



DEPARTAMENTO DE  
INGENIERÍA ELÉCTRICA  
UNIVERSIDAD DE CONCEPCIÓN

**ANÁLISIS Y ESTIMACIÓN DE CONCENTRACIÓN DE NaCl  
EN SOLUCIÓN ACUOSA MEDIANTE EL USO DE  
ESPECTROSCOPIA RAMAN**

POR

**Edgardo Andrés Pérez Foix**

Tesis presentada a la Escuela de graduados de la Universidad de Concepción para optar al grado de Magíster en Ingeniería Eléctrica.

**Profesor Guía**

Daniel Sbarbaro H, Ph.D.

Concepción,  
3 de marzo de 2026

© 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

La cuantificación de la salinidad en soluciones acuosas es un problema relevante en múltiples áreas de la ingeniería y la ciencia aplicada, debido a su impacto sobre propiedades fisicoquímicas y a la necesidad de monitoreo rápido y no destructivo en diversas condiciones operacionales. En particular, la determinación de la concentración de cloruro de sodio (NaCl) mediante espectroscopía Raman presenta un desafío fundamental: al separarse en átomos individuales que han ganado o perdido uno o más electrones, el NaCl no exhibe bandas Raman intensas en el rango espectral habitual, por lo que la estimación debe realizarse de manera indirecta a partir de la respuesta vibracional del solvente. En este contexto, la banda de estiramiento O–H del agua constituye una región espectral de alta sensibilidad, cuya forma varía sistemáticamente con la concentración salina y con la temperatura. Sin embargo, la marcada dependencia térmica de dicha banda introduce un confusor dominante que debe ser considerado explícitamente para lograr estimaciones robustas.

Esta tesis desarrolla y valida una metodología cuantitativa basada en espectroscopía Raman para estimar la concentración de NaCl en soluciones acuosas a partir de descriptores efectivos extraídos de la banda O–H, incorporando la temperatura como variable del modelo. Se prepararon soluciones estándar de NaCl en el rango de 0 a 5 g/100 mL y se adquirieron espectros Raman a distintas temperaturas controladas entre 10 y 50 °C, conformando una base de datos de entrenamiento y un conjunto independiente de validación externa. Los espectros fueron recortados en la región 3150–3550  $\text{cm}^{-1}$  y se parametrizaron mediante una descomposición en un número reducido de funciones Gaussianas, empleada como herramienta matemática para capturar redistribuciones sistemáticas de intensidad sin atribuir necesariamente poblaciones moleculares discretas a cada componente. A partir de las amplitudes obtenidas se formuló un modelo de regresión que relaciona cada descriptor con la concentración  $C$ , la

temperatura  $T$  y un término de interacción  $C \cdot T$ . Esta formulación permite derivar un esquema para estimar  $C$  desde un espectro observado bajo una temperatura conocida, favoreciendo la interpretabilidad y la trazabilidad del proceso de estimación.

El desempeño de la metodología se evaluó mediante validación interna y externa, cuantificando la exactitud a través de métricas estándar, tales como el error absoluto medio (MAE) y la raíz del error cuadrático medio (RMSE). Adicionalmente, se comparó el método propuesto con un enfoque de referencia basado en regresión por mínimos cuadrados parciales (PLS), analizando ventajas y limitaciones en términos de precisión, estabilidad frente a variaciones térmicas e interpretabilidad. En conjunto, los resultados muestran que la banda O–H del agua contiene información cuantitativa suficiente para estimar la concentración de NaCl dentro del dominio experimental estudiado, siempre que se incorpore explícitamente la temperatura. La estrategia propuesta contribuye con un marco metodológico robusto para la cuantificación indirecta de salinidad mediante Raman, estableciendo bases para futuras extensiones hacia rangos más amplios de concentración, mezclas multicomponente y aplicaciones en condiciones operacionales más complejas.

# Abstract

Indirect Raman quantification of dissolved salts in water is attractive because it can be performed with minimal sample preparation, yet it is challenging for electrolytes such as NaCl where concentration must be inferred from solvent-mediated spectral changes rather than from strong solute-specific bands. The Raman O–H stretching band is information-rich, but it is also strongly temperature-dependent, making temperature a dominant confounder for salinity estimation unless it is handled explicitly. Here, we develop and evaluate two distinct temperature-aware Raman routes for NaCl estimation in water (0–5 g/100 mL) using a common region of interest within the O–H band (3150–3550  $\text{cm}^{-1}$ ). Spectra were treated by dark subtraction and analyzed in absolute intensity (counts), without baseline correction, smoothing, or normalization.

The first route builds an interpretable estimator by representing the O–H sub-band with a global constrained three-Gaussian basis, extracting non-negative amplitudes, and performing bounded concentration inversion using the measured acquisition temperature. This formulation provides a compact set of physically interpretable features and enables concentration predictions that remain within the calibrated domain by construction. The second route addresses temperature effects through Individual Contribution Standardization (ICS), standardizing spectra to a reference temperature prior to Partial Least Squares regression ( $LV = 3$ ). Under both internal (in-sample) evaluation and external testing on spectra acquired on a different day at intermediate temperatures, both routes delivered consistently high agreement between predicted and reference concentrations. Beyond concentration accuracy, spectral overlay analyses supported physical fidelity of the Gaussian representation on both calibration and external examples, indicating that the model reproduces the O–H band structure rather than relying on compensation artifacts. The two pipelines exhibited complementary operational behavior: the inverse formulation enforces physically plausible outputs,

whereas the ICS–PLS route follows a standard chemometric workflow and can yield occasional out-of-range predictions typical of unconstrained linear regression.

This work provides a transparent framework for temperature-aware Raman estimation of NaCl concentration from the O–H band and demonstrates the value of reporting method comparisons under a shared ROI and validation protocol. The results highlight practical trade-offs between an interpretable, physically inverse model and a multivariate regression approach following explicit temperature standardization. These findings motivate temperature-aware method design and critical pipeline comparisons when indirect Raman quantification relies on water-band features.

# 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.

A mis amigos, Raúl López con quien vengo compartiendo desde los 15 años, al Kova, mi primer amigo de la carrera, Angelo y Yoyo, por ser parte esencial de esta etapa.

Este logro también les pertenece a ustedes. Gracias.

# Índice General

Resumen	I
Abstract	III
Agradecimientos	v
Índice de Figuras	IX
Índice de Tablas	XI
<b>1. Introducción</b>	<b>1</b>
1.1. Trabajos previos . . . . .	2
1.1.1. Principio físico de la dispersión Raman . . . . .	2
1.1.2. Espectroscopía Raman en medios acuosos: fundamentos para cuantificación . . . . .	3
1.1.3. Espectro Raman del agua y dependencia con temperatura . . . .	4
1.1.4. Efecto de la salinidad en el espectro Raman del agua . . . . .	6
1.1.5. Cuantificación Raman: enfoques directos e indirectos . . . . .	8
1.1.6. Factores instrumentales y metrológicos en cuantificación basada en O–H . . . . .	9
1.1.7. Casos límite . . . . .	10
1.1.8. Discusión . . . . .	10
1.2. Hipótesis de trabajo . . . . .	11
1.3. Objetivos . . . . .	11
1.3.1. Objetivo general . . . . .	11
1.3.2. Objetivos específicos . . . . .	11
1.4. Metodología . . . . .	12

---

1.4.1.	Preparación de soluciones de NaCl . . . . .	12
1.4.2.	Generación de base de datos con espectros Raman . . . . .	13
1.4.3.	Procesamiento espectral y extracción de descriptores . . . . .	13
1.4.4.	Validar el desempeño del modelo mediante estrategias de validación interna y externa . . . . .	13
1.4.5.	Comparar la metodología propuesta con enfoque de referencia Partial Least Square (PLS) . . . . .	14
1.5.	Alcances y limitaciones . . . . .	14
1.5.1.	Alcances . . . . .	14
1.5.2.	Limitaciones . . . . .	15
<b>2.</b>	<b>Desarrollo</b> . . . . .	<b>17</b>
2.1.	Calibración del espectrómetro . . . . .	17
2.1.1.	Calibración Absoluta en nanómetros [nm] . . . . .	19
2.1.2.	Calibración en Número de Onda [ $\text{cm}^{-1}$ ] . . . . .	20
2.2.	Preparación de soluciones estándar de NaCl . . . . .	25
2.3.	Adquisición de espectros Raman . . . . .	26
2.4.	Preprocesamiento de los espectros . . . . .	28
2.5.	Modelo de descomposición global de la banda O–H mediante funciones Gaussianas . . . . .	29
2.5.1.	Construcción de la matriz de intensidades Y . . . . .	30
2.5.2.	Definición de las funciones Gaussianas globales . . . . .	31
2.5.3.	Formulación del problema de optimización . . . . .	31
2.5.4.	Resolución del problema de optimización . . . . .	32
2.6.	Modelo de amplitudes . . . . .	35
2.6.1.	Formulación del modelo de amplitudes . . . . .	36
2.6.2.	Resolución del modelo de amplitudes . . . . .	38
2.7.	Modelo inverso para la estimación de concentración . . . . .	38
2.7.1.	Formulación del problema inverso . . . . .	40
2.7.2.	Solución analítica del problema inverso . . . . .	40
2.8.	Modelo Partial Least Square (PLS) con Individual Contribution Standardization (ICS) . . . . .	41
2.8.1.	Individual Contribution Standardization (ICS) . . . . .	41
2.8.2.	Formulación del modelo PLS para cuantificación . . . . .	42

---

<b>3. Resultados</b>	<b>46</b>
3.1. Conductividad como verificación experimental . . . . .	46
3.2. Resultados de descomposición global con funciones Gaussianas . . . . .	48
3.3. Efecto del NaCl y la temperatura en el espectro Raman . . . . .	51
3.4. Resultados de las amplitudes Gaussianas como variables latentes físicoquímicas . . . . .	52
3.5. Resultados de los modelos temperatura-concentración de las amplitudes	55
3.6. Resultados de entrenamiento del modelo inverso para la estimación de concentración . . . . .	60
3.7. Resultados de validación externa del modelo inverso para la estimación de concentración . . . . .	63
3.8. Resultados de entrenamiento del modelo ICS-PLS para la estimación de concentración . . . . .	68
3.8.1. Resultados de validación externa del modelo ICS-PLS . . . . .	71
3.9. Análisis de incertidumbre del sistema de estimación . . . . .	75
3.10. Discusión comparativa de los métodos de estimación de concentración .	76
<b>4. Conclusiones</b>	<b>79</b>
4.1. Conclusiones generales . . . . .	79
4.2. Aportaciones metodológicas y conceptuales . . . . .	80
4.3. Desempeño y alcance de los modelos de estimación . . . . .	81
4.4. Incertidumbre experimental y límites de desempeño . . . . .	81
4.5. Limitaciones del estudio . . . . .	82
4.6. Proyecciones y conclusión final . . . . .	82
<b>Bibliografía</b>	<b>88</b>

# Índice de Figuras

2.1. Configuración del cabezal Raman para calibración. . . . .	18
2.2. Curva de calibración entre localización del centro de distribución de un peak en número de píxel versus el valor de la línea de referencia atómica en nanómetros. . . . .	20
2.3. Gráfico superior, muestra los valores de los peak, en número de píxel, con respecto al centro de distribución Gaussiana en torno a un valor máximo entero. Gráfico inferior, transformación de calibración absoluta aplicada, para convertir el número de píxel en longitud de onda. . . . .	21
2.4. Espectros característicos de Raman Shift [ $\text{cm}^{-1}$ ] para los estándares de calibración de Azufre y Poliestireno. El valor de las líneas se encuentra ordenados en la Tabla 2.2 . . . . .	21
2.5. Curva de calibración del vector número de onda, ajustando un polinomio de 3er orden, para los pares ordenados de los peaks como estándar de referencia entre número de píxel y número de onda. . . . .	23
2.6. Gráfico superior, muestra los valores de los peak, en número de píxel, con respecto al centro de distribución Gaussiana en torno a un valor máximo entero. Gráfico inferior, transformación de calibración absoluta aplicada, para convertir el número de píxel en Número de onda [ $\text{cm}^{-1}$ ]. . . . .	23
2.7. Configuración experimental: (1) Fuente de luz láser, (2) soporte de cubeta, (3) cubeta, (4) controlador de temperatura, (5) espectrómetro HORIBA Jobin Yvon ISA HR-320, (6) computador portátil [Fuente propia].	27
2.8. Esquema experimental de la cubeta y su sistema porta cubeta. . . . .	28
2.9. Espectros de agua desionizada a diferentes temperaturas junto con la zona de interés para el recorte. . . . .	29

---

2.10. Diagrama de bloques del proceso de estimación de concentración a partir de espectros Raman. . . . .	45
3.1. Comportamiento entre la masa de NaCl (g) y conductividad eléctrica (mS/cm). . . . .	48
3.2. Ajuste de 3 funciones Gaussianas para espectro agua a 30 °C. . . . .	49
3.3. Dependencias espectrales principales. . . . .	52
3.4. Comportamiento de $A_1$ respecto a la concentración y temperatura. . . . .	53
3.5. Comportamiento de $A_2$ respecto a la concentración y temperatura. . . . .	54
3.6. Comportamiento de $A_3$ respecto a la concentración y temperatura. . . . .	54
3.7. Modelo de amplitudes para $A_1$ . . . . .	57
3.8. Modelo de amplitudes para $A_2$ . . . . .	58
3.9. Modelo de amplitudes para $A_3$ . . . . .	59
3.10. Comparación entre concentración real y concentración predicha del modelo inverso en el entrenamiento. . . . .	61
3.11. Espectro real vs espectro estimado con el modelo inverso en el entrenamiento para muestra de (0g/100 mL, 30°C). . . . .	62
3.12. Residuos de la estimación del modelo inverso en el entrenamiento. . . . .	62
3.13. Relación entre concentración real vs concentración predicha con el modelo inverso en validación externa. . . . .	64
3.14. Residuos de la estimación del modelo inverso en validación externa. . . . .	65
3.15. Espectro real vs espectro estimado con modelo inverso correspondiente a una muestra externa de 2g/100mL a 35°C. . . . .	66
3.16. Relación entre concentración real vs concentración predicha en entrenamiento del modelo ICS-PLS. . . . .	69
3.17. Residuos de la estimación en entrenamiento del modelo ICS-PLS. . . . .	70
3.18. $X$ -loadings $\mathbf{P}$ del modelo PLS (LV=3) para $T_{\text{ref}} = 30^\circ\text{C}$ en la región 3150–3550 $\text{cm}^{-1}$ . . . . .	70
3.19. Relación entre concentración real vs concentración predicha en validación externa del modelo ICS-PLS. . . . .	72
3.20. Residuos de la estimación en validación externa del modelo ICS-PLS. . . . .	73

# Índice de Tablas

2.1. Líneas espectrales de Mercurio (Hg), Argón (Ar) y Neón (Ne). . . . .	19
2.2. Líneas referenciales en Raman Shift [ $\text{cm}^{-1}$ ] y desviación estándar [ $\text{cm}^{-1}$ ]. Muestra mineral de Azufre 100% puro y RPC Thorlabs Polystyrene Raman Spectroscopy Calibration Sample, Average of FT-Ramanand Scanning (Spex 1403) Data Sets. . . . .	22
2.3. Conversión de parámetros espectroscópicos entre longitud de onda (nm) y número de onda ( $\text{cm}^{-1}$ ). Láser: 532,10 nm, rendija (slit): 50 $\mu\text{m}$ . . . . .	24
2.4. Masa de NaCl en cada solución preparada. . . . .	25
3.1. Conductividad medida para distintas masas de NaCl disueltas en solución.	47
3.2. Desempeño global del ajuste gaussiano. . . . .	49
3.3. Significado físico-químico de los parámetros de ajuste gaussiano. . . . .	50
3.4. Parámetros ajustados para los modelos de amplitudes $A_k(C, T)$ junto con sus métricas de calidad. . . . .	56
3.5. Interpretación de los coeficientes del modelo de interacción para $A_1(C, T)$ .	56
3.6. Interpretación de los coeficientes del modelo de interacción para $A_2(C, T)$ .	57
3.7. Interpretación de los coeficientes del modelo de interacción para $A_3(C, T)$ .	58
3.8. Desempeño global del modelo inverso en el entrenamiento. . . . .	61
3.9. Desempeño global del modelo inverso en validación externa. . . . .	64
3.10. Desempeño de espectro de 2g/100mL a 35°C estimado. . . . .	65
3.11. Resultados individuales de validación externa: concentración real, concentración estimada y temperatura de medición. . . . .	67
3.12. Desempeño global del modelo ICS-PLS en el entrenamiento. . . . .	69
3.13. Desempeño global del modelo ICS-PLS en validación externa. . . . .	72
3.14. Resultados individuales de validación externa para el modelo ICS-PLS .	74

# Capítulo 1

## Introducción

La cuantificación de sales disueltas en agua constituye un problema recurrente en ciencias e ingeniería, debido a su impacto directo en propiedades fisicoquímicas relevantes y en múltiples aplicaciones tecnológicas. En particular, la concentración de cloruro de sodio (NaCl) es un parámetro de interés en el monitoreo ambiental, el control de procesos, los sistemas industriales basados en soluciones acuosas y contextos donde la salinidad condiciona fenómenos de corrosión, estabilidad de materiales y rendimiento operacional. En este escenario, disponer de metodologías de medición rápidas, no destructivas y con potencial de implementación en línea resulta especialmente valioso.

La espectroscopía Raman es una técnica óptica que permite investigar modos vibracionales moleculares mediante la dispersión inelástica de la luz. Su carácter no destructivo, junto con la posibilidad de realizar mediciones en medios líquidos y de integrarse a configuraciones instrumentales compactas, la posiciona como una herramienta atractiva para la caracterización de soluciones acuosas. Sin embargo, el problema de cuantificar NaCl mediante Raman presenta una dificultad fundamental: al disociarse en iones monoatómicos,  $\text{Na}^+$  y  $\text{Cl}^-$  no exhiben bandas Raman intensas en el rango espectral usual. En consecuencia, la cuantificación no puede basarse en “picos del soluto” de forma directa, como ocurre en analitos Raman-activos, sino que debe apoyarse en señales indirectas asociadas al solvente.

En soluciones acuosas, la banda de estiramiento O–H del agua constituye una de las regiones espectrales de mayor intensidad y sensibilidad. La forma de esta banda —ancha y altamente solapada— refleja una distribución continua de microentornos de enlace

de hidrógeno. La presencia de solutos iónicos perturba dicha distribución mediante interacciones ión-agua, produciendo redistribuciones sistemáticas de intensidad dentro de la banda O–H. Este fenómeno abre una vía para inferir la concentración salina a partir de cambios en el contorno espectral del agua. No obstante, esta estrategia enfrenta un desafío crítico: la banda O–H es también fuertemente dependiente de la temperatura, lo que introduce un confusor dominante que puede degradar la estimación si no se controla o modela explícitamente.

A partir de esta problemática, el propósito de esta tesis es desarrollar y validar una metodología cuantitativa para estimar la concentración de NaCl en soluciones acuosas utilizando espectroscopía Raman, explotando la información contenida en la banda O–H del agua e incorporando la temperatura como variable explícita del modelo. El enfoque adoptado se basa en describir la banda O–H mediante un conjunto reducido de descriptores espectrales efectivos, obtenidos a través de una descomposición paramétrica en funciones Gaussianas. Es importante enfatizar que dicha descomposición se utiliza como una herramienta matemática orientada a capturar redistribuciones sistemáticas de intensidad y a estabilizar parámetros para el modelado, sin atribuir necesariamente una interpretación estructural única o poblacional a cada componente.

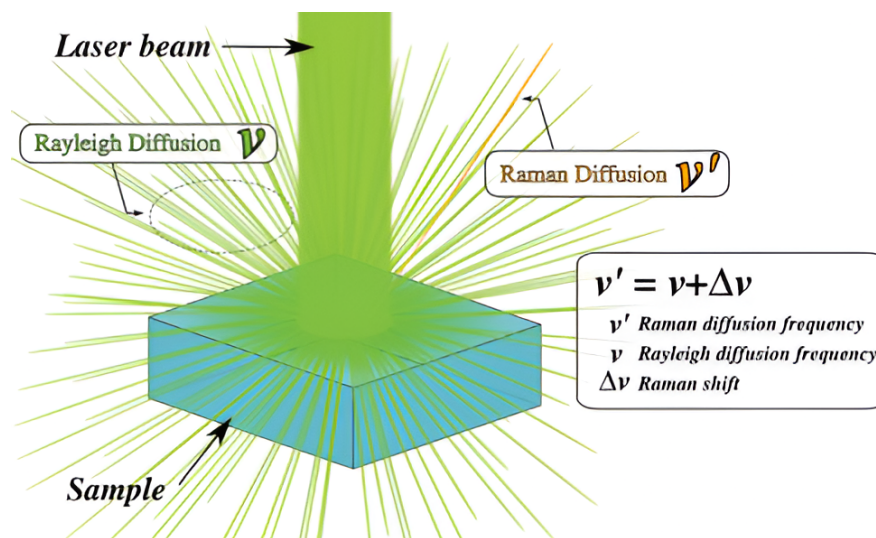
## **1.1. Trabajos previos**

### **1.1.1. Principio físico de la dispersión Raman**

La espectroscopía Raman se fundamenta en la dispersión inelástica de la luz por la materia. Cuando un haz láser monocromático interactúa con una muestra, la mayor parte de la radiación se dispersa de manera elástica (dispersión de Rayleigh), conservando su frecuencia. Sin embargo, una fracción mucho menor de la radiación interactúa con las vibraciones moleculares del sistema, produciendo una dispersión inelástica en la cual la frecuencia de la luz dispersada difiere de la incidente. Esta diferencia de frecuencia corresponde a la energía intercambiada con un modo vibracional y constituye el desplazamiento Raman [1].

En términos energéticos, la dispersión Raman puede describirse como un proceso de dos fotones que involucra un estado virtual: el fotón incidente induce una polarización

electrónica instantánea (no resonante) y el sistema reemite un fotón con energía distinta, quedando finalmente en un estado vibracional diferente al inicial. Si la molécula absorbe energía durante la interacción, el fotón dispersado tiene menor energía (menor frecuencia) y se observa una línea *Stokes*. Si, en cambio, la molécula se encontraba inicialmente en un estado vibracional excitado y cede energía al campo, el fotón dispersado aumenta su energía y se observa una línea *anti-Stokes* [1].



**Fig. 1.1:** Principio físico de la dispersión Raman [1].

La Figura 1.1 ilustra el fenómeno de dispersión de la luz en una muestra iluminada por un haz láser monocromático. La mayor parte de la radiación se dispersa elásticamente (difusión de Rayleigh, sin cambio de frecuencia  $\nu$ ), mientras que una fracción muy pequeña se dispersa inelásticamente, dando lugar a la radiación Raman con frecuencia  $\nu' = \nu \pm \Delta\nu$ . Este desplazamiento Raman ( $\Delta\nu$ ) corresponde a la energía de los modos vibracionales de las moléculas del material. La figura resume los fundamentos ópticos del efecto Raman, diferenciando los procesos Rayleigh y Raman y destacando la relación directa entre la energía vibracional y el desplazamiento expresado en  $\text{cm}^{-1}$ .

### 1.1.2. Espectroscopía Raman en medios acuosos: fundamentos para cuantificación

El uso de Raman como técnica analítica cuantitativa se sustenta en la relación entre la intensidad espectral y propiedades del sistema (concentración, temperatura,

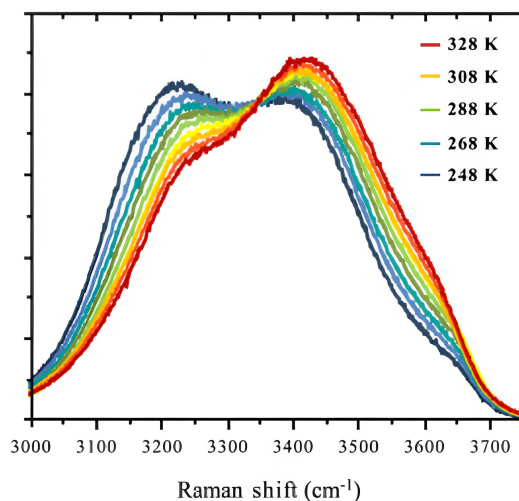
composición, estado físico), modulada por condiciones experimentales (longitud de onda de excitación, geometría de colección, enfoque, respuesta espectral del instrumento). En aplicaciones orientadas a cuantificación, la literatura enfatiza la necesidad de establecer protocolos de preprocesamiento, normalización y validación para mitigar variaciones no químicas, particularmente en monitoreo de procesos donde la temperatura y el entorno experimental cambian en el tiempo [1,2].

En el ámbito de medios acuosos, revisiones y capítulos orientados a aplicaciones destacan que Raman puede emplearse tanto para caracterización del solvente (agua) como para cuantificación de especies disueltas, con diferentes niveles de complejidad dependiendo de si el analito es Raman-activo o si la cuantificación debe realizarse de forma indirecta [1,3,4]. Esta distinción resulta crítica para sistemas salinos dominados por iones monoatómicos, donde la estrategia cuantitativa debe apoyarse en descriptores robustos del espectro del agua.

### **1.1.3. Espectro Raman del agua y dependencia con temperatura**

En agua líquida, la banda Raman del estiramiento O–H es intrínsecamente ancha y refleja una distribución continua de enlaces de hidrógeno. Diversos estudios experimentales muestran que cambios de temperatura inducen variaciones sistemáticas de la forma espectral (desplazamientos suaves, cambios de anchura y redistribución de intensidad), sin requerir la aparición de nuevas bandas discretas [5]. Esta sensibilidad convierte a la región O–H en un portador de información relevante para inferencia de variables termodinámicas, pero simultáneamente introduce la temperatura como un confusor mayor en esquemas de cuantificación por forma espectral.

La Figura 1.2 muestra los espectros Raman del agua pura registrados en el rango  $3000\text{--}3700\text{ cm}^{-1}$  a distintas temperaturas (248–328 K). Se observa una clara variación sistemática de la forma y la intensidad de la banda O–H: al aumentar la temperatura, el máximo principal se desplaza hacia números de onda mayores y la intensidad relativa del hombro de baja frecuencia disminuye. Este comportamiento refleja la ruptura progresiva de los enlaces de hidrógeno y la mayor proporción de moléculas en entornos menos coordinados. La figura ilustra así la alta sensibilidad de la banda O–H frente a la

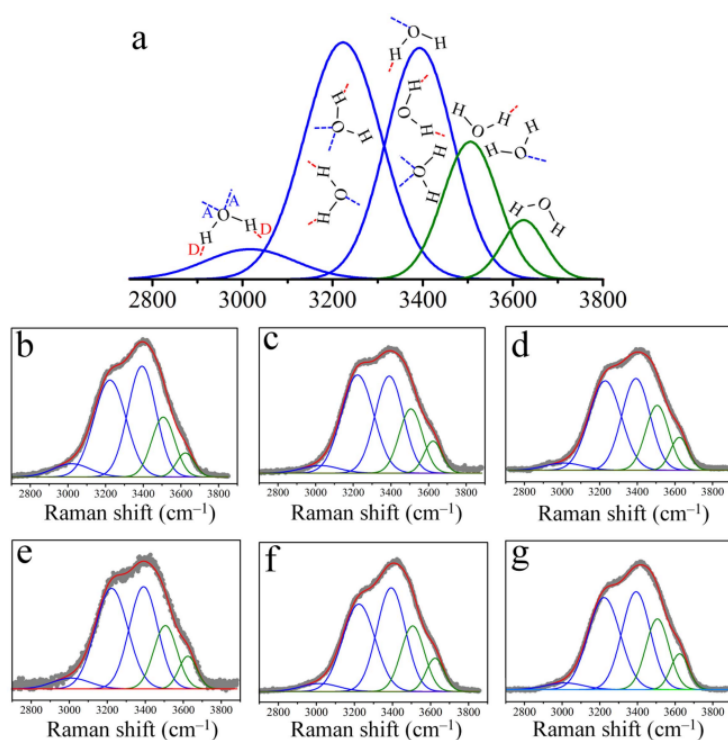


**Fig. 1.2:** Dependencia de la banda O–H del agua con la temperatura [2].

temperatura y su papel como descriptor térmico.

Desde una perspectiva moderna de espectroscopía vibracional, la banda O–H observada en espectros estacionarios se interpreta como un promedio sobre fluctuaciones rápidas del entorno local, lo que limita la lectura de la banda en términos de poblaciones estructurales estáticas [2]. Este marco respalda el uso de parámetros efectivos (por ejemplo, posiciones, amplitudes o anchos de componentes ajustados) como descriptores fenomenológicos útiles para modelado cuantitativo, evitando atribuciones ontológicas estrictas.

Una práctica recurrente en la literatura consiste en aproximar la banda O–H mediante combinaciones de funciones (p. ej., Gaussianas) para capturar redistribuciones internas de intensidad. En estudios de agua confinada, se han utilizado descomposiciones multicomponente para describir cambios sistemáticos asociados al entorno, si bien la interpretación de cada componente como especie discreta es debatida [6]. De manera análoga, análisis avanzados en mezclas agua–etanol emplean técnicas multivariantes para descomponer contribuciones espectrales y extraer tendencias con temperatura y composición, sin embargo, la reproducibilidad de la descomposición no implica unicidad física de los componentes [7]. En línea con la literatura sobre dinámica del agua [2], tales descomposiciones se consideran herramientas matemáticas para estabilizar descriptores espectrales, no evidencia directa de poblaciones moleculares discretas.



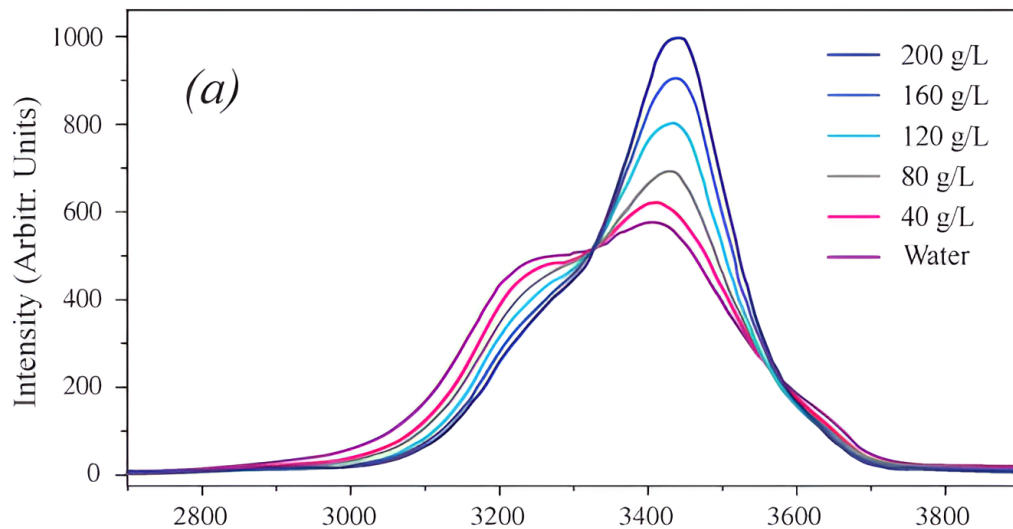
**Fig. 1.3:** Descomposición multicomponente de la banda O–H [8].

En la Figura 1.3, la parte superior (a) muestra una descomposición esquemática de la banda O–H del agua en varias contribuciones Gaussianas o lorentzianas, asociadas a diferentes entornos de enlace de hidrógeno (fuertemente ligados, débilmente ligados y libres). Las figuras (b–g) presentan ejemplos experimentales de ajustes realizados a espectros reales, donde la suma de las componentes (curva roja) reproduce con alta fidelidad la forma global de la banda observada. Este procedimiento permite cuantificar variaciones sutiles de intensidad y posición, utilizadas como descriptores espectrales para el modelado cuantitativo.

#### 1.1.4. Efecto de la salinidad en el espectro Raman del agua

La incorporación de sales en agua modifica la red de enlaces de hidrógeno a través de la hidratación iónica, afectando la distribución de entornos locales y, por tanto, la forma de la banda O–H. En cloruros, donde no aparecen bandas Raman intensas del soluto, el efecto cuantitativo se manifiesta principalmente como cambios suaves de forma y desplazamientos efectivos de subregiones de la banda. Estudios orientados

a altas salinidades y condiciones extendidas (incluyendo fases sólidas y transiciones) muestran que la región O–H continúa siendo el descriptor dominante, con sensibilidad a composición y estado físico [9].



**Fig. 1.4:** Efecto de la concentración salina en la banda O–H del agua [10].

La Figura 1.4 presenta la evolución del espectro Raman del agua para diferentes concentraciones de NaCl (0 a 200 g/L). A medida que aumenta la concentración salina, la banda O–H experimenta una disminución general de la intensidad y un cambio de forma asociado a la perturbación de la red de enlaces de hidrógeno. La redistribución de intensidad hacia frecuencias más altas indica un debilitamiento promedio de las interacciones H–O–H. Este patrón confirma que la salinidad afecta indirectamente la estructura del agua y, por tanto, puede utilizarse como variable de cuantificación a partir del espectro Raman.

Trabajos orientados al desarrollo de sensores ópticos han explorado la cuantificación en soluciones NaCl mediante descriptores espectrales asociados al agua, proponiendo esquemas de medición y calibración basados en cambios sistemáticos del espectro [10, 11]. En un contexto distinto, el estudio de inclusiones fluidas alojadas en cuarzo destaca que, aun cuando la señal de interés se encuentra en el O–H, factores como la polarización pueden inducir variaciones espectrales relevantes, lo que obliga a controlar condiciones ópticas o adoptar estrategias robustas de adquisición y extracción de descriptores [3]. Este resultado es especialmente importante para cualquier metodología cuantitativa

basada en forma espectral, ya que muestra que efectos instrumentales y geométricos pueden competir con el efecto químico de la salinidad.

Más allá de NaCl, el análisis Raman de soluciones salinas con otros iones permite evaluar tendencias generales de hidratación y perturbación del agua. Por ejemplo, estudios sobre cloruro y estructura del agua en rangos de temperatura extendidos evidencian cambios sistemáticos asociados a la interacción ión-agua [12]. Asimismo, estudios en sistemas LiCl y mezclas LiCl–MgCl<sub>2</sub> exploran espectros excedentes y espectros de hidratación como herramientas para aislar contribuciones asociadas al entorno de hidratación [13]. Estos antecedentes refuerzan la idea de que el mapeo entre forma espectral del O–H y concentración puede ser no trivial y dependiente del ion.

En sistemas binarios de sales, se ha reportado que esquemas univariados pueden fallar por efectos cruzados que alteran la banda O–H de forma no lineal, motivando el uso de modelos multivariantes y no lineales para capturar la interacción entre composición y respuesta espectral [14].

### 1.1.5. Cuantificación Raman: enfoques directos e indirectos

Cuando el analito posee bandas Raman intensas y distinguibles, la cuantificación puede abordarse mediante integración de picos y calibraciones univariadas o multivariantes, con validación frente a métodos de referencia. Ejemplos incluyen cuantificación de sales en eflorescencias sólidas mediante FT-Raman [15], cuantificación de formulaciones farmacéuticas [16] o detección/estimación de contaminantes en agua combinando Raman y SERS [17]. Revisiones sobre detección de nitratos mediante Raman también se sitúan en este régimen, donde la señal relevante corresponde a modos vibracionales del analito [18]. Estos casos son útiles como referencia metodológica, pero no son directamente transferibles a NaCl disuelto, donde la cuantificación debe ser indirecta.

Para analitos Raman-inactivos o con señal débil, la cuantificación se apoya en descriptores del espectro del solvente. En agua, la banda O–H permite construir observables (posición efectiva, centroides, áreas parciales, parámetros de ajuste) que varían con temperatura, composición iónica y concentración. Aplicaciones recientes reportan incluso mediciones simultáneas de temperatura y salinidad mediante descriptores Raman del agua, destacando la necesidad de tratar ambas variables de

forma conjunta [19]. En escenarios aplicados, el uso del O–H evidencia que la banda puede funcionar como sensor del entorno local, aunque el régimen biológico introduce otras fuentes de variabilidad [20].

La quimiometría proporciona un marco sistemático para relacionar espectros complejos con variables objetivo mediante modelos supervisados. En cuantificación Raman, estrategias como Partial Least Square (PLS) han demostrado utilidad al reducir dimensionalidad y enfatizar regiones informativas, especialmente cuando la respuesta espectral es solapada o presenta interferencias [21]. En paralelo, se reportan métodos basados en transformadas (por ejemplo, wavelets) como preprocesamiento para estabilizar el modelado cuantitativo en sistemas multicomponente [22]. Estos enfoques son relevantes como línea base comparativa en problemas de cuantificación indirecta, pero su interpretabilidad y extrapolación dependen fuertemente del dominio experimental.

El uso de redes neuronales u otros modelos no lineales se ha propuesto para capturar relaciones complejas entre forma espectral y composición, particularmente en mezclas salinas donde la no linealidad se vuelve dominante [14]. Revisiones recientes discuten el crecimiento de métodos de aprendizaje automático aplicados a Raman, destacando su potencial y también desafíos asociados a generalización, trazabilidad y explicabilidad [23]. Estos antecedentes se consideran relevantes para contextualizar el espacio metodológico, aun cuando el enfoque adoptado prioriza modelos explícitos.

### **1.1.6. Factores instrumentales y metrológicos en cuantificación basada en O–H**

Cuando la cuantificación descansa sobre cambios sutiles de forma espectral, la variabilidad instrumental puede ser del mismo orden que el efecto de interés. En mediciones de inclusiones fluidas en cuarzo, la polarización inducida por la interfaz puede modificar parámetros del O–H, exigiendo control geométrico y/o estrategias robustas de adquisición [3]. Este antecedente es particularmente útil para justificar la incorporación de procedimientos de control de calidad espectral, repetibilidad y evaluación de sensibilidad a alineación experimental.

Dado que la temperatura modifica sistemáticamente la banda O–H [2, 5], cualquier

esquema de cuantificación indirecta debe (i) estabilizar la temperatura experimental o (ii) incorporarla explícitamente en el modelo. La literatura orientada a monitoreo de procesos discute la necesidad de modelar variaciones espectrales inducidas por temperatura para evitar sesgos en estimación cuantitativa [24]. En el contexto de mediciones simultáneas de salinidad y temperatura, se vuelve explícito que ambas variables deben tratarse de manera simultánea para evitar ambigüedades [19].

### 1.1.7. Casos límite

Con el fin de delimitar el alcance de los modelos cuantitativos basados en O–H, es útil considerar casos límite donde el régimen físico cambia. En condiciones de nanoconfinamiento y realce superficial (SERS), se han reportado espectros compatibles con estructuras tipo hielo bajo condiciones ambientales, evidenciando que el entorno puede llevar a respuestas vibracionales cualitativamente distintas a las propiedades físicas aparentes del agua [25]. Asimismo, en condiciones de alta presión/temperatura, se han reportado reacciones o transformaciones en sistemas que involucran NaCl, las cuales no son representativas de soluciones acuosas convencionales pero sirven como frontera conceptual del comportamiento del sistema [26]. En el dominio de fases sólidas e hidratos, se discuten estructuras hidratadas de NaCl relevantes para condiciones planetarias o para estados metaestables, que nuevamente delimitan un régimen diferente al de las soluciones líquidas estudiadas en esta tesis [27, 28].

Adicionalmente, estudios teóricos (por ejemplo, DFT) sobre la disolución molecular de NaCl en agua aportan perspectivas mecanísticas sobre interacciones ion–agua, pero su conexión directa con espectros Raman de soluciones macroscópicas requiere cautela, dado que los modelos se construyen en escalas microscópicas y condiciones idealizadas [29].

### 1.1.8. Discusión

La literatura revisada converge en varios puntos: (i) la banda O–H del agua es altamente sensible al entorno y a variables termodinámicas, especialmente temperatura [2, 5]; (ii) en soluciones de cloruros y particularmente en NaCl, la información espectral útil para cuantificación surge principalmente de cambios inducidos en el agua [9, 10]; (iii) en

sistemas multicomponente, la relación entre forma espectral y concentración puede ser no lineal y requerir enfoques multivariantes/no lineales [14]; y (iv) la variabilidad instrumental y geométrica puede sesgar la estimación si no se controla, especialmente cuando los descriptores dependen de cambios sutiles en el O–H [3].

Sobre esta base, la presente tesis se posiciona en el marco de cuantificación indirecta basada en descriptores efectivos del O–H, adoptando una postura conservadora respecto a la interpretación estructural de descomposiciones multicomponente [2, 6, 7]. En particular, se busca un modelado cuantitativo que incorpore explícitamente variables físicas relevantes (concentración y temperatura), privilegiando robustez y precisión para la estimación de salinidad a partir de espectros Raman.

## 1.2. Hipótesis de trabajo

Existe una relación estable entre la concentración y la temperatura de una solución acuosa de NaCl que permite recuperar la concentración con un error acotado usando espectroscopía Raman a partir de descriptores de la banda O-H del agua e incorporando explícitamente la temperatura.

## 1.3. Objetivos

### 1.3.1. Objetivo general

Desarrollar y validar una metodología cuantitativa basada en espectroscopía Raman para estimar la concentración de NaCl en soluciones acuosas a partir de la banda O–H del agua, incorporando explícitamente la temperatura como variable del modelo para asegurar precisión y robustez metrológica.

### 1.3.2. Objetivos específicos

1. Generar una base de datos espectral Raman de soluciones acuosas de NaCl en función de la concentración y la temperatura, que permita el análisis y modelado

de la respuesta espectral asociada a la banda O-H del agua.

2. Implementar un flujo de preprocesamiento espectral y extraer descriptores robustos de la banda O-H mediante descomposición en funciones Gaussianas.
3. Formular y ajustar un modelo cuantitativo que relacione los descriptores espectrales con la concentración  $C$  y la temperatura  $T$
4. Validar el desempeño del modelo mediante estrategias de validación interna y externa, cuantificando la exactitud y la capacidad de generalización mediante métricas como el error absoluto medio (MAE) y la raíz del error cuadrático medio (RMSE).
5. Evaluar comparativamente el desempeño del modelo paramétrico propuesto frente a un enfoque multivariante de referencia basado en Partial Least Squares (PLS)).

## 1.4. Metodología

### 1.4.1. Preparación de soluciones de NaCl

Se prepararán soluciones acuosas de Cloruro de Sodio (NaCl) con concentraciones conocidas, abarcando un rango representativo que asegure variabilidad suficiente para construir y validar modelos de cuantificación. Las concentraciones oscilan entre 0g/100mL y 5g/100mL, distribuidas en intervalos regulares de 1g/100mL. Para la preparación de las muestras se utiliza:

- Cloruro de Sodio para análisis.
- Agua desionizada tipo I.
- 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.

### 1.4.2. 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: 60s.

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 substrerá el fondo espectral en cada cambio de temperatura y cambio de muestra para la obtención de espectros más limpios.

### 1.4.3. Procesamiento espectral y extracción de descriptores

Para el procesamiento espectral únicamente se aplicará un recorte espectral manteniendo el intervalo 3150–3550  $\text{cm}^{-1}$ , correspondiente a la región más informativa de la banda de estiramiento O–H del agua.

En esta metodología, no se aplicaron rutinas adicionales de corrección de línea base, normalización ni suavizado (e.g, Savitzky–Golay o FFT).

La banda O–H se modelará mediante una descomposición paramétrica global basada en  $K$  funciones Gaussianas con parámetros compartidos de posición y ancho para todo el conjunto de entrenamiento, permitiendo que las amplitudes varíen por espectro. Por balance entre error de reconstrucción y estabilidad paramétrica, se empleará  $K = 3$ .

### 1.4.4. Validar el desempeño del modelo mediante estrategias de validación interna y externa

El desempeño se evaluará mediante:

- **Validación interna:** evaluación sobre el conjunto de entrenamiento, verificando la consistencia del ajuste y el comportamiento de residuos.
- **Validación externa:** evaluación sobre un conjunto independiente de 30 espectros no utilizados en la calibración. Para estos espectros, se aplicará el mismo pipeline de preprocesamiento.

#### 1.4.5. Comparar la metodología propuesta con enfoque de referencia Partial Least Square (PLS)

Finalmente, se comparará el rendimiento del modelo propuesto frente al enfoque combinado de Individual Contribution Standardization (ICS) y Partial Least Square (PLS), analizando exactitud global, estabilidad de residuos y comportamiento en el rango de concentración.

### 1.5. Alcances y limitaciones

#### 1.5.1. Alcances

El presente trabajo se enmarca en el desarrollo y validación de una metodología cuantitativa para la estimación de concentración de NaCl en soluciones acuosas mediante espectroscopía Raman, utilizando como portador de información la banda de estiramiento O–H del agua. En particular, los alcances de la tesis se delimitan como sigue:

- **Sistema de estudio:** soluciones acuosas binarias NaCl–H<sub>2</sub>O, preparadas en condiciones de laboratorio y con concentraciones nominales en el rango definido experimentalmente (0 a 5 g/100 mL).
- **Dominio térmico:** mediciones realizadas con control de temperatura dentro del rango experimental considerado (10 a 50 °C), incorporando explícitamente  $T$  en el modelado para reducir sesgos por variación térmica.
- **Región espectral de análisis:** el modelado se centra en la banda O–H del agua, trabajando sobre un intervalo espectral acotado (3150–3550 cm<sup>-1</sup>), el cual

contiene la información más relevante para describir redistribuciones sistemáticas asociadas a la presencia de NaCl y a cambios de temperatura.

- **Extracción de descriptores:** parametrización del contorno O–H mediante descomposición en un número reducido de funciones Gaussianas, empleadas como descriptores efectivos (principalmente amplitudes) para el modelado concentración–temperatura.
- **Modelo cuantitativo e inversión:** formulación y ajuste de un modelo explícito que relaciona los descriptores con concentración  $C$  y temperatura  $T$  (incluyendo el término de interacción  $C \cdot T$ ), junto con un esquema invertible para estimar  $C$  a partir de un espectro observado y una temperatura conocida.
- **Evaluación de desempeño:** validación del método mediante estrategias de validación interna y externa, cuantificando exactitud y generalización con métricas estándar (MAE, RMSE y  $R^2$ ), y comparación con un enfoque de referencia basado en PLS.

### 1.5.2. Limitaciones

Aun cuando la metodología propuesta demuestra capacidad de estimación dentro del dominio experimental definido, existen limitaciones inherentes al alcance de esta tesis y a la naturaleza del problema de cuantificación indirecta basada en la banda O–H:

- **Dominio de validez acotado:** los modelos calibrados son válidos dentro de los rangos de concentración y temperatura muestreados. Fuera de dicho dominio (concentraciones mayores, temperaturas fuera de rango o cercanas a cambios de fase), el desempeño no está garantizado y requeriría recalibración o extensión del conjunto de entrenamiento.
- **Monosalinidad (sistema binario):** el estudio se restringe a NaCl como único soluto. En sistemas multicomponente o con iones adicionales, la banda O–H puede presentar efectos cruzados y no linealidades que podrían degradar la estimación si no se incorporan variables adicionales o modelos más complejos.
- **Dependencia instrumental y de configuración:** al basarse en cambios sutiles de forma espectral, el método puede ser sensible a variaciones instrumentales

(alineación, enfoque, estabilidad de potencia, respuesta espectral del sistema, condiciones ópticas de la cubeta). En consecuencia, su transferencia a otro equipo o configuración requiere verificar reproducibilidad y, potencialmente, recalibrar.

- **Interpretación de componentes:** la descomposición en Gaussianas se utiliza como herramienta matemática para estabilizar descriptores; no constituye evidencia directa de poblaciones moleculares discretas ni permite, por sí sola, asignaciones estructurales únicas del agua. Por lo tanto, la tesis prioriza cuantificación y reproducibilidad por sobre interpretación microscópica.
- **Requisitos de temperatura:** el esquema asume control de la temperatura al momento de estimar la concentración. Si  $T$  no se conoce o presenta gradientes significativos, la incertidumbre en la estimación de  $C$  puede aumentar de manera sistemática.
- **Cobertura de interferencias químicas:** no se evalúan interferencias por otras especies químicas (p. ej., materia orgánica, carbonatos, sulfatos, variaciones de pH u otras sales). En aplicaciones reales, tales interferencias podrían modificar el contorno O–H y requerir estrategias de control, detección de outliers o ampliación del modelo.
- **Condiciones físico-químicas no exploradas:** no se aborda el efecto de presión, confinamiento, presencia de burbujas, dispersión múltiple u otros regímenes no equivalentes al de soluciones líquidas macroscópicas en laboratorio, los cuales pueden producir respuestas espectrales cualitativamente distintas.

En conjunto, estos alcances y limitaciones indican que la metodología es adecuada como solución cuantitativa dentro del dominio experimental controlado y como base para futuras extensiones. En trabajos futuros, la generalización podría fortalecerse mediante: (i) ampliación del dominio  $C$ – $T$ , (ii) incorporación de mezclas salinas y potenciales interferentes, (iii) evaluación inter-instrumento y protocolos de estandarización, y (iv) validación en condiciones cercanas al entorno de aplicación.

# Capítulo 2

## Desarrollo

Esta sección describe el trabajo experimental y computacional realizado en la presente tesis, con el propósito de construir un modelo 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 y modelado espectral, y (6) Modelo con Individual Contribution Standardization (ICS) y Partial Least Square (PLS).

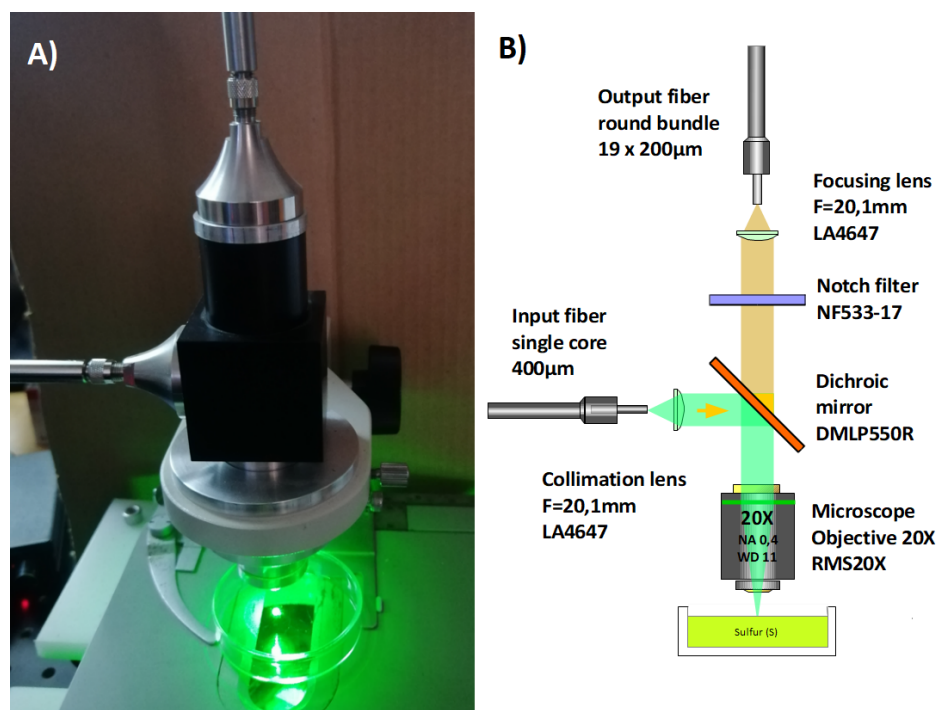
### 2.1. Calibración del espectrómetro

El montaje experimental utilizado en este estudio consistió en un sistema de espectroscopía Raman equipado con los siguientes elementos:

- Diodo láser de 532,10 nm con una potencia de 300 mW.
- Un porta cubeta de 10 x 10 x 40 mm, diseñado para ofrecer puertos de observación a 90° y 180°, con control de temperatura mediante un controlador modelo Quantum Northwest Qpod.
- Un espectrómetro HORIBA Jobin Yvon ISA HR-320, con una fibra óptica conectada al sistema para la adquisición de los espectros.
- Computadora para adquisición de espectros y análisis mediante el lenguaje de

programación Python.

El láser emite luz monocromática a 532,10 nm a través de una fibra hasta el cabezal Raman. Dentro del cabezal Raman, la luz láser se hace incidir a 45° directamente sobre un espejo dicroico, cuya propiedad reflectiva es total para 532 nm y transmisiva para longitudes de onda mayores a 535 nm. Posteriormente, la luz láser es reflejada hacia abajo cruzando un objetivo de microscopio de 20X el cual concentra la irradiancia láser directamente sobre la muestra. Esta configuración de cabezal Raman se utiliza para la medición de sustancias estándares de referencia como el Azufre y Poliestireno.



**Fig. 2.1:** Configuración del cabezal Raman para calibración.

Una vez que la radiación láser es enfocada sobre la muestra de azufre contenida en un envase de poliestireno, la excitación inducida genera transiciones roto-vibracionales que producen un leve corrimiento en la frecuencia de parte de los fotones por efecto Raman. Una fracción de esta radiación dispersada es colectada por el mismo objetivo del microscopio. La señal recolectada atraviesa el espejo dicroico, el cual permite únicamente la transmisión de longitudes de onda mayores a 535 nm. Posteriormente, el haz pasa por un filtro notch que suprime la radiación residual correspondiente a la línea láser.

A continuación, la señal es enfocada mediante una lente hacia una fibra óptica de análisis

de múltiples núcleos. En el extremo opuesto, dichos núcleos se disponen en un arreglo lineal que se alinea paralelamente con la rendija de entrada del espectrómetro HR-320 ( $50 \mu\text{m} \times 6 \text{mm}$ ).

Para llevar a cabo el análisis y reducción de los espectros Raman para el estudio, se adoptó el protocolo descrito por D. Hutsebaut, P. Vandenabele y L. Moens, *Analyst*, 2005,130, 1204-1214, “Evaluation of an accurate calibration and spectral standardization procedure for Raman spectroscopy” [30]. En el cual se describe un procedimiento analítico para la reproducibilidad, estabilidad e intercomparación de espectros Raman medidos, con respecto a estándares de referencia establecidos, tanto para calibración absoluta en nanómetros de longitud de onda, como también la conversión a número de onda  $\text{cm}^{-1}$ , dependiente de la estabilidad de la línea principal de excitación láser.

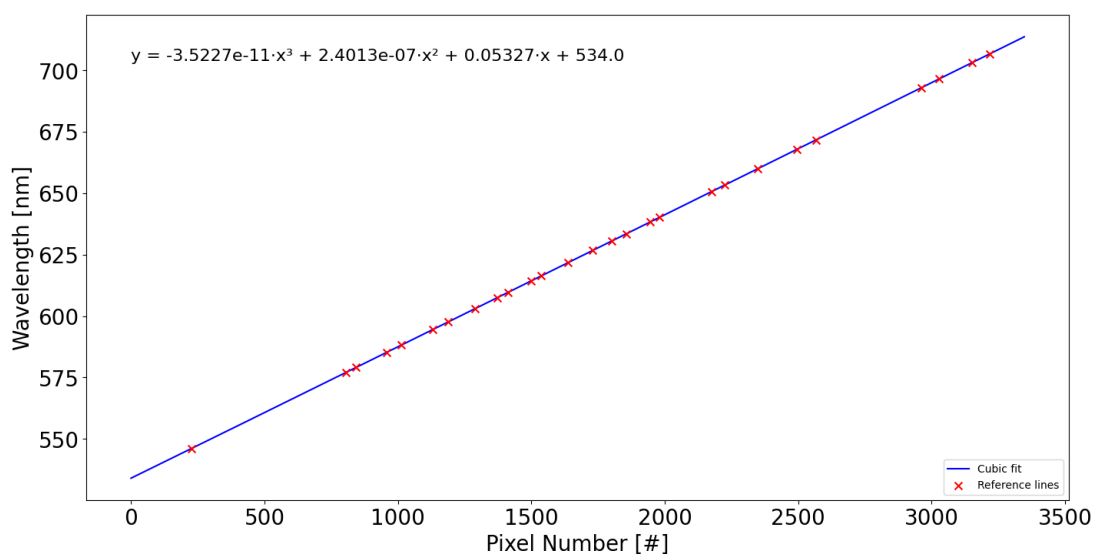
### 2.1.1. Calibración Absoluta en nanómetros [nm]

Utilizando lámparas espectrales de calibración de Mercurio [Hg], Argón [Ar] y Neón [Ne], se convierte desde el número de píxel de un espectro a nanómetros consiguiendo así, el vector de calibración en [nm] por medio de un ajuste polinomial de un orden tal que minimice el residual, típicamente de segundo o tercer orden. Las principales líneas de emisión espectral contenidas dentro de este rango están resumidas en la tabla 2.1.

**Tabla 2.1:** Líneas espectrales de Mercurio (Hg), Argón (Ar) y Neón (Ne).

Elemento	$\lambda$ (nm)	Elemento	$\lambda$ (nm)	Elemento	$\lambda$ (nm)
Hg	546,074	Ne	609,616	Ne	650,653
Hg	576,960	Ne	614,306	Ne	653,288
Hg	579,066	Ne	616,359	Ne	659,895
Ne	585,249	Ne	621,728	Ne	667,827
Ne	588,189	Ne	626,649	Ne	671,704
Ne	594,483	Ne	630,479	Ne	692,947
Ne	597,553	Ne	633,443	Ar	696,543
Ne	602,999	Ne	638,299	Ne	703,241
Ne	607,433	Ne	640,220	Ar	706,722

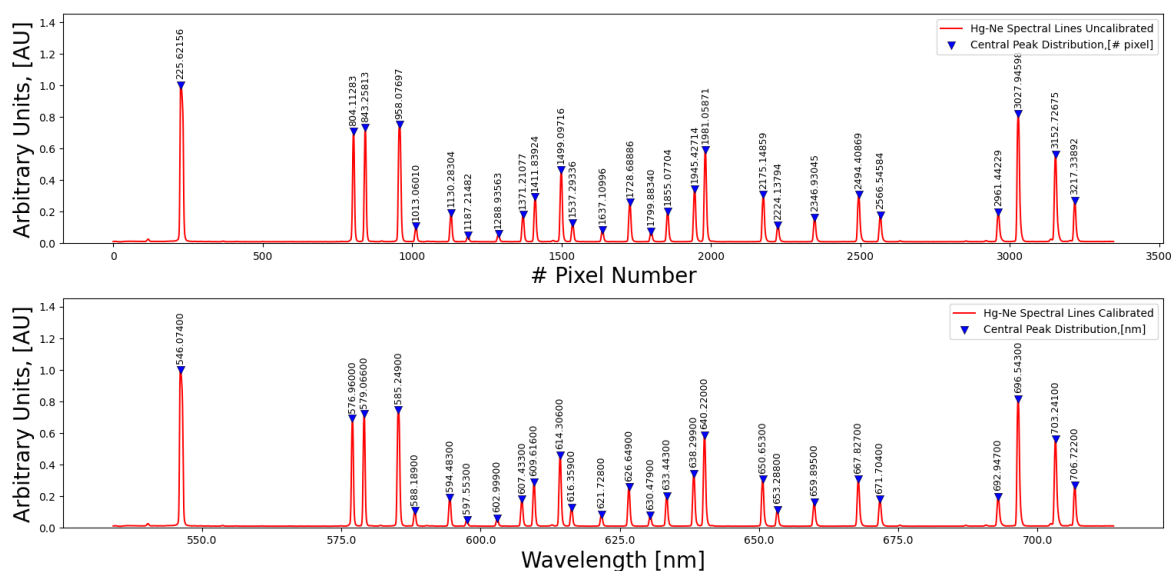
Esta etapa del protocolo consiste en identificar el centro de distribución de cada peak en número de píxel y mapearlo con respecto a las líneas espectrales atómicas de referencia en nanómetros por medio de una transformación analítica con un polinomio de orden superior. Tal como aparece en la Figura 2.2, donde cada punto del gráfico es un par ordenado entre un peak localizado en un número de píxel y su correspondiente línea de referencia. En los ejemplos de aplicación efectuados, si el mapeo entre líneas no tiene errores de interpretación de los espectros, la calibración será muy precisa con un error máximo estimado, menor a  $\pm 0,005\text{nm}$ . En la Figura 2.3 se muestra el resultado final aplicando la transformación encontrada, donde cada línea estimada se aproxima muy cercano al valor de la línea de referencia.



**Fig. 2.2:** Curva de calibración entre localización del centro de distribución de un peak en número de píxel versus el valor de la línea de referencia atómica en nanómetros.

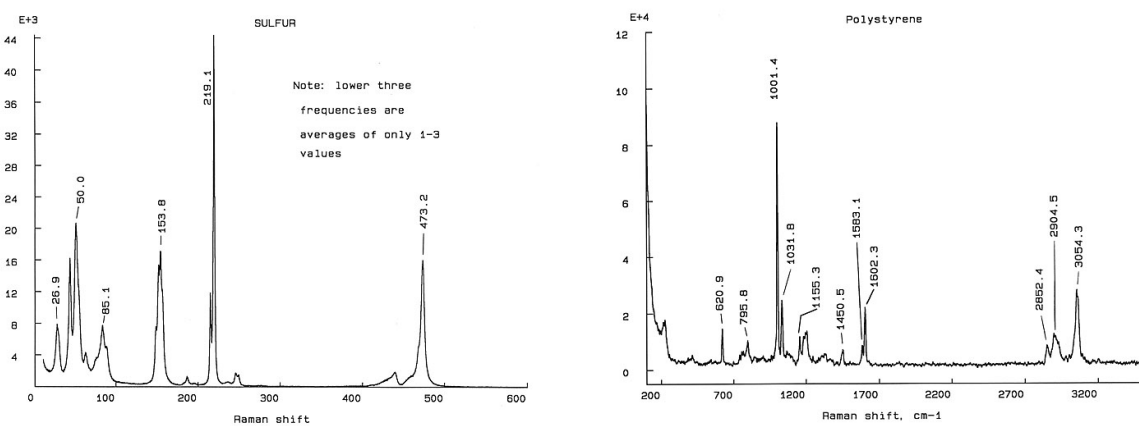
### 2.1.2. Calibración en Número de Onda [ $\text{cm}^{-1}$ ]

El segundo paso es convertir desde número de píxel a número de onda [ $\text{cm}^{-1}$ ], para ello se consideran muestras de referencia cuyos espectros Raman han sido medidos y catalogados en varias bases de datos disponibles para toda la comunidad de espectroscopía Raman. Las sustancias empleadas como referencia fueron el Azufre (S) y Poliestireno, cuyas líneas Raman espectrales y desviación estándar se resumen en la Tabla 2.2.



**Fig. 2.3:** Gráfico superior, muestra los valores de los peak, en número de píxel, con respecto al centro de distribución Gaussiana en torno a un valor máximo entero. Gráfico inferior, transformación de calibración absoluta aplicada, para convertir el número de píxel en longitud de onda.

El aspecto de los espectros Raman para estas dos sustancias de referencia tomado de la base de datos es el siguiente:



(a) Espectro Raman de Azufre.

(b) Espectro Raman de Poliestireno.

**Fig. 2.4:** Espectros característicos de Raman Shift [ $\text{cm}^{-1}$ ] para los estándares de calibración de Azufre y Poliestireno. El valor de las líneas se encuentra ordenados en la Tabla 2.2

**Tabla 2.2:** Líneas referenciales en Raman Shift [ $\text{cm}^{-1}$ ] y desviación estándar [ $\text{cm}^{-1}$ ]. Muestra mineral de Azufre 100% puro y RPC Thorlabs Polystyrene Raman Spectroscopy Calibration Sample, Average of FT-Ramanand Scanning (Spex 1403) Data Sets.

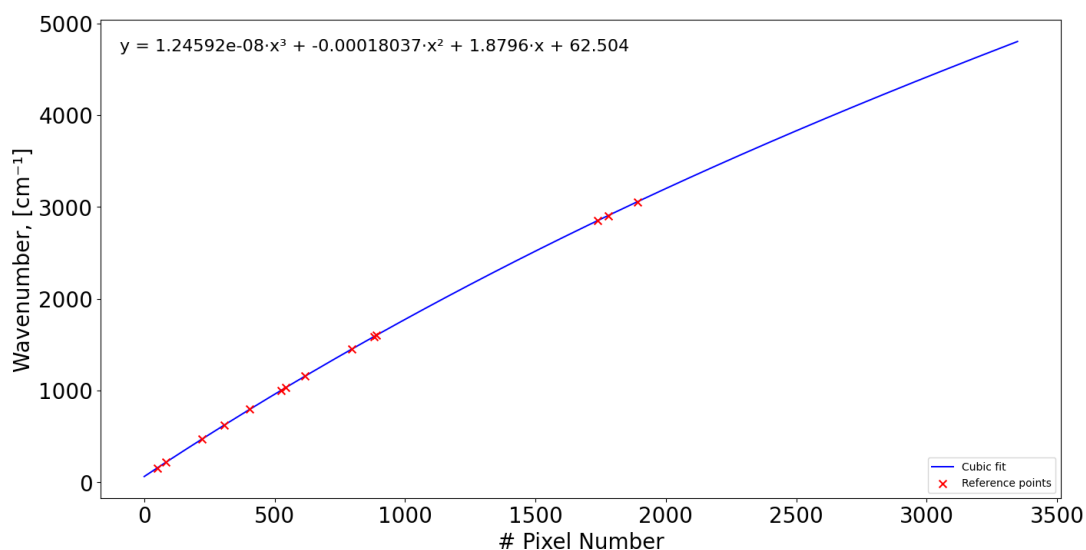
N° Línea	Azufre (S) [ $\text{cm}^{-1}$ ]	Poliestireno [ $\text{cm}^{-1}$ ]
(1)	$153,80 \pm 0,50$	$620,90 \pm 0,69$
(2)	$219,10 \pm 0,57$	$795,80 \pm 0,78$
(3)	$473,20 \pm 0,49$	$1001,40 \pm 0,54$
(4)		$1031,80 \pm 0,43$
(5)		$1155,30 \pm 0,56$
(6)		$1450,50 \pm 0,56$
(7)		$1583,10 \pm 0,86$
(8)		$1602,30 \pm 0,73$
(9)		$2852,40 \pm 0,89$
(10)		$2904,50 \pm 1,22$
(11)		$3054,30 \pm 1,36$

Tras el reconocimiento visual de los espectros Raman se consigue identificar cada peak con su valor referencial, por medio de su ubicación y característica del perfil, reasignando los pares coordenados para la calibración.

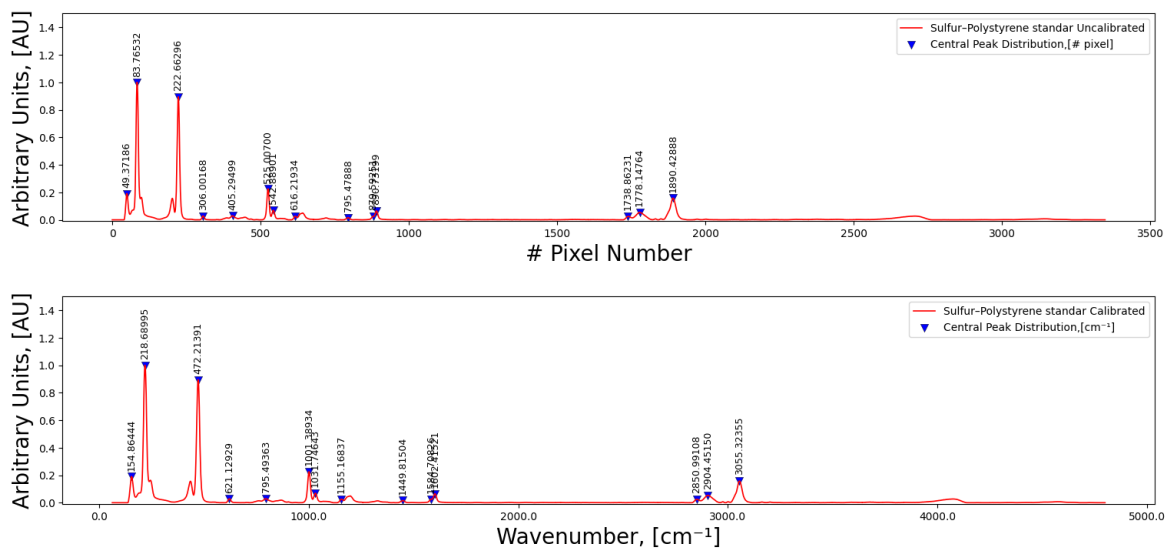
Con la realización de los pasos 1 y 2, se obtienen los outputs de los vectores de calibración de longitud de onda y número de onda, los cuales son exportados como datos recurrentes para la reducción de cualquier muestra desconocida en un proceso de medición.

Durante la etapa de calibración se verificó la correcta correspondencia entre el eje espectral del sistema y valores de referencia conocidos. La Fig 2.5 muestra la curva de calibración en número de onda obtenida de las líneas referenciales de azufre y poliestireno en concordancia con el número de píxeles del espectrómetro. Asimismo, la Fig 2.6 presenta la repetibilidad de la medición bajo condiciones controladas, evidenciando estabilidad en la posición y forma espectral. Por su parte, la Tabla 2.3 resume cuantitativamente los resultados de la calibración.

El proceso de calibración espectral se realizó en dos etapas complementarias con el



**Fig. 2.5:** Curva de calibración del vector número de onda, ajustando un polinomio de 3er orden, para los pares ordenados de los peaks como estándar de referencia entre número de píxel y número de onda.



**Fig. 2.6:** Gráfico superior, muestra los valores de los peak, en número de píxel, con respecto al centro de distribución Gaussiana en torno a un valor máximo entero. Gráfico inferior, transformación de calibración absoluta aplicada, para convertir el número de píxel en Número de onda [cm<sup>-1</sup>].

objetivo de asegurar la trazabilidad y exactitud del eje espectral del sistema.

**Tabla 2.3:** Conversión de parámetros espectroscópicos entre longitud de onda (nm) y número de onda ( $\text{cm}^{-1}$ ). Láser: 532,10 nm, rendija (slit): 50  $\mu\text{m}$ .

Parámetro	Rango espectral [nm]	Raman shift [ $\text{cm}^{-1}$ ]
$\Delta\lambda$	[533,98, 713,82] = 179,94 nm	[71,0, 4828,9] = 4757,9 $\text{cm}^{-1}$
FWHM	0,30 nm	8 $\text{cm}^{-1}$
Número de píxeles	3348	3348
Densidad ( $\Delta\lambda/N_{\text{pixel}}$ )	0,05 nm/píxel	1,42 $\text{cm}^{-1}$ /píxel
FWHM / densidad	5,58 píxeles	5,63 píxeles

En una primera etapa se efectuó la calibración en longitud de onda (nm) utilizando lámparas patrón de Neón (Ne) y Mercurio (Hg), cuyos picos espectrales presentan posiciones conocidas y bien tabuladas. Se identificaron los máximos experimentales correspondientes a cada línea espectral y se asoció su posición en píxeles del detector CCD con la longitud de onda nominal, ajustando una función de calibración que permite establecer la relación píxel–longitud de onda.

En una segunda etapa se verificó y refinó la calibración en términos de desplazamiento Raman ( $\text{cm}^{-1}$ ) mediante la medición de materiales patrón (azufre y poliestireno), cuyos modos vibracionales Raman presentan posiciones características conocidas. Esta etapa permite validar la consistencia de la conversión a número de onda Raman dentro del rango espectral de interés.

Una vez determinado el vector de calibración espectral, este se aplicó de manera uniforme a todos los espectros adquiridos en la base de datos experimental, garantizando que cada medición sea expresada en intensidad versus desplazamiento Raman ( $\text{cm}^{-1}$ ) con consistencia metrológica.

La calibración espectral se realizó al inicio de cada jornada experimental, verificando que no existieran desviaciones significativas en la correspondencia píxel–número de onda.

## 2.2. 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 sirvan 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 de alta precisión ( $\pm 0.1$  mg). Las masas pesadas fueron:

**Tabla 2.4:** Masa de NaCl en cada solución preparada.

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

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

La concentración de cloruro de sodio se expresó en unidades de g/100 mL, correspondientes a una concentración másica. Esta elección se encuentra directamente relacionada con el protocolo experimental de preparación de las muestras, el cual estuvo condicionado por el uso de frascos con un volumen nominal de 100 mL.

Las soluciones se prepararon pesando una masa controlada de NaCl mediante una balanza analítica y completando el volumen final a 100 mL, garantizando así condiciones reproducibles entre muestras. Cabe destacar que esta forma de expresar la concentración es equivalente, mediante una conversión directa, a concentraciones comúnmente reportadas en g/L, donde 1 g/100 mL corresponde a 10 g/L.

Durante la preparación se identificaron las siguientes fuentes de error:

- **Balanza analítica:** resolución de 0.1 mg.

- **Material volumétrico:** probeta con una tolerancia de  $\pm 1.0$  mL.
- **Variabilidad del conductivímetro:** observada experimentalmente a partir de las tres réplicas realizadas por muestra.

## 2.3. 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.

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

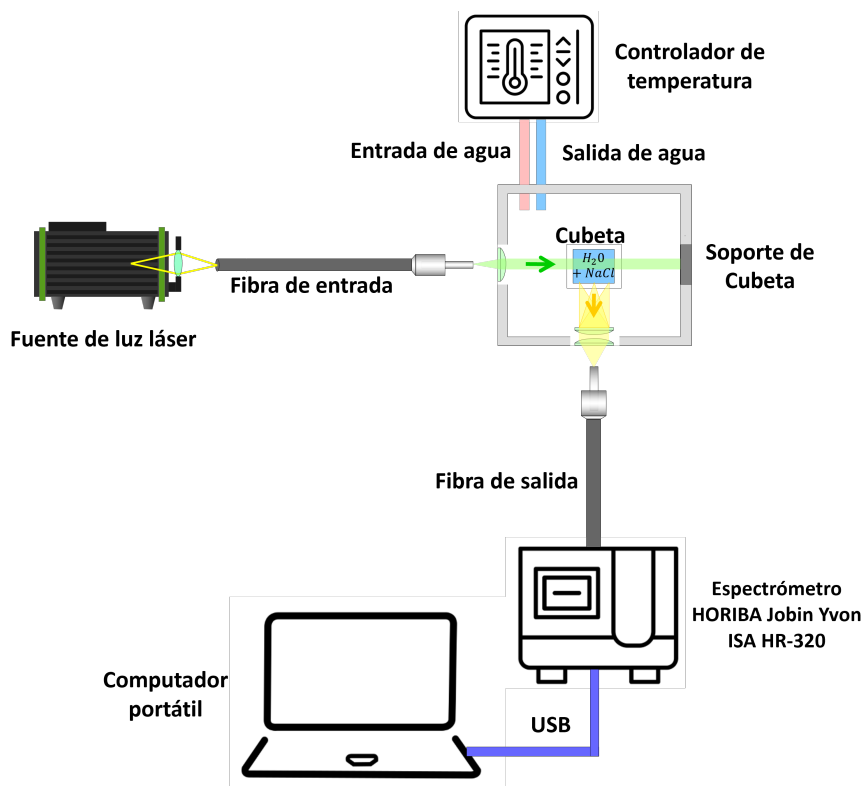
1. Establecer la temperatura en  $10\text{ }^{\circ}\text{C}$  y, con el láser apagado, registrar un espectro de ruido ( $t_{\text{int}} = 60\text{s}$ ) para su posterior sustracción.
2. Esperar 12 minutos para que la muestra estabilice su temperatura.
3. Encender el láser y capturar cuatro espectros por separado con un tiempo de integración ( $t_{\text{int}} = 60\text{s}$ ) a esa temperatura.
4. Repetir los pasos 1, 2 y 3 para 20, 30, 40 y  $50\text{ }^{\circ}\text{C}$ .

El tiempo de 12 minutos fue determinado empíricamente como suficiente para alcanzar estabilidad térmica. Junto a esto, 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. Cómo se tienen 5 temperaturas, se generó una base de datos de 120 espectros.

La Figura 2.7 muestra el setup con todos los componentes utilizados para realizar la adquisición de los espectros Raman.

Para realizar los experimentos de medición Raman en sistemas líquidos con dependencia térmica, se utilizó el porta cubetas modelo Quantum Northwest qpod con control de estabilización de temperatura modelo TC-1. El Porta cubetas soporta cubetas de cuarzo



**Fig. 2.7:** Configuración experimental: (1) Fuente de luz láser, (2) soporte de cubeta, (3) cubeta, (4) controlador de temperatura, (5) espectrómetro HORIBA Jobin Yvon ISA HR-320, (6) computador portátil [Fuente propia].

de dimensiones 10x10x40 mm, las cuales requieren que las 4 caras estén pulidas, con tapa de rosca y 3,5 ml de capacidad.

Esquemáticamente, la emisión láser de 300 mW a 532,10 nm se transporta por fibra óptica desde el cabezal láser MGL-U-532-300 mW hasta el puerto de entrada en el porta cubeta. El ingreso se hace como un haz que se propaga al interior de la celda de cuarzo. Dentro del medio de transmisión líquido, el haz láser produce la emisión estimulada de dispersión Raman en todas las direcciones posibles, donde un segundo puerto óptico de observación colecta la débil emisión Raman, haciendo que la imagen del láser dentro de la muestra quede acoplada ópticamente con la fibra de análisis, por medio de un doblete de lentes contra opuestos, como se observa en la Figura 2.8.

La Figura 2.8 muestra el esquema experimental de la cubeta y su sistema porta cubeta, el cual permite hacer espectroscopía de esparcimiento óptico por efectos de dispersión. La cubeta queda ópticamente accesible desde sus cuatro caras, en el cual son accesibles

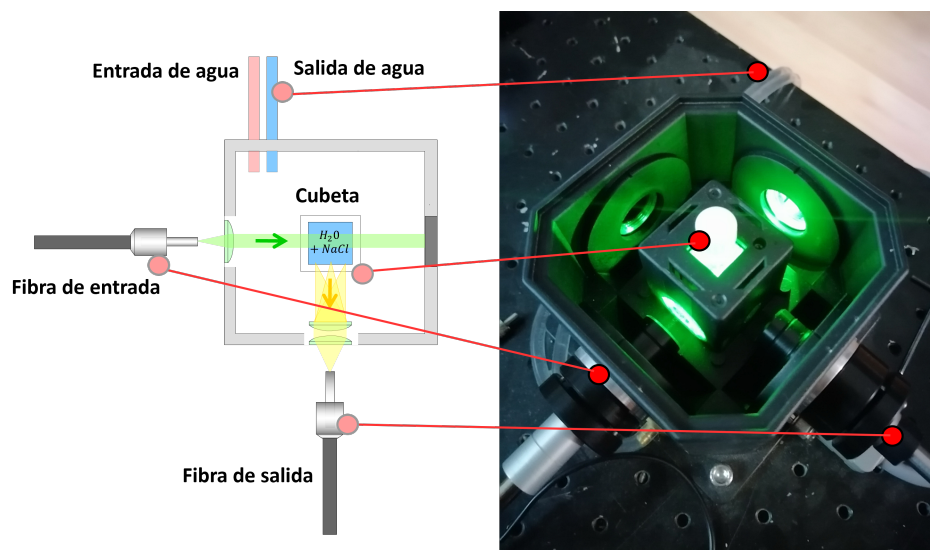


Fig. 2.8: Esquema experimental de la cubeta y su sistema porta cubeta.

por fibras ópticas. El sistema de porta cubeta además tiene un control de temperatura, que permitió estabilizar la temperatura en el rango entre 10-50°C.

## 2.4. Preprocesamiento de los espectros

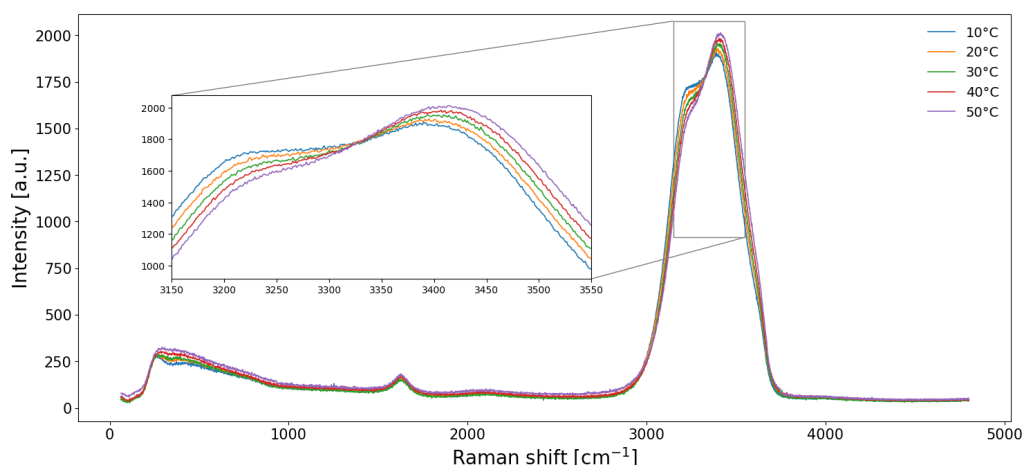
El objetivo del preprocesamiento espectral es preparar los espectros Raman obtenidos para su posterior análisis cuantitativo, en este caso se utilizó únicamente el recorte espectral, manteniendo las características físicas y químicas relevantes de la señal. No se utilizó corrección de línea base y ningún tipo de normalización o suavizado, como Savitzky–Golay o FFT. Esta decisión se adoptó con el propósito de mantener el flujo de procesamiento lo más parsimonioso posible, evitando transformaciones adicionales que pudiesen introducir dependencias artificiales entre espectros o alterar la redistribución relativa de intensidad dentro de la banda O–H. En consecuencia, se privilegió un esquema de procesamiento mínimo, con el fin de evaluar la capacidad cuantitativa intrínseca de la banda O–H bajo condiciones instrumentales controladas, manteniendo trazabilidad directa entre señal medida y descriptores modelados.

Cada espectro adquirido fue transformado desde el dominio de píxel al dominio espectral ( $\text{cm}^{-1}$ ), utilizando la curva de calibración descrita en la Etapa 3.1. Posteriormente, se extrajo el intervalo comprendido entre 3150 y 3550  $\text{cm}^{-1}$ , correspondiente a la región más relevante de la banda de estiramiento O–H.

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

- La concentración de iones disueltos (como  $\text{Na}^+$  y  $\text{Cl}^-$ ), 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  $\text{NaCl-H}_2\text{O}$ .



**Fig. 2.9:** Espectros de agua desionizada a diferentes temperaturas junto con la zona de interés para el recorte.

## 2.5. Modelo de descomposición global de la banda O–H mediante funciones Gaussianas

La estructura de la banda O–H es conocida por su naturaleza compleja, ancha y asimétrica. Diversos estudios han demostrado que esta banda puede descomponerse en un conjunto limitado de sub-bandas asociadas a poblaciones moleculares con distinto grado de estructuración del agua.

- Zonas más estructuradas (baja frecuencia,  $\sim 3200 \text{ cm}^{-1}$ ),

- Zonas menos estructuradas (alta frecuencia,  $\sim 3400\text{-}3550\text{ cm}^{-1}$ ).

El enfoque adoptado en esta tesis consiste en ajustar en conjunto los 120 espectros de la base de datos para un modelo de  $K$  Gaussianas compartidas, mismo centro de la función Gaussiana ( $\mu$ ) y mismo ancho de la campana ( $\sigma$ ) para todos los espectros pero amplitudes  $A_i$  distintas por espectro. Esto con la finalidad de capturar la variación de la concentración y la temperatura respecto a la variación de las amplitudes únicas de cada espectro.

### 2.5.1. Construcción de la matriz de intensidades $\mathbf{Y}$

Para comenzar con la descomposición en funciones Gaussianas se debe partir definiendo la matriz conjunta de todos los espectros.

Sea:

$S = 120$  : número total de espectros,

$\tilde{\nu}_j$  : desplazamiento Raman del punto  $j$ ,

$N$  : número de puntos espectrales entre  $3150$  y  $3550\text{ cm}^{-1}$ .

Con el fin de garantizar comparabilidad punto a punto entre espectros, se define un eje espectral común:

$$\mathbf{x} = (\tilde{\nu}_1, \tilde{\nu}_2, \dots, \tilde{\nu}_N),$$

tomado directamente del primer espectro disponible y cubriendo el intervalo  $3150\text{-}3550\text{ cm}^{-1}$  mediante  $N$  puntos equiespaciados.

Cada espectro se representa como un vector:

$$\mathbf{y}_s = (y_s(\tilde{\nu}_1), y_s(\tilde{\nu}_2), \dots, y_s(\tilde{\nu}_N)) \in \mathbb{R}^N.$$

Juntando todos los espectros en una matriz:

$$\mathbf{Y} = \begin{bmatrix} y_1(\tilde{\nu}_1) & \cdots & y_1(\tilde{\nu}_N) \\ \vdots & \ddots & \vdots \\ y_S(\tilde{\nu}_1) & \cdots & y_S(\tilde{\nu}_N) \end{bmatrix} \in \mathbb{R}^{S \times N}.$$

Esta matriz constituye la base del análisis, ya que concentra toda la información experimental en un formato unificado.

### 2.5.2. Definición de las funciones Gaussianas globales

El modelo supone que la banda O–H puede representarse como la superposición de  $K$  funciones Gaussianas globales:

$$G_k(\tilde{\nu}) = \exp\left[-\frac{(\tilde{\nu} - \mu_k)^2}{2\sigma_k^2}\right], \quad k = 1, \dots, K,$$

donde:

$\mu_k$  es el centro de la función Gaussiana,  $\sigma_k$  es su ancho,

y ambos parámetros son idénticos para todos los espectros.

Cada Gaussiana se discretiza evaluándola sobre el eje común, generando el vector:

$$G_k(x) = (G_k(\tilde{\nu}_1), \dots, G_k(\tilde{\nu}_N)).$$

Reuniendo todas las funciones Gaussianas como columnas se forma la matriz de diseño:

$$\mathbf{X} = \begin{bmatrix} G_1(\tilde{\nu}_1) & \cdots & G_K(\tilde{\nu}_1) \\ \vdots & \ddots & \vdots \\ G_1(\tilde{\nu}_N) & \cdots & G_K(\tilde{\nu}_N) \end{bmatrix} \in \mathbb{R}^{N \times K}.$$

### 2.5.3. Formulación del problema de optimización

El modelo reconstruye cada espectro como:

$$\hat{y}_s(\tilde{\nu}_j) = \sum_{k=1}^K A_{s,k} G_k(\tilde{\nu}_j). \quad (2.1)$$

donde

$$A_{s,k}$$

representa la amplitud de la Gaussiana  $k$  en el espectro  $s$ . En forma matricial:

$$\hat{Y} = A X^\top(\mu, \sigma), \quad (2.2)$$

donde

$$A \in \mathbb{R}^{S \times K}$$

contiene las amplitudes de cada Gaussiana en cada espectro.

El objetivo del método es encontrar los parámetros globales  $\{\mu_k, \sigma_k\}_{k=1}^K$  y las amplitudes específicas  $\{A_{s,k}\}$ , que minimicen el error cuadrático global:

$$\min_{\mu, \sigma, A} \sum_{s=1}^S \sum_{j=1}^N \left( Y_{s,j} - \sum_{k=1}^K A_{s,k} \exp \left[ -\frac{(\tilde{\nu}_j - \mu_k)^2}{2 \sigma_k^2} \right] \right)^2. \quad (2.3)$$

En su forma matricial:

$$\min_{\mu, \sigma, A} \|Y - A X^\top(\mu, \sigma)\|^2 \quad (2.4)$$

sujeto a:

$$A_{s,k} \geq 0, \quad \tilde{\nu}_1 \leq \mu_k \leq \tilde{\nu}_N, \quad \sigma_{\min} \leq \sigma_k \leq \sigma_{\max}.$$

Este problema combina variables lineales en  $A$  (si  $\mu$  y  $\sigma$  están fijos) y variables no lineales en  $\mu$  y  $\sigma$ . Esto es, un problema de optimización no lineal con restricciones.

#### 2.5.4. Resolución del problema de optimización

El problema descrito en la sección anterior conduce a un ajuste simultáneo de los parámetros globales  $\mu = (\mu_1, \dots, \mu_K)$ ,  $\sigma = (\sigma_1, \dots, \sigma_K)$  y de las amplitudes

específicas  $A = [A_{s,k}]$  que minimizan el error cuadrático entre la matriz de intensidades experimental  $Y$  y su reconstrucción Gaussiana

$$\hat{Y} = A X^\top(\mu, \sigma).$$

Dado el carácter no lineal del modelo y la presencia de múltiples mínimos locales, la obtención de una solución estable requiere una estrategia de inicialización y optimización cuidadosamente diseñada.

En esta subsección se describe el método completo de resolución, que se basa en tres etapas: (i) estimación inicial de parámetros globales, (ii) estimación inicial de amplitudes y (iii) optimización conjunta multi-inicio.

### (1) Semillas iniciales para $\mu_k$ y $\sigma_k$

Antes de resolver el problema completo, es necesario disponer de valores iniciales razonables para los centros y anchuras de las Gaussianas. Para ello:

1. Se calcula el espectro promedio  $\bar{y}(\tilde{\nu})$  a partir de los  $S$  espectros individuales.
2. Se ajusta un modelo de mezcla Gaussiana (GMM) sobre el eje espectral, ponderado según la intensidad del espectro promedio.
3. Las medias del GMM se utilizan como aproximación inicial para los centros  $\mu_k^{(0)}$ , mientras que sus varianzas generan estimaciones iniciales de las anchuras  $\sigma_k^{(0)}$ , respetando los límites físicos definidos.

Este procedimiento permite ubicar de forma automática las regiones de mayor estructura espectral, entregando puntos iniciales consistentes que favorecen la convergencia del ajuste.

### (2) Semillas iniciales para las amplitudes $A_{s,k}$

Con las aproximaciones iniciales  $\mu^{(0)}$  y  $\sigma^{(0)}$  se construye la matriz de diseño inicial

$$X_0 = X(\mu^{(0)}, \sigma^{(0)}).$$

Luego, para cada espectro  $y_s$  se estima un vector de amplitudes iniciales  $a_s^{(0)}$  resolviendo un problema de mínimos cuadrados no negativos (NNLS):

$$a_s^{(0)} = \arg \min_{a_s \geq 0} \|y_s - X_0 a_s\|_2^2.$$

Este paso entrega un conjunto inicial  $A^{(0)}$  que: (i) respeta la restricción física  $A_{s,k} \geq 0$  para todo  $s, k$ , y (ii) proporciona un punto de partida estable para la optimización global.

### (3) Optimización conjunta de todos los parámetros

Una vez definidas las semillas  $(A^{(0)}, \mu^{(0)}, \sigma^{(0)})$ , se resuelve el problema completo

$$\min_{A, \mu, \sigma} \|Y - A X^\top(\mu, \sigma)\|^2, \quad (2.5)$$

sujeto a las restricciones

$$\begin{aligned} A_{s,k} &\geq 0, & \forall s, k, \\ \tilde{\nu}_{\min} &\leq \mu_k \leq \tilde{\nu}_{\max}, & \forall k, \\ \sigma_{\min} &\leq \sigma_k \leq \sigma_{\max}, & \forall k. \end{aligned}$$

El problema se resuelve mediante un algoritmo de mínimos cuadrados no lineales restringidos, el cual ajusta simultáneamente todos los parámetros. Sin embargo, debido a la naturaleza multimodal de la función objetivo, esta optimización no garantiza por sí sola alcanzar el mínimo global.

Para robustecer el proceso, se adopta una estrategia multi-inicio, que consiste en:

- **Primer inicio:** se ejecuta la optimización a partir de las semillas  $(A^{(0)}, \mu^{(0)}, \sigma^{(0)})$ .
- **Reinicios perturbados:** se generan varias configuraciones adicionales  $(\mu^{(r)}, \sigma^{(r)})$  mediante pequeñas perturbaciones controladas de las semillas originales, y para cada una de ellas se resuelve nuevamente el problema completo.
- **Selección final:** de entre todas las soluciones obtenidas se elige aquella que presenta el menor error cuadrático total

$$\text{SSE}^{(r)} = \sum_{s=1}^S \sum_{j=1}^N (Y_{s,j} - \hat{Y}_{s,j}^{(r)})^2.$$

Este procedimiento asegura que la solución final no dependa de un único punto inicial y reduce significativamente el riesgo de converger a mínimos locales no representativos del comportamiento espectral.

### Resultado de la optimización

Tras completar el proceso multi-inicio, se obtiene el conjunto definitivo de parámetros

$$\mu_k^*, \quad \sigma_k^*, \quad A_{s,k}^*,$$

los cuales permiten reconstruir con alta fidelidad la banda OH en todos los espectros mediante

$$\hat{Y}_{s,j} = \sum_{k=1}^K A_{s,k}^* G_k(\tilde{\nu}_j; \mu_k^*, \sigma_k^*),$$

y constituyen una representación compacta y físicamente interpretable de la estructura espectral, que será utilizada posteriormente para el análisis de tendencias con concentración y temperatura.

## 2.6. Modelo de amplitudes

Para el modelado anterior se consideró  $K = 3$  por balance entre RMSE de ajuste y estabilidad de los parámetros, en línea con la literatura sobre descomposición de la banda OH del agua. Una vez obtenidas las amplitudes  $A_1, A_2, A_3$  asociadas a las tres sub-bandas Gaussianas de la banda O–H, surge la pregunta fundamental:

¿Cómo varía cada una de estas amplitudes cuando cambian la concentración de NaCl y la temperatura?

Responder esta pregunta es esencial porque las amplitudes codifican el peso relativo de diferentes subestructuras del agua. Por tanto, modelarlas matemáticamente en función de  $C$  y  $T$  permite entender y predecir cómo evoluciona la estructura del agua salina bajo diferentes condiciones termodinámicas.

Este análisis se realiza mediante un modelo concentración–temperatura, cuyo desarrollo se describe a continuación.

### 2.6.1. Formulación del modelo de amplitudes

Para cada uno de los 120 espectros experimentales conocemos:

- Concentración:  $C_i$ .
- Temperatura:  $T_i$ .
- Amplitudes Gaussianas:  $A_{1,i}$ ,  $A_{2,i}$ ,  $A_{3,i}$ .

Por conveniencia, para un componente cualquiera  $k \in \{1, 2, 3\}$ , definimos el vector de amplitudes:

$$\mathbf{y}^{(k)} = \begin{bmatrix} A_{k,1} \\ A_{k,2} \\ \vdots \\ A_{k,120} \end{bmatrix}.$$

Nuestro objetivo es encontrar una función matemática

$$A_k \approx f_k(C, T)$$

que explique estos valores y permita interpolar y predecir amplitudes para cualquier combinación  $(C, T)$  dentro del rango experimental.

La concentración y la temperatura afectan las amplitudes de forma dependiente e independiente.

Por ejemplo:

- aumentar la temperatura debilita los enlaces fuertes o estructurados de la banda O-H,
- aumentar la concentración reorganiza la red de puentes de hidrógeno,
- pero el efecto de la temperatura no es el mismo en soluciones diluidas que en soluciones concentradas.

Esto constituye una interacción física. De este modo, una amplitud se modela como:

$$A_k(C, T) \approx \beta_0^{(k)} + \beta_1^{(k)}T + \beta_2^{(k)}C + \beta_3^{(k)}(TC). \quad (2.6)$$

Este modelo incorpora:

- un término constante ( $\beta_0^{(k)}$ ),
- efectos individuales de  $T$  y  $C$ ,
- un término de interacción que permite que el efecto de una variable dependa del valor de la otra.

Para cada espectro, evaluamos los términos  $(1, T, C, TC)$ . Esto define el vector:

$$\phi(C_i, T_i) = \begin{bmatrix} 1 \\ T_i \\ C_i \\ T_i C_i \end{bmatrix}.$$

Apilando estos vectores para los 120 espectros se obtiene la matriz de diseño:

$$X = \begin{bmatrix} 1 & T_1 & C_1 & T_1 C_1 \\ 1 & T_2 & C_2 & T_2 C_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & T_{120} & C_{120} & T_{120} C_{120} \end{bmatrix} \in \mathbb{R}^{120 \times 4}.$$

Cada columna representa:

Columna	Significado físico
1	Nivel base de la amplitud
$T$	Efecto de la temperatura
$C$	Efecto de la concentración
$TC$	Interacción entre ambos

Con esta construcción, la ecuación del modelo se vuelve:

$$y^{(k)} \approx X \beta^{(k)},$$

donde

$$\beta^{(k)} = \begin{bmatrix} \beta_0^{(k)} \\ \beta_1^{(k)} \\ \beta_2^{(k)} \\ \beta_3^{(k)} \end{bmatrix}.$$

### 2.6.2. Resolución del modelo de amplitudes

Para estimar los coeficientes  $\beta^{(k)}$  se plantea el problema:

$$\beta^{(k)} = \arg \min_{\beta} \|y^{(k)} - X \beta\|^2.$$

Este es el clásico problema de mínimos cuadrados lineales, cuya solución óptima minimiza el error cuadrático de las amplitudes.

El resultado se obtiene mediante:

$$\beta^{(k)} = (X^T X)^{-1} X^T y^{(k)}. \quad (2.7)$$

y luego se evalúa el ajuste mediante  $R^2$  y RMSE.

## 2.7. Modelo inverso para la estimación de concentración

Tras construir el modelo directo que describe las amplitudes Gaussianas en función de la concentración y la temperatura, se plantea el problema inverso: dado un espectro (o sus amplitudes) y una temperatura conocida, estimar la concentración  $C$ .

Se conoce que cada espectro puede escribirse como:

$$I(\tilde{\nu}, C, T) = \sum_{i=1}^3 A_i(C, T) G_k(\tilde{\nu}), \quad (2.8)$$

donde:

- $G_k(\tilde{\nu})$  es la función Gaussiana número  $k$  evaluada en el desplazamiento Raman  $\tilde{\nu}$ ,
- $A_i(C, T)$  es la amplitud asociada a dicha sub-banda para las condiciones  $(C, T)$ .

Cada amplitud  $A_i(C, T)$  se modela mediante un polinomio lineal:

$$A_i(C, T) = b_{0,i} + b_{1,i} T + b_{2,i} C + b_{3,i} TC.$$

Insertando (2) en (1):

$$I(\tilde{\nu}, C, T) = \sum_{i=1}^3 G_k(\tilde{\nu}) (b_{0,i} + b_{1,i} T + b_{2,i} C + b_{3,i} TC).$$

Agrupando por términos independientes de  $C$  y términos dependientes de  $C$ . El espectro puede reescribirse como:

$$I(\tilde{\nu}, C, T) = \underbrace{\sum_{i=1}^3 G_k(\tilde{\nu}) (b_{0,i} + b_{1,i} T)}_{P(\tilde{\nu}, T)} + \underbrace{\sum_{i=1}^3 G_k(\tilde{\nu}) (b_{2,i} + b_{3,i} T) C}_{Q(\tilde{\nu}, T)}.$$

Donde se han definido las funciones:

### Parte independiente de la concentración

$$P(\tilde{\nu}, T) = \sum_{i=1}^3 G_k(\tilde{\nu}) (b_{0,i} + b_{1,i} T),$$

### Parte proporcional a la concentración

$$Q(\tilde{\nu}, T) = \sum_{i=1}^3 G_k(\tilde{\nu}) (b_{2,i} + b_{3,i} T).$$

De este modo, el espectro completo puede escribirse como:

$$I(\tilde{\nu}, C, T) = P(\tilde{\nu}, T) + Q(\tilde{\nu}, T) C. \tag{2.9}$$

### 2.7.1. Formulación del problema inverso

Dado un espectro observado  $I_{\text{obs}}(\tilde{\nu})$  y la temperatura  $T$ , buscamos el valor de  $C$  que haga que:

$$I_{\text{obs}}(\tilde{\nu}) \approx P(\tilde{\nu}, T) + C Q(\tilde{\nu}, T). \quad (5)$$

Planteamos un criterio de mínimos cuadrados:

$$J(C) = \sum_{j=1}^N [I_{\text{obs}}(\tilde{\nu}_j) - P(\tilde{\nu}_j, T) - C Q(\tilde{\nu}_j, T)]^2.$$

El objetivo es resolver:

$$C^* = \arg \min_C J(C).$$

Obsérvese que se trata de un problema unidimensional, por lo que puede resolverse de forma exacta.

### 2.7.2. Solución analítica del problema inverso

Derivando (5) respecto de  $C$  e igualando a cero:

$$Q(T)^\top [I_{\text{obs}} - P(T) - C^* Q(T)] = 0,$$

de donde se obtiene:

$$C^* = \frac{Q(T)^\top [I_{\text{obs}} - P(T)]}{Q(T)^\top Q(T)}. \quad (2.10)$$

Esta expresión constituye la solución óptima en el sentido de mínimos cuadrados.

## 2.8. Modelo Partial Least Square (PLS) con Individual Contribution Standardization (ICS)

### 2.8.1. Individual Contribution Standardization (ICS)

La temperatura modifica de forma sistemática los espectros Raman, incluso si la concentración del analito se mantiene fija, al cambiar la temperatura pueden variar tanto la intensidad como la forma de las bandas. En un modelo quimiométrico, esa variación térmica puede confundirse con variación de concentración. Para reducir este problema, se utiliza *Individual Contribution Standardization* (ICS) [24], cuyo propósito es tomar un espectro medido a una temperatura  $t_{\text{test}}$  y convertirlo en un espectro equivalente a una temperatura de referencia  $t_{\text{ref}}$ , de modo que la predicción de concentración sea menos sensible a cambios térmicos.

ICS aprende, a partir de datos de calibración, una representación llamada *response pattern*. Para explicar esto de manera intuitiva, para cada temperatura  $t_k$ , el método estima un patrón espectral  $\mathbf{S}(t_k)$  que actúa como una especie de “firma espectral” asociada a la concentración (o a cada especie en sistemas multicomponentes).

Formalmente, sea  $\mathbf{X}(t_k) \in \mathbb{R}^{N_k \times p}$  la matriz de espectros de calibración medidos a la temperatura  $t_k$  (con  $N_k$  muestras y  $p$  variables espectrales), y sea  $\mathbf{Y}(t_k) \in \mathbb{R}^{N_k \times m}$  la matriz de concentraciones conocidas. Se definen los patrones a cada temperatura mediante regresión directa:

$$\mathbf{S}(t_k) = \mathbf{Y}^+(t_k) \mathbf{X}(t_k), \quad (2.11)$$

donde  $(\cdot)^+$  es la pseudoinversa de Moore–Penrose. En palabras simples, usando los espectros y las concentraciones conocidas a  $t_k$ , se calcula el patrón  $\mathbf{S}(t_k)$  que mejor relaciona concentración con espectro en esa temperatura.

En esta tesis el método se aplica al caso unicomponente ( $m = 1$ ), donde  $\mathbf{Y}(t_k)$  corresponde a la concentración de NaCl. Por lo tanto,  $\mathbf{S}(t_k)$  se reduce a un único patrón espectral dependiente de la temperatura.

Una vez que se tienen patrones calculados en temperaturas discretas  $\{t_k\}$ , ICS necesita poder obtener un patrón también para temperaturas intermedias. Para ello, se modela explícitamente la dependencia con temperatura en cada variable espectral. Es decir,

para cada punto espectral  $j$  (cada columna del espectro), se ajusta una función suave de  $t$  sobre los valores observados de  $S_j(t_k)$ . En el trabajo original se utiliza típicamente un polinomio de segundo orden:

$$s_{l,j}(t) = a_{l,j} + b_{l,j} t + c_{l,j} t^2, \quad (2.12)$$

lo que permite estimar  $\mathbf{S}(t)$  para cualquier temperatura  $t$  dentro del rango de calibración. En otras palabras, ICS no “predice” temperatura; lo que hace es predecir cómo debería verse el patrón espectral a una temperatura dada.

Sea  $\mathbf{x}(t_{\text{test}}) \in \mathbb{R}^{1 \times p}$  el espectro medido de una muestra a temperatura  $t_{\text{test}}$ . ICS realiza la corrección en dos pasos:

1. Calcular una estimación auxiliar  $\hat{\mathbf{y}}$  (no es todavía la concentración final, es un coeficiente interno del método) usando el patrón en  $t_{\text{test}}$ .
2. Calcular la diferencia entre patrones:  $\mathbf{S}(t_{\text{ref}}) - \mathbf{S}(t_{\text{test}})$ , que representa cómo cambiaría la contribución espectral si la medición se hubiese hecho en  $t_{\text{ref}}$  en lugar de  $t_{\text{test}}$ . Esa diferencia se escala por  $\hat{\mathbf{y}}$  y se agrega al espectro medido, obteniendo el espectro estandarizado:

$$\mathbf{x}(t_{\text{test}} \rightarrow t_{\text{ref}}).$$

Con esto, dos espectros medidos a diferentes temperaturas quedan expresados como si ambos hubiesen sido medidos en la misma referencia  $t_{\text{ref}}$ , reduciendo la variabilidad térmica.

ICS es una etapa de preprocesamiento/estandarización: su función es reducir la componente de variación debida a temperatura. Una vez estandarizado el espectro, la concentración se estima con un modelo de calibración multivariante. En este trabajo se emplea PLS, aplicado sobre los espectros corregidos. Bajo este esquema, PLS se enfoca principalmente en la variación espectral atribuible a concentración, en vez de dedicar parte de su espacio latente a capturar efectos térmicos.

### 2.8.2. Formulación del modelo PLS para cuantificación

Se dispone de una base de datos con  $S$  espectros y sus concentraciones de referencia. Cada espectro está muestreado en  $N$  variables (intensidades a distintos números de onda).

$$X \in R^{S \times N}$$

es la matriz de espectros, donde cada fila corresponde a un espectro.

$$y \in R^{S \times 1}$$

vector de concentraciones reales.

El objetivo es construir un modelo que, dado un espectro nuevo  $\mathbf{x}$ , estime su concentración  $\hat{y}$ .

En espectros, las variables están fuertemente correlacionadas entre sí (colinealidad) y suelen ser numerosas ( $N$  grande). Si se intenta ajustar una regresión lineal directa

$$y \approx X\beta,$$

el modelo puede resultar inestable o presentar sobreajuste. La regresión por mínimos cuadrados parciales (PLS) evita este problema reduciendo la dimensionalidad a un número pequeño de variables resumen (componentes latentes) diseñadas específicamente para predecir la concentración.

PLS construye  $a$  componentes latentes, con  $a \ll N$ , que son combinaciones lineales del espectro y que se eligen para maximizar su relación con  $\mathbf{y}$ . Reemplaza el espectro original, descrito por muchas variables, por un conjunto reducido de *scores* o variables latentes:

$$\mathbf{T} \in \mathbb{R}^{S \times a},$$

donde cada columna de  $\mathbf{T}$  corresponde a una componente latente y cada fila contiene las coordenadas del espectro asociado en el espacio reducido.

Conceptualmente:

- $\mathbf{X}$  se aproxima mediante  $\mathbf{T}$ , capturando la estructura relevante del espectro.
- $\mathbf{y}$  se explica a partir de  $\mathbf{T}$ , capturando su relación con la concentración.

A diferencia del análisis de componentes principales (PCA), que busca únicamente explicar la varianza en  $\mathbf{X}$ , PLS selecciona cada componente de modo que la proyección

de  $\mathbf{X}$  sobre ella sea altamente informativa para predecir  $\mathbf{y}$ . En otras palabras, PLS identifica direcciones espectrales que varían sistemáticamente con la concentración.

Aunque el modelo PLS se construye en términos de componentes latentes, el resultado final es equivalente a una regresión lineal:

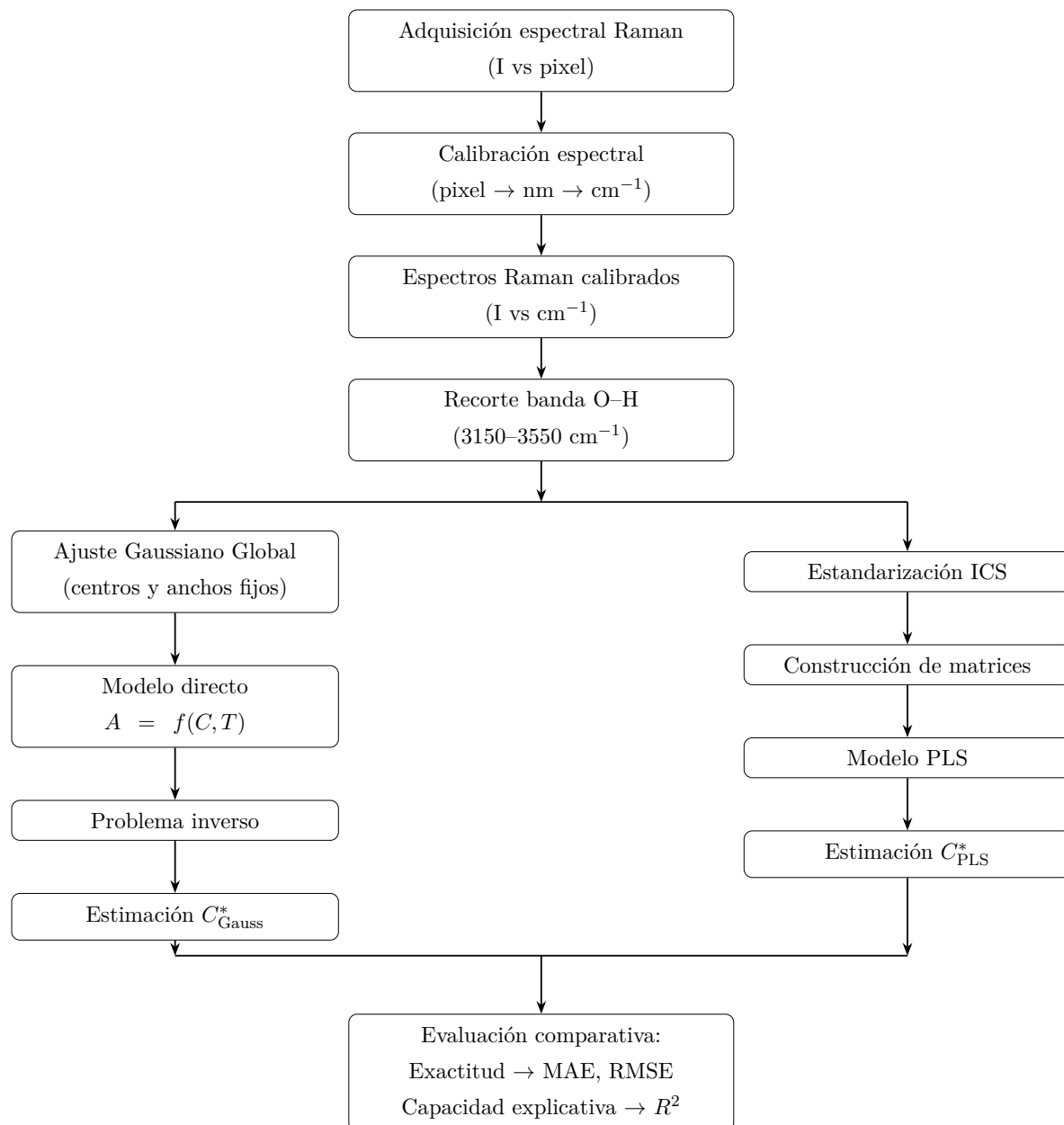
$$\hat{y} = \mathbf{x}\boldsymbol{\beta} + b_0, \quad (2.13)$$

donde  $\boldsymbol{\beta}$  representa los pesos asociados a las  $N$  variables espectrales y  $b_0$  es el intercepto. La diferencia con una regresión lineal convencional es que  $\boldsymbol{\beta}$  se estima de forma más robusta en presencia de colinealidad, ya que el ajuste se realiza en un espacio reducido de componentes latentes, en lugar de hacerlo directamente sobre variables espectrales altamente correlacionadas.

El principal hiperparámetro del modelo PLS es el número de componentes latentes  $a$ . Si  $a$  es demasiado pequeño, el modelo puede presentar subajuste; si  $a$  es demasiado grande, puede capturar ruido y sobreajustar. Por esta razón,  $a$  se ha seleccionado igual a 3, debido a que en el modelo de amplitudes utiliza las 3 amplitudes de las 3 funciones Gaussianas.

Con el fin de sintetizar el procedimiento desarrollado, la Figura 2.10 presenta el esquema metodológico completo del proceso de estimación de concentración a partir de espectros Raman. El flujo comienza con la adquisición espectral (I vs pixel), seguida de la calibración espectral que establece la correspondencia pixel–nm–cm<sup>-1</sup>, asegurando consistencia en el eje de desplazamiento Raman. Posteriormente, se selecciona la región de la banda O–H (3150–3550 cm<sup>-1</sup>), sobre la cual se realiza el análisis cuantitativo.

A partir de esta región, el proceso se bifurca en dos enfoques. En el modelo paramétrico propuesto se aplica un ajuste gaussiano global para extraer amplitudes, las cuales se modelan explícitamente como función de concentración y temperatura, formulando un problema inverso para estimar la concentración. En paralelo, el enfoque ICS–PLS emplea estandarización y modelado multivariante para la predicción directa. Finalmente, ambas estimaciones se comparan mediante métricas de exactitud (MAE, RMSE) y capacidad explicativa ( $R^2$ ), permitiendo evaluar cuantitativamente el desempeño relativo de cada metodología.



**Fig. 2.10:** Diagrama de bloques del proceso de estimación de concentración a partir de espectros Raman.

# Capítulo 3

## Resultados

### 3.1. 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 3.1 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 de la forma:

$$y = ax + b$$

donde:

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

El ajuste obtenido fue:

$$y = 12,69x + 7,88$$

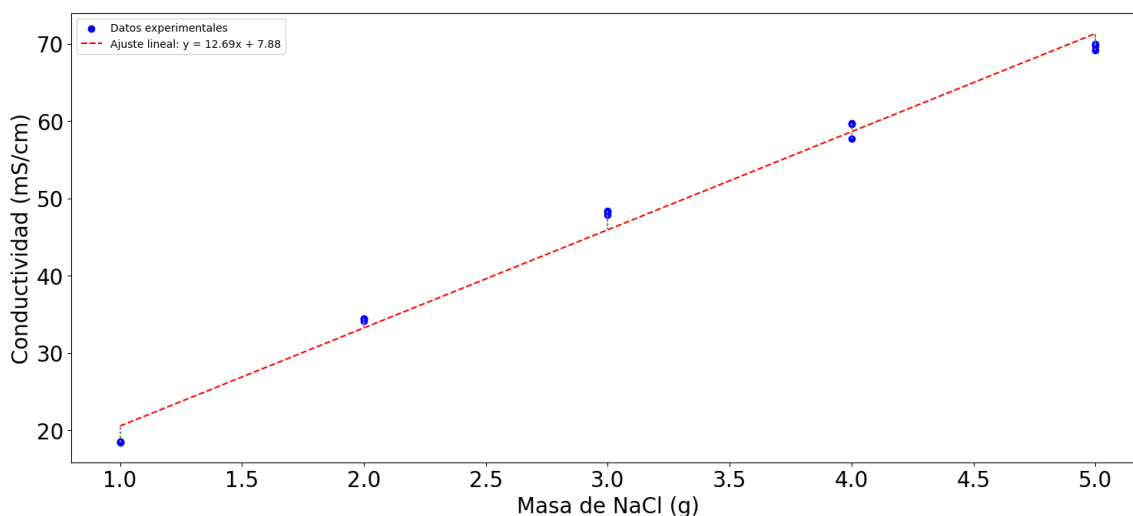
**Tabla 3.1:** Conductividad medida para distintas masas de NaCl disueltas en solución.

Masa de 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

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 3.1, lo que sugiere ausencia de sesgos.

El objetivo de este procedimiento no fue estimar concentración con conductividad, sino verificar linealidad y coherencia del contenido iónico. 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. 3.1:** Comportamiento entre la masa de NaCl (g) y conductividad eléctrica (mS/cm).

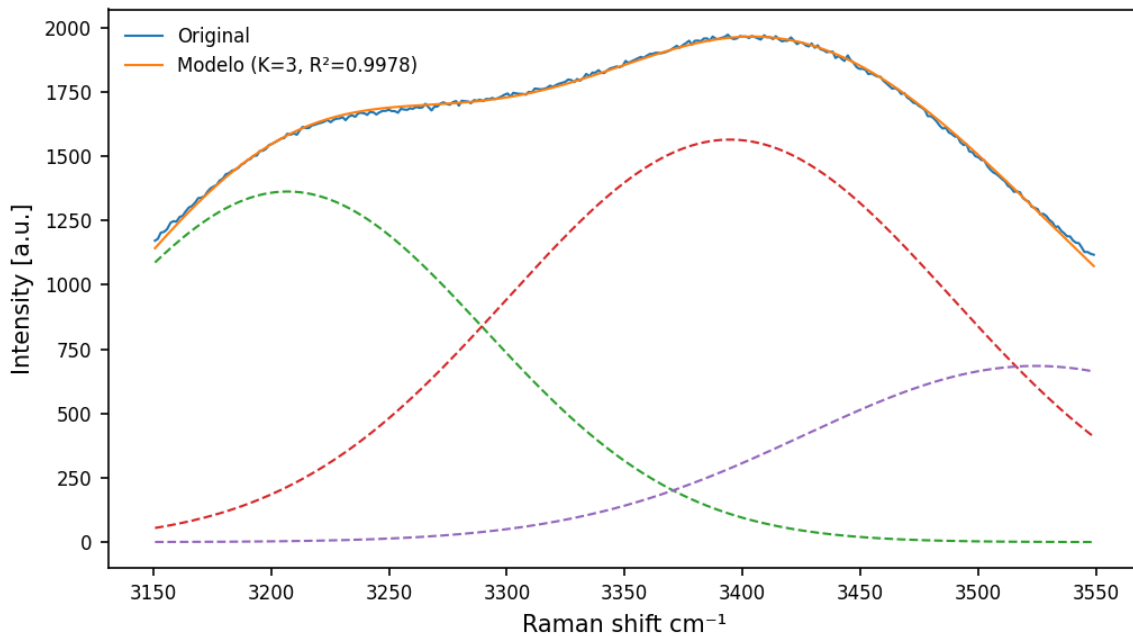
## 3.2. Resultados de descomposición global con funciones Gaussianas

Con el fin de caracterizar la banda O–H entre  $3150$  y  $3550\text{ cm}^{-1}$  y de extraer parámetros espectrales físicamente interpretables, se implementó el modelo de descomposición global descrito en el Capítulo 2. Este modelo asume que la forma espectral observada puede representarse como la suma lineal de un número reducido de funciones Gaussianas  $G_k(\nu)$ , donde cada Gaussiana captura un tipo de enlace O–H asociado al entorno del agua en presencia de NaCl.

A diferencia de un ajuste independiente por espectro, la descomposición global optimiza simultáneamente los centros  $\mu_k$  y anchos  $\sigma_k$  de las Gaussianas para todo el conjunto de 120 espectros, mientras que las amplitudes  $A_{i,k}$  se ajustan individualmente para cada espectro. Este esquema global garantiza consistencia física entre todas las mediciones y reduce los grados de libertad.

La Figura 3.2 muestra un ejemplo representativo del ajuste para un espectro real (0g/100 mL,  $30^\circ\text{C}$ ). En la gráfica se observa:

- la señal experimental (línea azul),



**Fig. 3.2:** Ajuste de 3 funciones Gaussianas para espectro agua a 30 °C.

- la reconstrucción del modelo con  $K = 3$  Gaussianas (línea naranja),
- y cada una de las contribuciones Gaussianas individuales (líneas segmentadas).

La superposición entre la curva experimental y la curva modelada es prácticamente perfecta. El modelo captura tanto la asimetría de la banda O–H como la transición suave entre el hombro de baja frecuencia ( $\sim 3200 \text{ cm}^{-1}$ ) y el máximo vibracional ( $\sim 3400 \text{ cm}^{-1}$ ).

La calidad del ajuste se evaluó mediante dos métricas:

**Tabla 3.2:** Desempeño global del ajuste gaussiano.

Métrica	Valor
$\text{RMSE}_g$	14,186
$R_g^2$	0,9971

Estas métricas se calculan considerando simultáneamente las 120 reconstrucciones espectrales generadas por el modelo global. El valor de  $R_g^2$  cercano a 1 indica que el modelo explica prácticamente toda la variabilidad espectral presente en la banda

O–H recortada. Por otro lado, el RMSE global de aproximadamente 14 unidades de intensidad es extremadamente bajo considerando que las intensidades típicas de la banda se encuentran entre 1000 y 3000 a.u.

En conjunto, estos resultados confirman que:

- Tres Gaussianas son suficientes para representar de manera compacta la banda O–H en soluciones NaCl–agua.
- El modelo global reproduce con gran fidelidad tanto la forma como la variación espectral entre las distintas condiciones de concentración y temperatura.
- Los parámetros  $(A_{i,k}, \mu_k, \sigma_k)$  obtenidos son robustos y físicamente interpretables, lo que habilita su uso en las siguientes etapas del trabajo, especialmente:
  - La construcción de modelos de amplitudes  $A_k(C, T)$ .
  - La inversión del modelo para predecir concentración.

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.

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

Parámetro	Significado físico-químico
$A_i$	Actúa como población relativa de moléculas de agua con cierto tipo de enlace.
$\mu_i$	Tipo de enlace O–H predominante (más débil $\rightarrow$ mayor $\mu_i$ ).
$\sigma_i$	Grado de desorden molecular; anchura de la distribución vibracional.

### 3.3. Efecto del NaCl y la temperatura en el espectro Raman

Antes de analizar las amplitudes Gaussianas obtenidas mediante la descomposición espectral global, es fundamental caracterizar la forma en que la temperatura y la concentración salina modifican directamente la señal Raman de la banda O–H. Esta observación preliminar permite comprender la estructura de la variabilidad espectral que posteriormente será capturada por los parámetros  $A_k(C, T)$ .

A continuación se presentan los dos análisis exploratorios clave:

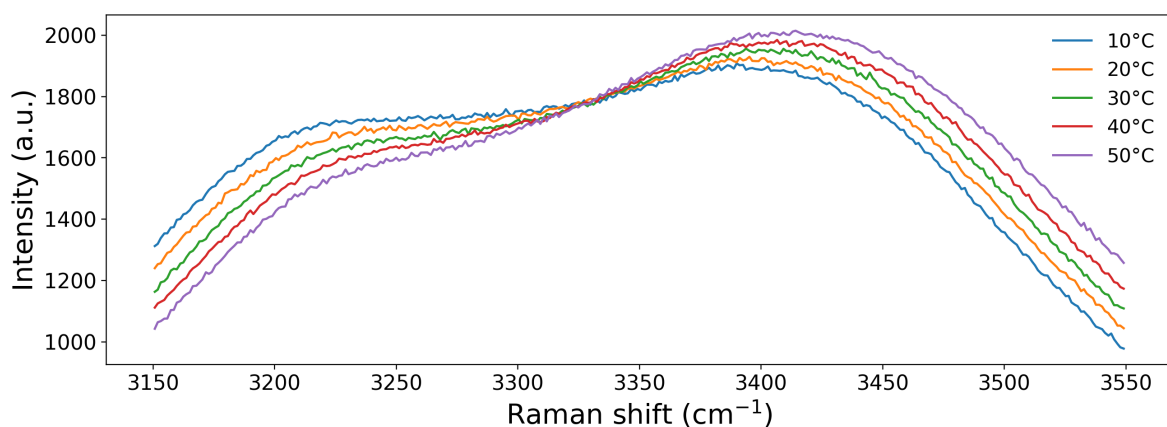
- La variación espectral a concentración fija y temperatura variable.
- La variación espectral a temperatura fija y concentración variable.

La Figura 3.3a muestra la dependencia espectral de la banda O–H para el agua a diferentes temperaturas entre 10 y 50 °C.

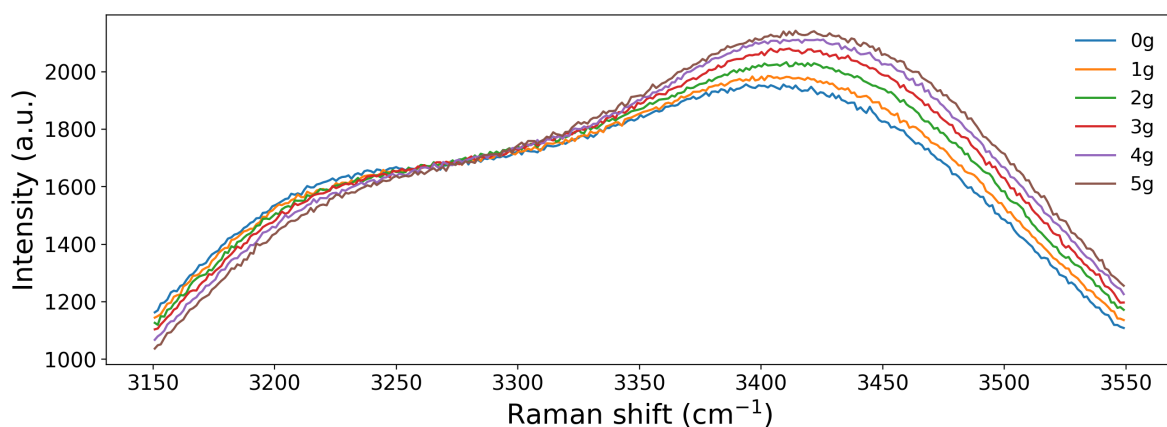
Los espectros muestran una disminución del hombro en baja frecuencia (3150–3250  $cm^{-1}$ ) y una intensificación en la zona de mayor frecuencia ( $> 3400 cm^{-1}$ ). Este patrón es característico del debilitamiento de la red de enlaces de hidrógeno al aumentar  $T$ .

La Figura 3.3b muestra los espectros correspondientes a 30 °C para concentraciones entre 0 y 5 g/100 mL de NaCl. A mayor concentración, el hombro de baja frecuencia (3150–3250  $cm^{-1}$ ) disminuye ligeramente en intensidad y en la zona de mayor frecuencia ( $> 3400 cm^{-1}$ ) existe un aumento en la intensidad, lo que sugiere una reducción de las poblaciones fuertemente estructuradas del agua.

Este comportamiento es consistente con la física molecular del agua: el aumento de temperatura y la concentración de NaCl induce un mayor desorden en la red de enlaces de hidrógeno, reduciendo la fracción de moléculas fuertemente ligadas y aumentando la contribución de estructuras menos coordinadas, que presentan mayor intensidad Raman.



(a) Espectros de agua en diferentes temperaturas.



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

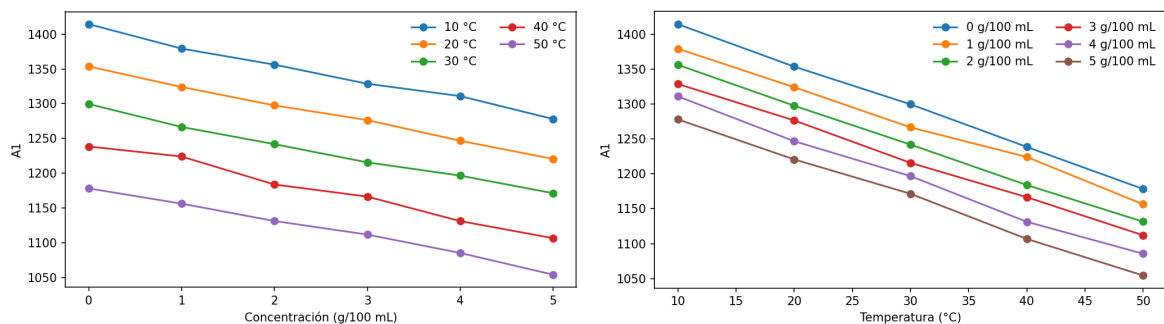
Fig. 3.3: Dependencias espectrales principales.

### 3.4. Resultados de las amplitudes Gaussianas como variables latentes fisicoquímicas

Como se describió en el Capítulo 2, la descomposición espectral global permitió representar la banda O–H entre 3150 y 3550  $cm^{-1}$  como la suma de tres contribuciones Gaussianas compartidas por todos los espectros de la base experimental. Esta representación produce, para cada espectro, un trío de amplitudes ( $A_1, A_2, A_3$ ) que actúan como variables latentes fisicoquímicas, reflejando la contribución relativa de tres subpoblaciones de moléculas de agua: estados altamente estructurados, estados intermedios y estados débilmente enlazados.

Con el fin de analizar la sensibilidad de estas amplitudes a los dos factores

experimentales —temperatura y concentración de NaCl— se evaluaron las relaciones  $A_k$  vs.  $T$  y  $A_k$  vs.  $C$  para cada componente. A partir de este análisis se identifican patrones sistemáticos y físicamente coherentes.



(a)  $A_1$  respecto a la concentración.

(b)  $A_1$  respecto a la temperatura.

**Fig. 3.4:** Comportamiento de  $A_1$  respecto a la concentración y temperatura.

Las Figuras 3.4a y 3.4b muestran la dependencia de  $A_1$  respecto a la concentración y la temperatura.

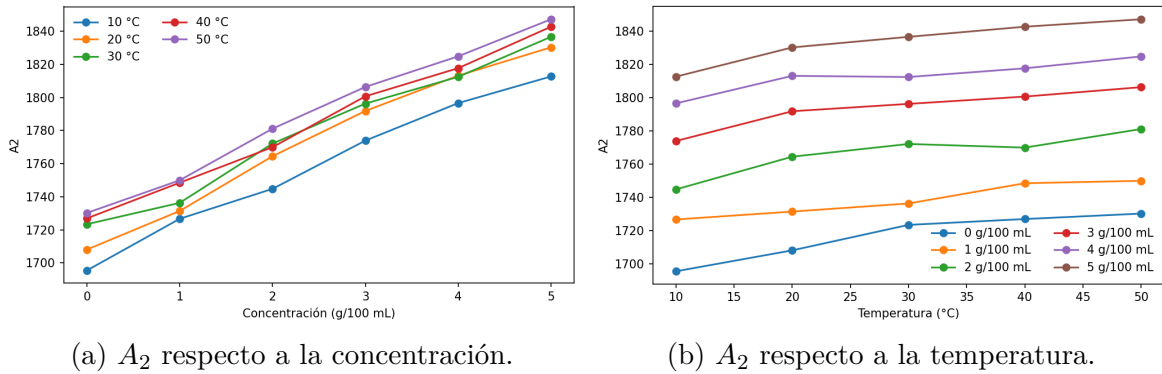
A todas las temperaturas,  $A_1$  exhibe una disminución monótonica al aumentar la concentración de NaCl. Este comportamiento espectral está asociado a estructuras de agua fuertemente coordinadas, típicas de redes de enlaces de hidrógeno más ordenadas. La incorporación de iones  $\text{Na}^+$  y  $\text{Cl}^-$  perturba dichos dominios estructurados, reduciendo progresivamente su población.

Para todas las concentraciones,  $A_1$  decrece de manera casi lineal con la temperatura. Este comportamiento es consistente con el debilitamiento térmico de la red de enlaces de hidrógeno: a mayor temperatura, disminuye la fracción de moléculas de agua en configuraciones altamente estructuradas.

$A_1$  es un descriptor dependiente de la temperatura y de la concentración. Refleja la fracción de agua en estados de orden elevado (O–H fuertemente enlazado), y sus variaciones son coherentes con la física molecular de soluciones acuosas.

Las Figuras 3.5a y 3.5b presentan el comportamiento de  $A_2$ , componente asociado a estados vibracionales intermedios del agua.

En todas las temperaturas,  $A_2$  aumenta de forma prácticamente lineal con la concentración salina. Este comportamiento es notablemente regular y consistente, lo que



(a)  $A_2$  respecto a la concentración. (b)  $A_2$  respecto a la temperatura.

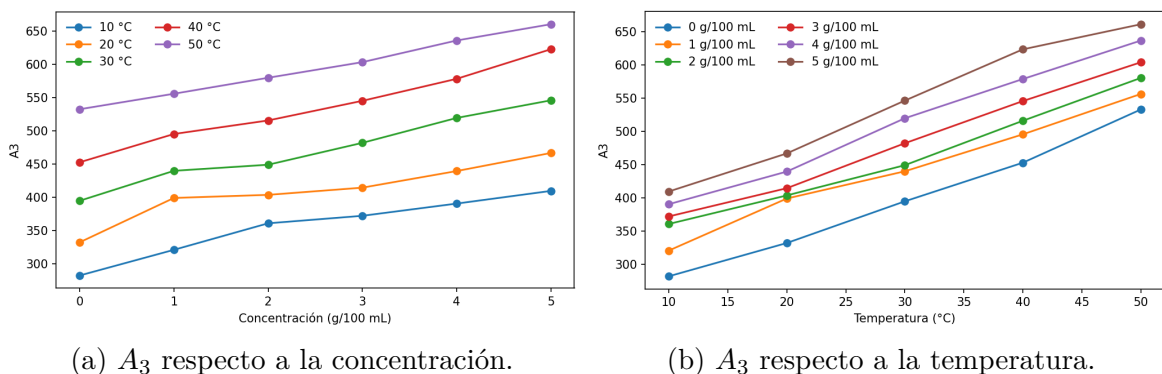
**Fig. 3.5:** Comportamiento de  $A_2$  respecto a la concentración y temperatura.

convierte a  $A_2$  en el mejor candidato entre las amplitudes Gaussianas para representar variaciones de concentración.

Este incremento indica que la presencia de iones produce una redistribución de la población de moléculas de agua hacia estados vibracionales intermedios, coherente con fenómenos de hidratación y ruptura parcial de la red de enlaces de hidrógeno.

A diferencia de  $A_1$  y  $A_3$ ,  $A_2$  muestra únicamente cambios leves con la temperatura. Sus curvas presentan una variabilidad mucho menor y no exhiben una tendencia sistemática pronunciada.

$A_2$  es un descriptor robusto de concentración, con mínima sensibilidad a la temperatura. Esta propiedad será crucial en el diseño del modelo de amplitudes  $A_k(C, T)$ .



(a)  $A_3$  respecto a la concentración. (b)  $A_3$  respecto a la temperatura.

**Fig. 3.6:** Comportamiento de  $A_3$  respecto a la concentración y temperatura.

Las Figuras 3.6a y 3.6b muestran el comportamiento de la amplitud  $A_3$ , asociada a poblaciones de moléculas de agua en estados menos coordinados o más libres.

$A_3$  aumenta significativamente con la temperatura, con una tendencia lineal clara para

todas las concentraciones. Este comportamiento refleja la transición térmica hacia configuraciones O–H más débiles, en concordancia con lo reportado en la literatura para la banda O–H a temperaturas elevadas.

$A_3$  también aumenta con la concentración de NaCl, aunque la magnitud del cambio es menor que en el caso de la temperatura. La presencia de iones promueve la ruptura de estructuras ordenadas, lo que incrementa la contribución relativa de estos estados menos enlazados.

$A_3$  es sensible tanto a la temperatura como a la concentración, pero la temperatura es la variable dominante.

### 3.5. Resultados de los modelos temperatura-concentración de las amplitudes

Una vez estabilizada la descomposición espectral global y habiéndose establecido que las amplitudes  $A_1$ ,  $A_2$  y  $A_3$  responden sistemáticamente a la temperatura y a la concentración de NaCl, el siguiente paso consistió en construir modelos cuantitativos que describieran dicha dependencia. Para ello se evaluó un modelo de amplitudes descrito en el capítulo 2, cuya forma general es:

$$A_k(C, T) = a_k + b_{T,k} T + c_{C,k} C + d_{TC,k} (T C),$$

donde  $k$  corresponde a cada uno de los tres componentes gaussianos de la banda O–H.

Este modelo incorpora tres elementos fundamentales:

- **Efectos principales:**
  - $b_{T,k}$ : sensibilidad de  $A_k$  a la temperatura.
  - $c_{C,k}$ : sensibilidad de  $A_k$  a la concentración de NaCl.
- **Término constante  $a_k$ :** nivel base de cada componente en ausencia de efectos.
- **Término de cruzado  $d_{TC,k}$ :** captura el hecho físico de que el efecto de la sal puede depender de la temperatura y, recíprocamente, que la respuesta térmica

puede modificarse por la salinidad.

**Tabla 3.4:** Parámetros ajustados para los modelos de amplitudes  $A_k(C, T)$  junto con sus métricas de calidad.

Amplitud	$R^2$	RMSE	a	$b_T$	$c_C$	$d_{TC}$
$A_1$	0.9952	6.3053	1465.3144	-5.6966	-26.9596	0.0392
$A_2$	0.9777	6.5755	1687.9414	0.8595	26.0902	-0.0448
$A_3$	0.9855	11.8188	224.6501	5.8835	25.6423	0.0600

A continuación se analiza cada amplitud como variable latente asociada a subpoblaciones específicas de enlaces O–H en el líquido.

**Componente  $A_1$ : Dependencia negativa con  $T$  y  $C$**

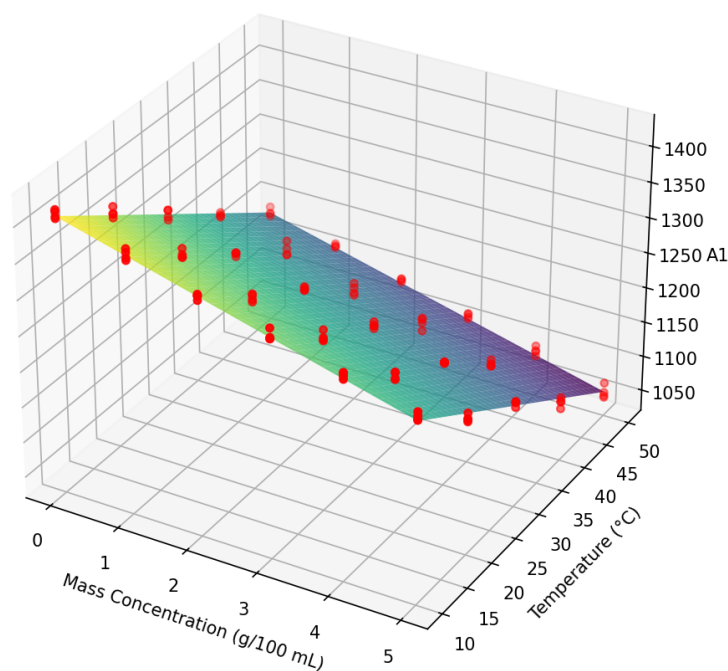
**Tabla 3.5:** Interpretación de los coeficientes del modelo de interacción para  $A_1(C, T)$ .

Coefficiente	Valor	Interpretación
$b_{T,1}$	-5,696	$A_1$ disminuye al aumentar la temperatura.
$c_{C,1}$	-26,959	$A_1$ disminuye al aumentar la concentración de NaCl.
$d_{TC,1}$	0,039	El efecto combinado $T \cdot C$ incrementa levemente el crecimiento.

El componente  $A_1$  corresponde a una subpoblación vibracional localizada en la región de O–H más fuertemente enlazados. El incremento de temperatura rompe parcialmente los puentes de hidrógeno, mientras que la presencia de sal perturba la red estructural del agua, reduciendo la contribución de estas estructuras altamente coordinadas.

La Figura 3.7 muestra el plano inclinado decreciente en ambas direcciones ( $T$  y  $C$ ). Esto sugiere una transición hacia estados menos estructurados.

**Componente  $A_2$ : Dependencia dominante con  $C$  y efecto térmico débil**



**Fig. 3.7:** Modelo de amplitudes para  $A_1$ .

**Tabla 3.6:** Interpretación de los coeficientes del modelo de interacción para  $A_2(C, T)$ .

Coeficiente	Valor	Interpretación
$b_{T,2}$	+0,859	$A_2$ aumenta levemente con la temperatura.
$c_{C,2}$	+26,090	$A_2$ aumenta de manera dominante con la concentración de NaCl.
$d_{TC,2}$	-0,044	El efecto combinado $T \cdot C$ atenúa parcialmente el crecimiento.

El componente  $A_2$  se asocia a una población de modos O–H intermedios (parcialmente enlazados). La presencia de sal favorece la ruptura local de estructuras, incrementando la fracción de estados menos coordinados; por ello,  $A_2$  crece sistemáticamente con la concentración.

La temperatura también promueve configuraciones menos coordinadas, pero su efecto es mucho menor en comparación con la influencia salina.

La Figura 3.8 muestra una superficie levemente curvada, crecimiento fuerte con  $C$  y una débil dependencia térmica.

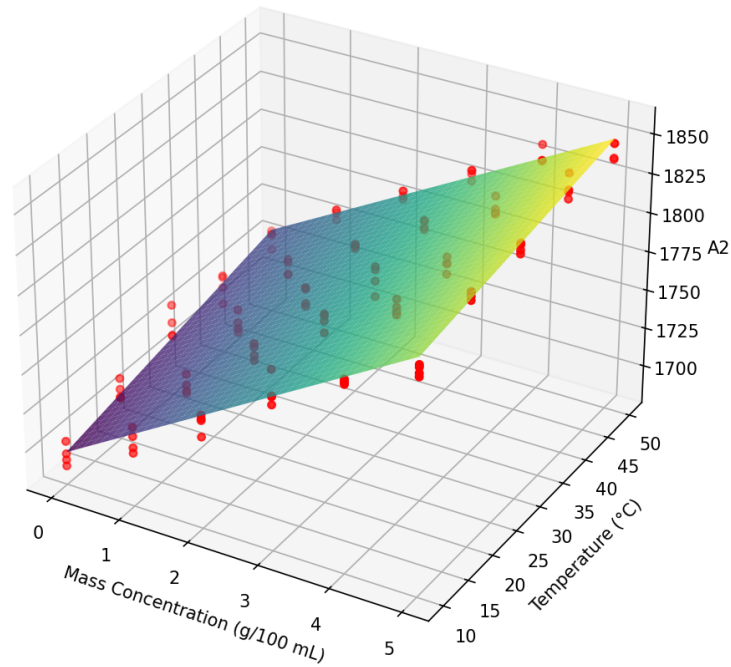


Fig. 3.8: Modelo de amplitudes para  $A_2$ .

**Componente  $A_3$ : Dependencia positiva con  $T$  y  $C$**

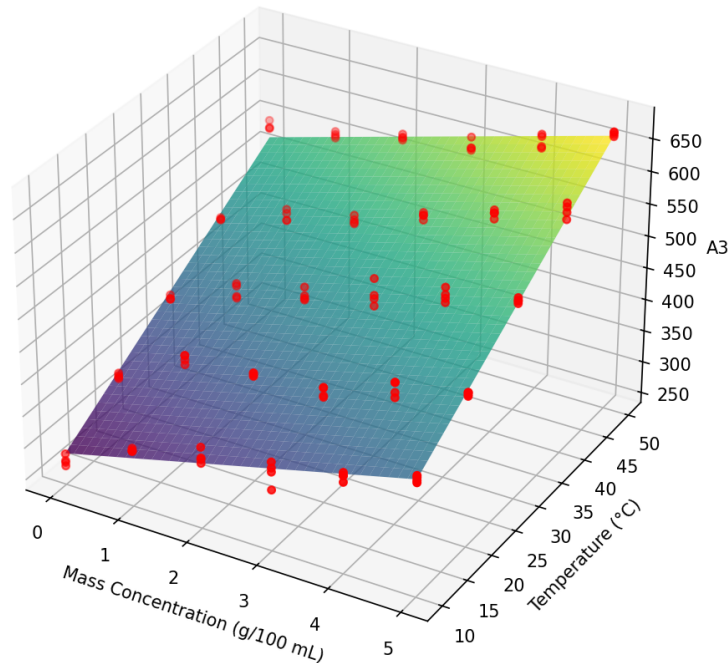
Tabla 3.7: Interpretación de los coeficientes del modelo de interacción para  $A_3(C, T)$ .

Coefficiente	Valor	Interpretación
$b_{T,3}$	+5,883	$A_3$ aumenta con la temperatura.
$c_{C,3}$	+25,642	$A_3$ aumenta también con la concentración de NaCl.
$d_{TC,3}$	+0,060	El aumento simultáneo de $T$ y $C$ intensifica levemente el crecimiento.

El componente  $A_3$  corresponde a modos O–H débilmente enlazados o casi libres. Tanto la temperatura como la sal empujan la red de agua hacia estados más abiertos y menos coordinados, lo que incrementa esta subpoblación.

Por lo tanto,  $A_3$  actúa como descriptor del grado de debilitamiento estructural del agua, en plena coherencia con los espectros Raman reportados para soluciones salinas.

La Figura 3.9 muestra una superficie ascendente en ambas direcciones con convexidad positiva. Esto indica que  $A_3$  captura los O–H más libres o menos coordinados.



**Fig. 3.9:** Modelo de amplitudes para  $A_3$ .

El término adicional  $T \cdot C$  permite capturar curvaturas suaves en las superficies, especialmente visibles en  $A_2$  y  $A_3$ .

El término de interacción mejora la fidelidad estructural del modelo, no produce sobreajuste (tiene valores bajos en comparación con los aportes de T y C) y aporta interpretabilidad fisicoquímica.

Los modelos de término cruzado permiten describir cuantitativamente el efecto conjunto de la temperatura y la concentración sobre las subpoblaciones vibracionales del agua, revelar relaciones fisicoquímicas coherentes con la literatura estructural del agua salina y generar superficies suaves, estables y robustas para la posterior etapa de inversión del modelo y estimación de concentración.

Este análisis confirma que las amplitudes  $A_1, A_2, A_3$  actúan como variables latentes

que capturan información estructural profunda de la red O–H, permitiendo su uso en cuantificación directa mediante técnicas de inversión o modelos híbridos.

### 3.6. Resultados de entrenamiento del modelo inverso para la estimación de concentración

Una vez construidos los modelos de interacción entre amplitudes y variables fisicoquímicas ( $C, T$ ), se evaluó el desempeño del método de inversión aplicado sobre el mismo conjunto de 120 espectros experimentales utilizados durante el ajuste del modelo global. Esta evaluación corresponde a una validación interna, cuyo objetivo es cuantificar la precisión del método de reconstrucción matemática sin aún someterlo a condiciones de generalización fuera del conjunto de entrenamiento.

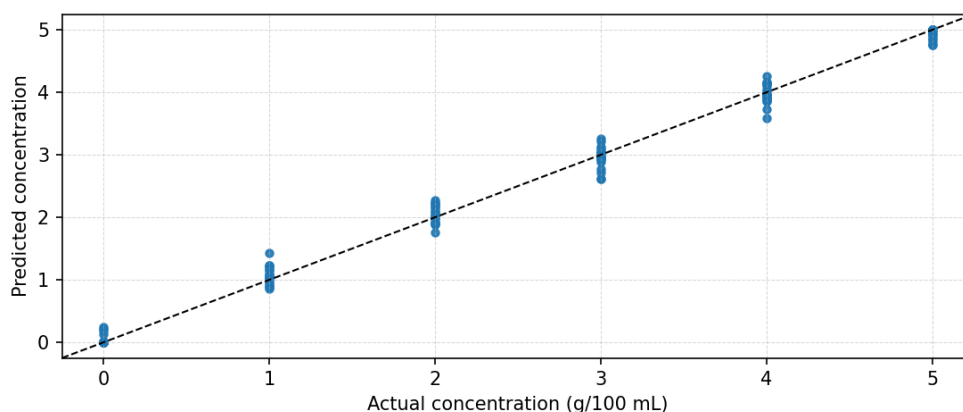
La descomposición espectral mediante Gaussianas globales se realizó exclusivamente utilizando el conjunto de entrenamiento, compuesto por 120 espectros, a partir del cual se estimaron y fijaron los parámetros globales de posición y ancho,  $(\mu_k, \sigma_k)$ .

Posteriormente, estos parámetros se mantuvieron constantes y fueron aplicados sin reajuste al conjunto de validación externa, estimándose únicamente las amplitudes asociadas a cada espectro externo. De este modo, la validación externa evalúa de forma genuina la capacidad de generalización del modelo, sin reutilización de información del conjunto de prueba durante la etapa de entrenamiento.

Los resultados obtenidos evidencian que la formulación basada en amplitudes Gaussianas como variables latentes permite recuperar la concentración de cada muestra con alta exactitud, así como reproducir coherentemente los espectros Raman originales.

La Figura 3.10 muestra la comparación entre la concentración real de cada muestra del dataset y la concentración estimada del modelo inverso.

El desempeño del modelo en el entrenamiento se observa en la Tabla 3.8



**Fig. 3.10:** Comparación entre concentración real y concentración predicha del modelo inverso en el entrenamiento.

**Tabla 3.8:** Desempeño global del modelo inverso en el entrenamiento.

Métrica	Valor
MAE	0,1007 g/100 mL
RMSE	0,1418 g/100 mL
$R^2$	0,9931

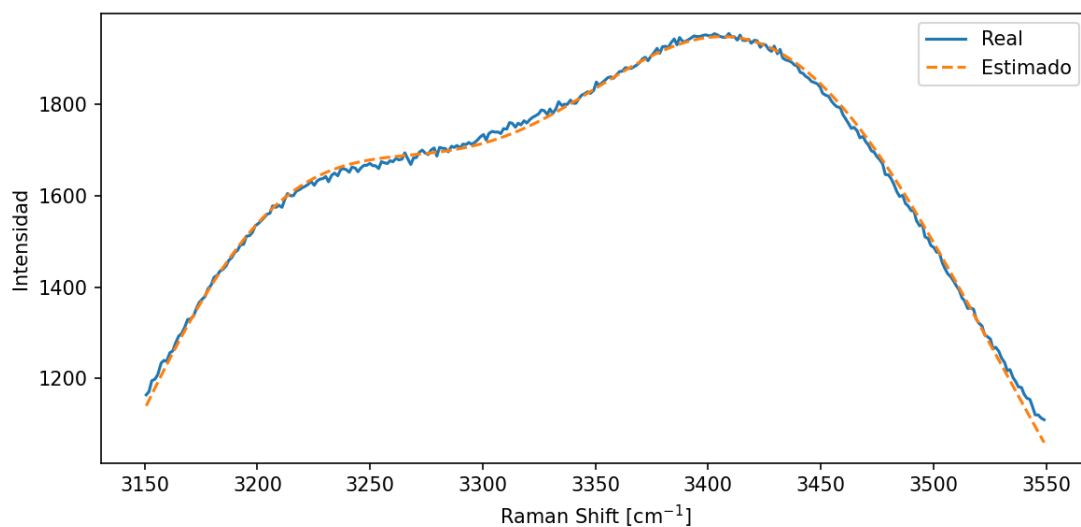
Esto confirma que el modelo de interacción  $A_k(C, T)$  es numéricamente estable y que las amplitudes espectrales contienen suficiente información para identificar la concentración dentro del dominio experimental utilizado.

Es importante destacar que estos resultados corresponden exclusivamente al conjunto de entrenamiento. La capacidad de generalización del modelo se evaluará posteriormente en la sección dedicada a soluciones externas/no vistas.

Más allá de la exactitud numérica en  $C$ , un requisito fundamental del modelo es que la concentración estimada permita reconstruir fielmente la banda O–H completa. Esto es particularmente relevante puesto que valida la consistencia físico–espectral del enfoque.

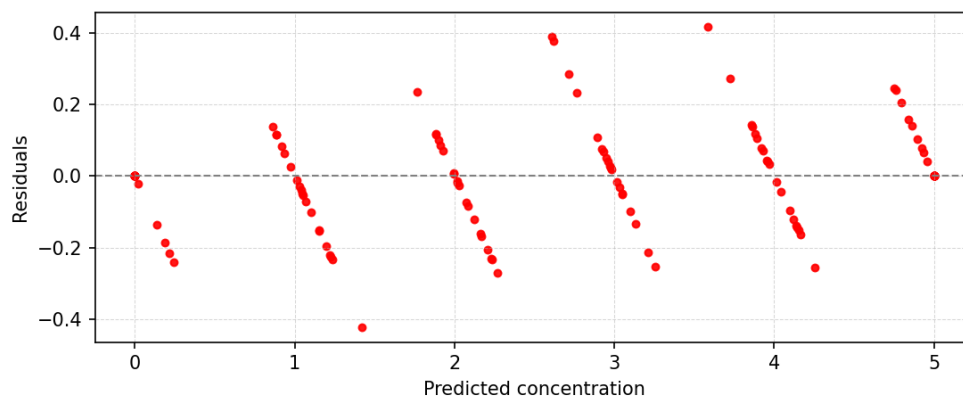
La Figura 3.11 ejemplifica esta verificación para una muestra representativa.

La diferencia entre espectro real y reconstruido es menor al 1 % del rango de intensidad, lo cual demuestra la consistencia del modelo inverso, que la estimación de concentración no introduce distorsiones en la representación espectral y que las amplitudes Gaussianas



**Fig. 3.11:** Espectro real vs espectro estimado con el modelo inverso en el entrenamiento para muestra de (0g/100 mL, 30°C).

actúan efectivamente como un conjunto reducido de coordenadas fisicoquímicas de la banda O–H.



**Fig. 3.12:** Residuos de la estimación del modelo inverso en el entrenamiento.

La Figura 3.12 muestra la distribución de residuos:

$$\varepsilon_i = C_{\text{pred},i} - C_{\text{real},i}.$$

Los resultados son consistentes con un método estable y libre de sesgo. No existe tendencia sistemática a sobreestimar o subestimar la concentración. Los residuos forman clústeres verticales correspondientes a las seis concentraciones nominales del estudio. La

forma de estas bandas es consecuencia del diseño experimental, no del modelo. Dentro de cada banda, la dispersión es reducida y simétrica. No se observan patrones que se correlacionen con valores específicos de  $T$ , lo cual confirma que el modelo de interacción eliminó correctamente el efecto térmico sobre las amplitudes Gaussianas.

Estos resultados validan el modelo en términos internos y permiten avanzar hacia la siguiente etapa: la validación externa, donde el método deberá estimar la concentración de espectros no utilizados en la construcción del modelo de interacción.

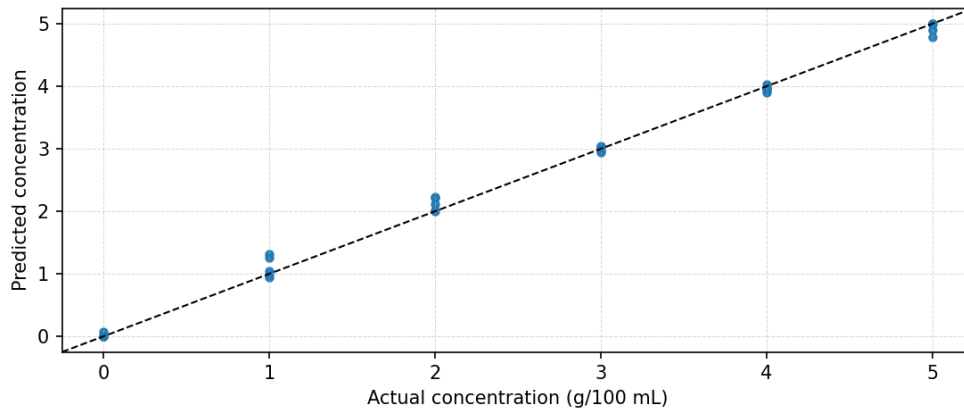
### 3.7. Resultados de validación externa del modelo inverso para la estimación de concentración

Con el objetivo de evaluar la capacidad de generalización del modelo de amplitudes Gaussianas  $A_k(C, T)$  y de su formulación inversa para la estimación de la concentración  $C$ , se llevó a cabo una validación externa utilizando espectros Raman no empleados durante la etapa de calibración. Esta base de datos generada para la validación externa cuenta con 30 espectros dentro del rango de concentraciones analizado, pero utilizando temperaturas conocidas y desconocidas. Cada espectro externo fue procesado mediante el mismo *pipeline* descrito en el Capítulo 3, incluyendo el recorte de la banda O–H, la reconstrucción espectral basada en Gaussianas globales y la estimación inversa de la concentración, considerando explícitamente la temperatura de adquisición  $T$ .

La relación entre la concentración real y la concentración estimada por el modelo para el conjunto de validación externa se muestra en la Figura 3.13:

En dicha figura se observa que la mayoría de los puntos experimentales se distribuyen cercanos a la recta ideal ( $\hat{C} = C$ ), lo que evidencia un comportamiento prácticamente lineal y no sesgado del modelo en todo el rango de concentraciones analizado.

La exactitud del modelo en el conjunto de validación externa se evaluó mediante el error absoluto medio (MAE) y la raíz del error cuadrático medio (RMSE), métricas que cuantifican la magnitud promedio de los errores de estimación en unidades físicas de concentración. Complementariamente, el coeficiente de determinación ( $R^2$ ) se utilizó como medida de la capacidad del modelo para explicar la variabilidad observada en los datos experimentales. Los resultados obtenidos se presentan en la Tabla 3.9



**Fig. 3.13:** Relación entre concentración real vs concentración predicha con el modelo inverso en validación externa.

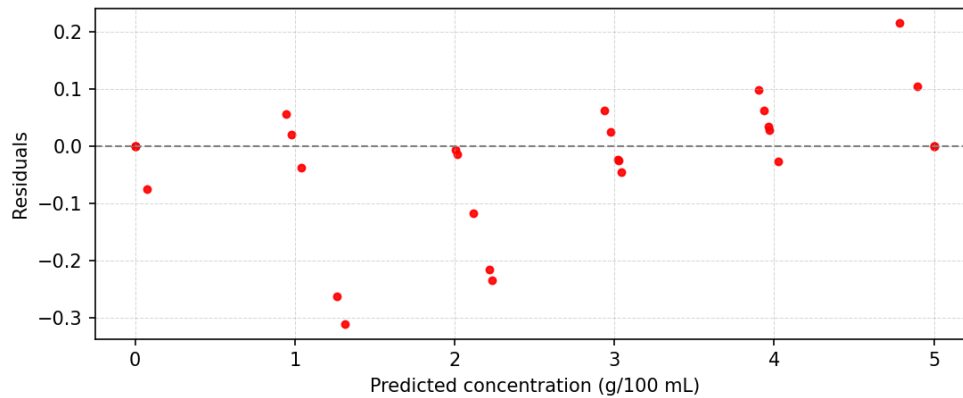
**Tabla 3.9:** Desempeño global del modelo inverso en validación externa.

Métrica	Valor
MAE	0,0699 g/100 mL
RMSE	0,1113 g/100 mL
$R^2$	0,9957

Estos valores confirman que el modelo presenta un alto poder predictivo en validación externa, con errores absolutos pequeños en relación con el rango total de concentraciones estudiado (0–5 g/100 mL) y una varianza explicada superior al 99 %. En particular, el bajo valor de MAE indica que, en promedio, la desviación entre la concentración real y la estimada es inferior a 0,1 g/100 mL, lo que resulta consistente con los requerimientos de precisión típicos en aplicaciones analíticas basadas en espectroscopía Raman.

Con el fin de evaluar la presencia de sesgos sistemáticos o heterocedasticidad, se analizó la distribución de los residuos definidos como la diferencia entre la concentración real y la concentración predicha, en función de la concentración estimada:

La Figura 3.14 de residuos muestra que estos se distribuyen de forma aproximadamente simétrica alrededor de cero, sin una tendencia clara creciente o decreciente con la concentración predicha. Este comportamiento indica la ausencia de sesgos sistemáticos globales en el modelo. Asimismo, no se observa un incremento marcado de la varianza



**Fig. 3.14:** Residuos de la estimación del modelo inverso en validación externa.

de los residuos con la concentración, lo que sugiere que el error es aproximadamente homocedástico en el rango estudiado.

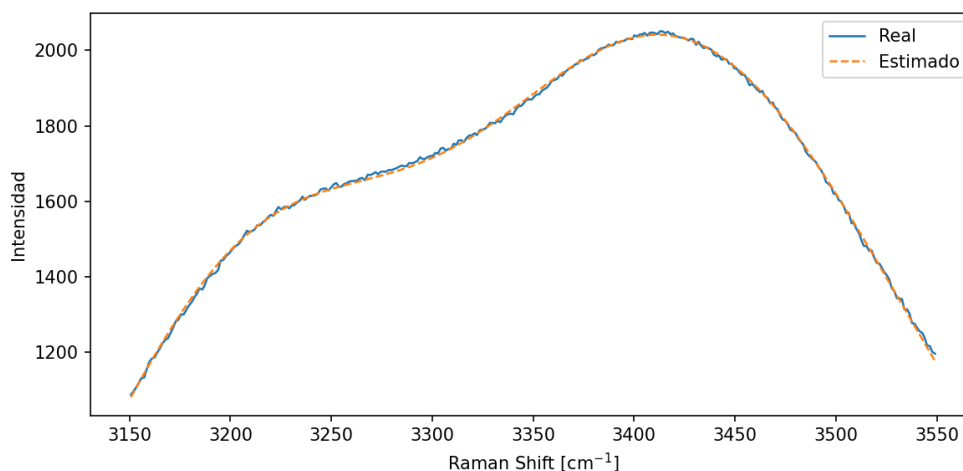
Se identifican algunos residuos de mayor magnitud en concentraciones bajas y en el extremo superior del rango, lo cual es consistente con lo observado en la relación real vs. predicha y con las limitaciones experimentales del dominio de calibración. No obstante, estos casos no dominan el comportamiento global y no comprometen la estabilidad del modelo.

Además del análisis cuantitativo de la concentración, se evaluó la coherencia espectral del modelo comparando espectros Raman medidos con aquellos reconstruidos a partir de la concentración estimada  $\hat{C}$  y la temperatura conocida  $T$ .

**Tabla 3.10:** Desempeño de espectro de 2g/100mL a 35°C estimado.

Métrica	Valor
$MAE_{\text{espectral}}$	5.43
$RMSE_{\text{espectral}}$	6.74
$R^2$	0,9993

La superposición entre el espectro experimental y el espectro reconstruido evidencia un ajuste prácticamente perfecto de la banda O–H, con errores espectrales pequeños y coeficiente de determinación  $R^2 = 0,9993$ . Este resultado confirma que el modelo no solo estima correctamente la concentración, sino que también preserva la forma espectral



**Fig. 3.15:** Espectro real vs espectro estimado con modelo inverso correspondiente a una muestra externa de 2g/100mL a 35°C.

completa, capturando adecuadamente la dependencia conjunta con la temperatura y la concentración del soluto.

En conjunto, los resultados de validación externa demuestran que el modelo de amplitudes Gaussianas presenta un desempeño robusto tanto desde el punto de vista cuantitativo, al estimar con precisión la concentración de NaCl en espectros no vistos, como desde el punto de vista estructural, al reproducir fielmente la forma de la banda O–H. La combinación de una alta correlación entre concentraciones reales y predichas, una distribución de residuos sin sesgos evidentes y una elevada coherencia espectral respalda la solidez del enfoque propuesto.

**Tabla 3.11:** Resultados individuales de validación externa: concentración real, concentración estimada y temperatura de medición.

$C_{\text{real}}$ [g/100 mL]	$\hat{C}$ [g/100 mL]	$T$ [°C]
0	0.000	15.0
0	0.000	20.0
0	0.000	25.0
0	0.000	35.0
0	0.074	45.0
1	1.311	15.0
1	1.262	25.0
1	1.037	35.0
1	0.944	40.0
1	0.980	45.0
2	2.234	15.0
2	2.215	20.0
2	2.117	25.0
2	2.006	35.0
2	2.014	45.0
3	3.045	15.0
3	3.023	25.0
3	3.025	35.0
3	2.976	40.0
3	2.938	45.0
4	3.902	15.0
4	3.938	20.0
4	3.966	25.0
4	4.027	35.0
4	3.972	45.0
5	4.785	15.0
5	4.896	25.0
5	5.000	35.0
5	5.000	40.0
5	5.000	45.0

Estos resultados consolidan al modelo de amplitudes como una herramienta confiable para la estimación de concentración salina en soluciones acuosas mediante espectroscopía Raman y proporcionan una base sólida para su comparación con métodos multivariantes alternativos, como el enfoque ICS–PLS.

### 3.8. Resultados de entrenamiento del modelo ICS–PLS para la estimación de concentración

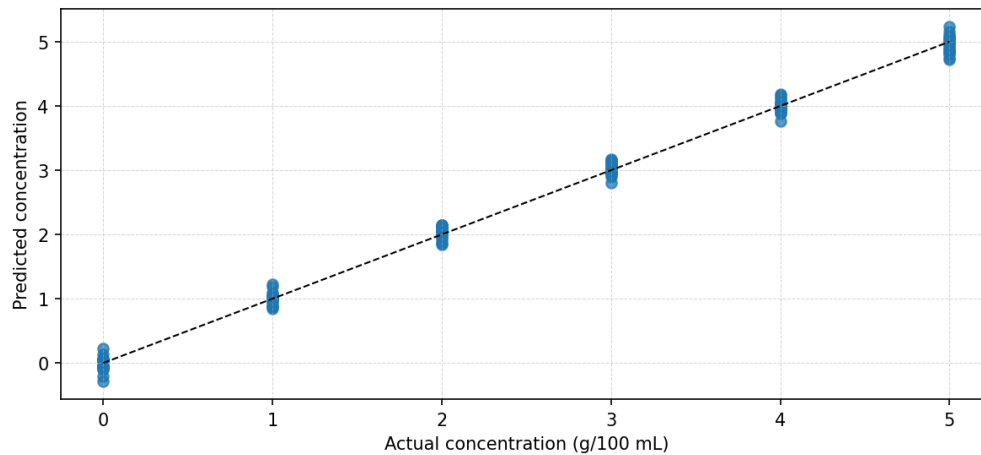
Con el fin de evaluar el desempeño del enfoque multivariante basado en *Individual Contribution Standardization* (ICS) combinado con *Partial Least Squares* (PLS), se analizó tanto el comportamiento del modelo durante la etapa de entrenamiento como su capacidad de generalización mediante validación externa. Este enfoque permite desacoplar explícitamente la variabilidad espectral inducida por la temperatura y concentrar la regresión multivariante únicamente en la información asociada a la concentración del analito.

El modelo PLS fue entrenado utilizando tres variables latentes ( $LV = 3$ ). Esta elección se realizó de forma deliberada con el objetivo de establecer una comparación directa y equitativa con el modelo inverso basado en la descomposición espectral en tres Gaussianas, el cual genera tres amplitudes espectrales ( $A_1, A_2, A_3$ ) como variables latentes físicas.

Cabe señalar que el número de variables latentes en PLS controla el compromiso sesgo–varianza del modelo. En pruebas exploratorias, al incrementar  $LV$  se observa una mejora significativa del rendimiento predictivo. Sin embargo, en el marco de esta comparación metodológica se fija  $LV=3$  para mantener una complejidad comparable entre ambos enfoques, evitando que el incremento de capacidad del modelo PLS introduzca una ventaja atribuible únicamente al aumento del número de componentes y no al mecanismo de modelado considerado.

De este modo, ambos enfoques emplean el mismo número de grados de libertad latentes, permitiendo evaluar sus desempeños predictivos bajo condiciones metodológicamente comparables.

En la Figura 3.16 se observa que los puntos experimentales se alinean estrechamente con



**Fig. 3.16:** Relación entre concentración real vs concentración predicha en entrenamiento del modelo ICS-PLS.

la recta identidad, sin evidencia de sesgos sistemáticos en el rango de concentraciones estudiadas.

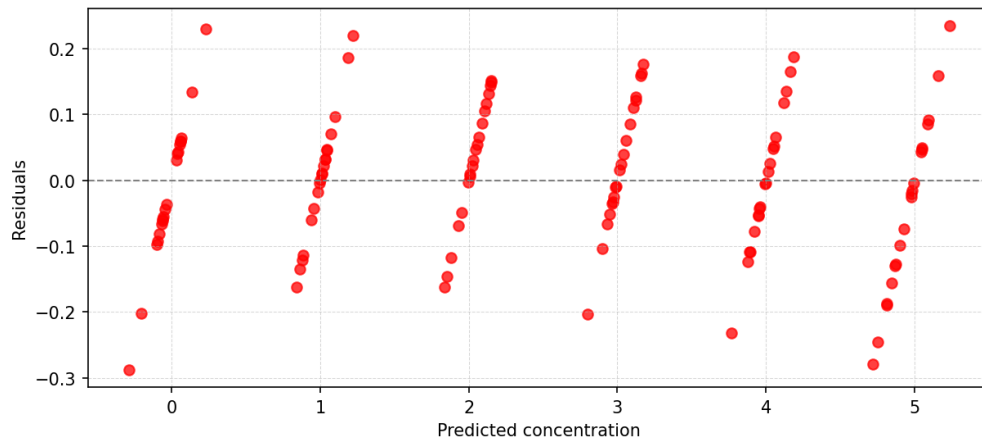
El desempeño global del modelo ICS-PLS durante la etapa de entrenamiento se evaluó mediante las métricas de error absoluto medio (MAE), raíz del error cuadrático medio (RMSE) y coeficiente de determinación ( $R^2$ ), obteniéndose los siguientes resultados:

**Tabla 3.12:** Desempeño global del modelo ICS-PLS en el entrenamiento.

Métrica	Valor
MAE	0,0870 g/100 mL
RMSE	0,1092 g/100 mL
$R^2$	0,9959

Estos valores indican un ajuste excelente del modelo en entrenamiento, con errores absolutos medios inferiores a 0,09 g/100 mL y una varianza explicada superior al 99,8 %. La elevada correlación entre las concentraciones reales y predichas refleja que, tras la corrección espectral mediante ICS, el modelo PLS es capaz de capturar de forma muy eficiente la relación entre la forma espectral corregida de la banda O-H y la concentración de NaCl.

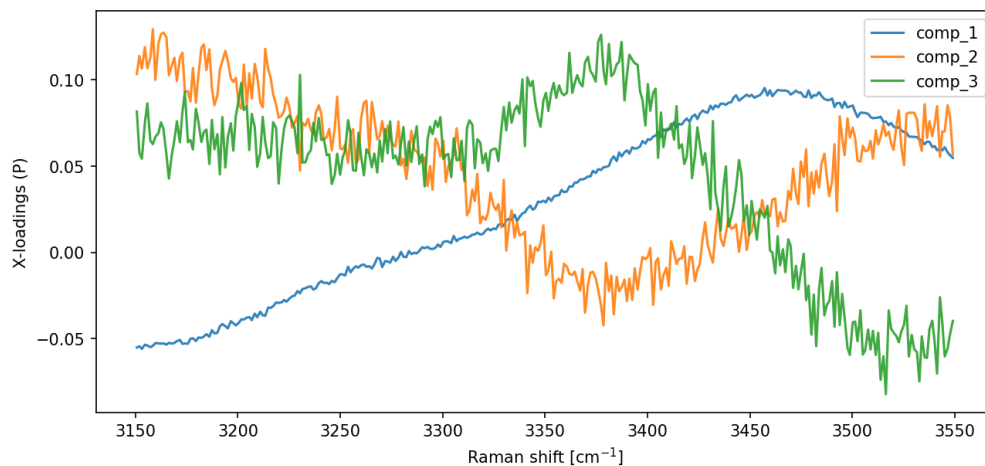
Para evaluar la estabilidad del modelo y la presencia de posibles sesgos locales, se analizó la distribución de los residuos en función de la concentración predicha.



**Fig. 3.17:** Residuos de la estimación en entrenamiento del modelo ICS-PLS.

La nube de puntos muestra residuos centrados alrededor de cero y una dispersión aproximadamente constante a lo largo de todo el rango de concentración, lo que sugiere un comportamiento homocedástico y la ausencia de tendencias sistemáticas. Este resultado confirma que la combinación ICS-PLS logra eliminar eficazmente la variabilidad térmica residual antes de la regresión multivariante.

Con el objetivo de identificar qué regiones espectrales de la banda O-H contribuyen a la modelación de la concentración, se analizan los loadings  $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$  del modelo PLS entrenado con  $LV = 3$ . En PLS, cada vector de loadings describe la estructura de  $\mathbf{X}$  capturada por la variable latente  $a$ , es decir, cómo se distribuye en el eje espectral la variación explicada por dicha componente cómo se observa en la Figura 3.18.



**Fig. 3.18:**  $X$ -loadings  $\mathbf{P}$  del modelo PLS ( $LV=3$ ) para  $T_{\text{ref}} = 30^\circ\text{C}$  en la región  $3150\text{--}3550\text{ cm}^{-1}$ .

El primer *loading* ( $\mathbf{p}_1$ ) presenta una variación suave y de gran escala a lo largo de la banda, con un incremento progresivo desde el extremo de menor número de onda hacia la región central–alta, seguido de una leve disminución hacia  $3550\text{ cm}^{-1}$ . Este comportamiento sugiere que la primera variable latente está asociada principalmente a cambios globales de forma en la banda O–H, más que a rasgos estrechos localizados.

El segundo *loading* ( $\mathbf{p}_2$ ) exhibe un patrón de contraste entre subregiones: valores positivos en el extremo de menor número de onda, un cambio de signo en torno a la región central, y una recuperación hacia valores positivos en la zona de mayor número de onda. Este tipo de estructura es consistente con variaciones diferenciales en la forma de la banda, donde el modelo captura redistribuciones relativas de intensidad entre distintas subzonas espectrales.

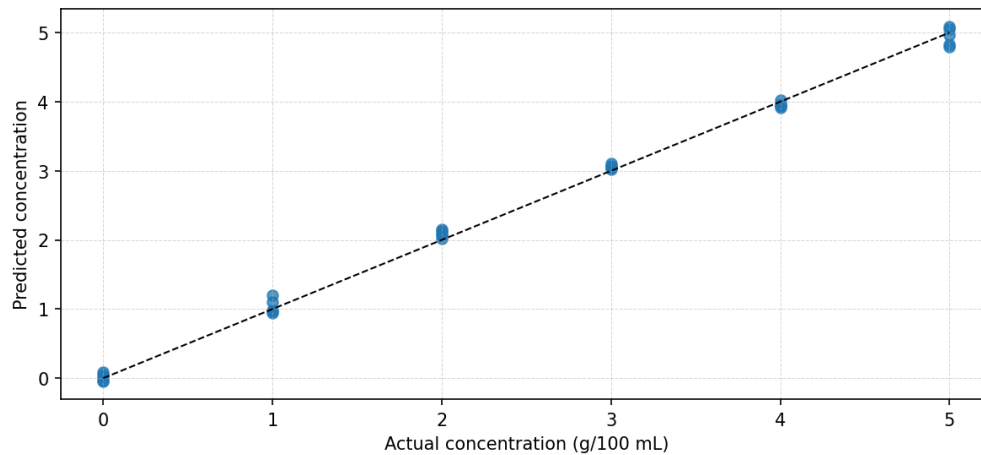
El tercer *loading* ( $\mathbf{p}_3$ ) muestra una estructura adicional con cambios más marcados en la región de mayor número de onda y mayor sensibilidad a variaciones locales, lo cual es esperable para componentes de orden superior que capturan efectos residuales de menor magnitud. Finalmente, se destaca que el signo de los *loadings* es arbitrario en PLS, por lo que la interpretación debe realizarse en términos relativos (forma y regiones dominantes) y no en el signo absoluto.

### 3.8.1. Resultados de validación externa del modelo ICS–PLS

La capacidad de generalización del modelo ICS–PLS se evaluó mediante un conjunto de espectros Raman no utilizados durante el entrenamiento, adquiridos a distintas combinaciones de concentración y temperatura. En esta etapa, los espectros fueron primero corregidos a una temperatura de referencia  $T_{\text{ref}} = 30\text{ °C}$  mediante ICS y posteriormente utilizados como entrada del modelo PLS entrenado.

La relación entre concentración real y predicha en validación externa muestra una buena alineación con la recta identidad, confirmando que el modelo conserva su capacidad predictiva frente a espectros no vistos y a variaciones de temperatura previamente corregidas mediante ICS.

El desempeño global obtenido en validación externa se resume a continuación:



**Fig. 3.19:** Relación entre concentración real vs concentración predicha en validación externa del modelo ICS-PLS.

**Tabla 3.13:** Desempeño global del modelo ICS-PLS en validación externa.

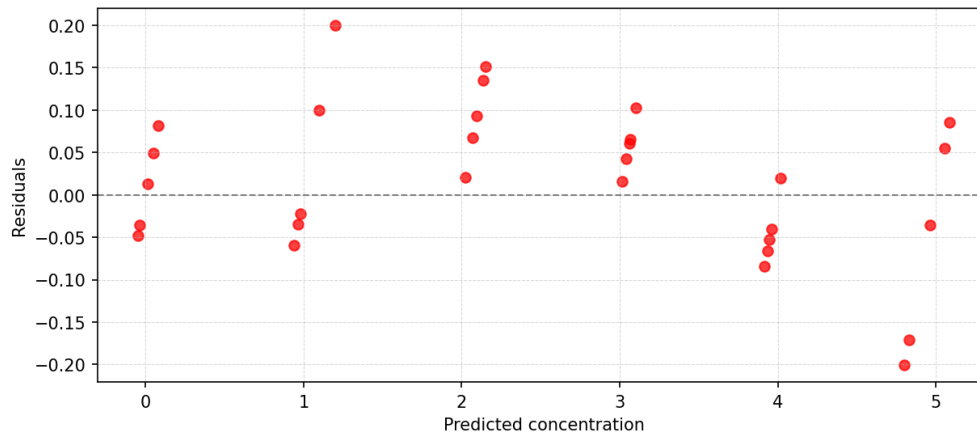
Métrica	Valor
MAE	0,0736 g/100 mL
RMSE	0,0895 g/100 mL
$R^2$	0,9973

Si bien los errores aumentan respecto a la etapa de entrenamiento —como es esperable en un escenario de validación externa— el modelo mantiene un alto nivel de precisión, con más del 99 % de la varianza explicada y errores absolutos medios del orden de 0,1 g/100 mL.

El análisis de residuos correspondiente a la validación externa se presenta en la Figura 3.20.

Los residuos se distribuyen de forma aproximadamente simétrica alrededor de cero, sin patrones evidentes de dependencia con la concentración predicha. Se observa una dispersión ligeramente mayor que en entrenamiento, particularmente en concentraciones extremas, lo cual es coherente con la propagación de errores introducida tanto por la corrección ICS como por la extrapolación moderada del modelo PLS fuera del conjunto de entrenamiento.

Los casos en que la concentración estimada alcanza exactamente el valor de 5 g/100 mL



**Fig. 3.20:** Residuos de la estimación en validación externa del modelo ICS-PLS.

o 0 g/100 mL corresponden a saturación del estimador debido a la restricción impuesta durante la resolución del problema inverso, y no a una pérdida de capacidad predictiva del modelo.

**Tabla 3.14:** Resultados individuales de validación externa para el modelo ICS-PLS

$C_{\text{real}}$ [g/100 mL]	$\hat{C}$ [g/100 mL]	$T$ [°C]
0.0000	-0.0503	15
0.0000	0.0229	20
0.0000	-0.0575	25
0.0000	-0.0096	35
0.0000	0.2640	45
1.0000	1.1608	15
1.0000	1.0474	25
1.0000	1.0006	35
1.0000	0.9724	40
1.0000	1.1059	45
2.0000	2.1126	15
2.0000	2.0609	20
2.0000	1.9655	25
2.0000	2.0304	35
2.0000	2.0936	45
3.0000	3.0658	15
3.0000	2.9283	25
3.0000	3.0405	35
3.0000	3.0239	40
3.0000	2.9910	45
4.0000	3.9667	15
4.0000	3.9474	20
4.0000	3.8474	25
4.0000	4.0832	35
4.0000	3.8951	45
5.0000	4.8758	15
5.0000	4.7863	25
5.0000	5.1553	35
5.0000	5.0392	40
5.0000	4.9406	45

### **3.9. Análisis de incertidumbre del sistema de estimación**

La incertidumbre total asociada a la estimación de la concentración de NaCl mediante espectroscopía Raman es el resultado de múltiples fuentes de error que se propagan a lo largo del proceso experimental y computacional. A continuación, se describen las principales contribuciones identificadas en este trabajo.

#### **Incertidumbre en la preparación de soluciones**

La concentración nominal de las soluciones fue determinada a partir de la masa de NaCl medida con una balanza analítica y del volumen fijado en 100 mL. La incertidumbre asociada a la balanza, junto con la tolerancia volumétrica del material de medición, introduce un error sistemático inicial en la concentración real de cada muestra, el cual se propaga a todas las etapas posteriores del análisis.

#### **Incertidumbre instrumental y de adquisición espectral**

El sistema Raman presenta fuentes de variabilidad asociadas a la estabilidad del láser, el ruido electrónico del detector, la resolución espectral finita y posibles fluctuaciones térmicas locales durante la medición. Estas contribuciones se reflejan en la dispersión observada entre réplicas espectrales adquiridas bajo condiciones nominalmente idénticas.

#### **Incertidumbre asociada al preprocesamiento y modelado espectral**

La descomposición global en Gaussianas introduce una incertidumbre adicional relacionada con el ajuste no lineal del modelo, particularmente en la estimación de las amplitudes espectrales. Si bien las posiciones y los anchos de las Gaussianas se mantienen fijos tras el entrenamiento, la variabilidad residual en las amplitudes refleja tanto ruido experimental como limitaciones inherentes al modelo paramétrico adoptado.

## **Incertidumbre en los modelos de estimación de concentración**

En el caso del modelo inverso, la incertidumbre está asociada a la propagación de errores desde las amplitudes estimadas hacia la solución del problema inverso. Para el modelo ICS-PLS, la incertidumbre adicional proviene del carácter estadístico del modelo multivariante y de su incapacidad para imponer restricciones físicas explícitas sobre la concentración estimada.

### **Incertidumbre global observada**

De forma integrada, estas fuentes de error se reflejan en los valores de MAE y RMSE obtenidos tanto en entrenamiento como en validación externa. Los resultados indican que la incertidumbre total del sistema se mantiene acotada dentro de márgenes aceptables para el rango de concentraciones estudiado, siendo consistente con aplicaciones de monitoreo y estimación cuantitativa en soluciones acuosas diluidas.

## **3.10. Discusión comparativa de los métodos de estimación de concentración**

La comparación entre el modelo inverso basado en descomposición espectral y el enfoque multivariante ICS-PLS debe interpretarse a la luz del análisis de incertidumbre presentado en la sección anterior. En efecto, las diferencias observadas en las métricas de desempeño no reflejan únicamente la capacidad intrínseca de cada método, sino también la forma en que cada uno gestiona, propaga o atenúa las distintas fuentes de incertidumbre presentes en el sistema experimental y computacional.

Desde el punto de vista del desempeño predictivo, ambos enfoques muestran una alta capacidad de estimación de la concentración de NaCl en el rango estudiado, con coeficientes de determinación superiores a 0,99 tanto en entrenamiento como en validación externa. No obstante, el modelo inverso presenta, en promedio, valores de MAE ligeramente inferiores, mientras que el método ICS-PLS exhibe un RMSE comparable o levemente menor en ciertas condiciones. Estas diferencias deben considerarse en relación con la incertidumbre experimental asociada a la preparación

de soluciones, la adquisición espectral y la dispersión observada entre réplicas, la cual establece un límite inferior práctico para el error alcanzable por cualquier modelo.

Un aspecto distintivo del modelo inverso es su capacidad para incorporar explícitamente restricciones físicas en el proceso de estimación, imponiendo el intervalo  $0 \leq C \leq 5$  g/100 mL durante la resolución del problema. Esta característica resulta particularmente relevante desde el punto de vista metrológico, ya que garantiza la plausibilidad física de las predicciones incluso en presencia de ruido espectral o incertidumbre en las amplitudes Gaussianas. En contraste, el modelo ICS-PLS, al tratarse de un enfoque lineal estadístico, carece de mecanismos internos para imponer dichas restricciones, lo que se manifiesta en la aparición ocasional de predicciones negativas o levemente fuera de rango, especialmente en condiciones de baja concentración o temperaturas elevadas. Este comportamiento no debe interpretarse como una falla del método, sino como una consecuencia directa de su naturaleza algebraica frente a la incertidumbre residual del sistema.

La interpretación de los resultados también difiere sustancialmente entre ambos enfoques. El modelo inverso permite una trazabilidad directa entre los cambios espectrales observados en la región O-H, las amplitudes Gaussianas y la concentración estimada, lo que facilita la interpretación físico-química del proceso de interacción ion-agua. Esta transparencia estructural contribuye a una mejor comprensión de cómo la incertidumbre espectral se propaga hacia la estimación final de la concentración. Por el contrario, el método ICS-PLS prioriza la eficiencia predictiva mediante la proyección del espectro corregido térmicamente sobre un espacio latente estadístico, sacrificando interpretabilidad física a cambio de simplicidad operativa.

Desde una perspectiva práctica, el enfoque ICS-PLS presenta ventajas en términos de implementación y velocidad computacional, especialmente en escenarios de monitoreo en línea o aplicaciones industriales donde se requiere una estimación rápida de la concentración y la temperatura puede ser compensada mediante un procedimiento de corrección espectral. Sin embargo, esta simplicidad viene acompañada de una mayor sensibilidad a la variabilidad no modelada y a la falta de restricciones físicas explícitas, aspectos que se vuelven relevantes cuando la incertidumbre instrumental o experimental aumenta.

En conjunto, los resultados obtenidos indican que ambos métodos son complementarios

más que excluyentes. El modelo inverso se perfila como una herramienta robusta y físicamente consistente para la estimación de concentración en contextos donde la interpretabilidad, el control de límites y la comprensión del fenómeno subyacente son prioritarios. Por su parte, el enfoque ICS–PLS constituye una alternativa eficaz y flexible para aplicaciones donde se privilegia la rapidez de implementación y el desempeño estadístico global, siempre que se acepte la necesidad de un tratamiento posterior de las predicciones fuera de rango.

Finalmente, al considerar explícitamente las fuentes de incertidumbre identificadas en este trabajo, se concluye que las diferencias de desempeño entre ambos métodos se encuentran, en gran medida, dentro de los márgenes esperables dados los límites experimentales del sistema. En este sentido, la elección del método más adecuado dependerá del equilibrio requerido entre interpretabilidad física, robustez frente a incertidumbre y requisitos operacionales específicos de la aplicación final.

# Capítulo 4

## Conclusiones

### 4.1. Conclusiones generales

En esta tesis se desarrolló y validó una metodología experimental–computacional para la estimación cuantitativa de la concentración de cloruro de sodio (NaCl) en soluciones acuosas a partir de espectros Raman adquiridos en la región del estiramiento O–H del agua, incorporando explícitamente el efecto de la temperatura. El enfoque propuesto integró espectroscopía Raman, modelado espectral físicamente fundamentado y técnicas de regresión, permitiendo abordar un problema de cuantificación tradicionalmente complejo debido a la naturaleza amplia y dominante de la señal del solvente.

Los resultados obtenidos demuestran que la banda Raman O–H, lejos de constituir una región espectral carente de información analítica, contiene una estructura interna rica y sistemática, evidenciada a través de descriptores espectrales que presentan relaciones reproducibles con la concentración y la temperatura. Esta información puede ser explotada con fines cuantitativos siempre que se utilice un tratamiento espectral adecuado y un modelo capaz de separar, organizar e interpretar dicha variabilidad de forma coherente con la física molecular del sistema.

## 4.2. Aportaciones metodológicas y conceptuales

La principal contribución de esta tesis radica en la formulación de un marco conceptual y metodológico que transforma un problema espectral altamente dimensional en un espacio latente de baja dimensión con significado físico explícito. En particular, se propuso una descomposición global de la banda O–H mediante un conjunto reducido de funciones Gaussianas compartidas, en la cual los parámetros espectrales de posición y ancho se consideran invariantes para todo el conjunto experimental, mientras que las amplitudes se permiten variar con la concentración y la temperatura.

Este enfoque permitió reducir drásticamente la dimensionalidad del problema espectral sin perder información relevante, preservando al mismo tiempo la interpretabilidad fisicoquímica del sistema. El análisis realizado confirmó que las amplitudes ( $A_1, A_2, A_3$ ) actúan como variables latentes físicas, asociadas a distintas subpoblaciones estructurales de enlaces O–H en el agua, y no simplemente como coeficientes matemáticos de ajuste.

En particular, se evidenció que una de estas amplitudes presenta una dependencia casi lineal con la concentración de NaCl y una sensibilidad limitada a la temperatura, lo que la convierte en un descriptor robusto y especialmente adecuado para la cuantificación. Este resultado constituye un aporte conceptual relevante, al demostrar que es posible extraer descriptores estructurales profundos de la red de enlaces de hidrógeno y utilizarlos directamente como coordenadas fisicoquímicas para la estimación cuantitativa.

Sobre esta base, se formuló y resolvió explícitamente el problema inverso de estimación de concentración, incorporando un modelo de interacción temperatura–concentración. Dicho modelo permitió no solo recuperar la concentración de forma precisa, sino también reconstruir coherentemente la forma espectral completa de la banda O–H, reforzando la consistencia físico–espectral del enfoque propuesto.

### 4.3. Desempeño y alcance de los modelos de estimación

Los resultados obtenidos muestran que el modelo inverso basado en amplitudes Gaussianas alcanza un desempeño elevado tanto en validación interna como externa, con errores absolutos y cuadráticos medios del orden de décimas de gramo por 100 mL y coeficientes de determinación superiores a 0,99 dentro del dominio experimental estudiado. Estos valores confirman que la información contenida en las amplitudes espectrales es suficiente para una estimación precisa y estable de la concentración de NaCl.

En paralelo, el enfoque multivariante basado en *Individual Contribution Standardization* combinado con *Partial Least Squares* (ICS-PLS) demostró ser una alternativa igualmente robusta desde el punto de vista predictivo, capaz de compensar eficazmente la variabilidad térmica y entregar resultados comparables en términos de precisión. La comparación entre ambos métodos puso de manifiesto que, si bien el modelo basado en Gaussianas ofrece una mayor interpretabilidad física y la posibilidad de imponer restricciones explícitas, el enfoque ICS-PLS destaca por su simplicidad operativa y flexibilidad, lo que lo hace atractivo para aplicaciones prácticas y automatizadas.

### 4.4. Incertidumbre experimental y límites de desempeño

Un aspecto central de este trabajo es la relación entre el desempeño alcanzado por los modelos y la incertidumbre inherente al sistema experimental. Las principales fuentes de error —asociadas a la preparación de soluciones, la adquisición espectral, la estabilidad instrumental y el ajuste del modelo— establecen un límite inferior práctico para la precisión alcanzable por cualquier método de estimación.

En este contexto, los valores de MAE y RMSE obtenidos deben interpretarse a la luz de dicha incertidumbre experimental combinada. Los resultados indican que ambos enfoques de estimación operan cerca de este límite práctico, lo que sugiere que mejoras adicionales en el desempeño requerirían no solo modelos más complejos, sino

también una reducción de la incertidumbre experimental mediante mejoras en el control instrumental, el protocolo de adquisición o la calidad espectral de los datos.

Esta observación refuerza la validez del enfoque adoptado y subraya que el desempeño alcanzado es consistente con los límites físicos y experimentales del sistema bajo estudio.

## 4.5. Limitaciones del estudio

A pesar de los resultados obtenidos, es importante reconocer las limitaciones del presente trabajo. El rango de concentraciones y temperaturas analizado es acotado, por lo que la extrapolación del modelo fuera de estos dominios no está garantizada. Asimismo, el estudio se centró exclusivamente en soluciones acuosas de NaCl, y la extensión directa de la metodología a otros solutos o sistemas multicomponentes requiere validaciones adicionales.

Por otra parte, aunque el modelo inverso mostró un comportamiento globalmente estable, se identificaron casos puntuales con errores mayores, lo que sugiere la conveniencia de incorporar mecanismos adicionales de control de calidad espectral o detección de valores atípicos en aplicaciones en línea. Finalmente, el análisis se basó en mediciones realizadas bajo condiciones controladas de laboratorio, sin abordar en profundidad efectos de largo plazo como deriva instrumental, ruido ambiental o cambios en la geometría de medición.

## 4.6. Proyecciones y conclusión final

A partir de los resultados obtenidos, se identifican diversas líneas de trabajo futuro, entre las que destacan la extensión del enfoque a sistemas multicomponentes, la implementación en esquemas de monitoreo en línea o en tiempo real, y la exploración de modelos híbridos que combinen la interpretabilidad física de las amplitudes Gaussianas con técnicas avanzadas de aprendizaje automático.

En conclusión, esta tesis demuestra que es posible estimar de manera precisa, consistente y físicamente interpretable la concentración de NaCl en soluciones acuosas a partir de espectros Raman del agua, incluso en presencia de variaciones térmicas significativas. El

uso de amplitudes espectrales con significado físico como variables latentes constituye un aporte conceptual central del trabajo, al permitir que la respuesta vibracional del solvente sea utilizada directamente como fuente de información cuantitativa. La combinación de modelado espectral fundamentado, análisis de incertidumbre y técnicas de estimación robustas posiciona este enfoque como una herramienta sólida y versátil, con alto potencial de transferencia a aplicaciones reales en análisis químico y control de procesos.

# Bibliografía

- [1] I. Durickovic, “Using raman spectroscopy for characterization of aqueous media and quantification of species in aqueous solution,” in *Applications of Molecular Spectroscopy to Current Research in the Chemical and Biological Sciences*, M. T. Stauffer, Ed. London: IntechOpen, 2016, ch. 19. [Online]. Available: <https://doi.org/10.5772/64550>
- [2] F. Perakis, L. Marco, A. Shalit, F. Tang, Z. Kann, T. Kühne, R. Torre, M. Bonn, and Y. Nagata, “Vibrational spectroscopy and dynamics of water,” *Chemical reviews*, vol. 116, 04 2016.
- [3] M. Baumgartner and R. Bakker, “Raman spectroscopy of pure h<sub>2</sub>o and nacl-h<sub>2</sub>o containing synthetic fluid inclusions in quartz - a study of polarization effects,” *Mineralogy and Petrology*, vol. 95, pp. 1–15, 01 2009.
- [4] 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>
- [5] S. Bashchenko, “On raman spectra of water, its structure and dependence on temperature,” *Semiconductor Physics Quantum Electronics and Optoelectronics*, vol. 14, pp. 77–79, 02 2011.
- [6] B. Ratajska-Gadomska and W. Gadomski, “Water structure in nanopores of agarose gel by raman spectroscopy,” *The Journal of Chemical Physics*, vol. 121, no. 24, pp. 12 583–12 588, 12 2004. [Online]. Available: <https://doi.org/10.1063/1.1826051>

- 
- [7] T. Dolenko, S. Burikov, S. Dolenko, A. Efitorov, I. Plastinin, V. Yuzhakov, and S. Patsaeva, “Raman spectroscopy of water-ethanol solutions: The estimation of hydrogen bonding energy and the appearance of clathrate-like structures in solutions,” *The journal of physical chemistry. A*, vol. 119, 10 2015.
- [8] H. Chen, H. Lin, H. Chen, F. Mai, Y. Liu, C. Lin, C. Chang, H. Tsai, and C. Yang, “Innovative strategy with potential to increase hemodialysis efficiency and safety,” *Scientific reports*, vol. 4, Mar. 2014, funding Information: The authors thank the National Science Council of ROC and Taipei Medical University for their financial support.
- [9] 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>
- [10] 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>
- [11] R. Claverie, M. D. Fontana, I. Đuričković, P. Bourson, M. Marchetti, and J.-M. Chassot, “Optical sensor for characterizing the phase transition in salted solutions,” *Sensors*, vol. 10, no. 4, pp. 3815–3823, 2010. [Online]. Available: <https://www.mdpi.com/1424-8220/10/4/3815>
- [12] Q. Hu and H. Zhao, “Understanding the effects of chlorine ion on water structure from a raman spectroscopic investigation up to 573k,” *Journal of Molecular Structure*, vol. 1182, pp. 191–196, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022286019300638>
- [13] H. Ge, Y. Zhao, H. Yang, and M. Wang, “Raman spectroscopy study for the systems (licl-h<sub>2</sub>o and licl-mgcl<sub>2</sub>-h<sub>2</sub>o): Excess spectra and hydration shell spectra,” *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 267, p. 120543, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1386142521011203>

- [14] J. A. Manrique, M. Veneranda, Y. Merino-Lomas, F. Rull, E. Charro, M. A. Gonzalez, J. M. Lopez, E. R. Gutiez, J. A. Sanz-Arranz, S. Maurice, and G. Lopez-Reyes, "Semi-quantification of binary saline solutions by raman spectroscopy: Implications for europa," *Journal of Chemometrics*, vol. 37, no. 9, p. e3440, 2023. [Online]. Available: <https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/cem.3440>
- [15] A. Broggi, E. Petrucci, M. P. Bracciale, and M. L. Santarelli, "Ft-raman spectroscopy for quantitative analysis of salt efflorescences," *Journal of Raman Spectroscopy*, vol. 43, no. 11, pp. 1560–1566, 2012. [Online]. Available: <https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/jrs.4153>
- [16] G. Rimsha, M. Shahbaz, M. I. Majeed, H. Nawaz, N. Rashid, M. W. Akram, I. Shabbir, K. Kainat, A. Amir, E. Sultan, M. Munir, and M. Imran, "Raman spectroscopy for the quantitative analysis of solid dosage forms of the active pharmaceutical ingredient of febuxostat," *ACS Omega*, vol. 8, no. 44, pp. 41 451–41 457, 2023. [Online]. Available: <https://doi.org/10.1021/acsomega.3c05243>
- [17] S. Almviva, F. Artuso, I. Giardina, A. Lai, and A. Pasquo, "Fast detection of different water contaminants by raman spectroscopy and surface-enhanced raman spectroscopy," *Sensors*, vol. 22, no. 21, 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/21/8338>
- [18] L. Luciani, A. Nocera, M. Raimondi, G. Ciattaglia, S. Spinsante, E. Gambi, and R. Galassi, "Raman spectroscopy for nitrate detection in water: A review of the current state of art," *ACS Measurement Science Au*, vol. 5, no. 4, pp. 443–460, 2025. [Online]. Available: <https://doi.org/10.1021/acsmesuresciau.5c00016>
- [19] C. J. Taylor, O. Kitzler, S. A. Curtis, B. R. Neimann, J. M. Dawes, J. E. Downes, D. J. Spence, and H. M. Pask, "Simultaneous measurement of water temperature and salinity using raman spectroscopy," *Opt. Continuum*, vol. 3, no. 6, pp. 909–920, Jun 2024. [Online]. Available: <https://opg.optica.org/optcon/abstract.cfm?URI=optcon-3-6-909>
- [20] T. Sugimura, S. Kajimoto, and T. Nakabayashi, "Label-free imaging of intracellular temperature by using the oh stretching raman band of water," *Angewandte*

- Chemie International Edition*, vol. 59, no. 20, pp. 7755–7760, 2020. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201915846>
- [21] C. Ma, L. Zhai, J. Ding, Y. Liu, S. Hu, T. Zhang, H. Tang, and H. Li, “Raman spectroscopy combined with partial least squares (pls) based on hybrid spectral preprocessing and backward interval pls (bipls) for quantitative analysis of four pahs in oil sludge,” *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 310, p. 123953, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1386142524001197>
- [22] 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>
- [23] 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>
- [24] Z.-P. Chen, J. Morris, and E. Martin, “Modelling temperature-induced spectral variations in chemical process monitoring,” *IFAC Proceedings Volumes*, vol. 37, no. 9, pp. 553–558, 2004, 7th IFAC Symposium on Dynamics and Control of Process Systems 2004 (DYCOPS -7), Cambridge, USA, 5-7 July, 2004. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1474667017318670>
- [25] D. Shin, J. Hwang, and W. Jhe, “Ice-vii-like molecular structure of ambient water nanomeniscus,” *Nature Communications*, vol. 10, 01 2019.
- [26] Y. Tian, X.-T. Liu, Y.-H. He, H.-F. Zhao, F. Jiang, D. Tan, and W. Xiao, “Raman evidences of chemical reaction of nacl-o<sub>2</sub> system at high pressure and high temperature,” *Gaoya Wuli Xuebao/Chinese Journal of High Pressure Physics*, vol. 31, pp. 692–697, 12 2017.
- [27] B. Journaux, A. Pakhomova, I. E. Collings, S. Petitgirard, T. B. Ballaran, J. M. Brown, S. D. Vance, S. Chariton, V. B. Prakapenka, D. Huang,

- 
- J. Ott, K. Glazyrin, G. Garbarino, D. Comboni, and M. Hanfland, “On the identification of hyperhydrated sodium chloride hydrates, stable at icy moon conditions,” *Proceedings of the National Academy of Sciences*, vol. 120, no. 9, p. e2217125120, 2023. [Online]. Available: <https://www.pnas.org/doi/abs/10.1073/pnas.2217125120>
- [28] R. Hamp, C. Salzmann, Z. Amato, M. Beaumont, H. Chinnery, P. Fawdon, T. Headen, P. Henry, L. Perera, S. Thompson, and M. Fox-Powell, “Metastable dihydrate of sodium chloride at ambient pressure,” *The Journal of Physical Chemistry Letters*, vol. 15, pp. 12 301–12 308, 12 2024.
- [29] M. A. Abdulsattar, A. H. Khalid, and Y. M. Abdul-Hussein, “Structural, vibrational and thermodynamic characteristics of nacl single-molecule dissolution in water: A dft study,” *Journal of Physics: Conference Series*, vol. 1818, no. 1, p. 012045, mar 2021. [Online]. Available: <https://doi.org/10.1088/1742-6596/1818/1/012045>
- [30] D. Hutsebaut, P. Vandenabeele, and L. Moens, “Evaluation of an accurate calibration and spectral standardization procedure for raman spectroscopy,” *Analyst*, vol. 130, pp. 1204–1214, 2005. [Online]. Available: <http://dx.doi.org/10.1039/B503624K>