



DEPARTAMENTO DE
INGENIERÍA ELÉCTRICA
UNIVERSIDAD DE CONCEPCIÓN

CARACTERIZACIÓN Y CUANTIFICACIÓN DE SOLUCIONES ACUOSAS UTILIZANDO ESPECTROS RAMAN

POR

Edgardo Andrés Pérez Foix

Memoria de Título presentada a la Facultad de Ingeniería de la Universidad de Concepción
para optar al título profesional de Ingeniero Civil en Telecomunicaciones.

Profesor Guía
Daniel Sbárbaro H, Ph.D.

Concepción,
11 de julio de 2025

© 2025 Edgardo Andrés Pérez Foix

© 2025 Edgardo Andrés Pérez Foix

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.

Para mis padres, quienes me apoyaron desde el primer día...

Resumen

Esta memoria de título tiene por objetivo desarrollar un modelo cuantitativo para estimar la concentración de Cloruro de Sodio (NaCl) en soluciones acuosas a partir de espectros Raman, incorporando el análisis de variaciones estructurales del agua en función de la temperatura y la salinidad. El trabajo combina técnicas experimentales de espectroscopía Raman con métodos computacionales de procesamiento espectral y modelado estadístico multivariante.

Se emplearon soluciones acuosas de NaCl preparadas con concentraciones conocidas y medidas espectroscópicas realizadas con un láser de 532 nm. Los espectros fueron preprocesados mediante corrección de línea base, filtrado por transformada rápida de Fourier (FFT) y recorte espectral al intervalo 2800–3800 cm^{-1} . Para caracterizar la banda O–H del agua, se aplicó un ajuste de tres funciones gaussianas, extrayendo nueve parámetros (posición, amplitud y ancho de cada pico) con significado físico-químico.

El modelado se realizó mediante regresión por mínimos cuadrados parciales (PLS por sus siglas en inglés Partial Least Squares), evaluando dos enfoques: uno basado en los espectros completos y otro utilizando los parámetros gaussianos. Los modelos fueron validados mediante métricas de error (RMSE, R^2), validación cruzada y pruebas con datos externos. Los resultados muestran que los espectros Raman reflejan de forma sistemática los efectos del NaCl y la temperatura sobre la red de enlaces de hidrógeno del agua, y que el modelo basado en parámetros gaussianos logra una predicción robusta y explicable de la concentración salina.

Abstract

This thesis aims to develop a quantitative model for estimating the concentration of NaCl in aqueous solutions using Raman spectroscopy, incorporating the analysis of water structural variations as a function of temperature and salinity. The work combines experimental Raman measurements with computational spectral processing and multivariate statistical modeling.

Aqueous NaCl solutions with known concentrations were prepared and analyzed using a 532nm laser Raman system. The spectra were preprocessed through baseline correction, high-frequency noise reduction using Fast Fourier Transform (FFT), and truncation to the 2800–3800 cm^{-1} range. To characterize the broad O–H stretching band of water, each spectrum was fitted to a sum of three Gaussian functions, yielding nine physically meaningful parameters (peak position, width, and amplitude).

Partial Least Squares (PLS) regression was used to build predictive models, comparing a full-spectrum approach against a reduced-variable model based on the Gaussian parameters. The models were validated using root mean square error (RMSE), coefficient of determination (R^2), cross-validation, and independent test data. Results show that Raman spectra systematically capture the effects of NaCl and temperature on the hydrogen-bond network of water, and that the Gaussian-parameter-based model achieves robust and interpretable salt concentration predictions.

Agradecimientos

Quiero expresar mi más profundo agradecimiento a todas las personas que me han acompañado y apoyado a lo largo de este camino académico y personal.

En primer lugar, a mis padres, Alfredo y Yolanda, por su amor incondicional, por enseñarme desde pequeño el valor del esfuerzo y por estar siempre presentes, aún en la distancia. A mis hermanos y hermanas, Sole, Yuyo, Juan Pablo y Rocío, por su apoyo constante, sus palabras de aliento y por compartir conmigo esta etapa, especialmente a Rocío, con quien viví todo mi período universitario y cuya compañía hizo de los días difíciles algo más llevadero.

A mi polola Bárbara, por estar a mi lado durante estos últimos dos años, en los que ha sido un pilar fundamental de compañía, paciencia y cariño. Gracias por creer en mí incluso cuando yo dudaba.

A mis amigos, Raúl López con quien vengo compartiendo desde los 15 años y la vida nos juntó en el mismo departamento de estudio, al Kova, mi primer amigo de la carrera. Angelo, Armant, Yoyo, Leo R, Efraín, Cristobal y Juan, por ser parte esencial de esta etapa. Por los espacios de desconexión, por las conversaciones filosóficas y los buenos momentos hicieron que este camino no fuera solitario.

Este logro también les pertenece a ustedes. Gracias.

Índice General

Resumen	I
Abstract	II
Agradecimientos	III
Índice de Figuras	VI
Índice de Tablas	VIII
1. Introducción	1
1.1. Trabajos previos	2
1.2. Discusión	4
1.3. Definición del problema	5
1.4. Objetivos	6
1.4.1. Objetivo general	6
1.4.2. Objetivos específicos	6
1.5. Metodología	7
1.5.1. Preparación de soluciones de NaCl	7
1.5.2. Generación de base de datos con espectros Raman	7
1.5.3. Procesamiento espectral	8
1.5.4. Análisis espectral para caracterización de las muestras	8
1.5.5. Desarrollo del modelo de regresión PLS	9
1.6. Alcances y Limitaciones	9
1.6.1. Alcances	9
1.6.2. Limitaciones	10
2. Marco Teórico	12

2.1. Introducción	12
2.2. Interacción de la luz con la materia	12
2.3. Fundamentos de la espectroscopía Raman	13
2.4. Espectro Raman del agua y efecto del NaCl	15
2.5. Análisis espectral mediante funciones Gaussianas	16
2.6. Preprocesamiento de datos espectrales	18
2.7. Fundamentos de la regresión Partial Least Squares (PLS)	19
3. Desarrollo	21
3.1. Calibración del espectrómetro	21
3.2. Preparación de soluciones estándar de NaCl	26
3.3. Adquisición de espectros Raman	28
3.4. Preprocesamiento de los espectros	29
3.5. Análisis espectral para caracterización de las muestras	32
3.6. Modelado con Partial Least Square (PLS)	34
4. Resultados	36
4.1. Conductividad como verificación experimental	36
4.2. Efecto del NaCl y la temperatura en el espectro Raman	38
4.3. Resultados del ajuste gaussiano	40
4.4. Evaluación del modelo PLS	41
4.5. Resumen de resultados	47
5. Conclusiones	49
5.1. Conclusiones generales	49
5.2. Trabajo a Futuro	50
Bibliografía	55

Índice de Figuras

2.1. Representación de dispersión Raman y Rayleigh [1].	14
2.2. Espectro Raman de disolución de NaCl con diferentes concentraciones [2].	16
2.3. Descomposición en funciones Gaussianas [3].	17
3.1. Espectro Raman del mercurio y del neón.	22
3.2. Espectro Raman del mercurio y del neón combinados.	22
3.3. Ajuste subpíxel de los 3 peaks del mercurio.	23
3.4. Curvas de calibración con lámparas de Hg y Ne.	24
3.5. Comparación de espectro en número de píxeles y calibrado en nm. . . .	24
3.6. Cálculo de FWHM.	25
3.7. Espectro de referencia azufre-polietileno estándar.	25
3.8. Curvas de calibración para desplazamiento Raman.	26
3.9. Comparación de espectro en número de píxel y calibrado en cm^{-1}	26
3.10. FWHM en cm^{-1}	27
3.11. Configuración experimental: (1) láser, (2) soporte de cubeta, (3) cubeta, (4) controlador de temperatura, (5) espectrómetro Almad de Thunder Optics, (6) computador portátil, (7) mesa óptica.	28
3.12. Comparación entre espectro completo y región O-H.	30
3.13. Filtrado FFT de los espectros.	31
4.1. Curva de calibración entre la masa de NaCl (g) y conductividad eléctrica (mS/cm).	38
4.2. Dependencias espectrales principales.	39
4.3. Ajuste gaussiano triple para agua a 50°C	41
4.4. Comparación RMSECV vs Número de componentes latentes entre ambos modelos.	42

4.5. Resultado entrenamiento regresión PLS.	43
4.6. Residuos de entrenamiento regresión PLS.	44

Índice de Tablas

3.1. Comparación de parámetros espectrales ajustados al Hg.	22
3.2. Masa pesada de NaCl en cada solución preparada.	27
3.3. Significado físico-químico de los parámetros de ajuste Gaussiano.	33
4.1. Conductividad medida para distintas masas de NaCl disueltas en solución.	37
4.2. Resultados de predicción para temperaturas incluidas en el entrenamientos	45
4.3. Resultados de predicción para temperaturas no incluidas en el entrenamiento	46

Capítulo 1

Introducción

El agua es un disolvente universal cuya estructura molecular presenta una compleja red de enlaces de hidrógeno, altamente sensible a perturbaciones como la temperatura o la presencia de solutos iónicos. Entre estos, el cloruro de sodio (NaCl) representa un electrolito de gran relevancia tanto en contextos industriales como biológicos y medioambientales. La cuantificación precisa de su concentración en solución acuosa es fundamental en diversas aplicaciones, como el monitoreo de procesos químicos, el control de calidad de alimentos y bebidas, o el análisis de aguas naturales y residuales.

Tradicionalmente, la determinación de salinidad se ha realizado mediante técnicas conductimétricas o métodos de titulación. No obstante, estos procedimientos suelen requerir contacto directo con la muestra, preparación previa o reactivos químicos, lo que limita su aplicabilidad en ciertos entornos. En este contexto, la espectroscopía Raman se presenta como una herramienta poderosa y no invasiva para el análisis de soluciones, gracias a su capacidad de proporcionar información estructural a nivel molecular sin necesidad de marcadores ni tratamientos adicionales.

En particular, la banda de estiramiento O–H del agua, observable en el rango de 2800–3800 cm^{-1} , es extremadamente sensible a cambios en la red de enlaces de hidrógeno provocados por la temperatura o por la interacción con iones disueltos. Esta sensibilidad hace posible inferir la salinidad de una muestra acuosa a partir del análisis espectral de dicha región. Sin embargo, esta tarea implica desafíos analíticos importantes debido a la superposición de señales y a la naturaleza altamente colineal de los datos espectrales.

Para enfrentar este problema, se propone en esta memoria de título un enfoque basado en la descomposición espectral mediante funciones gaussianas y en el modelado estadístico mediante regresión Partial Least Squares (PLS). Este enfoque permite transformar espectros Raman complejos en vectores de parámetros fisicoquímicamente interpretables y construir modelos de predicción robustos y explicables.

1.1. Trabajos previos

La espectroscopía Raman ha sido objeto de múltiples investigaciones que demuestran su utilidad en la caracterización de soluciones salinas bajo diferentes contextos y condiciones. Estos estudios destacan no solo las aplicaciones potenciales de esta técnica, sino también los desafíos que enfrenta.

En el contexto de la astrobiología y la exploración planetaria, Mason llevó a cabo un estudio exhaustivo del comportamiento espectral de sulfatos y cloruros en un rango de temperaturas que emula condiciones marcianas [4]. Este trabajo estableció una base de datos espectral única que resulta clave para futuras misiones planetarias, donde se requiere identificar y cuantificar especies químicas bajo condiciones extremas. De manera similar, Takahashi investigó la especiación de sílice disuelta en soluciones sometidas a alta presión y temperatura, demostrando cómo la espectroscopía Raman *in situ* puede revelar estructuras moleculares complejas [5].

La espectroscopía Raman también ha demostrado ser una herramienta eficaz para el monitoreo ambiental, especialmente en la determinación de salinidad y la identificación de contaminantes emergentes. Bao desarrolló un modelo espectral que correlaciona la intensidad Raman con la salinidad en soluciones de agua de mar, considerando la influencia de variaciones térmicas [2].

Las soluciones salinas presentan una alta complejidad espectral debido a la superposición de bandas Raman provenientes tanto de los iones disueltos como del agua. Estudios previos han demostrado que sales como Na_2SO_4 y K_2SO_4 generan espectros superpuestos que dificultan su correcta identificación, incluso utilizando bases de datos espectrales avanzadas [6]. Además, variaciones en factores externos como la temperatura y la salinidad alteran significativamente las intensidades espectrales, como lo evidencia un estudio de soluciones de agua de mar [2].

Recientemente, el uso de aprendizaje profundo ha permitido abordar estos desafíos con mayor eficacia. Redes neuronales convolucionales (CNN) y algoritmos de aprendizaje automático han sido aplicados para identificar patrones no lineales en mezclas multicomponentes, logrando una mayor precisión y robustez en la cuantificación de especies químicas [7, 8]. Asimismo, Sunil aplicó algoritmos de aprendizaje profundo en combinación con espectros Raman para la detección automatizada de microplásticos en agua [7].

El uso de modelos multivariados ha demostrado ser fundamental para correlacionar datos espectrales con concentraciones de especies químicas. Durickovic exploró la capacidad de la espectroscopía Raman para caracterizar fases acuosas y transiciones sólido-líquido, desarrollando curvas de calibración ajustadas mediante regresión [1].

La combinación de métodos multivariados con algoritmos de aprendizaje profundo ha revolucionado el análisis espectral. Yang implementó transformadas de onda continua (CWT) junto con CNN, logrando coeficientes de determinación superiores a 0.999 [9]. De manera similar, Sui diseñó un modelo que optimiza los hiperparámetros de redes neuronales, reduciendo errores sistemáticos [10]. Qi, por su parte, demostró el papel clave de algoritmos como CNN y máquinas de soporte vectorial (SVM) en la clasificación de compuestos a partir de huellas Raman [8].

A pesar de estos avances, los estudios también subrayan desafíos importantes. Qi y Guo identificaron problemas de sobreajuste y falta de interpretabilidad en los modelos, además de la necesidad de bases de datos espectrales más representativas [8].

Uno de los principales desafíos en espectros Raman es la mitigación del ruido y las interferencias. Fang revisó los avances en algoritmos de eliminación de ruido basados en aprendizaje profundo, destacando su impacto en el preprocesamiento espectral [11]. Zeng-Ping propuso un modelo ICS que corrige variaciones térmicas, mejorando la predicción en entornos variables [12]. En paralelo, Sbarbaro implementó un filtro de Kalman adaptativo para corregir señales afectadas por condiciones variables, mejorando la precisión de los modelos [13].

La superposición de bandas espectrales sigue siendo un desafío en mezclas complejas. Prieto-Taboada señaló las dificultades al diferenciar especies como Na_2SO_4 y K_2SO_4 [6], mientras que Baumgartner evidenció cómo la polarización y orientación cristalográfica afectan la interpretación espectral [3].

Finalmente, la integración de enfoques físicos y aprendizaje automático ha surgido como una estrategia prometedora. Wang analizó las bandas OH en presencia de aniones complejos, estableciendo correlaciones con propiedades fisicoquímicas como pH o conductividad [14].

1.2. Discusión

La cuantificación precisa de especies salinas en soluciones acuosas mediante espectroscopía Raman ha sido ampliamente reconocida como un desafío técnico y científico por múltiples razones. Si bien esta técnica ofrece ventajas importantes —como la no destrucción de la muestra, mínima preparación y posibilidad de análisis *in situ*—, su aplicación práctica en medios salinos enfrenta varias limitaciones que hasta la fecha no han sido completamente superadas [15, 1].

Uno de los principales vacíos identificados en la literatura es la escasa respuesta Raman de iones monovalentes como Na^+ y Cl^- . A diferencia de iones como SO_4^{2-} o NO_3^- , que presentan modos vibracionales Raman-activos intensos, el NaCl en solución no genera señales espectrales claras asociadas a sus propios enlaces. En consecuencia, su presencia se infiere de manera indirecta, principalmente a través de las perturbaciones inducidas en las bandas del agua (especialmente en la región del estiramiento O–H entre 2800–3800 cm^{-1}) [14]. Sin embargo, estas perturbaciones son sutiles y altamente dependientes de las condiciones del medio, lo cual dificulta la construcción de modelos robustos sin un tratamiento estadístico multivariado cuidadoso [9, 8].

En segundo lugar, la influencia de factores externos como la temperatura, la presión y el nivel de hidratación en las soluciones representa un obstáculo importante para la estabilidad y reproducibilidad de los espectros Raman. Estudios recientes han mostrado que variaciones de apenas unos grados Celsius pueden inducir desplazamientos en la intensidad relativa de bandas críticas [2], dificultando la cuantificación directa de la concentración salina. A pesar de que existen estrategias de corrección —como la inclusión de la temperatura como variable latente o el uso de métodos de estandarización espectral [12]— no hay consenso sobre cuál es el enfoque más eficiente y generalizable para soluciones simples como NaCl [11].

Además, aunque se han desarrollado modelos de regresión multivariada, muchos de

ellos se han centrado en mezclas más complejas o en sistemas planetarios simulados (como los que imitan condiciones en Europa o Marte), donde las concentraciones son más extremas o los solutos tienen mayores señales Raman [4, 5]. En cambio, los estudios centrados exclusivamente en NaCl en matrices acuosas puras y en condiciones ambientales variables siguen siendo limitados. No existe todavía una metodología sistemática y reproducible que permita modelar con precisión y robustez este sistema bajo un rango amplio de condiciones experimentales [15].

Aunque se reconoce que técnicas como la corrección de línea base, la derivación de espectros y la normalización son esenciales para mejorar la calidad del análisis, la literatura rara vez detalla o compara de forma sistemática cuál combinación de técnicas resulta más efectiva específicamente para soluciones salinas con bajo contraste espectral como NaCl.

Por último, aunque algunas investigaciones han explorado el uso de métodos de aprendizaje automático para el análisis espectral, aún existe una brecha entre la complejidad computacional de estas herramientas y su aplicabilidad práctica en ambientes de laboratorio o monitoreo en línea [7]. En este contexto, modelos más sencillos pero potentes como la regresión por mínimos cuadrados parciales (PLS) siguen siendo altamente relevantes, siempre que se acompañen de una caracterización experimental cuidadosa y una validación estadística robusta.

1.3. Definición del problema

A pesar de los avances recientes en espectroscopía Raman aplicada a soluciones salinas, persisten limitaciones significativas que obstaculizan su uso como técnica cuantitativa confiable para la estimación de concentraciones iónicas, particularmente en medios acuosos simples como las soluciones de NaCl. Las principales dificultades radican en la baja intensidad de las bandas asociadas a los iones disueltos, la superposición espectral con las bandas del agua y la alta sensibilidad de la señal a variables experimentales como temperatura, presión y estado de hidratación. Además, si bien existen estudios centrados en sales complejas o mezclas multicomponentes, se ha prestado relativamente poca atención a establecer modelos cuantitativos robustos y reproducibles específicamente para NaCl —una sal de alta relevancia en aplicaciones industriales y científicas— bajo

condiciones controladas.

La ausencia de modelos de referencia validados que permitan evaluar el rendimiento de técnicas avanzadas como la regresión por mínimos cuadrados parciales (PLS) también representa un vacío metodológico. Esta carencia impide comparar objetivamente distintas estrategias de modelado multivariado, lo que debilita la generalización de los resultados obtenidos en estudios previos. En este contexto, se plantea como problema central la necesidad de desarrollar una base experimental confiable y un modelo que permita cuantificar concentraciones de NaCl a partir de espectros Raman.

1.4. Objetivos

1.4.1. Objetivo general

Desarrollar un método basado en espectroscopía Raman para la caracterización y cuantificación de soluciones acuosas, identificando y analizando las variaciones espectrales inducidas por la composición química, concentración y condiciones físicas de las muestras.

1.4.2. Objetivos específicos

1. Preparar soluciones acuosas de NaCl con diferentes concentraciones.
2. Generar de una base de datos con espectros Raman.
3. Aplicar técnicas de preprocesamiento espectral para mejorar la calidad de los datos.
4. Analizar el espectro para caracterizar las muestras.
5. Implementar un modelo de regresión PLS para estimar las concentraciones de NaCl en función de la información espectral.
6. Evaluar el desempeño del modelo PLS utilizando métricas cuantitativas como RMSE, MAE y R^2 . Junto con la validación del modelo probando su capacidad de predicción con muestras en diferentes condiciones.

1.5. Metodología

La metodología adoptada para este estudio se estructura en función del cumplimiento secuencial de los objetivos específicos planteados, combinando procedimientos experimentales, adquisición espectral, procesamiento computacional y modelado estadístico. A continuación, se describe detalladamente cada etapa del proceso:

1.5.1. Preparación de soluciones de NaCl

Preparar soluciones acuosas de Cloruro de Sodio (NaCl) con concentraciones conocidas, abarcando un rango representativo que asegure variabilidad suficiente para construir y validar modelos de regresión. Las concentraciones oscilan entre 0g/100mL y 5g/100mL, distribuidas en intervalos regulares. Para la preparación de las muestras se utiliza:

- Cloruro de Sodio para análisis.
- Agua desionizada.
- Material volumétrico calibrado (balanzas analíticas y pipetas volumétricas).

En total se generarán 6 muestras las cuales serán etiquetadas como 0g/100mL, 1g/100mL, 2g/100mL, 3g/100mL, 4g/100mL y 5g/100mL. Se almacenarán en recipientes herméticos para evitar evaporación o contaminación.

1.5.2. Generación de base de datos con espectros Raman

Los espectros Raman de cada solución serán capturados utilizando un sistema de espectroscopía Raman configurado con parámetros constantes:

- Longitud de onda del láser: 532,10 nm.
- Potencia de excitación: 300mW.
- Tiempo de integración: 120s.

Se realizará una variación de temperatura a cada muestra las cuales serán de 10°, 20°, 30°, 40° y 50°celsius. Se obtendrán 4 espectros por temperatura obteniendo un total de 120 espectros. La celda que contiene cada muestra será de cuarzo y se enjuagará entre

mediciones para evitar contaminación. Se substraerá el fondo espectral en cada cambio de temperatura y cambio de muestra para la obtención de espectros más limpios.

1.5.3. Procesamiento espectral

Los espectros obtenidos serán preprocesados para extraer, transformar y/o reducir la información relevante contenida en el espectro.

- **Corrección de línea base:** para eliminar tendencias producidas por variaciones del fondo, si es necesario.
- **Normalización:** para corregir diferencias de intensidad absoluta y comparar señales relativas, si es necesario.
- **Recorte del espectro:** para trabajar solamente con la porción espectral con mayor relevancia e información.
- **Suavizado:** mediante filtros como Savitzky-Golay o filtros aplicados con Transformadas Rápidas de Fourier(FFT).

Estas etapas serán implementadas en Python y validadas gráficamente.

1.5.4. Análisis espectral para caracterización de las muestras

Una vez preprocesado el espectro, se puede aplicar una descomposición matemática de las bandas Raman en componentes individuales utilizando funciones gaussianas. Este procedimiento permite:

- Separar bandas superpuestas.
- Asignar cada componente a un modo vibracional específico.
- Estimar la posición (cm^{-1}), intensidad relativa y ancho de cada banda.

Esta etapa será implementada en Python y validadas gráficamente.

1.5.5. Desarrollo del modelo de regresión PLS

Se implementará un modelo de regresión *Partial Least Square* (PLS) utilizando como variables independientes los parámetros de Posición (μ_i), de amplitud (A_i) y anchura (δ_i) de las distribuciones Gaussianas encontradas, y como variable dependiente la concentración de NaCl. El modelo se construirá siguiendo estos pasos:

Definición de matrices

- $\mathbf{X} \in \mathbb{R}^{120 \times 9}$: matriz de predictores, donde cada fila corresponde a un espectro representado por sus 9 parámetros gaussianos.
- $\mathbf{y} \in \mathbb{R}^{120 \times 1}$: vector de concentraciones reales de NaCl (en g).

Componentes latentes

- La selección del número óptimo de componentes latentes se realiza mediante técnicas de validación cruzada, como *k-fold* o *leave-one-out*.

Validación del modelo

- Evaluación del error de predicción en conjuntos no utilizados durante el entrenamiento junto con métricas como RMSE y MAE.

1.6. Alcances y Limitaciones

1.6.1. Alcances

Este trabajo se centrará en la estimación cuantitativa de concentraciones de NaCl en soluciones acuosas utilizando espectros Raman y modelos de regresión multivariada, específicamente *Partial Least Square* (PLS). Los principales alcances de esta investigación son:

- **Construcción de una base de datos experimental:** Se obtendrá una colección de espectros Raman de soluciones de NaCl con concentraciones conocidas, bajo condiciones experimentales controladas y repetibles.

- **Implementación de un modelo predictivo de regresión:** Se desarrollará un modelo PLS capaz de predecir la concentración de NaCl a partir de la información espectral procesada, utilizando criterios de evaluación estadística robustos.
- **Identificación preliminar de variables espectrales relevantes:** Se analizarán las regiones del espectro que presentan mayor correlación con la concentración de NaCl, como base para interpretaciones físico-químicas futuras.

Este trabajo no pretende resolver todos los aspectos de la cuantificación multicomponente ni establecer un modelo definitivo para la implementación industrial. Más bien, su propósito es metodológico y exploratorio: demostrar la viabilidad del enfoque espectroscópico cuantitativo con un sistema salino sencillo como el NaCl.

1.6.2. Limitaciones

A pesar del diseño riguroso del estudio, existen ciertas restricciones que deben ser reconocidas explícitamente para acotar el alcance del trabajo y contextualizar sus resultados:

- **Sistema químico simplificado:** El estudio se limita exclusivamente a soluciones acuosas de NaCl. No se abordarán mezclas multicomponentes, ni se incluirán otras sales, lo que restringe la generalización de los modelos desarrollados.
- **Rango de concentraciones limitado:** Por razones de control experimental y linealidad espectral, se trabajará en un rango de concentraciones moderado (1g/100mL a 5g/100mL). Esto puede limitar la aplicabilidad del modelo en sistemas más concentrados o diluidos.
- **Condiciones ambientales controladas:** Las mediciones se realizarán en condiciones de laboratorio estables. No se considerarán variaciones de presión ni interferencias externas, aunque estas variables puedan tener un impacto importante en aplicaciones reales.
- **Sensibilidad instrumental y ruido espectral:** A pesar del procesamiento, el modelo podría verse afectado por ruido instrumental, intensidad del láser o variaciones del fondo que no siempre pueden eliminarse completamente.

Estas limitaciones no invalidan los objetivos del proyecto, sino que delimitan su contexto

de aplicación. Dejar constancia de ellas permite evaluar los resultados con criterio técnico y sienta las bases para futuras extensiones del estudio en el marco de una tesis de magíster o una investigación aplicada más amplia.

Capítulo 2

Marco Teórico

2.1. Introducción

La presente investigación se sitúa en la intersección entre la espectroscopía vibracional, la química analítica y la modelación estadística multivariada. Su propósito es desarrollar una estrategia confiable para estimar concentraciones salinas a partir de espectros Raman, con base en modelos matemáticos de regresión. Para ello, es necesario comprender los principios físicos que rigen la espectroscopía Raman, los desafíos inherentes al análisis de soluciones salinas, y las metodologías estadísticas utilizadas para extraer información cuantitativa de señales complejas.

2.2. Interacción de la luz con la materia

La radiación electromagnética puede experimentar diversos fenómenos al incidir sobre un material, entre ellos: reflexión, absorción, dispersión y emisión [16]. En la reflexión, la luz incide sobre una superficie y se desvía, conservando su frecuencia original. En la absorción, la energía del fotón se transfiere a la materia, excitando transiciones electrónicas, vibracionales o rotacionales, dependiendo de la energía involucrada. Por el contrario, la emisión ocurre cuando un sistema previamente excitado retorna a un estado de menor energía, liberando fotones; ejemplos comunes de este fenómeno son la fluorescencia y la fosforescencia [16].

La dispersión, por su parte, implica la redirección de la luz debido a interacciones con la materia. Esta puede ser elástica, cuando no hay intercambio neto de energía entre el fotón y el sistema (como en la dispersión Rayleigh), o inelástica, cuando el fotón pierde o gana energía (como en la dispersión Raman) [1, 16].

Estos mecanismos de interacción luz-materia dependen directamente de la estructura electrónica y vibracional del material, y constituyen la base de múltiples técnicas espectroscópicas utilizadas para el análisis molecular [15].

Cada tipo de movimiento molecular —electrónico, vibracional o rotacional— interactúa con fotones de diferentes regiones del espectro electromagnético. Por ejemplo, si la energía del fotón coincide con la diferencia entre dos niveles vibracionales, puede ser absorbido (absorción IR); o, en su defecto, dispersarse inelásticamente (efecto Raman), transfiriendo energía a ese modo vibracional [1, 16].

En conjunto, la interacción entre la luz y la materia abarca procesos fundamentales mediante los cuales los fotones pueden ser reflejados, absorbidos o dispersados, proporcionando una base teórica para las técnicas espectroscópicas que permiten caracterizar materiales a nivel molecular [15, 1].

2.3. Fundamentos de la espectroscopía Raman

La espectroscopía Raman se basa en la dispersión inelástica de la luz cuando un material es iluminado con una fuente monocromática intensa, típicamente un láser. La mayoría de los fotones incidentes se dispersan de forma elástica, sin cambiar su frecuencia (dispersión de Rayleigh); sin embargo, una fracción muy pequeña (~ 1 en 10^6) interactúa con los modos vibracionales de la muestra y se dispersa con una frecuencia diferente, fenómeno conocido como efecto Raman [1].

Cuando el fotón cede energía a un modo vibracional, su frecuencia disminuye (Stokes), mientras que si la molécula excitada transfiere energía al fotón, este se dispersa con mayor frecuencia (anti-Stokes). A temperatura ambiente, predominan las líneas Stokes debido a que la mayoría de las moléculas se encuentran en su estado vibracional fundamental, mientras que las anti-Stokes son menos intensas por la baja población de estados excitados [1].

Por su parte, la línea de Rayleigh, correspondiente a los fotones sin cambio de energía, suele ser mucho más intensa. Para que una vibración sea activa en Raman, debe implicar un cambio en la polarizabilidad de la molécula durante la vibración. Este criterio de selección difiere del de la espectroscopía infrarroja (IR), que requiere un cambio en el momento dipolar [1].

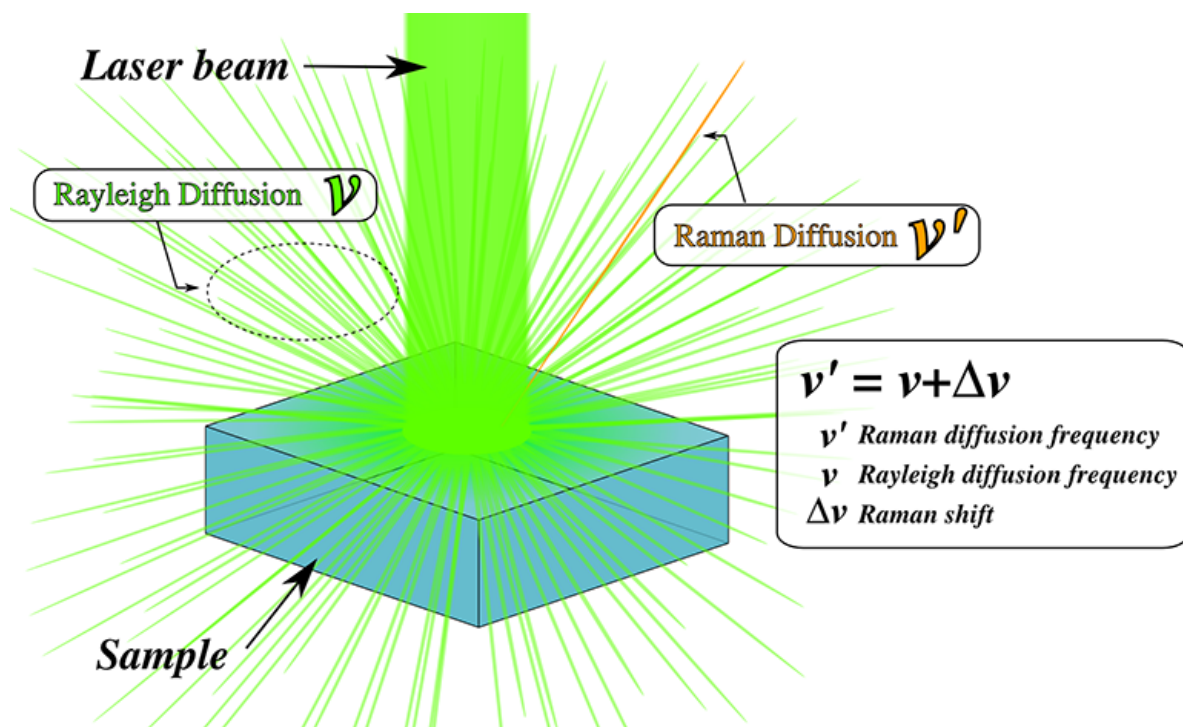


Fig. 2.1: Representación de dispersión Raman y Rayleigh [1].

De este modo, ambas técnicas son complementarias: vibraciones que son inactivas en IR pueden ser observadas en Raman, y viceversa. El agua, a pesar de tener un momento dipolar elevado, posee baja polarizabilidad, lo que la convierte en un dispersor débil en Raman. Paradójicamente, esto resulta ventajoso, ya que permite realizar análisis en soluciones acuosas sin una señal Raman dominante del disolvente [15, 1].

En resumen, el efecto Raman permite detectar vibraciones moleculares específicas mediante la dispersión inelástica, generando un “desplazamiento Raman” característico de cada enlace, y proporcionando información estructural a nivel molecular [15, 1].

2.4. Espectro Raman del agua y efecto del NaCl

El agua líquida presenta un espectro Raman caracterizado principalmente por una amplia y asimétrica banda de estiramiento O–H, ubicada entre 2800 y 3800 cm^{-1} , centrada alrededor de los 3350 cm^{-1} en condiciones ambientales. Esta forma ancha y asimétrica refleja la diversidad estructural del agua líquida, en la que las moléculas de H_2O establecen una red dinámica de puentes de hidrógeno. Algunas moléculas poseen enlaces O–H fuertemente cohesionados, que vibran a frecuencias más bajas ($\sim 3220 \text{ cm}^{-1}$), mientras que otras exhiben enlaces más débiles o “libres”, vibrando a frecuencias más altas ($\sim 3400 \text{ cm}^{-1}$) [15].

La posición y forma de esta banda Raman O–H son sensibles a variables como la temperatura y la presencia de solutos. Por ejemplo, al aumentar la temperatura, la banda se estrecha y se desplaza ligeramente hacia frecuencias mayores, debido a la ruptura parcial de puentes de hidrógeno y a la mayor movilidad de las moléculas, que vibran a frecuencias más altas [15].

De manera análoga, la adición de sales como el NaCl también provoca modificaciones estructurales detectables en el espectro Raman del agua. Diversos estudios han demostrado que al aumentar la concentración de NaCl en solución, la banda O–H se vuelve más estrecha y su perfil se transforma significativamente. En particular, se observa una disminución de la intensidad en la región de baja frecuencia ($<3220 \text{ cm}^{-1}$), atribuida a la reducción de enlaces O–H fuertes, y un aumento relativo en la región de alta frecuencia ($>3220 \text{ cm}^{-1}$), correspondiente a enlaces más débiles [15].

Estos cambios indican que los iones disueltos (Na^+ y Cl^-) interactúan con las moléculas de agua mediante solvatación, lo que debilita algunos de los puentes de hidrógeno más fuertes y favorece estados vibracionales menos ligados. Como resultado, el espectro Raman del agua es altamente sensible tanto a la salinidad como a la temperatura, ya que ambos factores modifican la red de enlaces de hidrógeno [15].

Esta sensibilidad espectral ha sido utilizada con fines analíticos. Por ejemplo, al calibrar el corrimiento de la banda O–H o sus variaciones de forma, es posible determinar con buena precisión la concentración de sal disuelta en una muestra. En esta investigación, nos centramos en la región espectral entre 2800 y 3800 cm^{-1} , correspondiente al estiramiento O–H, ya que allí se manifiestan de manera clara los efectos estructurales

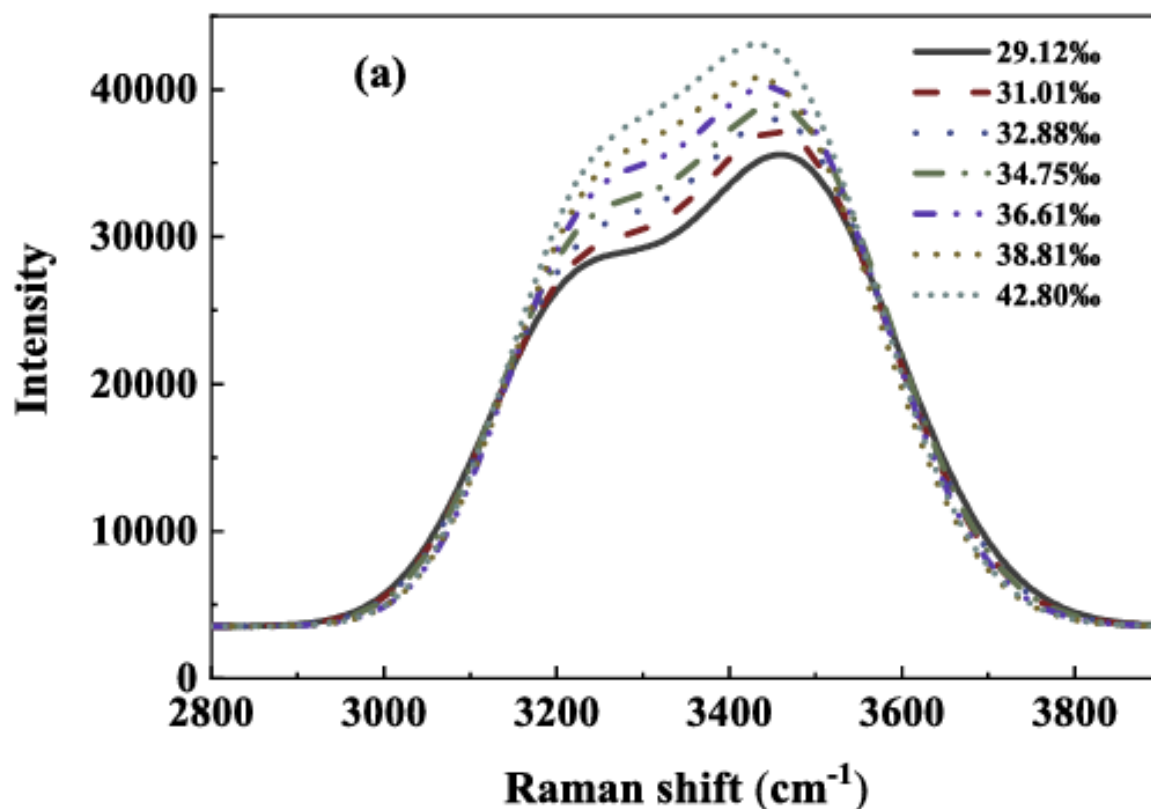


Fig. 2.2: Espectro Raman de disolución de NaCl con diferentes concentraciones [2].

inducidos por el NaCl, evitando al mismo tiempo interferencias provenientes de otras bandas no relevantes [15].

2.5. Análisis espectral mediante funciones Gaussianas

Para obtener una interpretación detallada de la banda O–H del agua observada en el espectro Raman, es habitual descomponer dicha banda en componentes Gaussianas, cada una representando diferentes entornos de enlace de hidrógeno. Esta estrategia se fundamenta en que la banda amplia y heterogénea registrada experimentalmente puede modelarse como una superposición de contribuciones de distintas subpoblaciones de moléculas, cada una con una frecuencia vibracional promedio y cierta dispersión [3].

Desde el punto de vista matemático, la suma de funciones Gaussianas (σ ,

alternativamente, Lorentzianas) permite reproducir con precisión la forma observada del espectro Raman del agua [3]. Típicamente se emplean dos o más Gaussianas: una centrada alrededor de 3200 cm^{-1} , correspondiente a moléculas con fuertes enlaces de hidrógeno, que forman estructuras altamente cohesionadas similares al hielo; y otra entre 3400 y 3450 cm^{-1} , que representa moléculas con enlaces más débiles o incompletos, es decir, agua “libre” o de menor coordinación [1]. En algunos casos, se añade una tercera componente hacia los 3600 cm^{-1} para representar estiramientos O–H no enlazados.

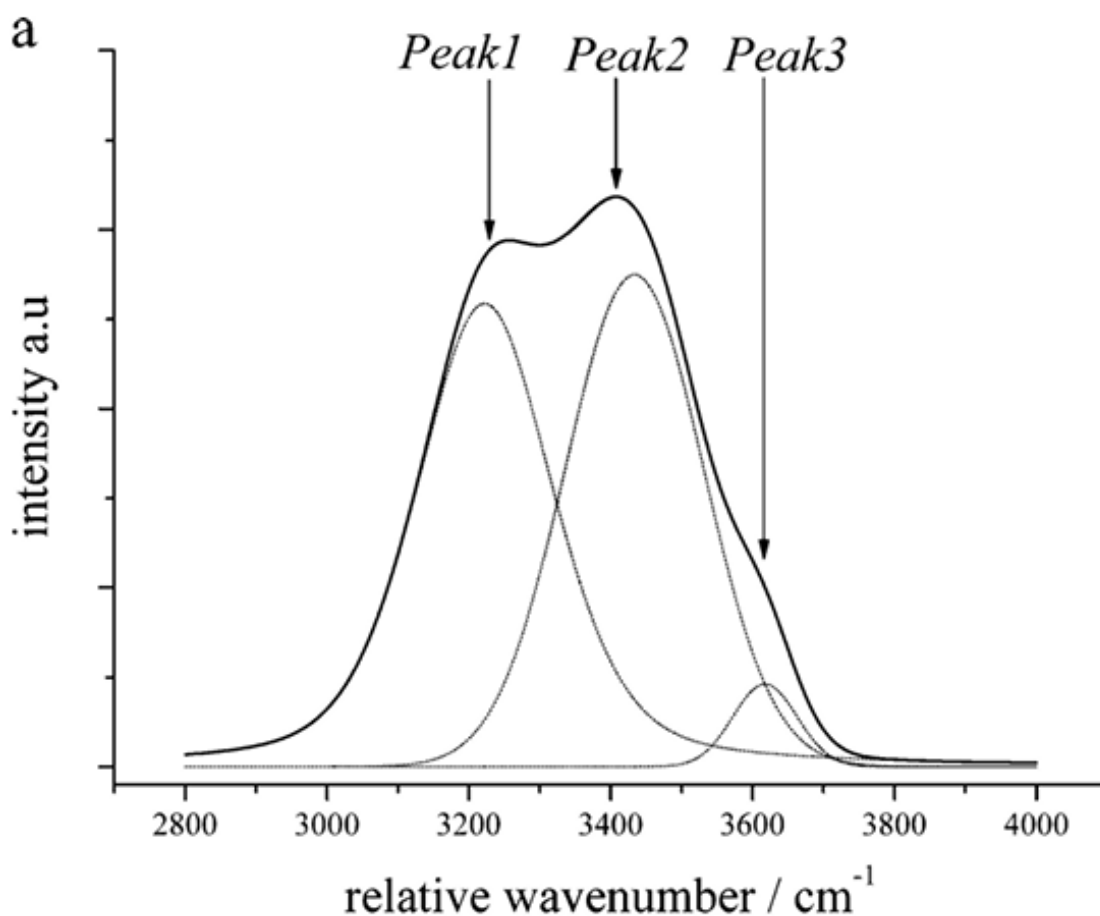


Fig. 2.3: Descomposición en funciones Gaussianas [3].

Cada parámetro ajustado en las funciones Gaussianas tiene un significado fisicoquímico relevante. La posición central de la Gaussiana indica la energía vibracional típica del entorno molecular correspondiente: una frecuencia menor implica enlaces más fuertes, mientras que una mayor indica una vibración menos confinada [1]. El ancho de la Gaussiana (por ejemplo, el ancho a mitad de altura, FWHM) refleja la dispersión de entornos o las fluctuaciones dinámicas del sistema: un valor más amplio sugiere mayor

heterogeneidad estructural, mientras que un valor estrecho indica uniformidad en los enlaces. Finalmente, el área bajo la curva representa la población relativa de moléculas asociadas a ese entorno de enlace; así, el área normalizada de cada componente se puede interpretar como la fracción de enlaces de cierto tipo presentes en la muestra [1, 3, 14].

Estos parámetros espectrales pueden correlacionarse con condiciones físico-químicas externas, como temperatura o concentración salina. En particular, se ha observado que al aumentar la concentración de NaCl, el área de la componente centrada en 3200 cm^{-1} disminuye, mientras que crece la de 3450 cm^{-1} . Este comportamiento sugiere una disminución de las moléculas con enlaces de hidrógeno fuertes y un aumento de aquellas con interacciones más débiles, como consecuencia de la solvatación de los iones disueltos [1].

Por tanto, la descomposición espectral mediante funciones Gaussianas ofrece una herramienta poderosa para traducir parámetros espectrales en información sobre la microestructura del agua. En esta investigación, empleamos este enfoque para ajustar la banda O–H y evaluar sistemáticamente la evolución de la posición, ancho y área de las sub-bandas en función de la salinidad. De este modo, es posible establecer una conexión directa entre los cambios observados en el espectro Raman y las interacciones moleculares inducidas por el NaCl.

2.6. Preprocesamiento de datos espectrales

Antes de aplicar modelos cuantitativos, los espectros Raman brutos deben ser preprocesados con el fin de mejorar su calidad analítica y asegurar la comparabilidad entre muestras. El primer paso en este proceso es la corrección de línea base, cuyo objetivo es eliminar el fondo instrumental o fluorescente que se superpone a la señal Raman real. Esta corrección se realiza identificando regiones libres de bandas Raman (donde sólo hay fondo) y ajustando una línea base —ya sea recta o un polinomio de bajo orden— que se sustrae del espectro completo. El resultado es un espectro “nivelado”, en el que las bandas Raman oscilan alrededor de una intensidad base cercana a cero [15, 1].

Luego se realiza la normalización del espectro, que busca corregir las diferencias de intensidad global originadas por factores ajenos a la composición química (como

fluctuaciones en la intensidad del láser, diferencias en el volumen de muestra o trayectoria óptica). Entre los métodos más comunes se encuentra la normalización al área total (integral del espectro) o a la intensidad máxima [9, 1].

Un tercer paso esencial es el suavizado o filtrado del ruido. Dado que los espectros Raman, especialmente en bajas concentraciones, pueden contener ruido aleatorio, se emplean métodos de procesamiento digital para mejorar la relación señal/ruido sin distorsionar las bandas espectrales. Una estrategia efectiva es el filtrado por transformada de Fourier, que convierte el espectro al dominio de frecuencias (no confundir con unidades de cm^{-1}). En este dominio, las señales Raman aparecen como componentes de baja frecuencia (formas amplias), mientras que el ruido aleatorio se manifiesta en componentes de alta frecuencia. Suprimiendo o atenuando estas últimas (por ejemplo, mediante umbrales de corte) y aplicando la transformada inversa, se obtiene un espectro suavizado que conserva las características esenciales.

Este método es funcionalmente equivalente a aplicar un filtro pasa-bajos digital. Además, se puede recurrir al método clásico de Savitzky–Golay, que consiste en ajustar localmente un polinomio sobre segmentos del espectro para reemplazar cada punto por el valor suavizado, preservando la posición de los picos [1].

Finalmente, se realiza la selección de la ventana espectral de interés, en este caso entre 2800 y 3800 cm^{-1} . Este recorte excluye regiones sin señal relevante o dominadas por ruido, reduciendo así la dimensionalidad de los datos a la porción espectral más sensible a la concentración de NaCl [15].

2.7. Fundamentos de la regresión Partial Least Squares (PLS)

Dado que un espectro Raman contiene cientos o miles de variables (intensidades en distintos desplazamientos) que están altamente correlacionadas entre sí, la regresión por mínimos cuadrados parciales (PLS, por sus siglas en inglés) se presenta como una herramienta ideal para modelar cuantitativamente la concentración de analitos como el NaCl. Este enfoque pertenece a la familia de métodos de variables latentes: en lugar de emplear todos los predictores originales de forma directa, PLS proyecta los datos de

entrada \mathbf{X} (matriz de variables espectrales) y la variable de salida \mathbf{Y} (concentración) en un espacio de menor dimensión [17].

En este nuevo espacio, se definen componentes latentes o factores que son combinaciones lineales de las variables originales de \mathbf{X} , seleccionadas para maximizar la covarianza conjunta con \mathbf{Y} . A diferencia del análisis de componentes principales (PCA), que solo optimiza la varianza de \mathbf{X} sin considerar \mathbf{Y} , la regresión PLS busca aquellas direcciones en los datos que mejor explican las variaciones en la variable dependiente [17].

Usualmente, bastan unos pocos factores para capturar la mayor parte de la variabilidad relevante. Además, estos factores pueden interpretarse químicamente al analizar sus pesos espectrales: por ejemplo, un factor PLS podría estar asociado a la intensidad de una banda Raman que varía con la concentración de NaCl. Esto permite vincular el modelo matemático con fenómenos fisicoquímicos reales, aunque los factores sean abstracciones de múltiples variables.

Un aspecto esencial del modelado PLS es su validación. Para evitar el sobreajuste, se emplea validación cruzada o conjuntos de prueba independientes, con el objetivo de determinar el número óptimo de componentes a utilizar. El procedimiento común es entrenar modelos con diferentes números de factores y elegir aquel que minimice el error de validación, usualmente medido como RMSECV (*Root Mean Square Error of Cross-Validation*) [17].

Capítulo 3

Desarrollo

Esta sección describe el trabajo experimental y computacional realizado en la presente memoria de título, con el propósito de construir un modelo preliminar que permita estimar la concentración de NaCl en solución acuosa a partir de espectros Raman. El desarrollo se estructura en seis etapas principales: (1) calibración del espectrómetro, (2) preparación de soluciones, (3) adquisición de espectros Raman, (4) preprocesamiento de datos espectrales, (5) análisis espectral, y (6) modelado y evaluación con Partial Least Square (PLS).

3.1. Calibración del espectrómetro

El primer objetivo es transformar las unidades entregadas por el espectrómetro (intensidad vs. número de píxel) a un eje de longitud de onda absoluto en nanómetros (nm). Se emplearon dos lámparas de emisión; Lámpara de mercurio (Hg) y Lámpara de neón (Ne). Ambas fuentes fueron medidas por separado y luego sus espectros fueron combinados para generar un espectro conjunto de calibración.

En la figura 3.1 se puede observar en color rosado el espectro Raman de la lámpara de mercurio y en color azul el de la lámpara de neón en (intensidad vs. número de píxel). Posteriormente, se alinearon y combinaron en un solo espectro de referencia 3.2.

Cada pico de emisión fue ajustado individualmente mediante una interpolación parabólica y un ajuste Gaussiano. Esto permitió determinar la posición central de cada

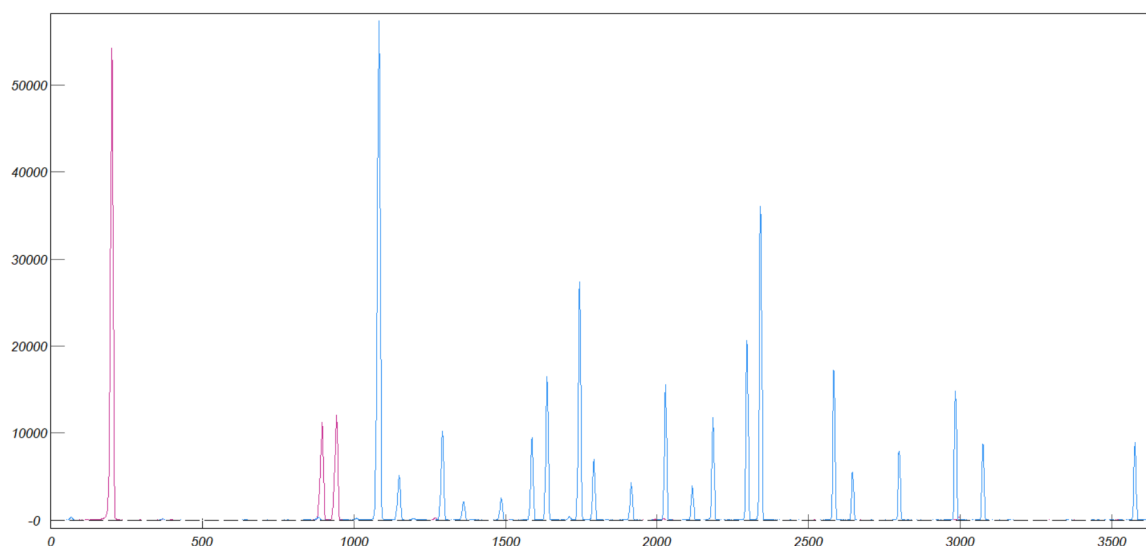


Fig. 3.1: Espectro Raman del mercurio y del neón.

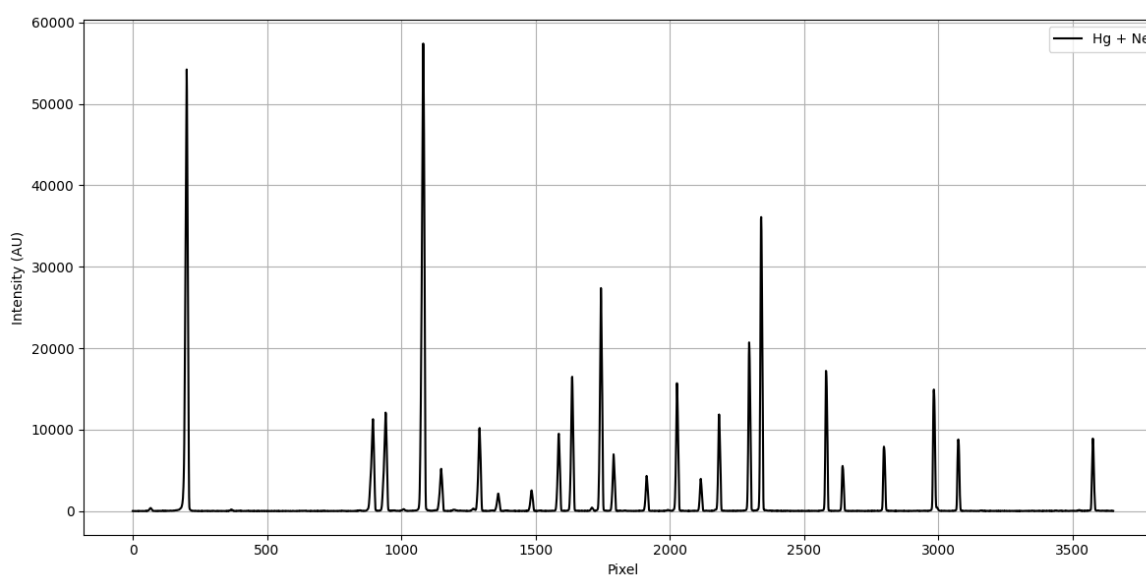


Fig. 3.2: Espectro Raman del mercurio y del neón combinados.

pico en unidades subpíxel 3.3.

Tabla 3.1: Comparación de parámetros espectrales ajustados al Hg.

Pico original	Intensidad original	Pico parabólico	Int. parabólica	Pico Gaussiano	Int. Gaussiana
201.0	54233.299	201.019609	52587.541	201.026064	52921.523
942.0	12073.303	942.234182	11789.891	942.255582	11852.266
895.0	11270.327	894.797844	11018.740	894.787044	11082.017

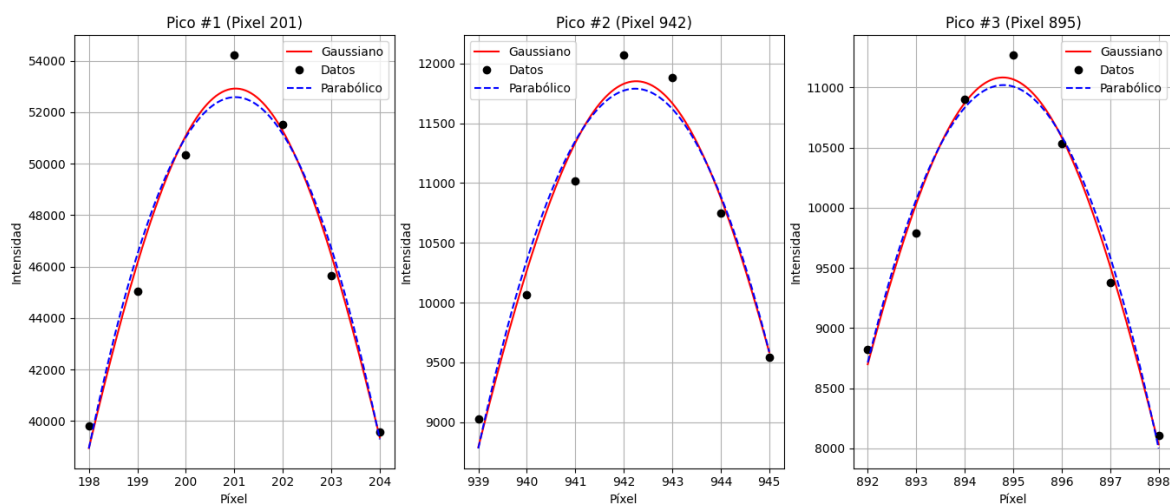


Fig. 3.3: Ajuste subpíxel de los 3 peaks del mercurio.

Se identificaron manualmente las longitudes de onda de emisión conocidas para cada línea espectral (según las lámparas de Hg y Ne), las cuales se asociaron a sus respectivos subpíxeles.

Se probaron tres modelos de regresión; Lineal, Cuadrático y Cúbico. La Figura 3.4 muestra el ajuste de cada modelo sobre los datos experimentales. El modelo cúbico presentó el menor error cuadrático medio ($RMSE = 0.0019$ nm), por lo que fue seleccionado como el modelo óptimo para la calibración del eje de píxeles a nm. Aplicando el modelo cúbico ajustado, se transformó el eje de píxeles del espectro a longitud de onda en nm. Esta transformación se visualiza en la Figura 3.5.

Para evaluar la resolución instrumental, se midió el ancho completo a media altura (FWHM) de un pico representativo. Como se muestra en la figura 3.6, el valor de FWHM obtenido fue de 0.51 nm, lo que caracteriza la resolución espectral efectiva del sistema en esa región.

Una vez calibrado el espectrómetro en longitud de onda absoluta (nm), es necesario convertir la escala del eje horizontal a desplazamiento Raman (shift en cm^{-1}).

Se partió de un espectro de referencia (azufre-polietileno estándar), al cual se aplicó una corrección de línea base utilizando un algoritmo de 'baseline als' 3.7.

Se identificaron manualmente los principales picos Raman presentes en el espectro, cada

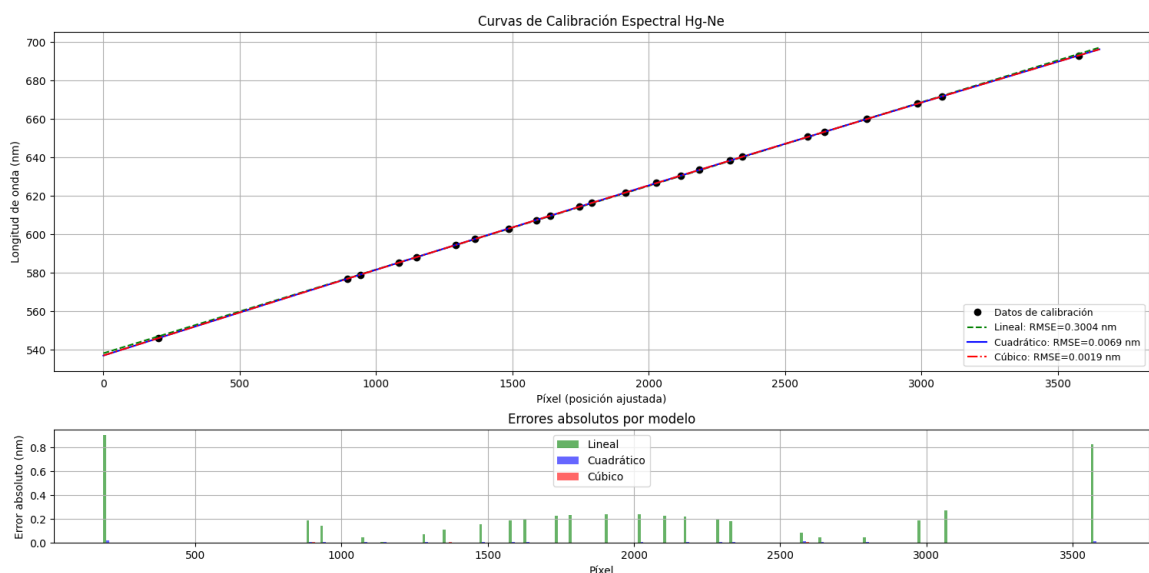


Fig. 3.4: Curvas de calibración con lámparas de Hg y Ne.

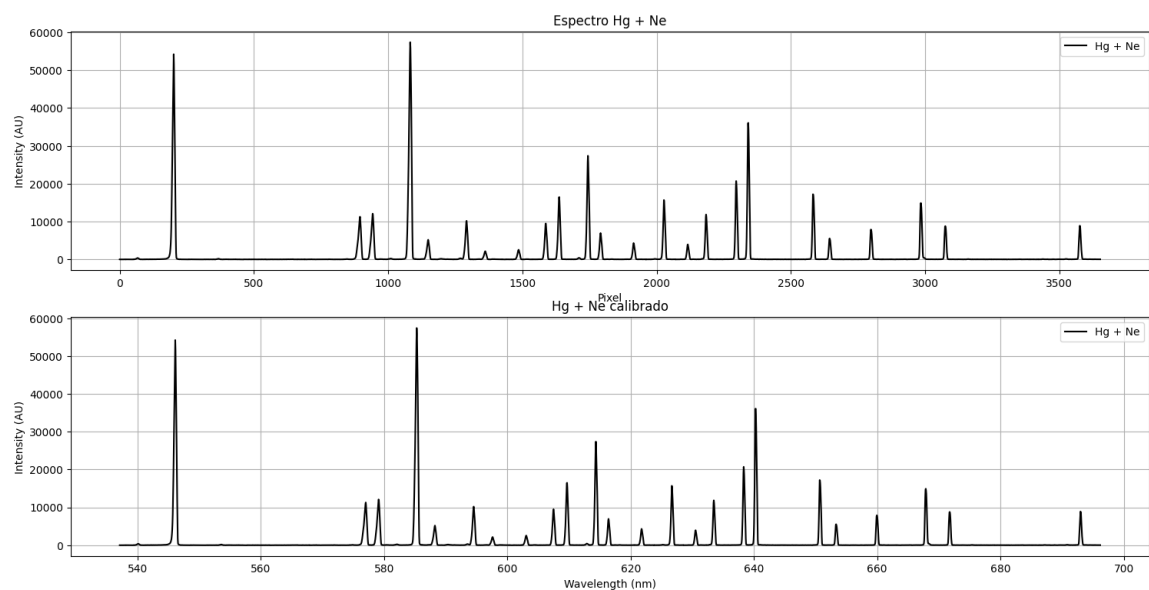


Fig. 3.5: Comparación de espectro en número de píxeles y calibrado en nm.

uno fue ajustado por un modelo Gaussiano para localizar el subpíxel. A cada pico se le asignó el valor teórico en cm^{-1} . Con los pares (píxel, desplazamiento Raman), se generó una curva de calibración ajustando polinomios de distintos grados. La figura 3.8 muestra las curvas de calibración. La elegida fue el polinomio cúbico con un $RMSE = 0,99\text{cm}^{-1}$. El modelo cúbico fue utilizado para convertir todo el eje de píxeles del espectro corregido a unidades de desplazamiento Raman, obteniendo así el espectro calibrado final 3.9.

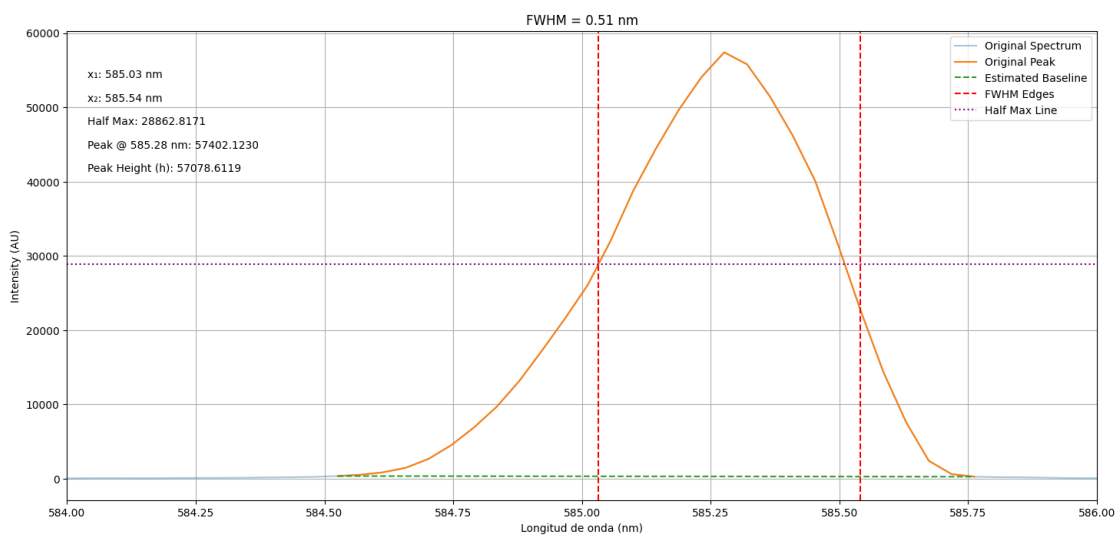


Fig. 3.6: Cálculo de FWHM.

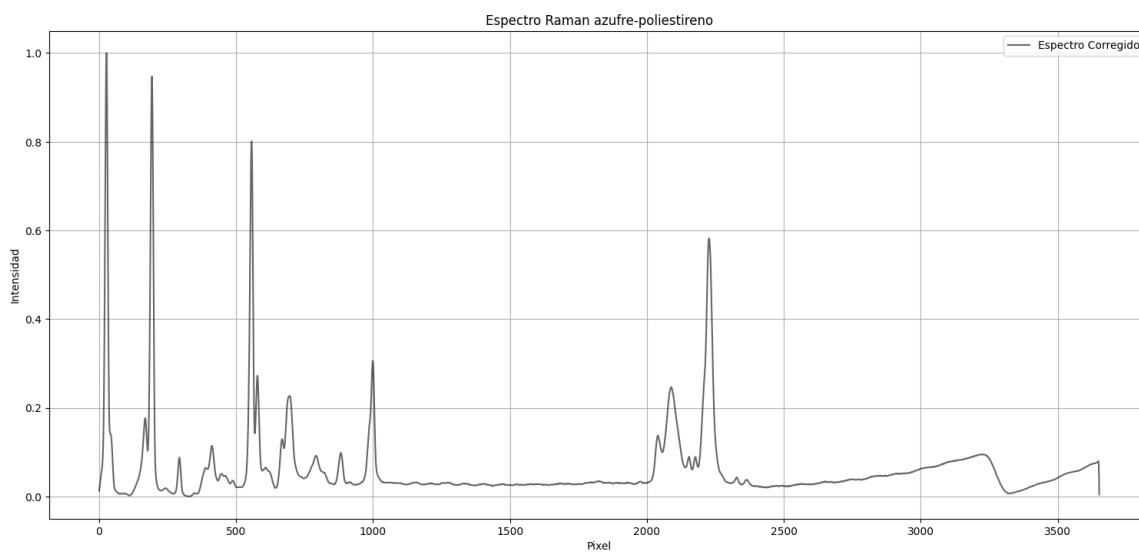


Fig. 3.7: Espectro de referencia azufre-polietileno estándar.

Además se evaluó el ancho completo a media altura (FWHM) del pico en 1000.99 cm^{-1} , resultando en un valor de 18.44 cm^{-1} , ver figura 3.10.

Gracias a este proceso, se logró establecer una escala de referencia en unidades absolutas en nanómetros y en unidades Raman (cm^{-1}).

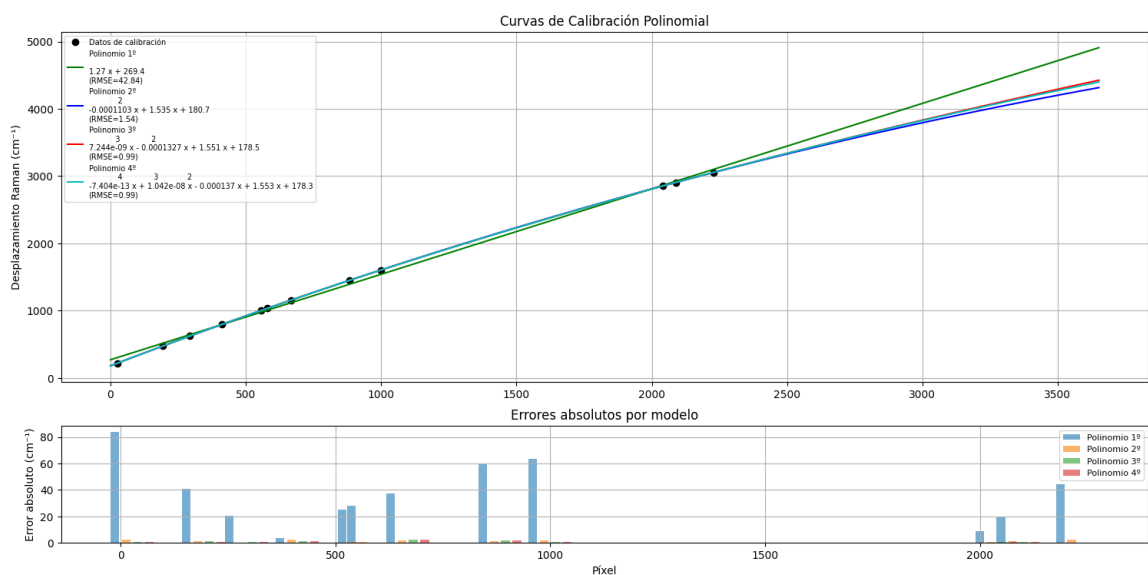


Fig. 3.8: Curvas de calibración para desplazamiento Raman.

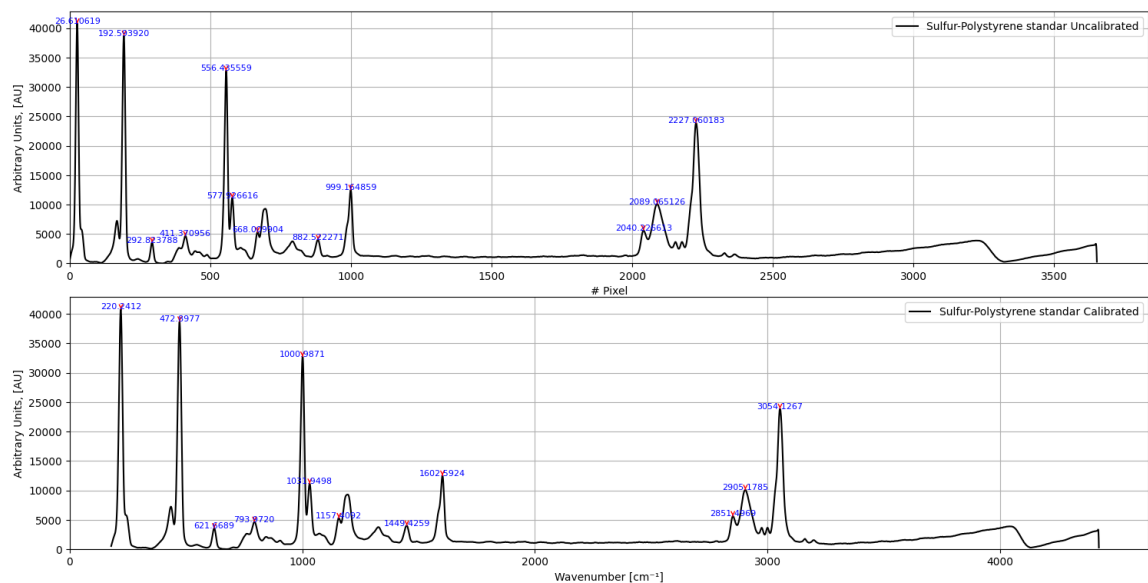


Fig. 3.9: Comparación de espectro en número de píxel y calibrado en cm^{-1} .

3.2. Preparación de soluciones estándar de NaCl

El objetivo de esta etapa fue preparar soluciones acuosas de cloruro de sodio (NaCl) con concentraciones conocidas, empleando procedimientos controlados y cuantificables, que sirvieran como base para la adquisición de espectros Raman y la posterior modelación de la concentración. La masa de NaCl fue determinada utilizando una balanza analítica

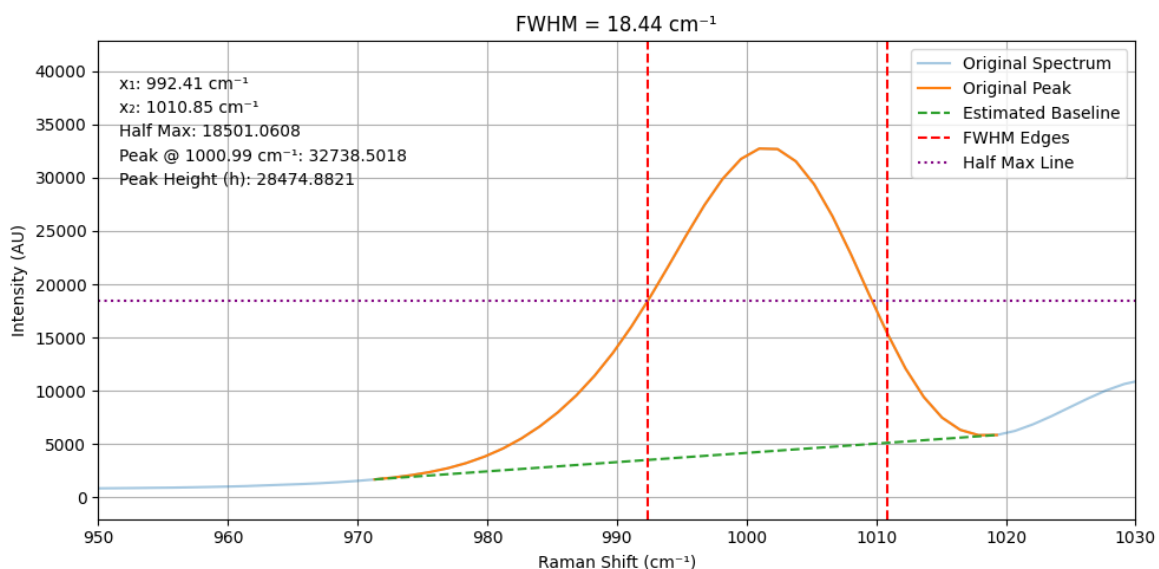


Fig. 3.10: FWHM en cm⁻¹.

de alta precisión (± 0.1 mg). Las masas pesadas fueron:

Solución	Masa de NaCl [g]
1 g	1.0004
2 g	2.0003
3 g	3.0003
4 g	4.0006
5 g	5.0000

Tabla 3.2: Masa pesada de NaCl en cada solución preparada.

Cada masa fue transferida a un vaso de precipitados limpio y seco. Cada muestra fue disuelta en 100 mL de agua desionizada, medida cuidadosamente con una probeta graduada (precisión ± 1.0 mL). La homogeneización se realizó mediante agitación ultrasonido para asegurar la completa disolución del soluto. Además, se preparó una muestra de agua desionizada pura como referencia (0 g de NaCl).

Durante la preparación se identificaron las siguientes fuentes de error:

- **Balanza analítica:** resolución de 0.1 mg.
- **Material volumétrico:** probeta con una tolerancia de ± 1.0 mL.
- **Pureza del NaCl:** impurezas.

- **Variabilidad del conductivímetro:** observada experimentalmente a partir de las tres réplicas realizadas por muestra.

3.3. Adquisición de espectros Raman

El objetivo en esta etapa es obtener espectros Raman representativos de cada solución salina a distintas temperaturas controladas, con condiciones de adquisición constantes, para su posterior análisis y modelado cuantitativo.

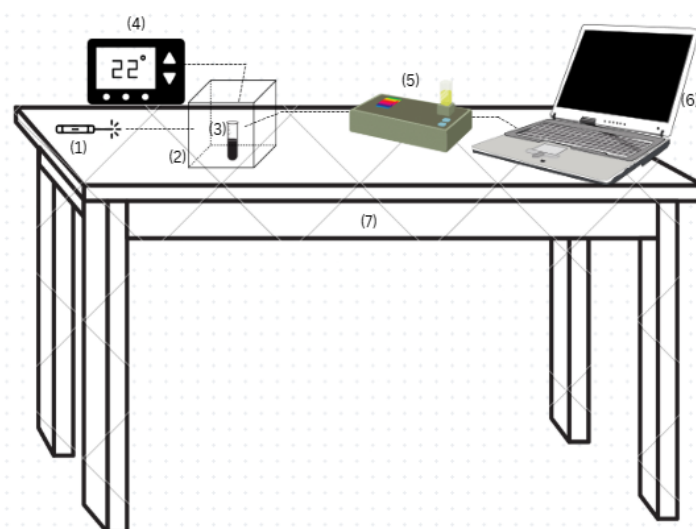


Fig. 3.11: Configuración experimental: (1) láser, (2) soporte de cubeta, (3) cubeta, (4) controlador de temperatura, (5) espectrómetro Almad de Thunder Optics, (6) computador portátil, (7) mesa óptica.

La disposición experimental (Fig. 3.11) consta de un láser de alta estabilidad que opera a 532,10 nm. El haz incide sobre una cubeta de cuarzo alojada en un soporte de cubeta termo-controlada; el espectrómetro Almad de Thunder Optics se sitúa a 90° respecto al haz incidente. El espectrómetro cubre el intervalo 537,08–696,18 nm con resolución $\Delta\lambda \approx 0,043$ nm. Para suprimir retrodispersión del láser se emplea un filtro de 533 nm.

Para cada muestra (agua desionizada, 1 g, 2 g, 3 g, 4 g y 5 g de NaCl), se realizaron mediciones bajo el siguiente protocolo:

1. Establecer la temperatura en 10 °C y, con el láser apagado, registrar un espectro de ruido ($t_{\text{int}} = 120$ s) para su posterior sustracción.

2. Encender el láser y capturar cuatro espectros a esa temperatura.
3. Repetir el paso anterior a 20, 30, 40 y 50 °C.

Para prevenir contaminación cruzada entre muestras, la cubeta fue enjuagada tres veces con la disolución correspondiente antes de cada nueva medición.

Para cada una de las 6 soluciones, se obtuvieron 4 espectros por temperatura. Como se tienen 5 temperaturas, se generó una base de datos de 120 espectros.

3.4. Preprocesamiento de los espectros

El objetivo del preprocesamiento espectral es optimizar los espectros Raman obtenidos para su posterior análisis cuantitativo, mediante recorte espectral y atenuación del ruido de alta frecuencia, manteniendo las características físicas y químicas relevantes de la señal.

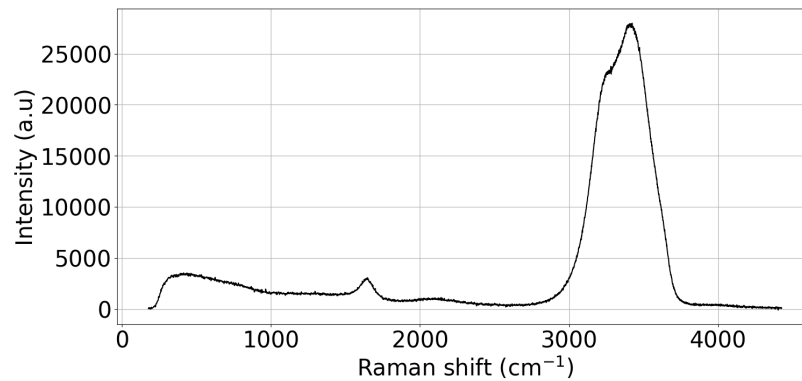
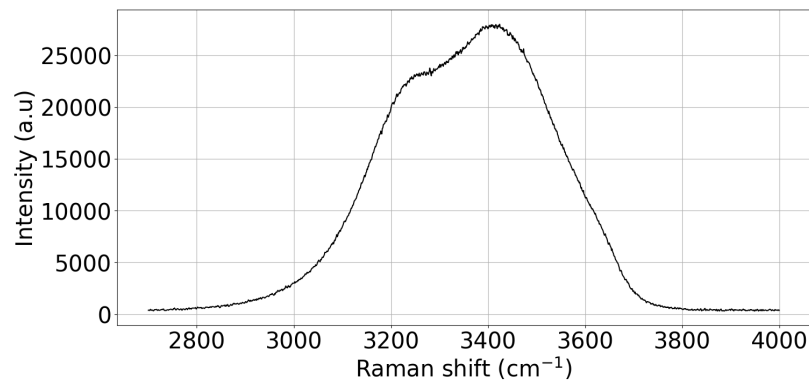
Cada espectro adquirido fue transformado desde el dominio de píxel al dominio espectral absoluto (cm^{-1}), utilizando la curva de calibración descrita en la Etapa 1. Posteriormente, se extrajo el intervalo comprendido entre 2800 y 3800 cm^{-1} , correspondiente a la banda de estiramiento O–H.

Esta banda es de particular interés por ser altamente sensible a:

- La concentración de iones disueltos (como Na^+ y Cl^-), debido a la perturbación que estos inducen en la red de enlaces de hidrógeno del agua.
- La temperatura, que influye sobre la intensidad y desplazamiento de las subbandas vibracionales a través del debilitamiento de los enlaces de hidrógeno y el aumento del desorden molecular.

El recorte del espectro no solo reduce la dimensionalidad del problema, sino que enfoca el análisis en la región espectral con mayor información relevante para el sistema $\text{NaCl-H}_2\text{O}$.

Para reducir el ruido de alta frecuencia sin afectar significativamente la información vibracional útil del espectro, se aplicó el siguiente procedimiento:

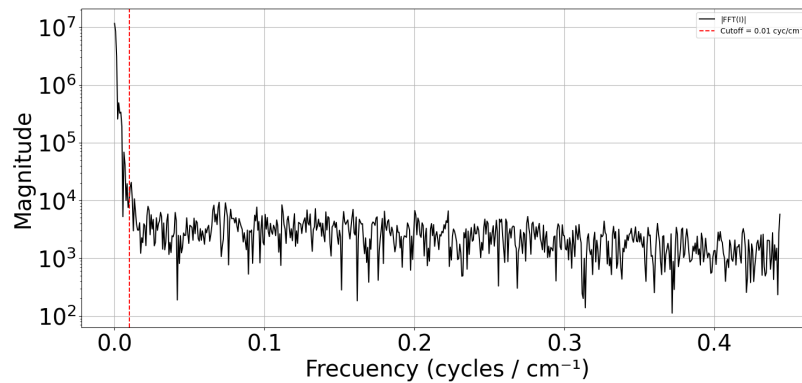
(a) Espectro crudo en cm^{-1} de agua pura a 50°C .

(b) Banda O–H recortada.

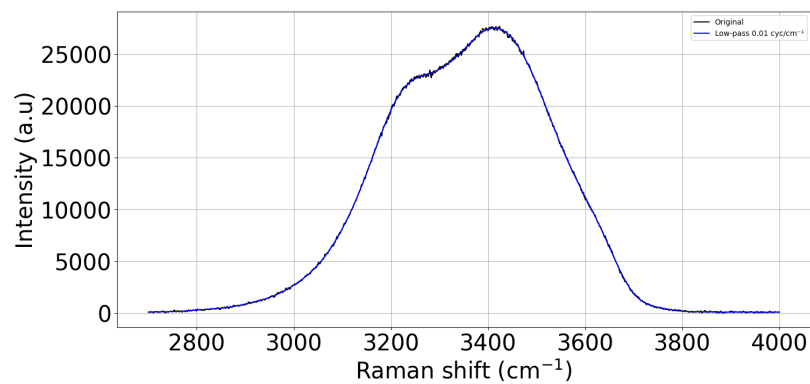
Fig. 3.12: Comparación entre espectro completo y región O–H.

1. **Transformada directa (FFT):** cada espectro fue transformado al dominio de la frecuencia mediante FFT, lo que permitió identificar la contribución de cada componente frecuencial, ver figura 3.13a.
2. **Filtrado por umbral:** se fijó un umbral de amplitud en 0,01. Todos los coeficientes de frecuencia con amplitud superior a este valor fueron suprimidos, es decir, eliminados del espectro de Fourier. Esta operación corresponde a un filtro pasa-bajos.
3. **Reconstrucción (IFFT):** se aplicó la transformada inversa de Fourier (IFFT) para volver al dominio espectral. El resultado, mostrado en la figura 3.13b, es una comparación de la versión suavizada del espectro vs. el espectro original, con reducción eficaz del ruido y preservación de las formas del espectro.

A diferencia de técnicas basadas en suavizado por convolución (Savitzky–Golay), el



(a) Espectro de Fourier.



(b) Espectro original vs. espectro filtrado.

Fig. 3.13: Filtrado FFT de los espectros.

filtrado por FFT no distorsiona las posiciones de los picos.

- Permite controlar explícitamente el contenido frecuencial que se conserva, lo que es especialmente útil en análisis espectrales donde la forma y el ancho de banda son importantes.
- Este preprocesamiento permitió generar una base de datos depurada y consistente, en la que cada espectro se encuentra centrado en la banda vibracional de interés y libre de componentes asociadas al ruido.

3.5. Análisis espectral para caracterización de las muestras

El objetivo es descomponer la banda de estiramiento (O–H) en componentes interpretables desde el punto de vista físico-químico, con el fin de caracterizar el efecto del NaCl sobre la estructura del agua y generar un conjunto de variables resumidas para el posterior modelado estadístico.

La banda O–H en espectros Raman de soluciones acuosas presenta una forma ancha y asimétrica, producto de la superposición de distintas configuraciones de enlaces de hidrógeno. Esta banda puede interpretarse como la suma de tres contribuciones Gaussianas asociadas a diferentes subpoblaciones de moléculas de agua:

- $\sim 3220 \text{ cm}^{-1}$ → moléculas fuertemente enlazadas,
- $\sim 3400 \text{ cm}^{-1}$ → moléculas con enlaces parcialmente rotos,
- $\sim 3600 \text{ cm}^{-1}$ → moléculas débilmente enlazadas o casi libres.

Cada espectro fue ajustado como la suma de tres funciones Gaussianas:

$$I(\lambda_i) = \sum_{i=1}^3 A_i \cdot \exp\left(-\frac{(\lambda_i - \mu_i)^2}{2\sigma_i^2}\right) \quad (3.1)$$

donde:

- A_i : amplitud (relacionada con la población relativa de la subpoblación i),
- μ_i : posición central (tipo de enlace O–H),
- σ_i : anchura (dispersión espectral, relacionada con el desorden estructural).

Parámetros iniciales del ajuste:

$$\begin{aligned} \mu_i^{(0)} &= \{3200, 3400, 3600\} \text{ cm}^{-1} \\ \sigma_i^{(0)} &= \{50, 70, 40\} \text{ cm}^{-1} \end{aligned}$$

Las amplitudes iniciales $A_i^{(0)}$ fueron tomadas directamente como la intensidad experimental en cada $\mu_i^{(0)}$.

Queremos minimizar la función de error cuadrático entre los valores observados $I(\lambda_i)$ y los valores ajustados $I_{\text{modelo}}(\lambda_i; \boldsymbol{\theta})$:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^3 [I(\lambda_i) - I_{\text{modelo}}(\lambda_i; \boldsymbol{\theta})]^2 \quad (3.2)$$

Donde $\boldsymbol{\theta}$ es el vector de parámetros del modelo:

$$\boldsymbol{\theta} = [A_1, \mu_1, \sigma_1, A_2, \mu_2, \sigma_2, A_3, \mu_3, \sigma_3] \quad (3.3)$$

Este método busca de forma iterativa el conjunto de parámetros que minimiza la suma de errores al cuadrado.

Es importante señalar que el significado de cada función Gaussiana no deriva del ajuste matemático en sí, sino de cómo dicha componente se comporta frente a cambios controlados del entorno (concentración salina, temperatura).

Esta descomposición espectral permite interpretar directamente cómo el entorno (presencia de NaCl, temperatura) afecta la red de enlaces de hidrógeno del agua.

Parámetro	Significado físico-químico
A_i	Población relativa de moléculas de agua con cierto tipo de enlace.
μ_i	Tipo de enlace O–H predominante (más débil \rightarrow mayor μ_i).
σ_i	Grado de desorden molecular, anchura de distribución vibracional

Tabla 3.3: Significado físico-químico de los parámetros de ajuste Gaussiano.

Esta parametrización reduce cada espectro de cientos de puntos a 9 variables numéricas con significado físico-químico claro, lo que facilita el análisis estadístico y entrenamiento del modelo, mejora la interpretabilidad del modelo predictivo y permite estudiar de forma sistemática los efectos de la salinidad y temperatura sobre el agua.

3.6. Modelado con Partial Least Square (PLS)

El objetivo es construir un modelo estadístico robusto capaz de predecir la concentración de NaCl en disolución acuosa a partir de información espectral resumida en nueve parámetros físicos, utilizando la técnica de regresión PLS.

A partir de cada uno de los 120 espectros Raman, se obtuvieron los parámetros del ajuste a tres funciones Gaussianas, generando un vector de características de 9 dimensiones para cada espectro:

$$x = [A_1, \mu_1, \sigma_1, A_2, \mu_2, \sigma_2, A_3, \mu_3, \sigma_3],$$

La matriz $\mathbf{X} \in \mathbb{R}^{120 \times 9}$ resultante contiene las observaciones (filas) y las variables predictoras (columnas) de entrada al modelo.

La variable dependiente \mathbf{y} corresponde a la masa de NaCl disuelta (en gramos) para cada espectro, con los siguientes niveles:

$$y \in \{0, 1, 2, 3, 4, 5\} \text{ g}$$

Cada nivel aparece replicado 20 veces, producto de los 4 espectros por temperatura en 5 temperaturas.

La técnica PLS fue seleccionada por su capacidad de:

- Manejar relaciones lineales y multicolinealidad entre predictores,
- Construir componentes latentes t_k que maximizan la covarianza entre \mathbf{X} y \mathbf{y} ,
- Producir un modelo interpretable incluso con pocos predictores y muestras.

Para evitar el sobreajuste y evaluar el desempeño generalizado del modelo, se aplicó una validación cruzada por predicción cruzada (*cross-validation*) utilizando las siguientes métricas:

- **RMSECV** (*Root Mean Square Error of Cross Validation*),
- R^2 **ajustado**,

- Gráficos de real vs. predicho.

Se comparó el rendimiento del modelo entrenado bajo dos condiciones:

1. Usando **el espectro completo en cm^{-1} sin preprocesar**,
2. Usando únicamente **los 9 parámetros Gaussianos**.

Capítulo 4

Resultados

4.1. Conductividad como verificación experimental

Como control complementario, se midió la conductividad eléctrica de cada solución utilizando un conductivímetro calibrado. Se tomaron tres mediciones por muestra, junto con la temperatura correspondiente, para evaluar la reproducibilidad. La tabla 4.1 muestra los resultados obtenidos.

Los valores de conductividad mostraron un comportamiento altamente lineal con respecto a la masa de NaCl disuelta. Se ajustó un modelo de regresión lineal simple de la forma:

$$y = ax + b \tag{4.1}$$

donde:

- y : conductividad [mS/cm],
- x : masa de NaCl [g].

El ajuste obtenido fue:

$$y = 12,69x + 7,88 \tag{4.2}$$

Masa NaCl [g]	Conductividad [mS/cm]	Temperatura [°C]
1	18.54	26.1
1	18.47	26.8
1	18.52	27.0
2	34.50	26.9
2	34.50	27.0
2	34.20	27.0
3	48.40	27.3
3	48.30	27.7
3	47.90	27.8
4	59.60	26.3
4	59.70	26.7
4	57.80	26.8
5	69.20	25.6
5	70.00	25.7
5	69.80	25.8

Tabla 4.1: Conductividad medida para distintas masas de NaCl disueltas en solución.

con un coeficiente de determinación de $R^2 = 0,9911$, lo que indica una relación prácticamente perfecta entre masa y conductividad en el intervalo analizado.

El error cuadrático medio global (RMSE) fue de 1.7013 mS/cm, y los residuos no mostraron patrones sistemáticos, tal como se observa en la Figura 4.1, lo que sugiere ausencia de sesgos.

Este resultado cumple dos propósitos esenciales:

1. Verifica empíricamente que la masa pesada y disuelta de NaCl corresponde de manera confiable con el contenido iónico real en solución.
2. Sirve como control independiente, sin recurrir a técnicas espectroscópicas, para validar la coherencia interna del experimento.

Dado que la conductividad depende directamente del número de iones libres en la disolución, este resultado confirma que no existieron errores significativos durante la preparación de las muestras, ni fenómenos de precipitación o contaminación cruzada.

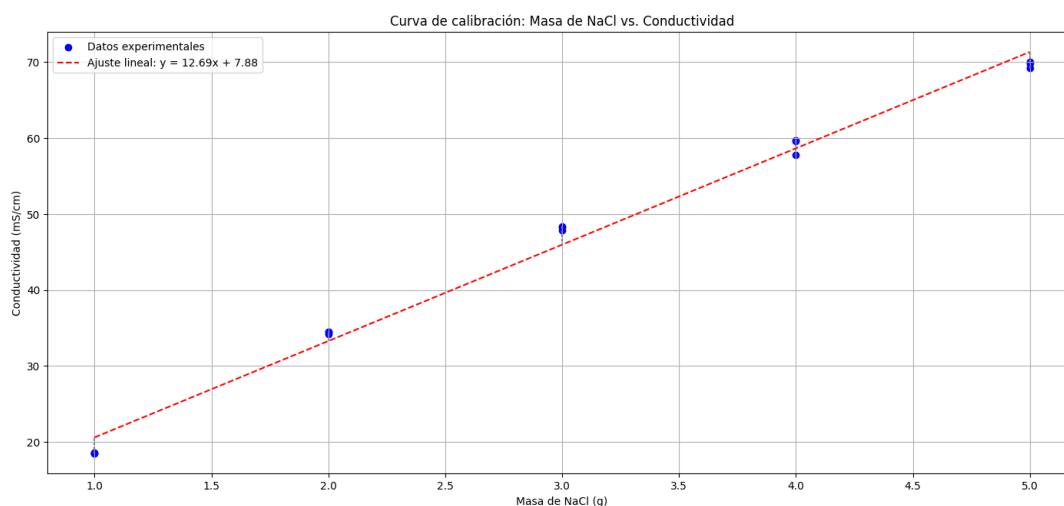


Fig. 4.1: Curva de calibración entre la masa de NaCl (g) y conductividad eléctrica (mS/cm).

4.2. Efecto del NaCl y la temperatura en el espectro Raman

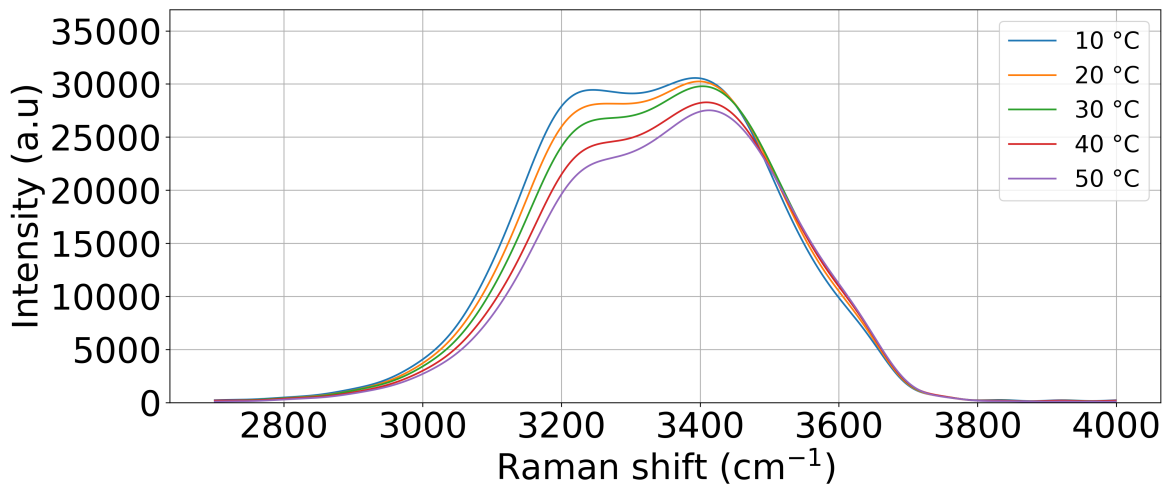
La región comprendida entre 2800 y 3800 cm^{-1} del espectro Raman corresponde a la banda de estiramiento O–H, la cual es particularmente sensible a las condiciones fisicoquímicas del entorno. En esta sección se analiza cómo la temperatura y la concentración de NaCl modifican esta banda, utilizando espectros representativos.

En la Figura 4.2a se muestran los espectros Raman del agua desionizada a cinco temperaturas (10 °C a 50 °C). A medida que la temperatura aumenta:

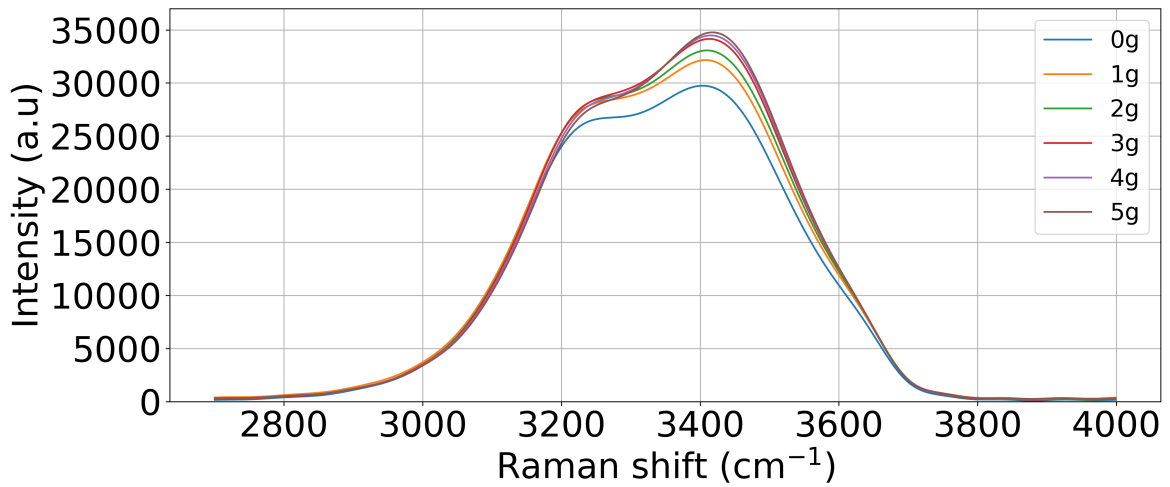
- Se observa una disminución de la intensidad total de la banda, en especial en la región de $\sim 3220\text{ cm}^{-1}$.
- El pico máximo sufre un leve desplazamiento hacia frecuencias más altas.
- Aparece una mayor contribución relativa en la región de $\sim 3600\text{ cm}^{-1}$.

Este comportamiento es consistente con la literatura y refleja el efecto destructivo del calor sobre la red de enlaces fuertes del agua líquida.

La Figura 4.2b muestra espectros Raman a temperatura constante de 50 °C para



(a) Espectros de agua pura en diferentes temperaturas.



(b) Espectros a 30° celsius con diferentes concentraciones.

Fig. 4.2: Dependencias espectrales principales.

soluciones con concentraciones crecientes de NaCl (0 a 5 g). Los efectos observados son:

- Disminución progresiva de la señal en $\sim 3220 \text{ cm}^{-1}$.
- Aumento en la región de $\sim 3600 \text{ cm}^{-1}$.
- La banda se ensancha y pierde definición.

Tanto el aumento de temperatura como la presencia de NaCl provocan:

- Reducción de la señal asociada a enlaces O–H fuertes,

- Desplazamiento hacia configuraciones más débiles o libres,
- Mayor dispersión espectral y aplanamiento de la banda.

Sin embargo, la naturaleza de los mecanismos es distinta:

- La temperatura actúa de forma térmica y global sobre todas las moléculas,
- El NaCl introduce efectos locales y específicos sobre las moléculas vecinas a los iones.

Esto refuerza la necesidad de considerar ambos factores en los modelos de predicción, para evitar sesgos o confusiones entre efectos espectrales similares pero de distinto origen.

4.3. Resultados del ajuste gaussiano

Con el objetivo de capturar la estructura interna de la banda de estiramiento O–H ($2800\text{--}3800\text{ cm}^{-1}$), cada espectro Raman fue ajustado a la suma de tres funciones gaussianas, basadas en la hipótesis de que esta banda está compuesta por tres subpoblaciones estructurales de moléculas de agua.

Se utilizó un algoritmo de mínimos cuadrados no lineales para ajustar los nueve parámetros de forma simultánea y minimizar el error cuadrático medio (RMSE) entre el espectro experimental y la suma de gaussianas.

La Figura 4.3 muestra un ejemplo representativo del ajuste sobre un espectro suavizado. Se aprecian las tres componentes gaussianas individuales (curvas punteadas) y su suma total (curva roja), que se superpone con alta precisión al espectro original (curva negra). El coeficiente de determinación fue $R^2 = 0,9989$, y el error cuadrático medio del ajuste fue de 353.2 unidades de intensidad, lo que indica un ajuste altamente preciso.

Este tipo de ajuste se realizó sobre los 120 espectros del conjunto de datos, generando para cada uno un vector de 9 parámetros físico-químicos:

$$\mathbf{x} = [\mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3, A_1, A_2, A_3]$$

Recordando que cada parámetro extraído tiene un significado físico-químico asociado:

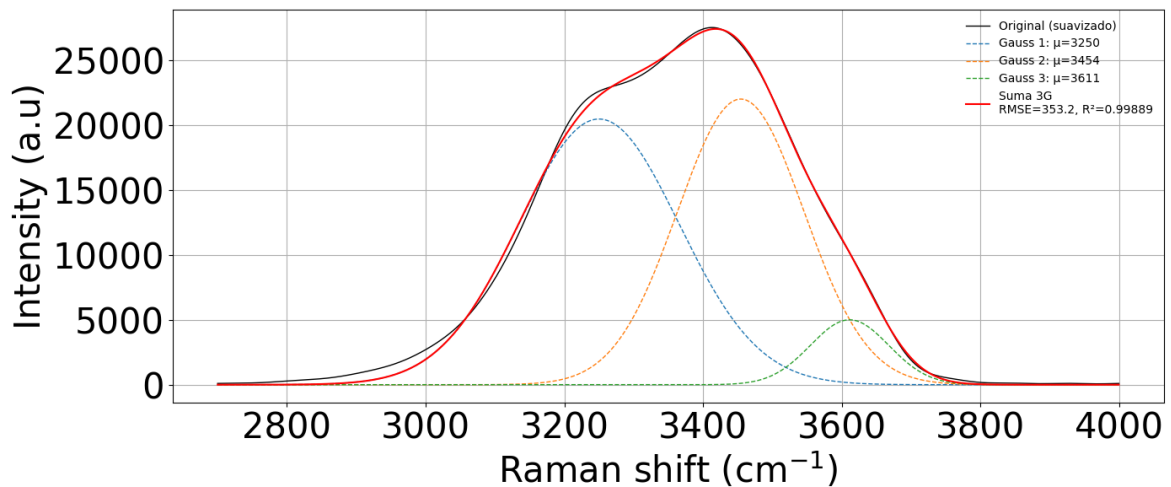


Fig. 4.3: Ajuste gaussiano triple para agua a 50 °C.

- μ_i : describe el tipo de enlace O–H dominante (desde enlaces fuertes hasta débiles).
- σ_i : refleja el grado de desorden estructural o dispersión espectral de cada subpoblación.
- A_i : está relacionado con la población relativa de moléculas de agua en esa configuración.

Este enfoque ofrece una descripción compacta pero rica en información estructural del entorno acuoso. Además, permite reducir la dimensionalidad del problema de modelado sin perder interpretabilidad, lo que es especialmente útil para técnicas estadísticas como *Partial Least Squares* (PLS).

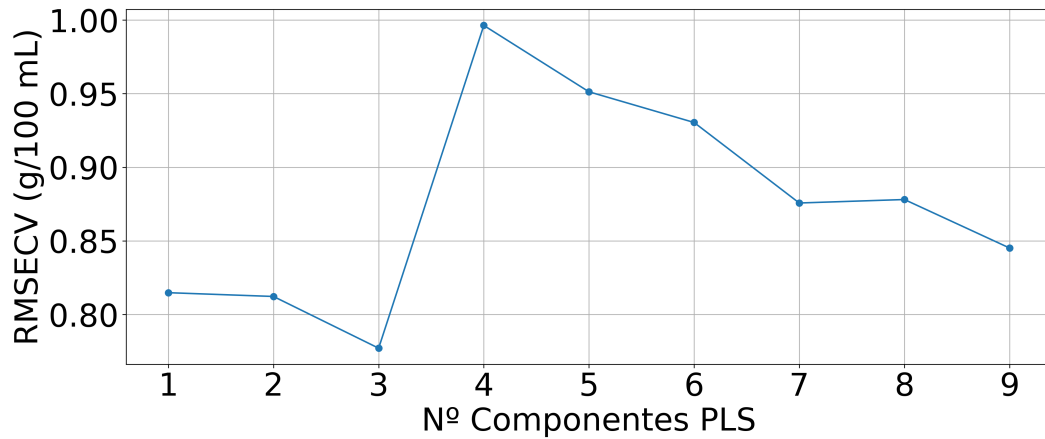
4.4. Evaluación del modelo PLS

Para estimar la concentración de NaCl a partir de los espectros Raman, se construyeron y evaluaron dos modelos de regresión (PLS), comparando dos formas de representar los datos espectrales:

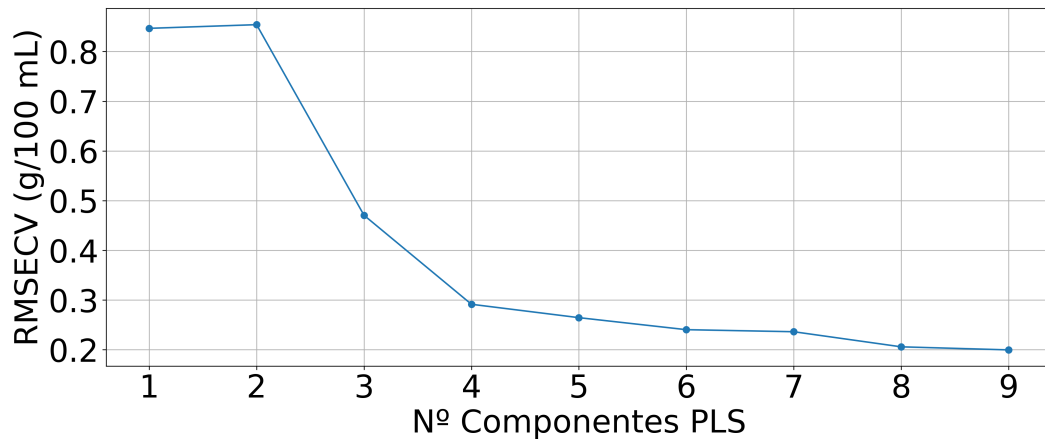
- **Modelo 1** — Entrenado con el espectro crudo completo, es decir, utilizando todos los puntos del espectro original desde 0 hasta 4200 cm^{-1} (ver Figura 3.12a), sin recorte ni extracción de características.
- **Modelo 2** — Entrenado con los 9 parámetros obtenidos del ajuste gaussiano

sobre la banda O–H (ver Figura 4.3), que condensan la información estructural de las moléculas de agua.

Ambos modelos se evaluaron mediante validación cruzada y métricas como el *RMSECV* (*Root Mean Square Error of Cross-Validation*), los residuos de predicción y las curvas de concentración real vs. predicha.



(a) Todo el espectro.



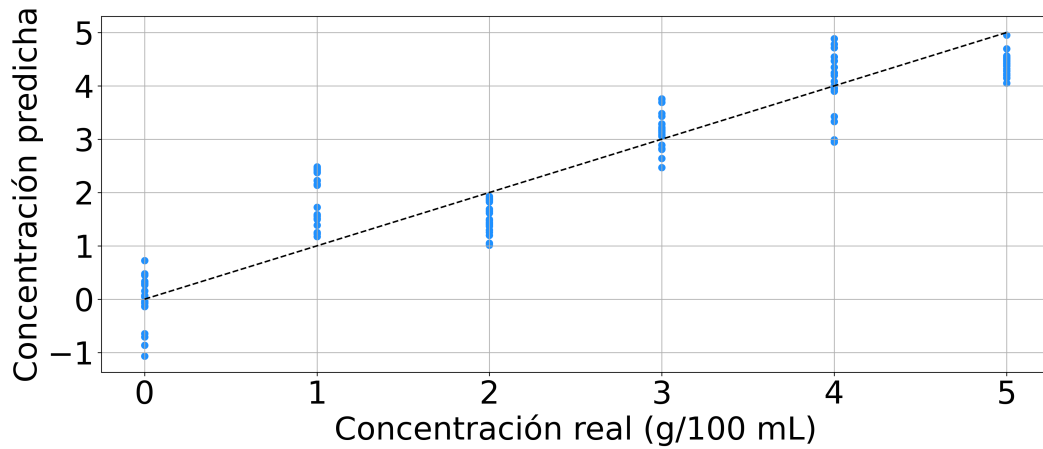
(b) RMSE con descomposición en gaussianas.

Fig. 4.4: Comparación RMSECV vs Número de componentes latentes entre ambos modelos.

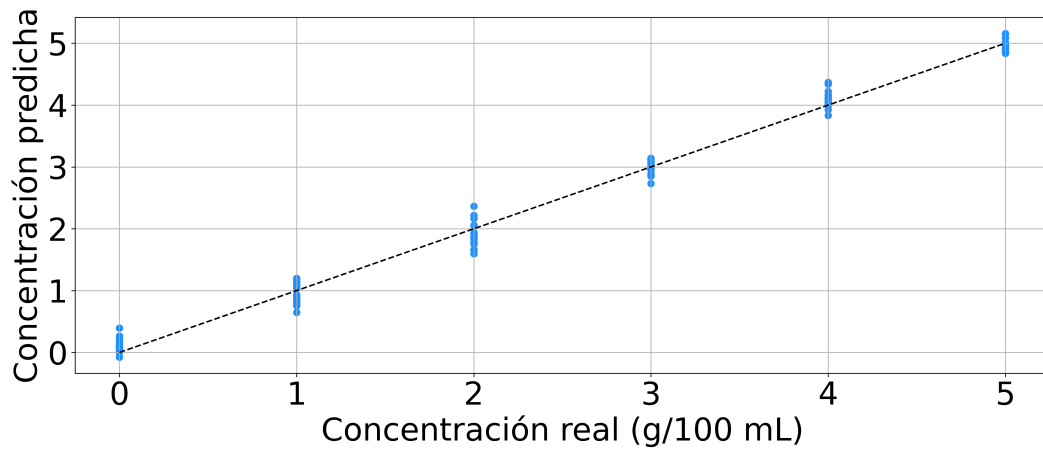
La Figura 4.4a muestra el RMSECV obtenido al entrenar el modelo PLS con el espectro crudo completo. Se observa un error mínimo de aproximadamente 0.78 g/100 mL utilizando 3 componentes latentes.

En comparación, la Figura 4.4b muestra el RMSECV para el modelo con 9 parámetros gaussianos, el cual alcanza un error mínimo de solo 0.20 g/100 mL

utilizando 9 componentes. Esto representa una mejora sustancial en precisión, con una representación espectral mucho más reducida.



(a) Todo el espectro.



(b) Descomposición en gaussianas.

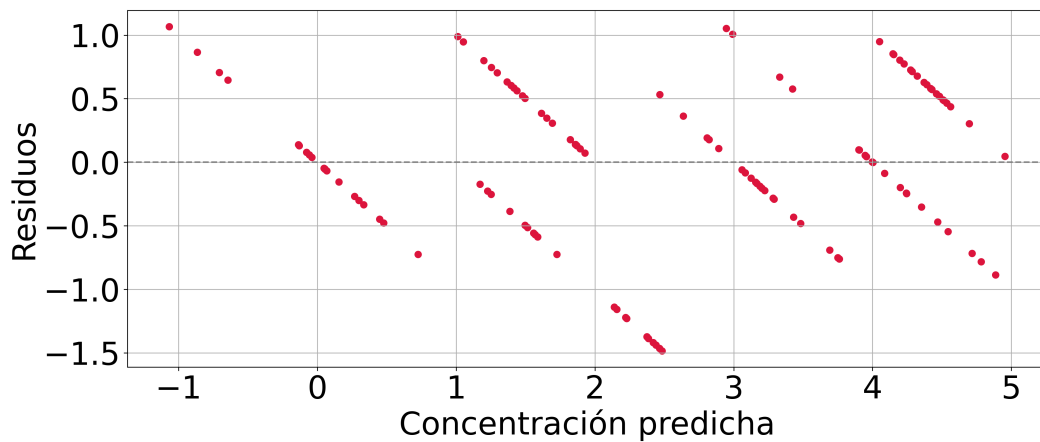
Fig. 4.5: Resultado entrenamiento regresión PLS.

Las Figuras 4.5a y 4.5b comparan la concentración real y predicha para ambos enfoques:

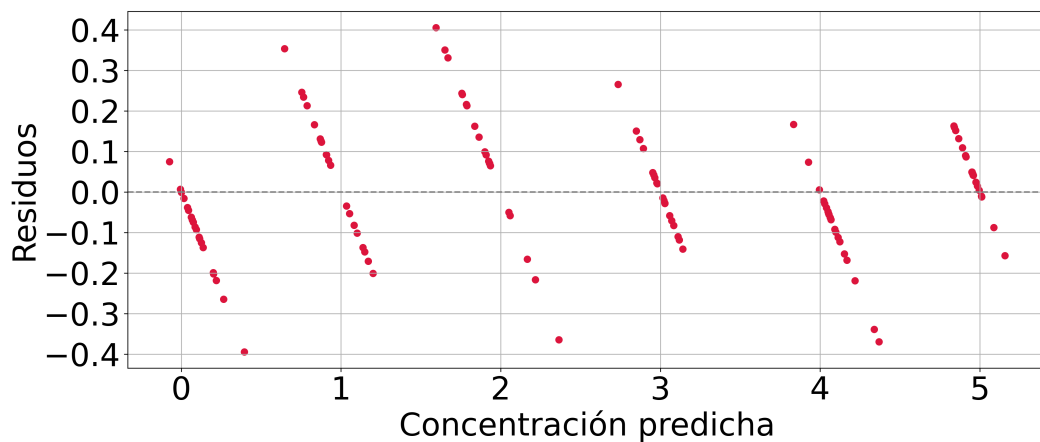
- El modelo basado en el espectro crudo muestra predicciones con una mayor dispersión y una evidente subestimación en concentraciones altas.
- El modelo basado en las tres gaussianas entrega predicciones alineadas con la diagonal ideal, lo que refleja una mejor capacidad de generalización.

El comportamiento de los residuos también confirma esta tendencia (ver Figuras 4.6a y 4.6b):

- El modelo basado en el espectro crudo presenta residuos grandes, distribuidos de



(a) Todo el espectro.



(b) Descomposición en gaussianas.

Fig. 4.6: Residuos de entrenamiento regresión PLS.

forma asimétrica y estructurada, lo que sugiere sobreajuste o la presencia de ruido no controlado.

- En contraste, el modelo construido con parámetros gaussianos muestra residuos pequeños, simétricos y centrados en cero, lo que indica un modelo bien ajustado y generalizable.

Estos resultados demuestran que utilizar el espectro completo en su forma cruda no garantiza un mejor desempeño predictivo. Aunque contiene más información, también incluye regiones espectrales irrelevantes o ruidosas (por ejemplo, en el intervalo 0–2800 cm^{-1}), lo que introduce colinealidad y dificulta el entrenamiento del modelo.

En cambio, la extracción de parámetros fisicoquímicos significativos mediante el ajuste

de gaussianas sobre la banda O–H proporciona una representación estructurada y eficiente, que mejora el desempeño del modelo PLS.

Para evaluar la capacidad de generalización del modelo, se construyó una nueva base de datos compuesta por espectros no utilizados en el entrenamiento, correspondientes a las mismas concentraciones de NaCl, pero obtenidos a temperaturas iguales y distintas a las consideradas previamente. Es importante destacar que el modelo PLS fue entrenado sin incluir la temperatura como variable predictora. El objetivo de esta validación fue doble; (i) medir el desempeño real del modelo ante condiciones nuevas, (ii) observar el efecto de la temperatura cuando no es explícitamente considerada en el modelado.

Tabla 4.2: Resultados de predicción para temperaturas incluidas en el entrenamientos

Conc. real (g/mL)	Conc. predicha (g/mL)	Temp (°C)
0	0.2182058	10
0	0.01561326	20
0	0.03855896	40
1	1.034822	20
1	1.053242	40
1	1.101336	50
2	2.364131	10
2	1.935113	20
2	1.787417	30
3	3.020735	10
3	3.118458	20
3	3.028319	30
4	3.878426	20
4	4.188728	40
4	4.058993	50
5	4.955894	30
5	5.087454	40
5	5.010401	50

Tabla 4.3: Resultados de predicción para temperaturas no incluidas en el entrenamiento

Conc. real (g/100 mL)	Conc. predicha (g/100 mL)	Temp (°C)
0	0.240287	25
0	0.274182	25
0	0.356284	45
0	0.735832	45
1	0.709065	15
1	0.409384	15
1	0.682559	35
1	0.677749	35
2	1.694319	15
2	1.727948	15
2	2.108687	5
2	1.835823	5

De la tabla 4.2 se concluye que el modelo muestra un rendimiento excelente cuando predice espectros tomados a las mismas temperaturas utilizadas en el entrenamiento. Esto refleja que los parámetros gaussianos permiten una representación robusta dentro del dominio espectral aprendido.

Por otro lado, en la tabla 4.3 se infiere que el rendimiento del modelo se degrada significativamente cuando se enfrenta a temperaturas no incluidas en el entrenamiento. Se observa que aparecen desviaciones más marcadas (sobreestimaciones en agua pura o subestimaciones en NaCl).

Esto demuestra que, aunque el modelo no incorpora la temperatura explícitamente, los espectros Raman sí contienen información térmica implícita. Por lo tanto, la temperatura influye en los parámetros gaussianos, afectando la precisión cuando el modelo extrapola fuera de su dominio térmico.

4.5. Resumen de resultados

Los resultados obtenidos a lo largo de este estudio permiten establecer con claridad la relación entre la composición salina de una disolución acuosa y su respuesta espectral Raman, así como la efectividad de diferentes estrategias de modelado para su cuantificación.

Las disoluciones de NaCl preparadas en el laboratorio fueron validadas mediante mediciones de conductividad eléctrica, evidenciando una relación altamente lineal con la masa de sal disuelta ($R^2 = 0,9911$, RMSE = 1,7013 mS/cm), lo que confirma su correcta preparación.

Se observó que tanto el aumento de la temperatura como la concentración de NaCl afectan significativamente la banda de estiramiento O–H (2800–3800 cm^{-1}), generando desplazamientos, aplanamiento y ensanchamiento de dicha banda.

Ambos factores provocan una disminución en la contribución de los enlaces fuertemente enlazados (alrededor de 3220 cm^{-1}) y un aumento relativo de los componentes débilmente enlazados (alrededor de 3600 cm^{-1}).

Cada espectro fue ajustado mediante la suma de tres funciones gaussianas, generando un vector de nueve parámetros con interpretación estructural. Esta parametrización permitió condensar significativamente la información espectral, manteniendo la coherencia fisicoquímica y reduciendo la dimensionalidad del problema de forma efectiva.

Se compararon dos modelos de regresión PLS: uno entrenado con el espectro crudo completo y otro basado en los nueve parámetros obtenidos del ajuste gaussiano.

El modelo construido a partir de parámetros gaussianos mostró un rendimiento superior, evidenciado por un menor error de validación cruzada (RMSECV 0,2g/100 mL frente a 0,78 g/100 mL), un mejor ajuste entre valores reales y predichos, y residuos más bajos, simétricos y sin sesgos aparentes.

Además se evaluaron espectros obtenidos a temperaturas no incluidas en el conjunto de entrenamiento (5, 15, 25, 35 y 45°C). El modelo mostró un alto desempeño dentro del dominio térmico aprendido ($R^2 = 0,973$); sin embargo, presentó fallas significativas al extrapolar, evidenciando errores extremos en combinaciones no vistas de temperatura.

Estos resultados ponen de manifiesto la necesidad de incorporar la temperatura como una variable explícita en futuros modelos, a fin de mejorar su capacidad generalizadora y robustez predictiva.

Capítulo 5

Conclusiones

5.1. Conclusiones generales

Los resultados obtenidos en esta investigación permiten concluir que se cumplió el objetivo principal planteado: construir un modelo preliminar capaz de estimar la concentración de NaCl en solución acuosa a partir de espectros Raman, mediante el uso de técnicas de procesamiento espectral y regresión multivariante. Asimismo, se respondió satisfactoriamente al problema central del estudio: cómo extraer información cuantitativa fiable desde la forma espectral del agua en presencia de sal, sin recurrir a modelos fisicoquímicos complejos ni requerir entrenamiento sobre un espacio multiespectral completo.

El modelo propuesto —basado en la parametrización de la banda de estiramiento O–H del espectro Raman como suma de tres funciones gaussianas— permitió reducir cada espectro a un vector de nueve parámetros significativos (posición, amplitud y anchura de cada subbanda). Esta representación mejoró la interpretabilidad del modelo y aumentó su precisión y eficiencia computacional en comparación con el uso del espectro crudo. Además, encapsuló información estructural relevante sobre los estados de enlace de hidrógeno del agua y su reorganización frente a la presencia de sal y variaciones de temperatura.

La aplicación de regresión Partial Least Squares (PLS) sobre esta representación gaussiana resultó en un error de validación cruzada (RMSECV) de 0,20g/100mL, con

un coeficiente de determinación $R^2 > 0,97$ al evaluar el modelo con espectros nuevos dentro del rango térmico aprendido. Estos resultados demuestran que el modelo es estadísticamente confiable.

No obstante, el análisis también evidenció limitaciones importantes: al aplicar el modelo a espectros tomados a temperaturas no incluidas durante el entrenamiento, su precisión disminuyó considerablemente ($R^2 = 0,79$), lo cual confirma que la temperatura influye de forma no trivial sobre la forma espectral del agua y, por ende, sobre los parámetros del modelo. Esta observación refuerza la necesidad de considerar la temperatura como una variable explícita en futuros desarrollos, especialmente si se busca aplicar el modelo en condiciones no controladas.

Desde una perspectiva disciplinar, este trabajo aporta una metodología replicable y científicamente fundamentada para la cuantificación de especies disueltas mediante espectroscopía Raman. Aporta, además, evidencia empírica sobre la utilidad de la descomposición gaussiana de la banda O–H como descriptor fisicoquímico y como puente entre la espectroscopía estructural y el modelado estadístico. En este sentido, contribuye al avance de las técnicas de análisis espectral aplicadas a sistemas acuosos y salinos, con potencial de extensión a otras aplicaciones en química analítica, monitoreo de procesos y control de calidad.

5.2. Trabajo a Futuro

A partir de las limitaciones detectadas y de las oportunidades metodológicas y tecnológicas identificadas durante la investigación, se sugieren las siguientes direcciones para continuar y profundizar esta línea de estudio:

1. Inclusión explícita de la temperatura en el modelo

Uno de los hallazgos clave de este estudio fue la influencia significativa de la temperatura sobre la forma espectral de la banda O–H. Esta variable afecta la posición, anchura y amplitud de las subpoblaciones asociadas a los distintos estados de enlace de hidrógeno.

2. Extensión del diseño experimental

Se recomienda ampliar la resolución térmica y química del conjunto de entrenamiento con el fin de mejorar la capacidad de interpolación del modelo y reducir errores asociados a regiones de baja densidad de datos. En cuanto a la temperatura, se sugiere incluir niveles intermedios (por ejemplo, cada 5°C), lo que permitiría capturar con mayor precisión las variaciones continuas del espectro asociadas a cambios térmicos.

Respecto a la concentración, se propone incorporar valores intermedios (por ejemplo, 0,5; 1,5; 2,5g/100mL, etc.), con el objetivo de mejorar la continuidad del modelo y suavizar las predicciones en regiones limítrofes.

Adicionalmente, aumentar el número de réplicas por condición experimental permitiría estimar de forma más robusta la varianza instrumental y biológica del sistema, mejorando la confiabilidad estadística del modelo.

3. Exploración de modelos no lineales y de aprendizaje profundo

Si bien el modelo PLS demostró ser efectivo y robusto, su capacidad se restringe a la modelación de relaciones lineales. Una línea de trabajo podría en explorar y comparar su desempeño con algoritmos más flexibles, capaces de capturar interacciones no lineales complejas. Entre estos enfoques se encuentran:

- **Support Vector Regression (SVR)**
- **Random Forests**
- **Redes neuronales multicapa (MLP)**
- **Redes neuronales convolucionales (CNN)**, aplicadas directamente sobre el espectro crudo

Estos modelos podrían mejorar la capacidad predictiva al capturar relaciones más complejas entre los parámetros gaussianos y la concentración de NaCl. En particular, el uso de redes neuronales convolucionales permitiría prescindir del paso de parametrización, siempre que se disponga de una base de datos lo suficientemente amplia y diversa.

4. Aplicación a otros solutos y mezclas iónicas

El enfoque desarrollado en este estudio es potencialmente generalizable a otras soluciones acuosas. Una línea de investigación natural consiste en extender el análisis a diferentes tipos de solutos, incluyendo otras sales inorgánicas, como KCl, MgCl₂ o Na₂SO₄ o sistemas binarios o ternarios (mezclas salinas).

Bibliografía

- [1] I. Durickovic, *Using Raman Spectroscopy for Characterization of Aqueous Media and Quantification of Species in Aqueous Solution*, 10 2016, pp. 405–427.
- [2] D. Bao, D. Hua, H. Qi, and J. Wang, “Investigation of a raman scattering spectral model for seawater containing a composite salt solute,” *Opt. Express*, vol. 30, no. 5, pp. 6713–6725, Feb 2022. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-30-5-6713>
- [3] M. Baumgartner and R. Bakker, “Raman spectroscopy of pure h₂o and nacl-h₂o containing synthetic fluid inclusions in quartz - a study of polarization effects,” *Mineralogy and Petrology*, vol. 95, pp. 1–15, 01 2009.
- [4] D. P. Mason and M. E. Elwood Madden, “Raman spectroscopy of high salinity brines and ices,” *Icarus*, vol. 372, p. 114759, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0019103521004127>
- [5] N. Takahashi, T. Tsujimori, S. Kamada, and M. Nakamura, “In-situ raman spectroscopic analysis of dissolved silica structures in na₂co₃ and naoh solutions at high pressure and temperature,” *Contributions to Mineralogy and Petrology*, vol. 177, 03 2022.
- [6] N. Prieto-Taboada, O. Laserna, I. Martínez-Arkarazo, M. Olazabal, and J. Madariaga, “Raman spectra of the different phases in the caso₄-h₂o system,” *Analytical chemistry*, vol. 86, 09 2014.
- [7] M. Sunil, N. Pallikkavaliyaveetil, M. N, A. Gopinath, S. Chidangil, S. Kumar, and J. Lukose, “Machine learning assisted raman spectroscopy: A viable approach for the detection of microplastics,” *Journal of Water Process Engineering*, vol. 60, p.

- 105150, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2214714424003829>
- [8] Y. Qi, D. Hu, Y. Jiang, Z. Wu, M. Zheng, E. X. Chen, Y. Liang, M. A. Sadi, K. Zhang, and Y. P. Chen, “Recent progresses in machine learning assisted raman spectroscopy,” *Advanced Optical Materials*, vol. 11, no. 14, p. 2203104, 2023. [Online]. Available: <https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/adom.202203104>
- [9] S.-W. Yang, Y. Xie, J.-Z. Liu, D. Zhang, J. Huang, and P. Liang, “A novel method for quantitative determination of multiple substances using raman spectroscopy combined with cwt,” *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 317, p. 124427, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1386142524005936>
- [10] A. Sui, Y. Deng, Y. Wang, and J. Yu, “A deep learning model designed for raman spectroscopy with a novel hyperparameter optimization method,” *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 280, p. 121560, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1386142522007090>
- [11] S. Fang, S. Wu, Z. Chen, C. He, L. L. Lin, and J. Ye, “Recent progress and applications of raman spectrum denoising algorithms in chemical and biological analyses: A review,” *TrAC Trends in Analytical Chemistry*, vol. 172, p. 117578, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0165993624000608>
- [12] Z.-P. Chen, J. Morris, and E. Martin, “Modelling temperature-induced spectral variations in chemical process monitoring,” vol. 37, 07 2004, pp. 553–558.
- [13] D. Sbarbaro, T. Johansen, and J. Yañez, “Adaptive kalman filter for on-line spectroscopic sensor corrections,” 07 2023, pp. 2109–2114.
- [14] X. Wang, W. Hu, and I. M. Chou, “Raman spectroscopic characterization on the OH stretching bands in NaCl-Na₂CO₃-Na₂SO₄-CO₂-H₂O systems: Implications for the measurement of chloride concentrations in fluid inclusions,” *Journal of Geochemical Exploration*, vol. 132, pp. 111–119, Jan. 2013.

- [15] I. Đuričković, M. Marchetti, R. Claverie, P. Bourson, J.-M. Chassot, and M. D. Fontana, “Experimental study of nacl aqueous solutions by raman spectroscopy: Towards a new optical sensor,” *Applied Spectroscopy*, vol. 64, no. 8, pp. 853–857, 2010, PMID: 20719047. [Online]. Available: <https://doi.org/10.1366/000370210792080984>
- [16] D. A. Skoog, F. J. Holler, and S. R. Crouch, *Principios de análisis instrumental*, 6th ed. México, D.F.: Cengage Learning Editores, 2008, traducción de *Principles of Instrumental Analysis*, 6th ed.
- [17] K. Kumar, “Partial least square (pls) analysis: Most favorite tool in chemometrics to build a calibration model,” *Resonance*, vol. 26, pp. 429–442, 03 2021.