                 | 2.42              | 9.63              |
|                    | PLS VNIR     | 2.9577             | 13.7174                 | 0.6993                       | 15.88                 | 1.83              | 7.29              |
|                    | PLS-SVR SWIR | 2.6295             | 12.1951                 | 0.7623                       | 14.12                 | 2.06              | 8.2               |
|                    | PLS SWIR     | 2.9745             | 13.7931                 | 0.6958                       | 15.97                 | 1.82              | 7.25              |
| Etapas             | Modelo       | RMSE <sub>cv</sub> | NRMSE <sub>cv</sub> (%) | R <sup>2</sup> <sub>cv</sub> | REP <sub>cv</sub> (%) | RPD <sub>cv</sub> | RER <sub>cv</sub> |
| CV                 | PLS-SVR MLDF | 2.3363             | 10.8342                 | 0.8124                       | 12.54                 | 2.32              | 9.23              |
|                    | PLS MLDF     | 2.9287             | 13.5685                 | 0.7052                       | 15.72                 | 1.85              | 7.37              |
|                    | PLS-SVR VNIR | 2.4824             | 11.5075                 | 0.7882                       | 13.33                 | 2.18              | 8.69              |
|                    | PLS VNIR     | 3.2759             | 15.1976                 | 0.6311                       | 17.59                 | 1.65              | 6.58              |
|                    | PLS-SVR SWIR | 2.8269             | 13.1062                 | 0.7253                       | 15.18                 | 1.91              | 7.63              |
|                    | PLS SWIR     | 3.287              | 15.2439                 | 0.6286                       | 17.65                 | 1.65              | 6.56              |
| Etapas             | Modelo       | RMSE <sub>v</sub>  | NRMSE <sub>v</sub> (%)  | R <sup>2</sup> <sub>v</sub>  | REP <sub>v</sub> (%)  | RPD <sub>v</sub>  | RER <sub>v</sub>  |
| Validación externa | PLS-SVR MLDF | 2.3422             | 11.5207                 | 0.8046                       | 13.28                 | 2.28              | 8.68              |
|                    | PLS MLDF     | 2.6398             | 12.9870                 | 0.7518                       | 14.97                 | 2.03              | 7.7               |
|                    | PLS-SVR VNIR | 2.5942             | 12.7714                 | 0.7603                       | 14.71                 | 2.06              | 7.83              |
|                    | PLS VNIR     | 3.2559             | 16.0256                 | 0.6224                       | 18.46                 | 1.64              | 6.24              |
|                    | PLS-SVR SWIR | 3.1101             | 15.3139                 | 0.6555                       | 17.63                 | 1.72              | 6.53              |
|                    | PLS SWIR     | 2.4729             | 12.1655                 | 0.7822                       | 14.02                 | 2.16              | 8.22              |
| Etapas             | Modelo       | RMSE <sub>p</sub>  | NRMSE <sub>p</sub> (%)  | R <sup>2</sup> <sub>p</sub>  | REP <sub>p</sub> (%)  | RPD <sub>p</sub>  | RER <sub>p</sub>  |
| Predicción         | PLS-SVR MLDF | 2.4581             | 10.1523                 | 0.802                        | 13.01                 | 2.27              | 9.85              |
|                    | PLS MLDF     | 2.8391             | 11.7233                 | 0.7358                       | 15.02                 | 1.96              | 8.53              |
|                    | PLS-SVR VNIR | 3.5032             | 14.4718                 | 0.5978                       | 18.54                 | 1.59              | 6.91              |
|                    | PLS VNIR     | 4.2691             | 17.6367                 | 0.4027                       | 22.59                 | 1.31              | 5.67              |
|                    | PLS-SVR SWIR | 3.2728             | 13.5135                 | 0.649                        | 17.32                 | 1.7               | 7.4               |
|                    | PLS SWIR     | 3.4785             | 14.3678                 | 0.6034                       | 18.41                 | 1.6               | 6.96              |

### 5.13. Análisis de desempeño para la estimación de %Cu

La Tabla 3 muestra en detalle las métricas de desempeño obtenidas por los distintos modelos desarrollados para predecir la concentración de cobre (%Cu), abarcando tres configuraciones de entrada espectral (VNIR, SWIR y MLDF) y dos estrategias de modelado (PLS lineal y PLS–SVR no lineal). Las métricas reportadas ***RMSE***, ***NRMSE***, ***R<sup>2</sup>***, ***REP%***, ***RPD*** y ***RER*** permiten evaluar el ajuste, la precisión relativa y la capacidad predictiva de cada modelo en diferentes etapas de validación: calibración, validación cruzada tipo Leave-One-Out (LOO), validación externa y conjunto de predicción.

De estas etapas, el conjunto de **predicción** formado por el 20% de las muestras reservadas exclusivamente para evaluación final representa una estimación directa y rigurosa del comportamiento esperado del modelo frente a datos realmente no vistos durante su construcción. Por ello, este conjunto ofrece la referencia más objetiva al momento de contrastar la calidad predictiva de las distintas configuraciones.

#### 5.13.1 Desempeño %Cu con fusión de datos (MLDF)

En este escenario, el modelo **PLS–SVR MLDF** se posiciona como el más efectivo, alcanzando en la etapa de predicción un ***RMSE<sub>P</sub>*** de **1.16**, un coeficiente de determinación de **0.87**, un ***REP%*** de **4.49**, ***RPD*** de **2.81** y un ***RER*** de **9.20**. Estos valores reflejan una combinación favorable de precisión y robustez, superando ampliamente los umbrales comúnmente aceptados para modelos fiables en contextos de análisis mineralógico.

La versión lineal del mismo enfoque, **PLS MLDF**, presenta un rendimiento algo menor: ***RMSE<sub>P</sub>*** de **1.48**, ***R<sup>2</sup><sub>P</sub>*** de **0.79**, y ***RER*** de **7.17**. Aunque el modelo sigue siendo competitivo, los resultados evidencian la limitación del enfoque lineal para capturar relaciones espectrales no lineales presentes entre las señales VNIR y SWIR, especialmente cuando estas son combinadas. La diferencia de aproximadamente 0.31 en ***RMSE<sub>P</sub>*** y casi 2 puntos en ***RER*** entre ambos modelos refleja una ganancia sustantiva en rendimiento atribuible a la incorporación de no linealidad mediante SVR.

Desde la perspectiva del ***NRMSE<sub>P</sub>***, este modelo presenta un valor de **10.87%**, lo que indica un error bajo en relación con el rango composicional del sistema, confirmando su adecuada capacidad para aplicaciones cuantitativas.

La versión lineal PLS MLDF presenta un  $NRMSE_P$  de 13.95%, que, si bien sigue dentro de un rango aceptable, evidencia una menor precisión relativa. La diferencia entre ambos modelos refleja que la incorporación de no linealidad mediante SVR no solo reduce el error absoluto, sino que también mejora su comportamiento relativo respecto a la variabilidad del sistema, reforzando la ventaja del enfoque híbrido.

### 5.13.2 Desempeño %Cu con sensores individuales (VNIR y SWIR)

En los modelos construidos exclusivamente con espectros VNIR, se observa nuevamente una mejora clara al incorporar SVR. El modelo **PLS–SVR VNIR** logra un  $RMSE_P$  de **1.59**,  $R^2$  de **0.76** y  $RER_P$  de **6.70**, mientras que su contraparte lineal **PLS VNIR** obtiene valores inferiores ( $RMSE_P$  de **1.66**,  $R^2_P$  de **0.73**,  $RER_P$  de **6.40**). Si bien las diferencias aquí son más estrechas que en MLDF, el comportamiento es consistente y reafirma la capacidad del modelo no lineal para ajustarse mejor a patrones espectrales complejos.

Por su parte, en el dominio SWIR se constata la mayor brecha de rendimiento entre modelos. El modelo **PLS–SVR SWIR** mejora significativamente respecto a **PLS SWIR** con un  $RMSE_P$  de **1.91** versus **2.87**, y un  $RER_P$  de **5.54** frente a **3.71**, evidenciando que el dominio SWIR, por sí solo, es menos informativo para estimar %Cu, pero puede ser potenciado mediante técnicas no lineales. En este caso, el enfoque lineal PLS demuestra limitaciones claras, con métricas que apenas alcanzan umbrales mínimos de aceptabilidad.

Desde la perspectiva del error relativo, el análisis mediante  $NRMSE$  permite reforzar estas observaciones. En el caso del **VNIR**, el modelo **PLS–SVR** presenta un  $NRMSE_P$  de **14.93%**, manteniéndose dentro del rango de error bajo, aunque cercano a su límite superior, mientras que el modelo **PLS** alcanza un **15.63%**, indicando un desempeño levemente inferior. En contraste, en el dominio **SWIR** los valores aumentan considerablemente, alcanzando **18.05%** para **PLS–SVR** y **26.95%** para **PLS**, lo que evidencia una pérdida significativa de precisión relativa cuando se utiliza esta región de forma aislada. Estos resultados confirman que, si bien la incorporación de no linealidad mejora sistemáticamente el desempeño en ambos dominios, la principal limitación radica en la cantidad y calidad de la información espectral disponible, siendo el **VNIR** más informativo que el **SWIR** para la estimación de %Cu.

En este contexto, el análisis sugiere que la mejora observada en los modelos fusionados no se debe únicamente al tipo de algoritmo, sino principalmente a la integración de información espectral complementaria.

#### **5.14. Análisis de desempeño para la estimación de %Calcopirita**

La Tabla 4 presenta los resultados de predicción para %Calcopirita bajo las distintas combinaciones de sensores (VNIR, SWIR, MLDF) y enfoques de modelado (PLS y PLS–SVR).

##### **5.14.1 Desempeño %Calcopirita con fusión de datos (MLDF)**

El modelo **PLS–SVR MLDF** nuevamente ofrece el mejor rendimiento global, con un  $RMSEP$  de **3.27**,  $R^2_p$  de **0.90** y un  $RER_p$  de **10.5**, lo que indica una predicción precisa y generalizable. Su contraparte lineal (**PLS MLDF**) muestra un desempeño considerablemente inferior ( $RMSEP$  de **5.48**,  $R^2_p$  de **0.72**,  $RER_p$  de **6.27**), resaltando los beneficios de la no linealidad en la regresión.

Desde la perspectiva del error relativo, el modelo **PLS–SVR MLDF** presenta un  $NRMSEP$  de 9.52%, lo que indica un error bajo respecto al rango composicional de la calcopirita, confirmando su alta capacidad predictiva. En contraste, el modelo **PLS MLDF** alcanza un  $NRMSEP$  de 15.95%, situándose en el umbral de aceptabilidad y evidenciando una pérdida de precisión relativa significativa. Esta diferencia refuerza que la incorporación de no linealidad mediante SVR no solo mejora el ajuste global, sino que también reduce el error en términos relativos al rango del sistema.

##### **5.14.2 Desempeño %Calcopirita con sensores individuales (VNIR y SWIR)**

En configuraciones con sensores individuales, **VNIR** supera a **SWIR** en ambos enfoques. El modelo **PLS–SVR VNIR** obtiene un  $RER_p$  de **8.95**, frente a **4.99** del modelo **PLS–SVR SWIR**. En general, los modelos **PLS–SVR** mantienen una ventaja consistente sobre los modelos PLS en todos los dominios, aunque esta ventaja se reduce cuando los datos espectrales son menos informativos, como en el caso de **SWIR**.

En las diferencias observadas entre sensores individuales. En el caso del **VNIR**, el modelo PLS–SVR alcanza un  $NRMSE_P$  de 11.17%, indicando un error bajo y confirmando que esta región espectral contiene información relevante para la estimación de la calcopirita. Sin embargo, su contraparte lineal presenta un  $NRMSE_P$  de 16.89%, evidenciando una disminución en la precisión relativa. Por su parte, en el dominio SWIR se observan los mayores errores relativos, con valores de  $NRMSE_P$  de 20.04% para PLS–SVR y 22.62% para PLS, lo que indica un desempeño limitado cuando esta región se utiliza de forma aislada.

Estos resultados confirman que, si bien el modelo no lineal mejora consistentemente el desempeño en ambos dominios, la principal limitación radica en la información espectral disponible, siendo el VNIR más informativo que el SWIR para la estimación de calcopirita. En consecuencia, la mejora observada en el enfoque MLDF se explica principalmente por la integración de información espectral complementaria, más que por el tipo de modelo utilizado de forma aislada.

### 5.15. Análisis de desempeño para la estimación de %Pirita

La Tabla 5 muestra los resultados obtenidos para %Pirita. Como en los casos anteriores, el foco se sitúa en el conjunto de **predicción**, que permite evaluar el comportamiento real de los modelos ante muestras completamente nuevas.

#### 5.15.1 Desempeño %Pirita con fusión de datos (MLDF)

El modelo **PLS–SVR MLDF** alcanza el mejor rendimiento general, con un  $RMSE_P$  de **2.46**,  $R^2_p$  de **0.80** y un  $RER_P$  de **9.85**. En términos relativos, este modelo presenta un  $NRMSE_P$  de 10.15%, lo que indica un error bajo en relación con el rango composicional de la pirita y confirma su adecuada capacidad de generalización. En comparación, el modelo **PLS MLDF** presenta un rendimiento notablemente inferior ( $RMSE_P$  de **2.83**,  $R^2_p$  de **0.74**,  $RER_P$  de **8.53**), con un  $NRMSE_P$  de 11.72%, evidenciando una menor precisión relativa. Estos resultados confirman que la incorporación de no linealidad mediante SVR, junto con la integración de información espectral, permite capturar de mejor forma la variabilidad de la pirita en muestras complejas.

### 5.15.2 Desempeño %Pirita con sensores individuales (VNIR y SWIR)

A diferencia de los minerales anteriores, los modelos para %Pirita presentan un comportamiento más limitado al utilizar sensores individuales. En el modelo **PLS–SVR SWIR** se obtiene un  $RMSEP$  de **3.27** y  $R^2_p$  de **0.65**, con un  $NRMSEP$  de 13.51%, lo que indica un error moderado en relación con el rango de variación superando a su versión lineal, pero quedando lejos del desempeño observado con fusión MLDF. Los modelos basados en **VNIR**, en especial PLS, muestran el peor comportamiento: **PLS VNIR** alcanza apenas un  $R^2_p$  de **0.40** y un  $RER_p$  de **5.67**, lo que indica una pobre capacidad de predicción.

el modelo **PLS–SVR VNIR** presenta un  $NRMSEP$  de 14.47%, mientras que el modelo lineal **PLS VNIR** alcanza un  $NRMSEP$  de 17.64%, reflejando un error elevado y una baja capacidad predictiva. En general, los valores de NRMSE superiores a ~15% evidencian dificultades para modelar adecuadamente la variabilidad de la pirita utilizando información espectral individual, lo que refuerza la necesidad de integrar múltiples fuentes espectrales.

### 5.16. Visualización del Desempeño Predictivo: Mejores Modelos para %Cu, %Calcopirita y %Pirita

A continuación, se presentan las gráficas de los mejores modelos para predecir %Cu, %Calcopirita y %Pirita. Estas muestran la relación entre los valores reales y predichos en el conjunto de predicción, incluyendo las métricas de desempeño ( $RMSE$ ,  $R^2$ ,  $REP\%$ ,  $RPD$ ,  $RER$ ) y los intervalos de incertidumbre obtenidos con Jackknife+.

Estas visualizaciones permiten apreciar gráficamente el ajuste, la precisión y la fiabilidad alcanzada por los modelos seleccionados. Las gráficas correspondientes al resto de las configuraciones evaluadas (modelos alternativos y sensores individuales) se incluyen en el Anexo para referencia complementaria.

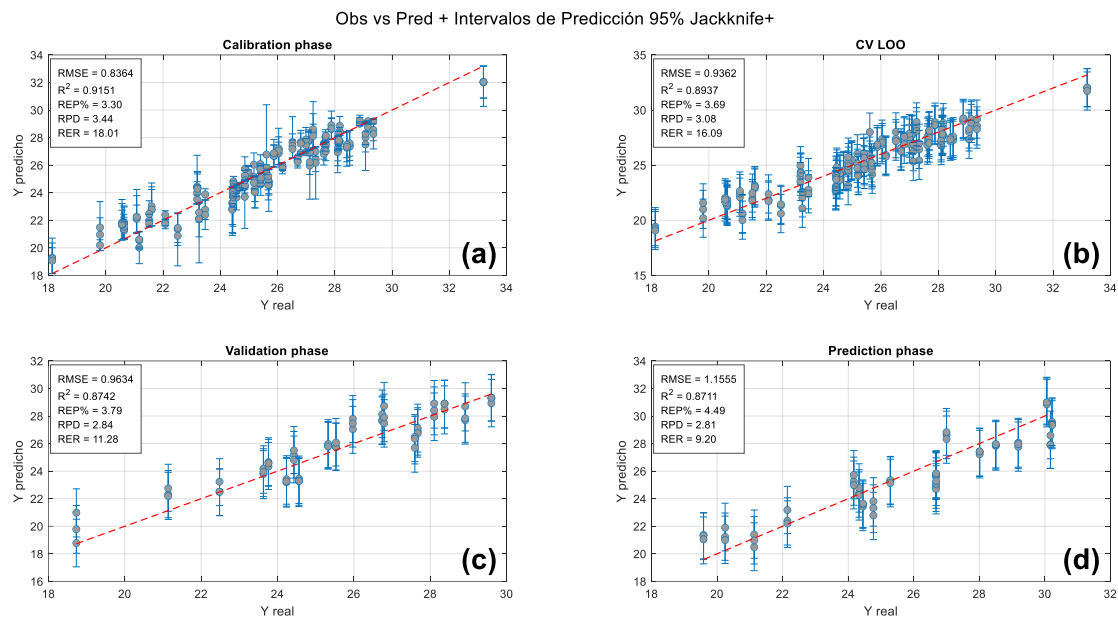


**Figura 5.16. Desempeño del modelo PLS-SVR (MLDF) para %Cu en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.**

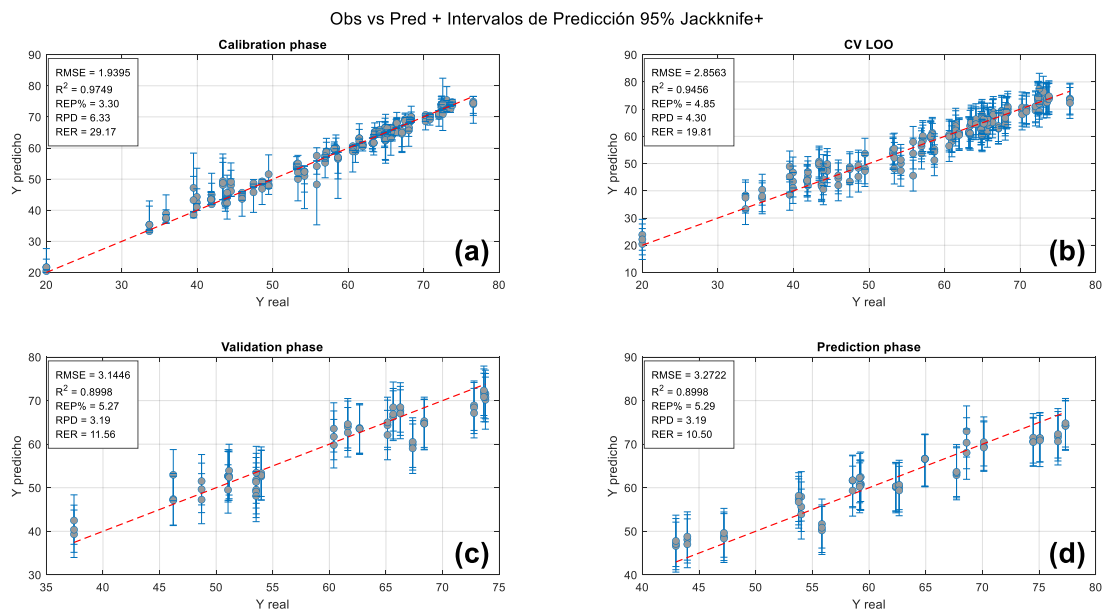


**Figura 5.17. Desempeño del modelo PLS-SVR (MLDF) para %Calcopirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.**

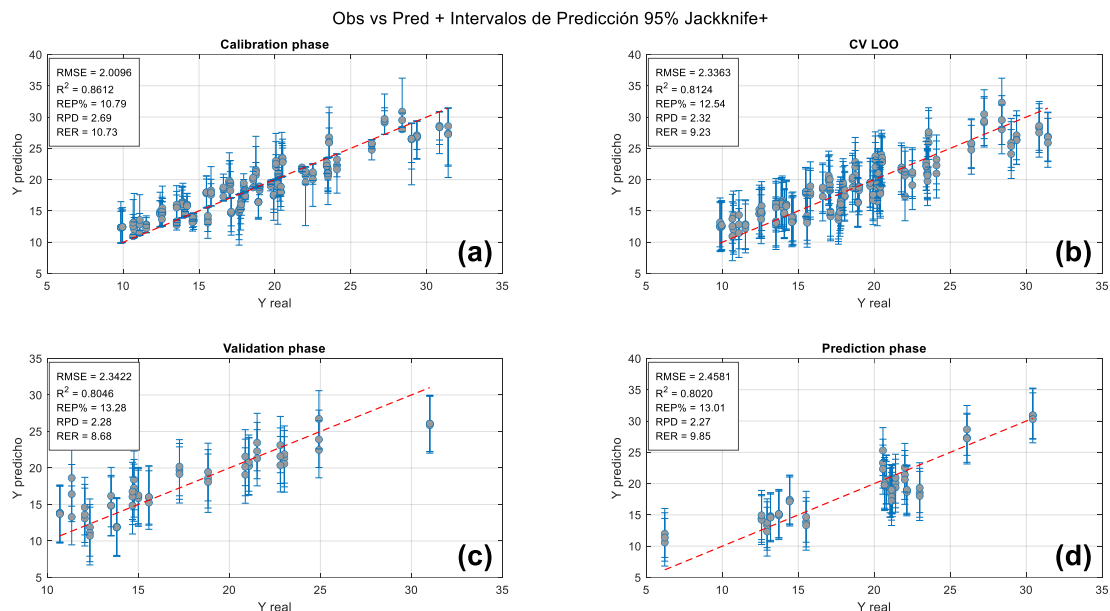


**Figura 5.18. Desempeño del modelo PLS-SVR (MLDF) para %Pirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.**

Las Figuras 5.15, 5.17 y 5.18 ilustran el desempeño del modelo PLS-SVR utilizando la configuración MLDF para predecir el contenido de %Cu, %Calcopirita y %Pirita, respectivamente, en las cuatro etapas de evaluación (calibración, validación cruzada, validación externa y predicción). En todos los casos, las gráficas muestran una fuerte correlación entre los valores reales y los valores predichos, con pendientes próximas a la diagonal ideal y dispersión controlada de los residuales, especialmente en las primeras dos fases. Para %Cu y %Calcopirita, los resultados son especialmente robustos, con valores altos de  $R^2$  y bajos errores (RMSE y REP%), lo que reafirma la capacidad del modelo para capturar con precisión sus patrones espectrales, incluso con variabilidad entre muestras.

En el caso de la Pirita, aunque el modelo conserva una tendencia general positiva, la dispersión en torno a la línea de identidad es visiblemente mayor, sobre todo en validación externa y predicción. Esto sugiere una mayor dificultad del modelo para generalizar adecuadamente para este mineral, debido a su menor representatividad en la mezcla mineralógica y a su menor asociación directa con las bandas espectrales dominantes.

No obstante, el modelo aún logra métricas aceptables, manteniendo niveles razonables de precisión y coherencia. Cabe señalar que las gráficas presentadas corresponden únicamente al modelo PLS-SVR con MLDF, por ser el de mejor rendimiento general; las visualizaciones del resto de los modelos evaluados pueden encontrarse en el Anexo correspondiente.

#### **5.17. Comparacion general de todos los modelos**

Al observar globalmente el rendimiento de los modelos en las tres variables objetivo, se confirma una tendencia clara: los modelos que incorporan regresión no lineal (PLS–SVR) superan sistemáticamente a los modelos lineales tradicionales (PLS) en todas las configuraciones de sensor y etapas de evaluación. Esta ventaja es especialmente evidente cuando se combina con estrategias de fusión de datos como MLDF, que permiten integrar información espectral complementaria de los rangos VNIR y SWIR. En todas las variables, los modelos PLS–SVR MLDF logran las mejores métricas de desempeño, tanto en precisión (RMSE, REP%) como en capacidad explicativa ( $R^2$ ), destacándose también en indicadores normalizados como RPD y RER. Esta superioridad no se limita a la etapa de calibración, sino que se mantiene en la validación cruzada, validación externa y, lo más importante, en el conjunto de predicción, el cual representa de manera más realista el comportamiento del modelo frente a datos completamente nuevos.

Por otro lado, el análisis comparativo entre las variables objetivo también proporciona información clave sobre la facilidad de modelado espectral en función de la naturaleza mineralógica del sistema. Tanto el %Cu como la calcopirita se modelaron con alta precisión y estabilidad, evidenciando valores elevados de  $R^2$  y RPD, y errores relativamente bajos en todas las etapas. Esto se explica en parte porque la calcopirita es la fase mineral dominante en el sistema, lo que refuerza su señal espectral, y porque el cobre se encuentra en forma casi exclusiva como parte de la calcopirita, lo que confiere al %Cu una representación más robusta y menos ambigua en el espectro. Esta correspondencia directa entre la señal espectral y la propiedad objetivo facilita enormemente el modelado y se refleja en la notable similitud de comportamiento entre ambas variables, especialmente en configuraciones MLDF.

En cambio, la predicción de %Pirita resultó más desafiante. Aunque la pirita estuvo presente en las muestras, su rol como fase mineral secundaria y su mayor variabilidad en distribución, textura y granulometría introdujeron complejidad adicional en su modelado.

Esto se traduce en un descenso generalizado en las métricas de desempeño, especialmente para los modelos lineales y aquellos que utilizan espectros SWIR de forma aislada. Incluso los mejores modelos PLS–SVR con fusión MLDF, si bien logran mantener un desempeño aceptable, muestran mayor error y menor capacidad explicativa comparado con las otras dos variables. Esta diferencia resalta cómo la estructura físico-química y mineralógica del sistema, junto con el diseño del modelo, influye directamente en la predictibilidad espectral de cada variable, evidenciando la necesidad de enfoques no lineales y fusionados para abordar adecuadamente sistemas complejos y multivariados.

### **5.18. Análisis transversal del desempeño y capacidad de generalización.**

Más allá de las diferencias puntuales entre configuraciones y variables objetivo, el análisis conjunto de las cuatro etapas de evaluación permite extraer conclusiones relevantes respecto a la estabilidad y capacidad de generalización de los modelos desarrollados. En las tres variables analizadas (%Cu, %Calcopirita y %Pirita), los modelos PLS–SVR con fusión de datos MLDF muestran una degradación progresiva y controlada del desempeño al transitar desde calibración hacia predicción, evidenciada por incrementos moderados en ***RMSE*** y descensos graduales en ***R<sup>2</sup>***, ***RPD*** y ***RER***. Este comportamiento es indicativo de modelos bien regularizados, capaces de generalizar adecuadamente frente a datos no vistos, y contrasta con los enfoques mono-sensor y lineales, donde la pérdida de desempeño entre etapas es más abrupta.

La validación cruzada tipo LOO actúa como una etapa intermedia coherente entre calibración y validación externa, confirmando que las configuraciones seleccionadas no dependen de particiones favorables del conjunto de datos. En particular, los modelos MLDF mantienen métricas consistentes entre CV y validación externa, lo que refuerza la robustez de la estrategia de fusión a nivel medio y la efectividad del proceso de optimización de hiperparámetros. En contraste, los modelos basados únicamente en SWIR muestran una mayor sensibilidad al cambio de dominio, especialmente en validación externa y predicción, lo que sugiere una menor capacidad intrínseca de este rango espectral para capturar de forma aislada la variabilidad asociada al contenido de cobre y minerales sulfurados.

Desde una perspectiva comparativa, la ventaja del enfoque no lineal (PLS–SVR) se vuelve más evidente a medida que aumenta la complejidad del problema. Para %Cu, donde la relación espectral es más directa, PLS y PLS–SVR presentan comportamientos cercanos en calibración y CV, aunque SVR mantiene ventajas claras en predicción. En %Calcopirita, esta diferencia se amplifica, y en

%Pirita se vuelve crítica, confirmando que la no linealidad resulta esencial para modelar relaciones espectrales más débiles, solapadas o indirectas. Estos resultados demuestran que la combinación de fusión MLDF y regresión no lineal no solo mejora el ajuste, sino que constituye un factor clave para asegurar estabilidad y aplicabilidad práctica en escenarios reales.

## 6. Conclusiones

---

### 6.1. Sumario

El presente trabajo abordó la cuantificación mineralógica y elemental de concentrados de cobre mediante el uso de imágenes hiperespectrales adquiridas en los rangos VNIR (400–1000 nm) y SWIR (900–1700 nm), integradas a través de una estrategia de fusión de datos a nivel medio (Mid-Level Data Fusion, MLDF). Se desarrolló una metodología completa que abarcó la adquisición controlada de datos hiperespectrales mediante cámaras de línea, la calibración espectral y radiométrica del sistema, el preprocesamiento avanzado de los espectros (Savitzky–Golay, MSC y autoscaling), la selección de variables relevantes mediante criterios VIP y la construcción de modelos de regresión multivariada.

Sobre esta base, se implementaron y compararon enfoques de modelado lineales (PLS) y no lineales (PLS-SVR), considerando configuraciones mono-sensor (VNIR y SWIR) y multiesensor (VNIR+SWIR mediante MLDF), aplicados a la estimación de %Cu, %Calcopirita y %Pirita. Los modelos fueron evaluados mediante un esquema de validación riguroso que incluyó calibración, validación cruzada Leave-One-Out, validación externa y un conjunto de predicción independiente, incorporando además intervalos de incertidumbre Jackknife+ para la evaluación de la confiabilidad de las predicciones individuales. **Los resultados evidenciaron que la incorporación de no linealidad mediante PLS-SVR mejora de forma consistente el desempeño respecto a PLS, mientras que la integración de información espectral mediante fusión MLDF permite capturar de manera más completa la variabilidad del sistema.** En conjunto, este enfoque permitió demostrar la **estabilidad, capacidad de generalización y ventajas comparativas** tanto de la regresión no lineal como de la fusión multiesensor en problemas de cuantificación mineralógica.

Si bien el estudio se desarrolló bajo condiciones de laboratorio controladas y con un conjunto de muestras de tamaño acotado especialmente en el caso de minerales secundarios como la pirita, la metodología propuesta fue concebida desde su diseño con criterios de escalabilidad y transferencia a escenarios operacionales reales. El uso de sensores hiperespectrales de línea, junto con estrategias de preprocesamiento, fusión de datos a nivel medio y modelos de regresión computacionalmente eficientes, resulta plenamente compatible con sistemas de adquisición continua y procesamiento en tiempo real o cuasi en tiempo real.

**Esta característica es particularmente relevante considerando que la mejora observada no depende únicamente del modelo, sino también de la calidad y complementariedad de la información espectral utilizada.** En este sentido, los resultados obtenidos no solo validan la viabilidad técnica del enfoque para la caracterización cuantitativa de concentrados de cobre, sino que **permiten identificar los factores que controlan el desempeño predictivo.** En particular, se observó que **la fusión de datos VNIR–SWIR genera mejoras sistemáticas en comparación con el uso de sensores individuales,** evidenciadas por la **reducción del error relativo (NRMSE)** y el aumento de métricas como **RPD y RER.** Por otra parte, **la incorporación de modelos no lineales (PLS-SVR) actúa como un mecanismo de refinamiento,** permitiendo capturar relaciones espectrales complejas y mejorar el ajuste de los modelos. De este modo, se concluye que **el mayor impacto en la calidad de las predicciones proviene de la integración de información espectral complementaria,** mientras que **el uso de modelos no lineales potencia este efecto.** En conjunto, **la combinación de fusión de datos y modelado no lineal constituye la estrategia más robusta** para la cuantificación mineralógica en sistemas complejos, estableciendo una base metodológica sólida para su futura implementación en entornos industriales.

## **6.2. Conclusiones**

A partir del desarrollo metodológico y del análisis detallado de los resultados obtenidos en las distintas etapas de evaluación, es posible establecer un conjunto de conclusiones que dan respuesta directa tanto al objetivo general como a los objetivos específicos planteados en esta investigación. Estas conclusiones se fundamentan en la comparación sistemática entre enfoques de regresión lineales y no lineales, configuraciones mono-sensor y multiesensor, y variables objetivo de distinta complejidad mineralógica, permitiendo evaluar de forma objetiva el grado de cumplimiento de los objetivos propuestos y el alcance de la metodología desarrollada.

1. Se desarrolló una metodología experimental completa, reproducible y coherente para la cuantificación mineralógica y elemental de concentrados de cobre, integrando preparación de muestras con referencia mineralógica tipo *gold standard*, adquisición hiperespectral controlada y modelado quimiométrico, estableciendo una base sólida para la evaluación comparativa de modelos y configuraciones espectrales.
2. El análisis exploratorio y el preprocesamiento espectral permitieron caracterizar la variabilidad de los datos y reducir efectos físicos no deseados, validando el uso combinado de técnicas

como filtrado Savitzky–Golay, corrección de dispersión (MSC), autoescalado y selección de variables mediante VIP como etapa fundamental previa al modelado.

3. En relación con el objetivo de evaluación en dominios espectrales individuales, se estableció que los modelos construidos sobre información VNIR presentan, en general, un mejor desempeño que aquellos basados únicamente en SWIR, evidenciando diferencias en la relevancia de cada región espectral para la cuantificación de las variables estudiadas.
4. La implementación de estrategias de fusión de datos a nivel medio (MLDF) permitió integrar de forma efectiva la información espectral proveniente de VNIR y SWIR, generando representaciones más completas del sistema y capturando la complementariedad entre ambas regiones espectrales.
5. La comparación entre configuraciones mono-sensor y multiesensor evidenció que la fusión de datos (MLDF) produce **mejoras sistemáticas en el desempeño predictivo**, reflejadas en la reducción del error relativo (NRMSE) y el aumento de métricas como RPD y RER, confirmando que la integración de información espectral constituye un factor determinante en la calidad de las predicciones.
6. En cuanto a la influencia del tipo de modelo de regresión, se demostró que los modelos no lineales (PLS-SVR) presentan un desempeño consistentemente superior a los modelos lineales (PLS), especialmente en escenarios donde las relaciones entre variables espectrales y composición presentan comportamientos no lineales.
7. No obstante, el análisis conjunto de resultados indica que **el mayor impacto en el desempeño predictivo proviene de la incorporación de información espectral complementaria mediante fusión de datos**, mientras que el uso de modelos no lineales actúa como un mecanismo de refinamiento del ajuste, potenciando pero no reemplazando el efecto de la integración multiesensor.
8. Para la estimación de %Cu y %Calcopirita, se obtuvieron los mejores resultados del estudio, con modelos robustos y estables en el conjunto de predicción, lo que se explica por la fuerte relación mineralógica entre ambas variables y la predominancia de fases cupríferas en las muestras analizadas.
9. La cuantificación de %Pirita representó el mayor desafío debido a su carácter de fase secundaria y menor expresión espectral; sin embargo, la combinación de fusión MLDF y regresión no lineal permitió mejorar significativamente el desempeño respecto a

configuraciones individuales, evidenciando la importancia de integrar múltiples fuentes de información en sistemas complejos.

10. La consistencia observada entre calibración, validación cruzada, validación externa y predicción confirma que los modelos desarrollados presentan una adecuada capacidad de generalización, validando el esquema de evaluación y optimización implementado.
11. En conjunto, la metodología propuesta cumple el objetivo general de la investigación y establece un marco robusto, escalable y con alto potencial de transferencia a contextos operacionales, donde la combinación de adquisición hiperespectral, fusión de datos multiesensor y modelado no lineal se posiciona como una estrategia efectiva para la cuantificación mineralógica en concentrados de cobre.

### 6.3. Trabajo Futuro

A partir de los resultados obtenidos y de la metodología desarrollada en este trabajo, se identifican diversas líneas de investigación y desarrollo que permitirían ampliar, consolidar y transferir los aportes alcanzados hacia escenarios de mayor complejidad y relevancia industrial.

Una primera línea de trabajo futuro corresponde a la **extensión de la metodología a condiciones operacionales reales**, evaluando el desempeño de los modelos en sistemas de adquisición continua o cuasi-en tiempo real. Esto incluye la implementación del esquema hiperespectral en flujos dinámicos de concentrado, considerando variaciones propias del proceso tales como cambios en granulometría, humedad, velocidad de transporte y condiciones ambientales, con el fin de validar la robustez del enfoque en entornos industriales.

En segundo lugar, resulta relevante **incrementar la diversidad y representatividad del conjunto de muestras**, incorporando concentrados provenientes de distintos yacimientos, con mayor variabilidad mineralógica y textural. Esto permitiría fortalecer la capacidad de generalización de los modelos y evaluar su comportamiento frente a escenarios más heterogéneos, particularmente en la cuantificación de minerales secundarios como la pirita.

Otra línea de investigación consiste en **explorar esquemas de fusión multiesensor más avanzados**, integrando información proveniente de técnicas complementarias como Raman o LIBS, así como evaluar estrategias de fusión a distintos niveles (low-, mid- y high-level).

Esto podría contribuir a mejorar la cuantificación de fases minerales con señal espectral débil o altamente solapada, ampliando el alcance del sistema de caracterización.

Finalmente, una línea natural de trabajo futuro corresponde a la **implementación y validación de la metodología propuesta en escenarios operacionales reales**, integrando el sistema de adquisición hiperspectral y los modelos desarrollados como herramientas de apoyo al monitoreo y control de procesos metalúrgicos. En este contexto, resulta de interés evaluar el desempeño del enfoque bajo condiciones dinámicas de operación y su capacidad de adaptación frente a cambios en la composición mineralógica del proceso, avanzando hacia sistemas de caracterización espectral continua y de utilidad práctica en la industria minera.

**Una implementación industrial del sistema propuesto podría realizarse en la línea de transporte del concentrado seco posterior a la etapa de secado y previa a su alimentación al horno flash. Esta ubicación permitiría disponer de una superficie representativa del material que ingresa al proceso de fusión, manteniendo condiciones más estables de humedad y facilitando la instalación de sistemas de iluminación controlada y protección ambiental para la cámara hiperspectral.**

# Referencias

- [1] Comisión Chilena del Cobre (COCHILCO). *Proyección de la Producción de Cobre en Chile, Período 2024–2034*. Dirección de Estudios y Políticas Públicas, DEEP 01/2025, Gobierno de Chile, 2025.
- [2] Comisión Chilena del Cobre (COCHILCO). *Mercado de Concentrados e Industria de Fundiciones*. Dirección de Estudios y Políticas Públicas, DEEP 09/2025, Gobierno de Chile, julio 2025.
- [3] Departamento de Ingeniería de Minas, Universidad de Chile, *MI4100 – Fusión (Tecnología): Apuntes de clase*, material docente, 2011. [En línea]. Disponible en: [https://www.ucursos.cl/ingenieria/2011/1/MI4100/1/material\\_docente/bajar?id=340612](https://www.ucursos.cl/ingenieria/2011/1/MI4100/1/material_docente/bajar?id=340612)
- [4] Schlesinger, M. E., King, M. J., Sole, K. C., Davenport, W. G. *Extractive Metallurgy of Copper*, 5th ed. Elsevier, Oxford, UK, 2011.
- [5] Davenport, W. G., King, M. J., Schlesinger, M. E., Biswas, A. K. *Extractive Metallurgy of Copper*. Capítulo: *Teniente Converter*. Elsevier, Oxford, UK, 2011.
- [6] CSIRO. Gottlieb, P., Wilkie, G., Sutherland, D., Ho-Tun, E., Suthers, S., Perera, K., Jenkins, B., Spencer, S., & Butcher, A. (2000). *Using quantitative electron microscopy for process mineralogy applications*. **JOM**, 52(4), 24–25.
- [7] Bish, D. L., & Post, J. E. (1993). *Quantitative mineralogical analysis using the Rietveld full-pattern fitting method*. **American Mineralogist**, 78(9–10), 932–940.
- [8] Lotter, N. O., Whiteman, E., & Bradshaw, D. J. (2011). *Modern process mineralogy: An integrated approach to flowsheet design and optimisation*. **Minerals Engineering**, 24(12), 1229–1237. <https://doi.org/10.1016/j.mineng.2011.06.008>
- [9] Habashi, F. (1997). *Handbook of Extractive Metallurgy*, Vol. II. Wiley-VCH, Weinheim, Germany.
- [10] L. Ni, H. Xu and X. Zhou, "Mineral Identification and Mapping by Synthesis of Hyperspectral VNIR/SWIR and Multispectral TIR Remotely Sensed Data With Different Classifiers," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 3155-3163, 2020, doi: 10.1109/JSTARS.2020.2999057.

- [11] Koerting, F., Asadzadeh, S., Hildebrand, J. C., Savinova, E., Kouzeli, E., Nikolakopoulos, K., Lindblom, D., Koellner, N., Buckley, S. J., Lehman, M., Schläpfer, D., & Micklethwaite, S. (2024). VNIR-SWIR Imaging Spectroscopy for Mining: Insights for Hyperspectral Drone Applications. *Mining*, 4(4), 1013-1057. <https://doi.org/10.3390/mining4040057>
- [12] Shishin, D.; Tripathi, N.; Sineva, S.; Jak, E. *Thermodynamic Modeling and Research for Processing Complex Concentrate Blends in Custom Copper Smelters for Maximum Revenue. Processes*, 2024, 12(12), 2820. DOI: 10.3390/pr12122820.
- [13] Fodor, M.; Matkovits, A.; Benes, E. L.; Jókai, Z. *The Role of Near-Infrared Spectroscopy in Food Quality Assurance: A Review of the Past Two Decades. Foods*, 2024, 13(21), 3501. DOI: 10.3390/foods13213501.
- [14] M. E. Schlesinger, M. J. King, K. C. Sole, and W. G. Davenport, *Extractive Metallurgy of Copper*, 5th ed. Amsterdam, The Netherlands: Elsevier, 2011.
- [15] Wills, B. A., & Finch, J. A. (2016). *Wills' mineral processing technology* (8th ed.). Butterworth-Heinemann
- [16] Coelho, P. A., Sandoval, C., Álvarez, J., Sanhueza, I., Godoy, C., Torres, S., & Sbarbaro, D. (2019). Automatic near-infrared hyperspectral image analysis of copper concentrates. *IFAC-PapersOnLine*, 52(14), 94–98. <https://doi.org/10.1016/j.ifacol.2019.09.170>
- [17] Rencher, Alvin C.; Christensen, William F. (2012), "Chapter 10, Multivariate regression Section 10.1, Introduction", *Methods of Multivariate Analysis*, Wiley Series in Probability and Statistics, vol. 709 (3rd ed.), John Wiley & Sons, p. 19, ISBN 9781118391679.
- [18] P. Rajkumar, N. Wang, G. Elmasry, G.S.V. Raghavan, Y. Gariepy, Studies on banana fruit quality and maturity stages using hyperspectral imaging, *Journal of Food Engineering*, Volume 108, Issue 1, 2012, Pages 194-200, ISSN 0260-8774, <https://doi.org/10.1016/j.jfoodeng.2011.05.002>
- [19] B. Li and L. Sun, "Estimation of Mixed Noise Parameters in Hyperspectral Images Based on Super-pixel Segmentation and Multiple Linear Regression," 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP), Nanjing, China, 2020, pp. 93-96, doi: 10.1109/ICSIP49896.2020.9339315.

- [20] I.T. Jolliffe, A note on the use of principal components in regression, *Journal of the Royal Statistical Society: Series A C.* 31 (1982) 300e303.
- [21] Svante Wold, Michael Sjöström, Lennart Eriksson, PLS-regression: a basic tool of chemometrics, *Chemometrics and Intelligent Laboratory Systems*, Volume 58, Issue 2, 2001, Pages 109-130, ISSN 0169-7439, [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
- [22] Irina Torres, José Manuel Amigo, Chapter 2.8 - An overview of regression methods in hyperspectral and multispectral imaging, Editor(s): José Manuel Amigo, *Data Handling in Science and Technology*, Elsevier, Volume 32, 2019, Pages 205-230, ISSN 0922-3487, ISBN 9780444639776, <https://doi.org/10.1016/B978-0-444-63977-6.00010-9>.
- [23] Fundamental of statistics, H. Lohninger, statistics, data analysis and chemometrics, PLS [http://www.statistics4u.info/fundstat\\_eng/dd\\_pls.html](http://www.statistics4u.info/fundstat_eng/dd_pls.html)
- [24] Acosta Medina, C. D., Castaño Duque, G. A., & Figueroa Flórez, J. A. (2016). Modelos de Regresión PLS Aplicados a Variables Educativas. *Scientia Et Technica*, 21(3),254-263.[fecha de Consulta 10 de Abril de 2024]. ISSN: 0122-1701. Recuperado de: <https://www.redalyc.org/articulo.oa?id=84950585008>.
- [25] de Jong, S. (1993). *SIMPLS: An alternative approach to partial least squares regression*. **Chemometrics and Intelligent Laboratory Systems**, 18(3), 251–263. [https://doi.org/10.1016/0169-7439\(93\)85002-X](https://doi.org/10.1016/0169-7439(93)85002-X).
- [26] Marini, F., Bucci, R., Magrì, A. L. & Magrì, A. D. (2008). *Artificial neural networks in chemometrics: History, examples and perspectives*. **Microchemical Journal**, 88(2). <https://doi.org/10.1016/j.microc.2007.11.008>.
- [27] Howard B. Demuth, Mark H. Beale, Orlando De Jess, and Martin T. Hagan. 2014. *Neural Network Design* (2nd. ed.). Martin Hagan, Stillwater, OK, USA.

- [28] Tuşa, L., Khodadadzadeh, M., Contreras, C., Rafiezadeh Shahi, K., Fuchs, M., Gloaguen, R., & Gutzmer, J. (2020). *Drill-Core Mineral Abundance Estimation Using Hyperspectral and High-Resolution Mineralogical Data*. *Remote Sensing*, **12**(7), 1218. <https://doi.org/10.3390/rs12071218>.
- [29] Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32. Disponible en: <https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>.
- [30] Louppe, Gilles. (2014). Understanding Random Forests: From Theory to Practice. 10.13140/2.1.1570.5928.
- [31] Tuşa, L., Khodadadzadeh, M., Contreras, C., Rafiezadeh Shahi, K., Fuchs, M., Gloaguen, R., & Gutzmer, J. (2020). Drill-Core Mineral Abundance Estimation Using Hyperspectral and High-Resolution Mineralogical Data. *Remote Sensing*, *12*(7), 1218. <https://doi.org/10.3390/rs12071218>.
- [32] Rodger, A., & Laukamp, C. (2022). Quantitative Geochemical Prediction from Spectral Measurements and Its Application to Spatially Dispersed Spectral Data. *Applied Sciences*, *12*(1), 282. <https://doi.org/10.3390/app12010282>.
- [33] Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, *29*(5), 1189-1232.
- [34] Y. Gao, Y. Liu, C. Wang, *et al.*, “3D mineral exploration Cu–Zn targeting with multi-source geoscience datasets in the Weilasituo–Bairendaba district, Inner Mongolia, China,” **Frontiers in Earth Science**, vol. 11, 2023. <https://doi.org/10.3389/feart.2023.1102640>.
- [35] Van Huffel, S. (1991). The Generalized Total Least Squares Problem: Formulation, Algorithm and Properties. In: Golub, G.H., Van Dooren, P. (eds) *Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms*. NATO ASI Series, vol 70. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-75536-1\\_54](https://doi.org/10.1007/978-3-642-75536-1_54).
- [36] H. Pu, Z. Chen, B. Wang, and W. Xia, “Constrained Least Squares Algorithms for Nonlinear Unmixing of Hyperspectral Imagery,” **IEEE Transactions on Geoscience and Remote Sensing**, vol. 53, no. 3, pp. 1287–1303, Mar. 2015. <https://doi.org/10.1109/TGRS.2014.2336858>.

- [37] Awad, M., Khanna, R. (2015). Support Vector Regression. In: Efficient Learning Machines. Apress, Berkeley, CA. [https://doi.org/10.1007/978-1-4302-5990-9\\_4](https://doi.org/10.1007/978-1-4302-5990-9_4)
- [38] W. Wei, Y. Zhang, H. Guo, Y. Liu, and Z. Chen, "Estimation of the dolomite content of carbonate rock outcrops based on spectral knowledge and machine learning," **Frontiers in Earth Science**, vol. 12, 2024, art. no. 1401026. <https://doi.org/10.3389/feart.2024.1401026>
- [39] C. Yuan, Y. Wang, Z. Li, and X. Zhang, "Revealing Shortwave Infrared Spectral Footprints of Mineralization: From Field Sampling to Hyperspectral Imagery," **SSRN Electronic Journal**, 2024. <https://doi.org/10.2139/ssrn.4762430>
- [40] Wani A. A. (2025). Comprehensive review of dimensionality reduction algorithms: challenges, limitations, and innovative solutions. *PeerJ. Computer science*, 11, e3025. <https://doi.org/10.7717/peerj-cs.3025>.
- [41] Cashman, D., Keller, M., Jeon, H., Kwon, B. C., & Wang, Q. (2025). *A critical analysis of the usage of dimensionality reduction in four domains* (arXiv:2503.08836). arXiv. <https://doi.org/10.48550/arXiv.2503.08836>
- [42] Tange RI, Rasmussen MA, Taira E, Bro R. Benchmarking support vector regression against partial least squares regression and artificial neural network: Effect of sample size on model performance. *Journal of Near Infrared Spectroscopy*. 2017;25(6):381-390. doi:[10.1177/0967033517734945](https://doi.org/10.1177/0967033517734945)
- [43] J. Lv and Liankui Dai, "Application of Partial Least Squares Support Vector Machines (PLS-SVM) in Spectroscopy Quantitative Analysis," 2006 6th World Congress on Intelligent Control and Automation, Dalian, China, 2006, pp. 5228-5232, doi: 10.1109/WCICA.2006.1713389.
- [44] C. Gaobo and C. Xiufang, "Combining partial least squares regression and least squares support vector machine for data mining," 2011 International Conference on E-Business and E-Government (ICEE), Shanghai, China, 2011, pp. 1-4, doi: 10.1109/ICEBEG.2011.5881755

- [45] L. Haoguang and P. Yan, "Study on Soft Sensing Technology of Penicillin Fermentation Based on PLS and SVR," 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA), Kristiansand, Norway, 2020, pp. 980-984, doi: 10.1109/ICIEA48937.2020.9248321
- [46] S. Mishra and P. Khanna, "Brain Age Estimation using PCA and PLS components for Healthy controls," 2020 IEEE 4th Conference on Information & Communication Technology (CICT), Chennai, India, 2020, pp. 1-6, doi: 10.1109/CICT51604.2020.9312057
- [47] Rodrigo Fuentes, Danny Luarte, Claudio Sandoval, Ashwin Kumar Myakalwar, Jorge Yáñez, Daniel Sbarbaro, Data fusion of Laser Induced Breakdown Spectroscopy and Diffuse Reflectance for improved analysis of mineral species in copper concentrates, *Minerals Engineering*, Volume 173, 2021, 107193, ISSN 0892-6875, <https://doi.org/10.1016/j.mineng.2021.107193>
- [48] Qi Wang, Jianting Xiao, Ying Li, Yuan Lu, Jinjia Guo, Ye Tian, Lihui Ren, Mid-level data fusion of Raman spectroscopy and laser-induced breakdown spectroscopy: Improving ores identification accuracy, *Analytica Chimica Acta*, Volume 1240, 2023, 340772, ISSN 0003-2670, <https://doi.org/10.1016/j.aca.2022.340772>
- [49] F. Rivas, F. Pérez, C. Sandoval, I. Sanhueza, B. Sepúlveda, J. Yáñez, and S. Torres, "Copper concentrate dual-band joint classification using reflectance hyperspectral images in the VIS-NIR and SWIR bands," *Appl. Opt.* 62, 2970-2977 (2023).
- [50] Rodrigo Fuentes, Danny Luarte, Claudio Sandoval, Ashwin Kumar Myakalwar, Jonnathan Alvarez, Jorge Yáñez, Daniel Sbarbaro, Laser-Induced Breakdown Spectroscopy and Hyperspectral Imaging Data Fusion for improved Mineralogical Analysis of Copper Concentrates, *IFAC-PapersOnLine*, Volume 55, Issue 21, 2022, Pages 85-90, ISSN 2405-8963, <https://doi.org/10.1016/j.ifacol.2022.09.248>
- [51] Olivieri, Alejandro. (2018). Introduction to Multivariate Calibration: A Practical Approach. <https://doi.org/10.1007/978-3-319-97097-4>

- [52] Martens, H. and Geladi, P. (2006). Multivariate Calibration. In Encyclopedia of Statistical Sciences (eds S. Kotz, C.B. Read, N. Balakrishnan, B. Vidakovic and N.L. Johnson). <https://doi.org/10.1002/0471667196.ess1105.pub2>
- [53] Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least-squares procedures. *Analytical Chemistry*, 36(8), 1627–1639. <https://doi.org/10.1021/ac60214a047>.
- [54] R. W. Schafer, "What Is a Savitzky-Golay Filter? " In *IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 111-117, July 2011, <https://doi.org/10.1109/MSP.2011.941097>.
- [55] Isaksson, T., & Næs, T. (1988). The Effect of Multiplicative Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy. *Applied Spectroscopy*, 42(7), 1273–1284. <https://doi.org/10.1366/0003702884429869>
- [56] Nørgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB. Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Applied Spectroscopy*. 2000;54(3):413-419. doi:[10.1366/0003702001949500](https://doi.org/10.1366/0003702001949500).
- [57] Shariati-Rad, M. and Hasani, M. (2010), Selection of individual variables versus intervals of variables in PLSR. *J. Chemometrics*, 24: 45-56. <https://doi.org/10.1002/cem.1266>
- [58] Ezenarro, J. and Schorn-García, D. (2025), How Are Chemometric Models Validated? A Systematic Review of Linear Regression Models for NIRS Data in Food Analysis. *Journal of Chemometrics*, 39: e70036. <https://doi.org/10.1002/cem.70036>
- [59] Esbensen, Kim & Geladi, Paul & Larsen, Anders. (2014). The RPD myth.... *NIR news*. 25. 24. 10.1255/nirn.1462.
- [60] Fearn, T. (2002). Assessing calibrations: SEP, RPD, RER and R2. *NIR news*, 13(6), 12-14.
- [61] Fred Sigernes, Mikko Syrjäsoo, Rune Storvold, João Fortuna, Mariusz Eivind Grøtte, and Tor Arne Johansen, "Do it yourself hyperspectral imager for handheld to airborne operations," *Opt. Express* 26, 6021-6035 (2018)
- [62] **McCreery, R. L. (2000).** *Raman Spectroscopy for Chemical Analysis*. John Wiley & Sons. (Cap. 4: Instrumentation). (Serie Chemical Analysis Vol. 157). [ISBN 0471252875]

- [63] **Ocean Insight. (2024).** *Wavelength Calibration – Quick Start Guide (MNL-1015 Rev A)*. Ocean Optics, Inc
- [64] **Joseph, G. (2015).** *Building Earth Observation Cameras*. CRC Press (Taylor & Francis). DOI: 10.1201/b18022.
- [65] **Sigernes, F. (2018).** *Basic Hyper Spectral Imaging*. Lecture notes, NTNU. Recuperado de [http://kho.unis.no/doc/TTK20/BASIC\\_HYPER\\_SPECTRAL\\_IMAGING.pdf](http://kho.unis.no/doc/TTK20/BASIC_HYPER_SPECTRAL_IMAGING.pdf)
- [66] Ramos, Violeta & Fernandes, Isabel & Rollinson, Gavyn & Andersen, Jens & Menéndez, Esperanza & Santos Silva, António. (2024). Automated Mineralogical Analysis (QEMSCAN®) as a Complementary Petrographic Tool in Assessing the Potential Alkaline Release from Aggregates. 10.1007/978-3-031-59419-9\_15.
- [67] Gottlieb, P., Wilkie, G., Sutherland, D., Ho-Tun, E., Suthers, S., Perera, K., ... Rayner, J. (2000). *Using quantitative electron microscopy for process mineralogy applications*. JOM, 52(4), 24–25. <https://doi.org/10.1007/s11837-000-0126-9>.
- [68] Thermo Fisher Scientific. (2018). *QEMSCAN® automated mineralogy: Technical overview and applications*. Thermo Fisher Scientific, Instrumentation Manual.
- [69] SGS Minerals Services. (2015). *Mineralogy capabilities: QEMSCAN and MLA*. SGS Technical Bulletin, SGS MIN-357.
- [70] CSIRO. (2020). *QEMSCAN®: Quantitative Evaluation of Minerals by Scanning Electron Microscopy*. Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia. <https://www.csiro.au/en/work-with-us/industries/mining-resources/qemscan>.
- [71] Flores, L. 2006. QEMSCAN. Evaluación Cuantitativa de Mineralogía, Imágenes BSE + Energía dispersiva de Rayos X. Generalidades del Método y Bibliografía. Informe (inédito). SGS Minerals S.A., Mineralogía Avanzada: 16 pp. Santiago
- [72] IDS Imaging Development Systems GmbH. (s. f.). *IDS µEye color/VNIR sensor – Datasheet AB00624*. IDS Imaging Development Systems. [https://es.ids-imaging.com/IDS/pdfsheets\\_pdf.php?sku=AB00624](https://es.ids-imaging.com/IDS/pdfsheets_pdf.php?sku=AB00624).

- [73] Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012, NIPS 2012* (pp. 2951-2959). (Advances in Neural Information Processing Systems; Vol. 4).
- [74] Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2016). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1), 148-175. Article 7352306. <https://doi.org/10.1109/JPROC.2015.2494218>
- [75] Omar Khater, Ali Khater, Ashar Seif Al-Nasr, Samir Abozyd, Bassem Mortada, Yasser M. Sabry, Advancing near-infrared spectroscopy: A synergistic approach through Bayesian optimization and model stacking, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, Volume 318, 2024, 124492, ISSN 1386-1425, <https://doi.org/10.1016/j.saa.2024.124492>
- [76] Rina Foygel Barber. Emmanuel J. Candès. Aaditya Ramdas. Ryan J. Tibshirani. "Predictive inference with the jackknife+." *Ann. Statist.* 49 (1) 486 - 507, February 2021. <https://doi.org/10.1214/20-AOS1965>
- [77] Kim, B., Xu, C., & Barber, R. F. (2020). *Predictive inference is free with the Jackknife+-after-Bootstrap*. arXiv preprint arXiv:2002.09025. <https://arxiv.org/abs/2002.09025>.
- [78] Metso, "Smelting and Converting," Metso. [Online]. Available: <https://www.metso.com/metals-refining/solutions/smelting-and-converting/>.

## A. Anexo Tablas y gráficos asociados

### A.1. Tablas de hiperparámetros asociados

Este anexo presenta las tablas de hiperparámetros óptimos obtenidos para los modelos PLS y PLS-SVR aplicados a la estimación de %Cu, %Calcopirita y %Pirita, considerando distintas configuraciones espectrales (VNIR, SWIR y fusión de datos MLDF VNIR+SWIR). Estos valores corresponden a las configuraciones que maximizaron el desempeño predictivo durante el proceso de optimización y validación, y se incluyen con fines de reproducibilidad, trazabilidad metodológica y comparación entre enfoques mono-sensor y multiesensor.

**Tabla A.1 Hiperparámetros Óptimos para la Estimación de %Cu según Modelo y Configuración Espectral.**

| Modelo %Cu       | Model   | BoxConstraint | Epsilon | KernelScale | nCompVNIR | nCompSWIR | SavGolVNIR | SavGolSWIR | kVNIR | kSWIR | blockGamma |
|------------------|---------|---------------|---------|-------------|-----------|-----------|------------|------------|-------|-------|------------|
| MLDF (VNIR+SWIR) | PLS-SVR | 38.288        | 0.4017  | 1.1587      | 9         | 8         | 7          | 37         | 216   | 302   | 0.5443     |
| MLDF (VNIR+SWIR) | PLS     | —             | —       | —           | 9         | 7         | 5          | 23         | 223   | 266   | 0.83126    |
| SWIR             | PLS-SVR | 14.779        | 0.39675 | 0.55529     | —         | 8         | —          | 6          | —     | 187   | —          |
| SWIR             | PLS     | —             | —       | —           | —         | 11        | —          | 35         | —     | 344   | —          |
| VNIR             | PLS-SVR | 26.758        | 0.37257 | 0.638       | 7         | —         | 5          | —          | 159   | —     | —          |
| VNIR             | PLS     | —             | —       | —           | 7         | —         | 5          | —          | 154   | —     | —          |

**Tabla A.2 Hiperparámetros Óptimos para la Estimación de %Calcopirita según Modelo y Configuración Espectral.**

| Modelo %Calco    | Model   | BoxConstraint | Epsilon | KernelScale | nCompVNIR | nCompSWIR | SavGolVNIR | SavGolSWIR | kVNIR | kSWIR | blockGamma |
|------------------|---------|---------------|---------|-------------|-----------|-----------|------------|------------|-------|-------|------------|
| MLDF (VNIR+SWIR) | PLS-SVR | 25.639        | 0.1365  | 0.5543      | 8         | 4         | 63         | 57         | 214   | 478   | 0.6705     |
| MLDF (VNIR+SWIR) | PLS     | —             | —       | —           | 9         | 5         | 16         | 13         | 210   | 199   | 0.52848    |
| SWIR             | PLS-SVR | 66.586        | 0.50362 | 0.9033      | —         | 9         | —          | 20         | —     | 329   | —          |
| SWIR             | PLS     | —             | —       | —           | —         | 5         | —          | 29         | —     | 343   | —          |
| VNIR             | PLS-SVR | 81.839        | 0.36768 | 0.60245     | 5         | —         | 21         | —          | 295   | —     | —          |
| VNIR             | PLS     | —             | —       | —           | 6         | —         | 5          | —          | 133   | —     | —          |

**Tabla A.3 Hiperparámetros Óptimos para la Estimación de %Pirita según Modelo y Configuración Espectral.**

| Modelo %Pirita   | Model   | BoxConstraint | Epsilon | KernelScale | nCompVNIR | nCompSWIR | SavGolVNIR | SavGolSWIR | kVNIR | kSWIR | blockGamma |
|------------------|---------|---------------|---------|-------------|-----------|-----------|------------|------------|-------|-------|------------|
| MLDF (VNIR+SWIR) | PLS-SVR | 78.733        | 0.45598 | 0.86084     | 8         | 7         | 71         | 5          | 153   | 317   | 0.79564    |
| MLDF (VNIR+SWIR) | PLS     | —             | —       | —           | 5         | 7         | 5          | 12         | 203   | 228   | 0.59344    |
| SWIR             | PLS-SVR | 43.115        | 0.38112 | 0.52457     | —         | 6         | —          | 37         | —     | 344   | —          |
| SWIR             | PLS     | —             | —       | —           | —         | 10        | —          | 47         | —     | 320   | —          |
| VNIR             | PLS-SVR | 94.852        | 0.4674  | 0.36759     | 5         | —         | 6          | —          | 181   | —     | —          |
| VNIR             | PLS     | —             | —       | —           | 5         | —         | 5          | —          | 51    | —     | —          |

## A.2. Graficas de los modelos

Este anexo presenta las gráficas de desempeño de los modelos PLS y PLS-SVR para la estimación de %Cu, %Calcopirita y %Pirita, considerando las configuraciones VNIR, SWIR y MLDF (VNIR+SWIR). Las figuras incluyen los resultados de calibración, validación cruzada (LOO), validación externa y predicción, incorporando cuando corresponde intervalos de incertidumbre Jackknife+. Su inclusión tiene como objetivo complementar los resultados cuantitativos y facilitar la comparación visual del desempeño y la robustez de los modelos.

### PLS-SVR-VNIR %Cu:

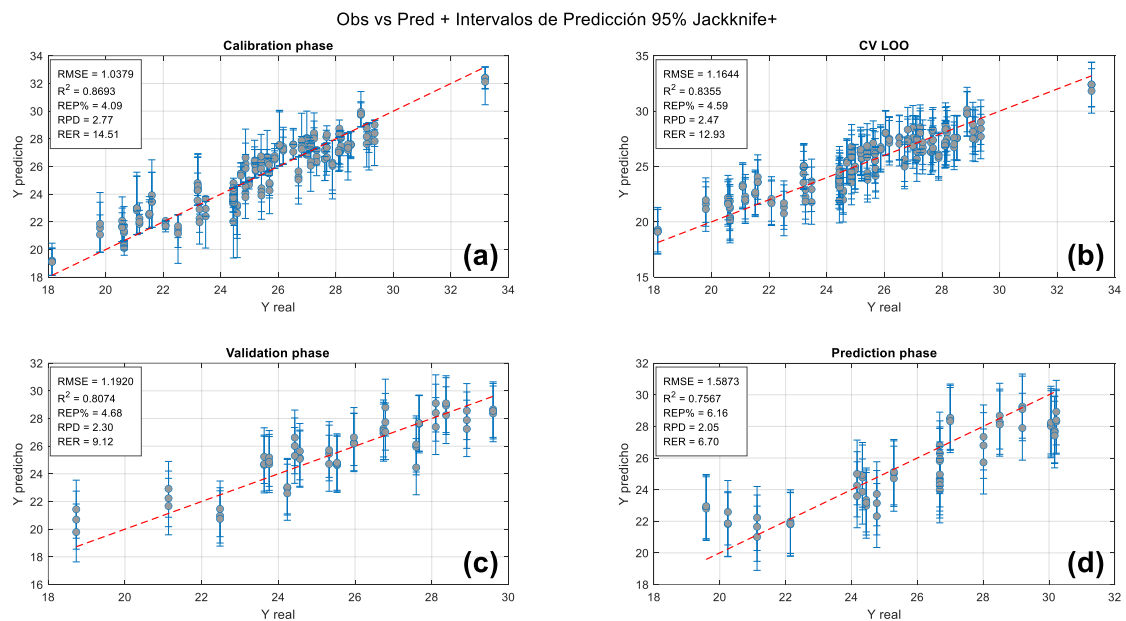


**Figura A.1** Desempeño del modelo PLS-SVR (VNIR) para %Cu en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

## PLS-SVR-SWIR %Cu:

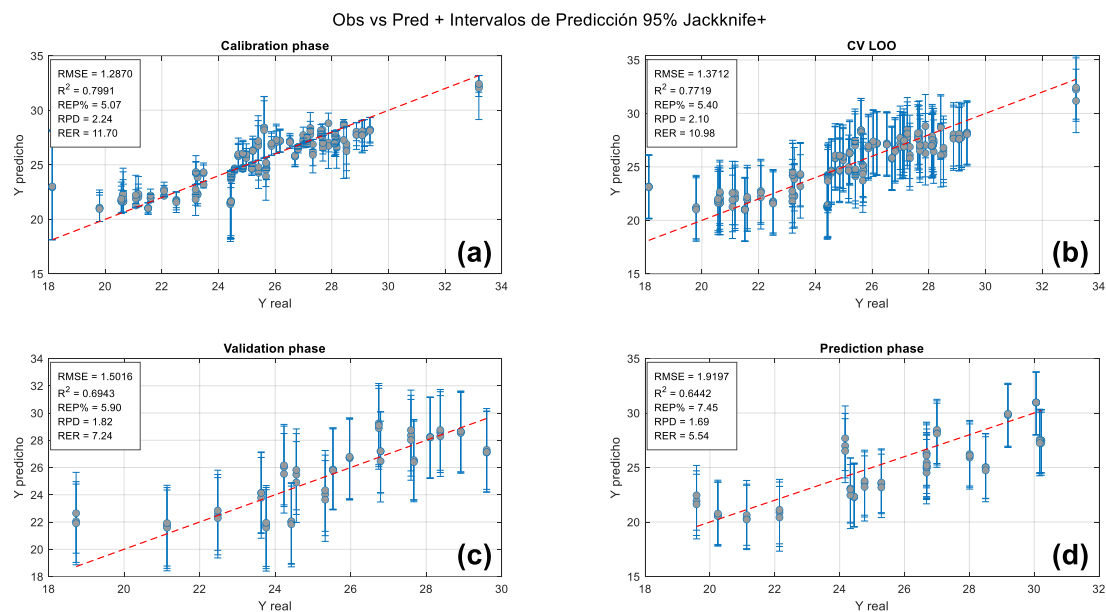


**Figura A.2** Desempeño del modelo PLS-SVR (SWIR) para %Cu en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

## PLS-MLDF %Cu:

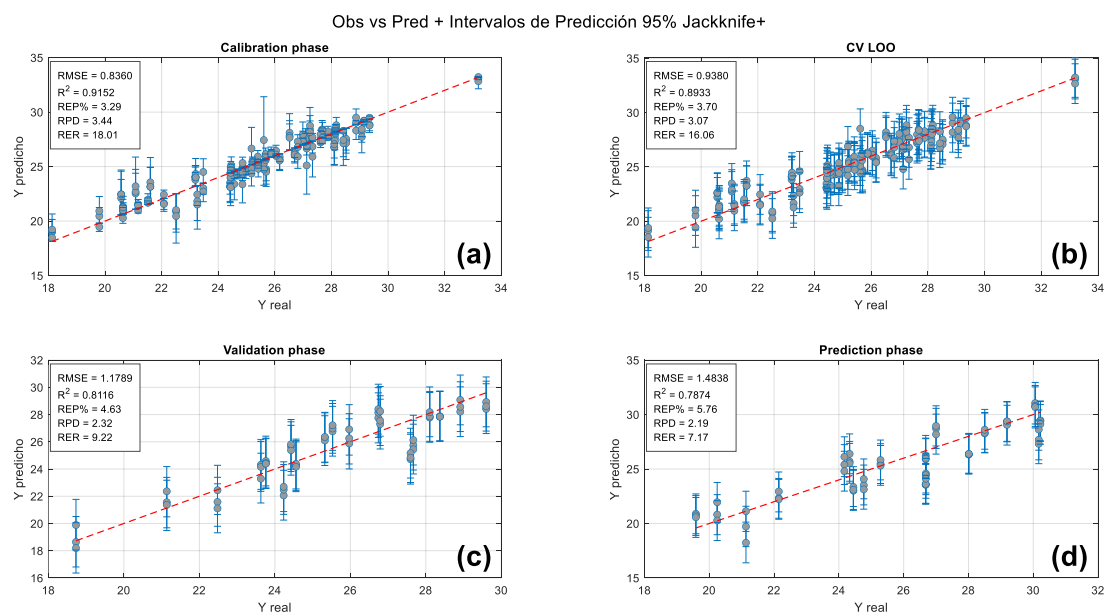


**Figura A.3** Desempeño del modelo PLS (MLDF) para %Cu en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

## PLS-VNIR %Cu:

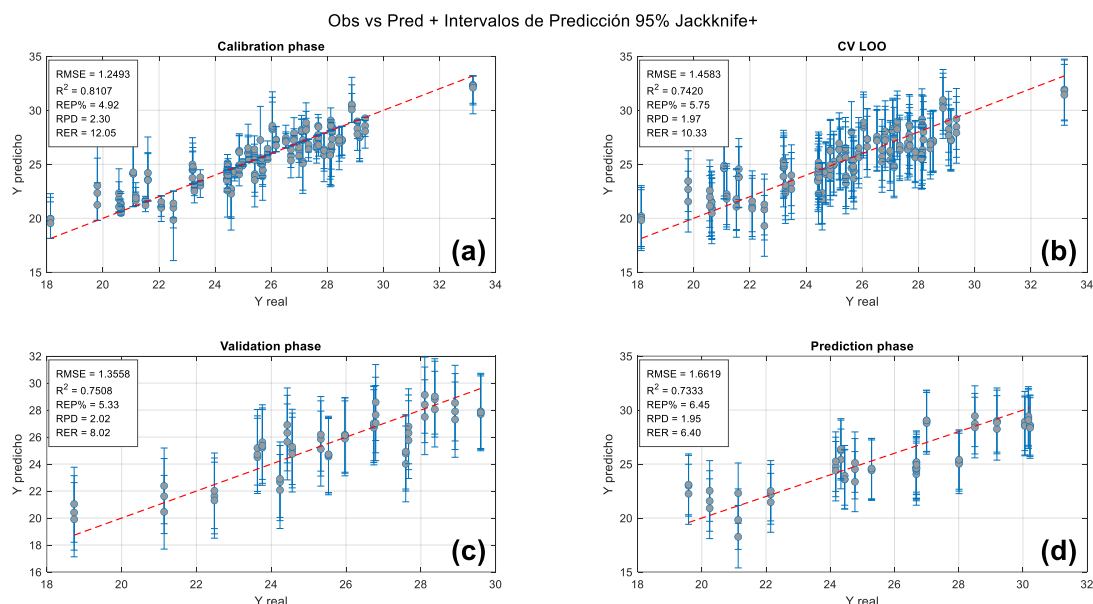


**Figura A.4** Desempeño del modelo PLS (VNIR) para %Cu en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

## PLS-SWIR %Cu:

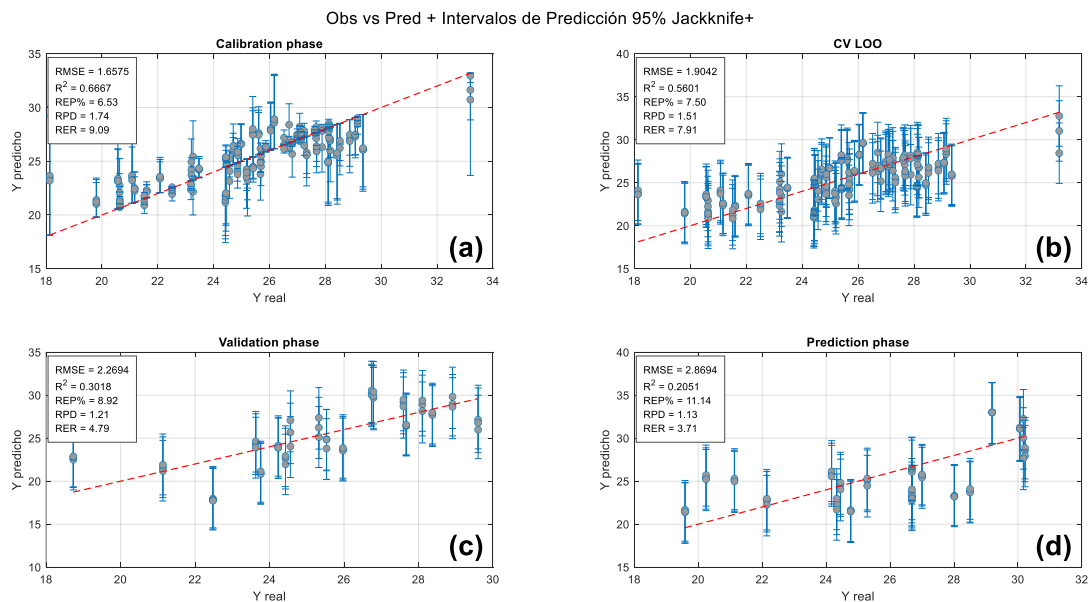


**Figura A.5** Desempeño del modelo PLS (SWIR) para %Cu en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

## PLS-SVR-VNIR %Calcopirita

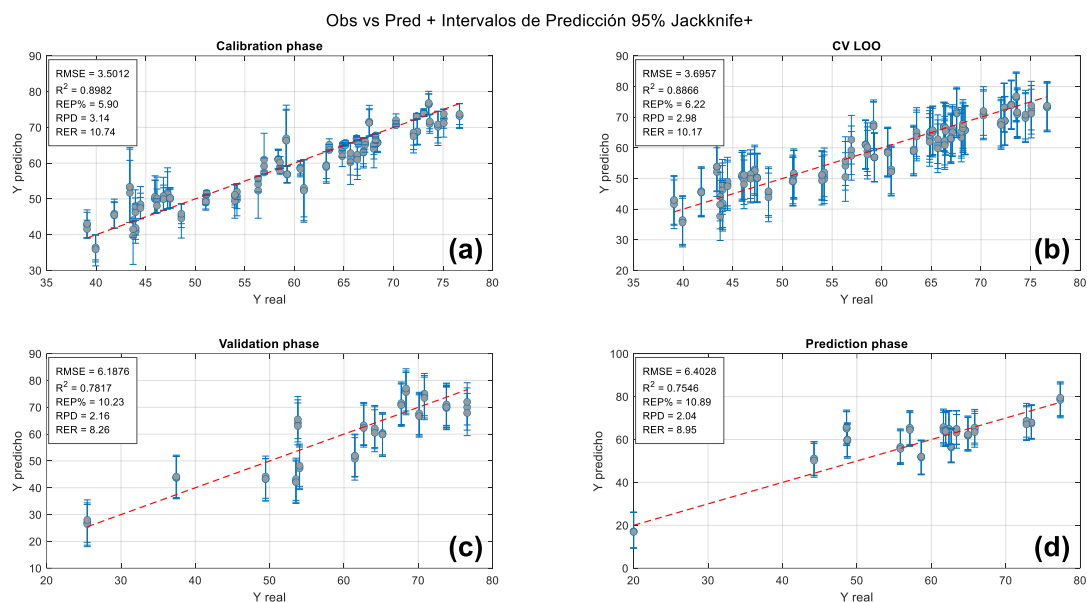


**Figura A.6** Desempeño del modelo PLS-SVR (VNIR) para %Calcopirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

## PLS-SVR-SWIR %Calcopirita:

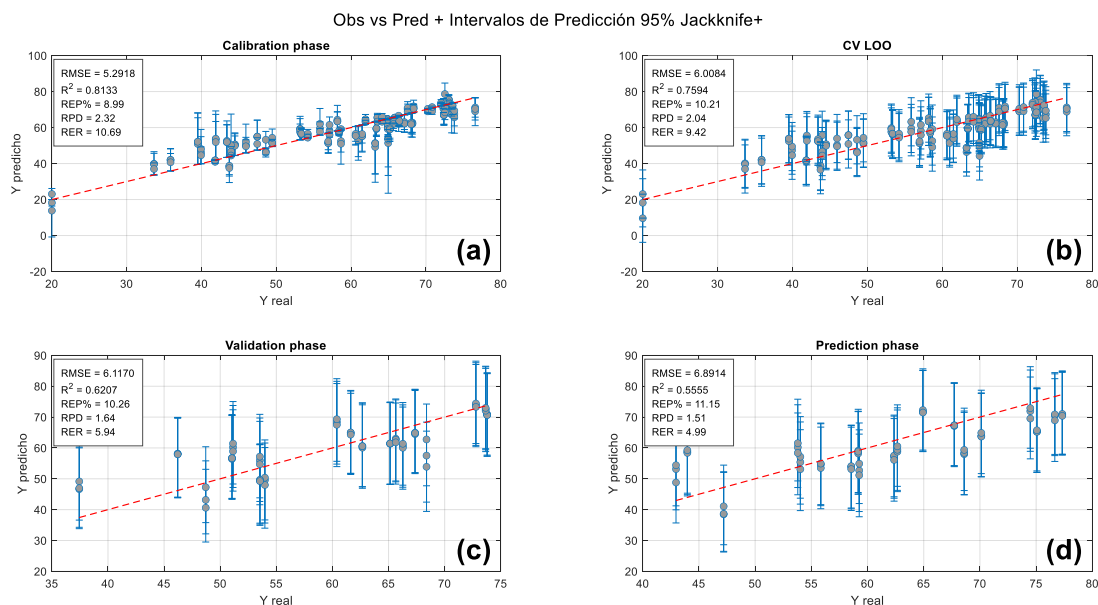


**Figura A.7** Desempeño del modelo PLS-SVR (SWIR) para %Calcopirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

## PLS-MLDF %Calcopirita:

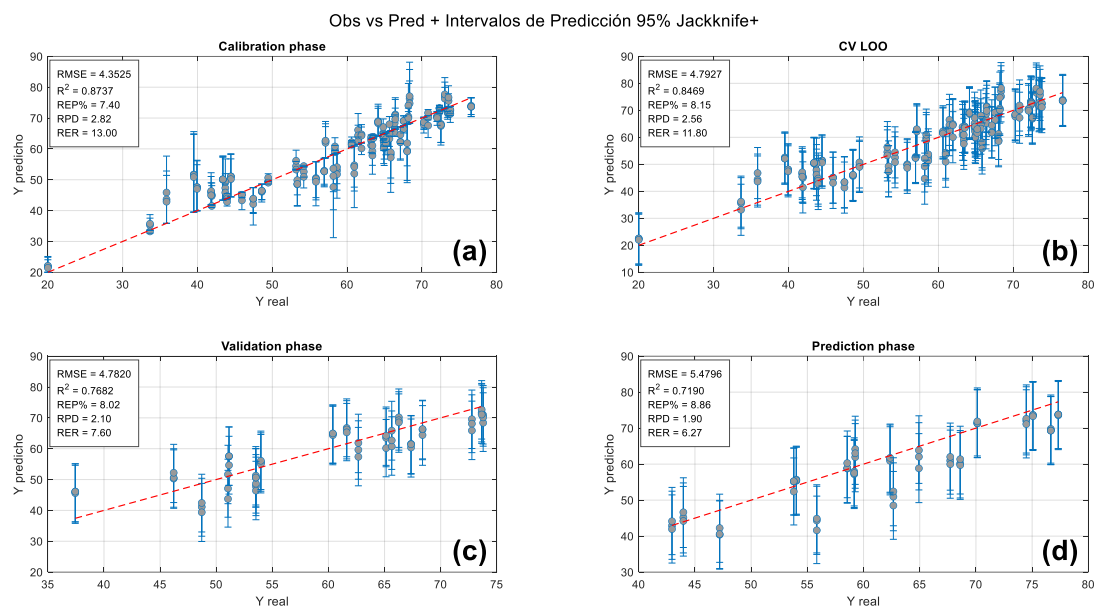


**Figura A.8** Desempeño del modelo PLS (MLDF) para %Calcopirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

## PLS VNIR %Calcopirita:

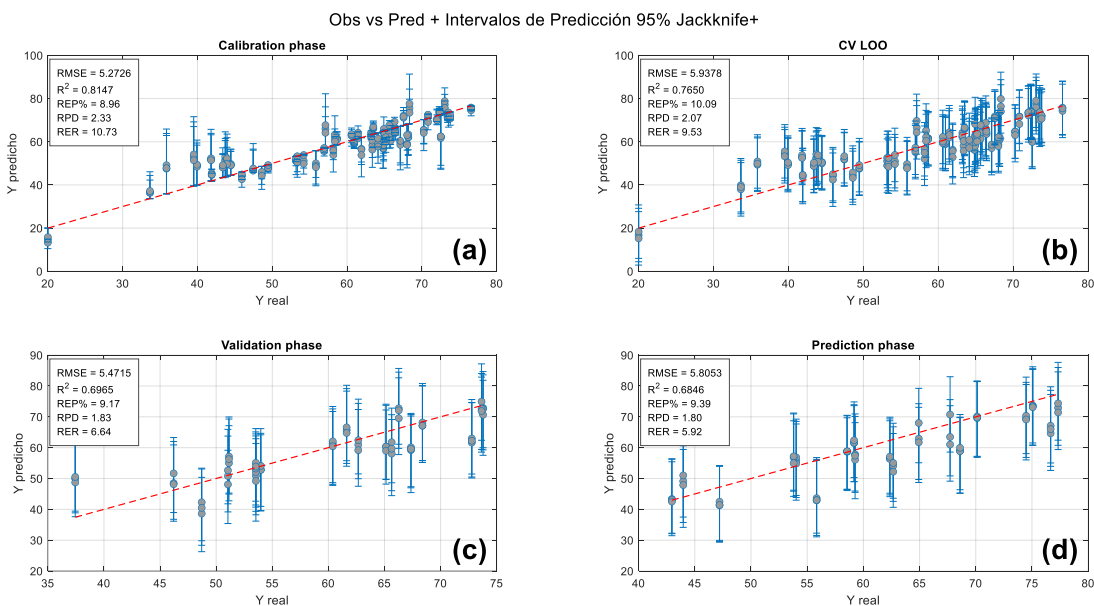


**Figura A.9** Desempeño del modelo PLS (VNIR) para %Calcopirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

## PLS SWIR %Calcopirita:

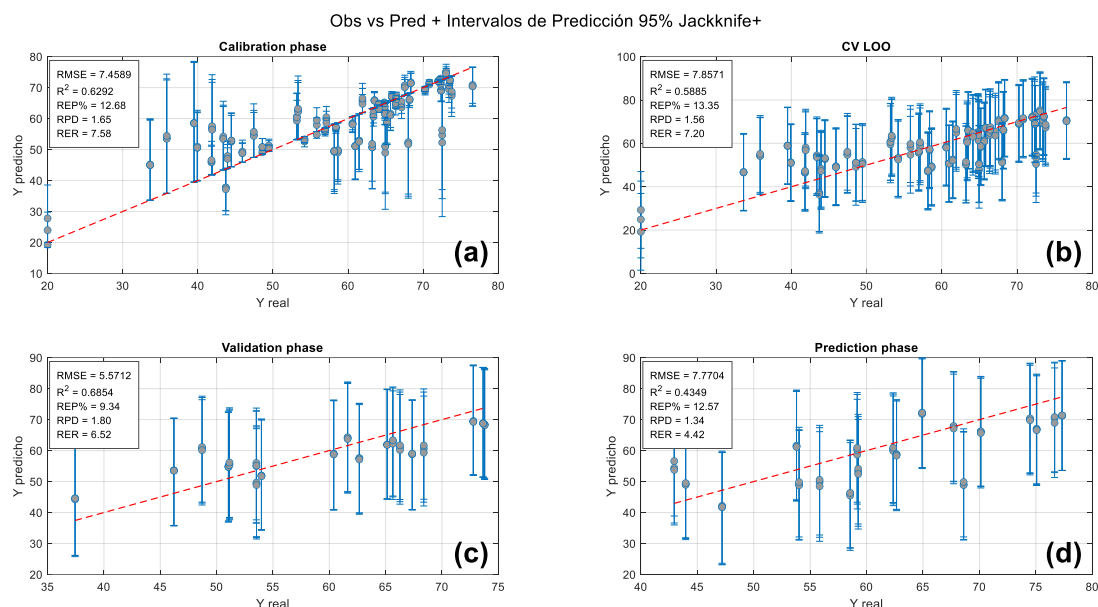


**Figura A.10** Desempeño del modelo PLS (SWIR) para %Calcopirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

## PLS-SVR-VNIR Pirita:

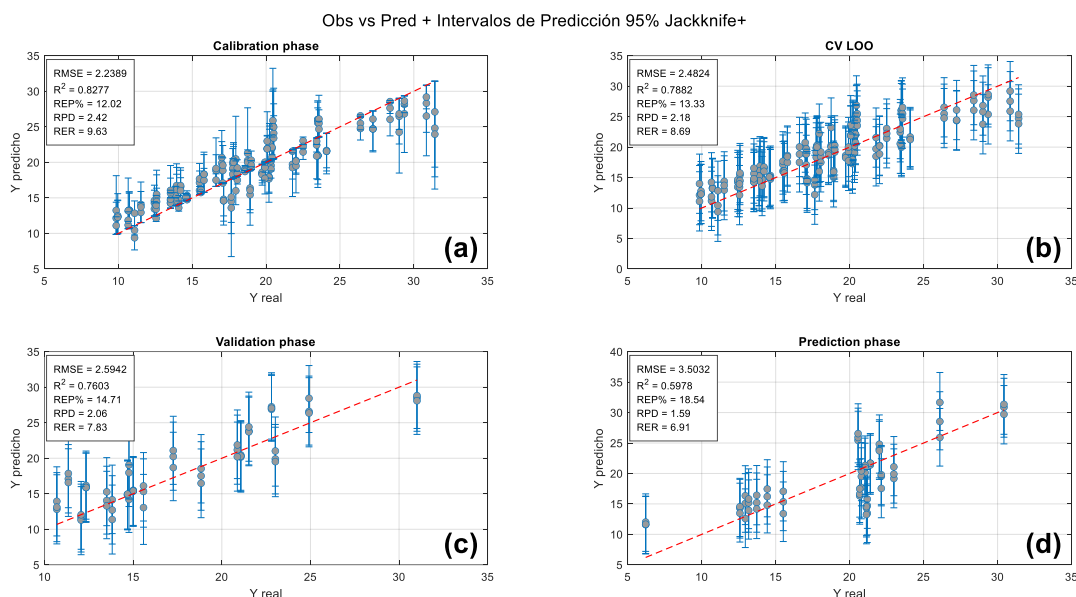


**Figura A.11** Desempeño del modelo PLS-SVR (VNIR) para %Pirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

## PLS-SVR-SWIR Pirita:

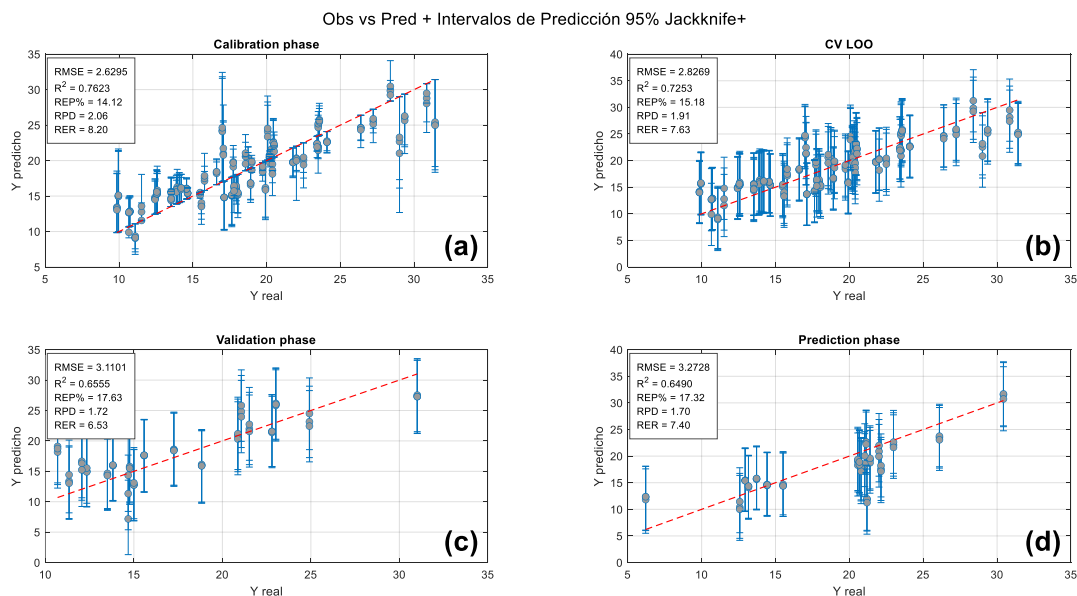


**Figura A.12** Desempeño del modelo PLS-SVR (SWIR) para %Pirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

## PLS-MLDF Pirita:

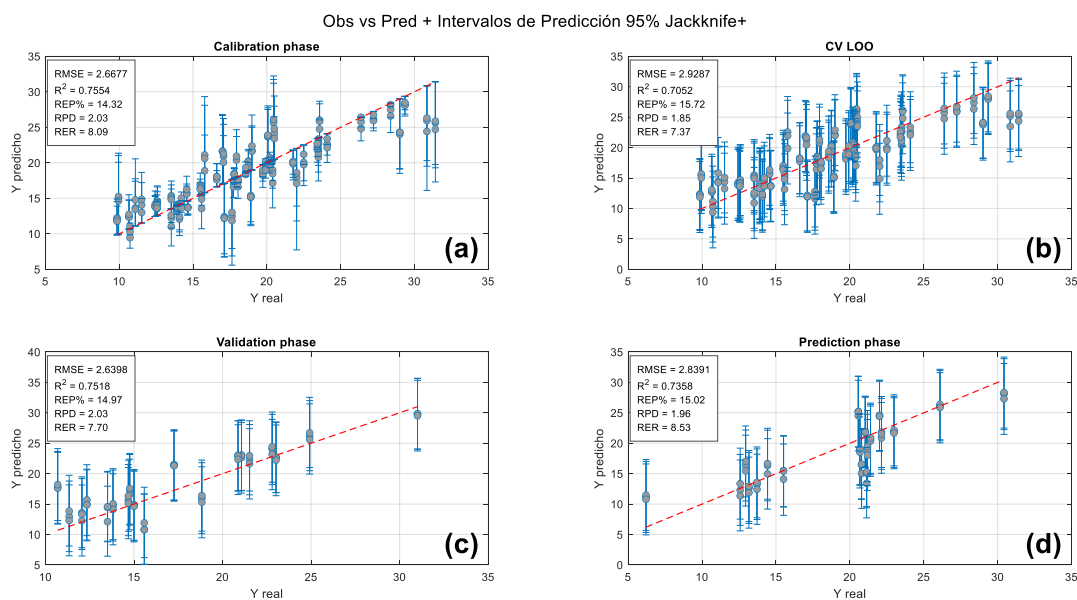


**Figura A.13** Desempeño del modelo PLS (MLDF) para %Pirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

## PLS-VNIR Pirita:

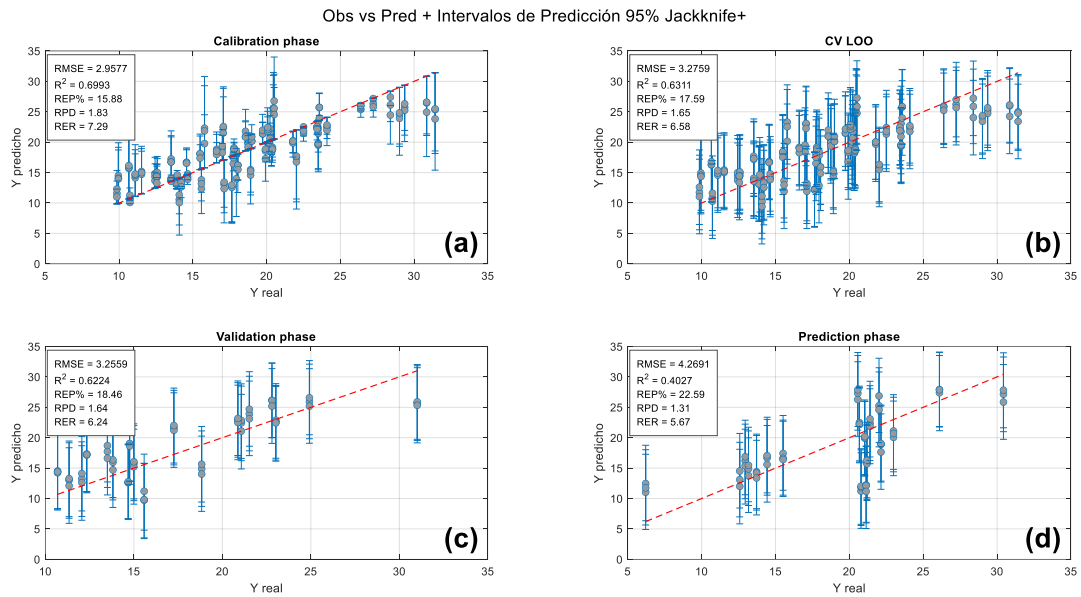


**Figura A.14** Desempeño del modelo PLS (VNIR) para %Pirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

## PLS-SWIR Pirita:

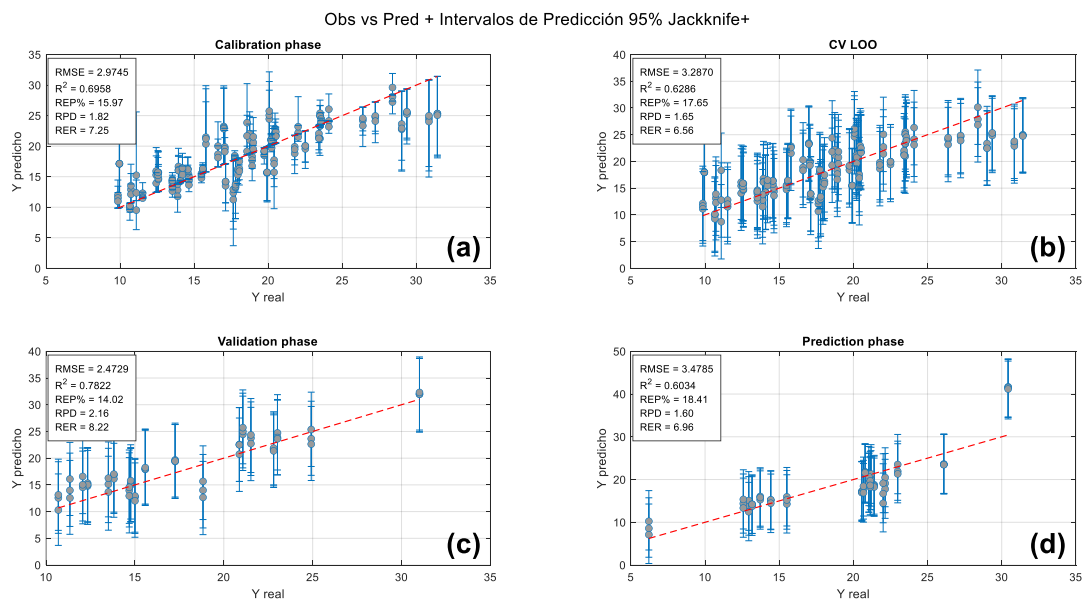


**Figura A.15** Desempeño del modelo PLS (SWIR) para %Pirita en: (a) Calibración, (b) Validación cruzada (LOO), (c) Validación externa, y (d) Predicción. Las barras indican los intervalos Jackknife+.

### A.3. Datasheet de los sensores VNIR y SWIR

En este anexo se presentan los extractos de las fichas técnicas (datasheets) de los sensores utilizados en el sistema de adquisición hiperespectral. Estas especificaciones permiten contextualizar las capacidades espectrales, espaciales y radiométricas de cada cámara, así como justificar su selección para la caracterización mineralógica y elemental de concentrados de cobre. En particular, se incluyen parámetros relevantes como el tipo de sensor, resolución espacial, profundidad radiométrica y respuesta espectral, los cuales influyen directamente en la calidad de los datos hiperespectrales adquiridos.

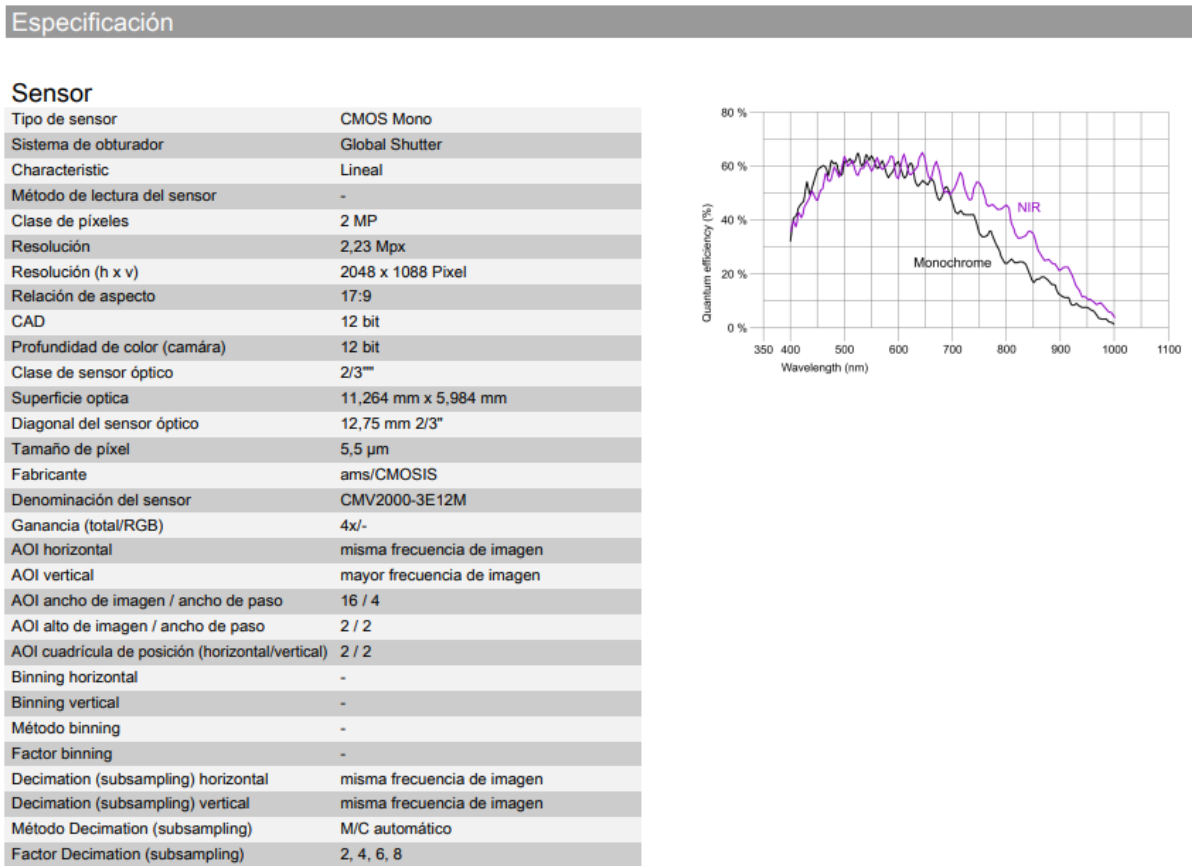


**Figura A.16** Extracto del datasheet del sensor CMOS monocromático utilizado en la cámara hiperespectral VNIR, incluyendo sus principales especificaciones técnicas y la respuesta espectral cuántica en el rango visible–infrarrojo cercano.

## PIKA IR+ HYPERSPECTRAL CAMERA



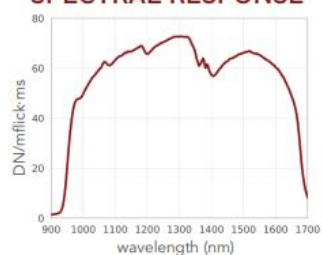
The Pika IR+ is a high-performance hyperspectral camera for the Near-Infrared (NIR) spectral range (900 – 1700 nm). It has high spatial and spectral resolutions and is often used for infrared laboratory applications.

The Pika IR+ can be used in our **airborne**, **laboratory**, and **outdoor** hyperspectral systems, as well as standalone or integrated into your system with our **programming guidance document**.

### FEATURES

- Spectral Range: 900 – 1700 nm
- 640 Spatial Pixels Per Line
- 336 Spectral Channels Per Line
- High Performance (5.6 nm FWHM spectral resolution)

### SPECTRAL RESPONSE



### ACTUAL DATA

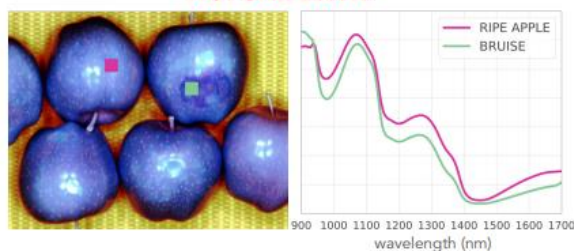


**Figura A.17** Extracto del datasheet de la cámara hiperespectral Pika IR+, mostrando el rango espectral SWIR (900–1700 nm), sus principales características técnicas y ejemplos de respuesta espectral.