



VNIVERSITAT
DE VALÈNCIA

FACULTAT DE QUÍMICA

Desarrollo de nuevas estrategias en cromatografía líquida para el análisis de muestras complejas

**Memoria para alcanzar el Grado de Doctor en Química dentro
del Programa de Doctorado en Química (RD 1999/2011)
presentada por:**

Adrián Gisbert Alonso

Directores:

Dr. José Ramón Torres Lapasió

Dra. María Celia García Álvarez-Coque

Burjassot, Valencia, Abril 2023.



D. JOSÉ RAMÓN TORRES LAPASÍO, Profesor Titular de Universidad, y Dña. MARÍA CELIA GARCÍA ÁLVAREZ COQUE, Catedrática de Universidad, adscritos al Departamento de Química Analítica de la Universidad de Valencia,

CERTIFICAN

Que la presente Memoria, “Desarrollo de nuevas estrategias en cromatografía líquida para el análisis de muestras complejas”, constituye la Tesis Doctoral de

D. ADRIÁN GISBERT ALONSO

Asimismo, certifican haber dirigido y supervisado tanto los distintos aspectos del trabajo, como su redacción.

Y para que conste a los efectos oportunos, firmamos la presente en Burjassot, a veinte de abril de dos mil veintitrés.

José Ramón
Torres Lapasíó

María Celia García
Álvarez-Coque

AGRADECIMIENTOS

Quisiera agradecer a todas aquellas personas que, directa o indirectamente, han contribuido a que se hiciera realidad la presente Memoria de Tesis Doctoral.

En primer lugar, deseo expresar mi agradecimiento al Grupo FUSCHROM (Estudios Fundamentales en Cromatografía) de la Universitat de València, por haber puesto a mi disposición todos los medios necesarios para llevar a cabo los estudios de investigación que dan lugar a los diferentes capítulos expuestos a continuación. En especial, a mis directores de Tesis Doctoral, los profesores José Ramón Torres Lapasió y María Celia García Álvarez-Coque. Mi relación con ellos comenzó durante la elaboración del Trabajo Final del Grado de Química. Tanto la calidad de la investigación como el trato recibido por ambos fue excelente, animándome así a continuar con ellos durante mis estudios en el Máster en Técnicas Experimentales en Química, de la Universitat de Valencia, y posteriormente, a iniciar los estudios de Doctorado que concluyen en esta Memoria de Tesis.

Ambos directores han mostrado siempre total disponibilidad para el correcto desarrollo de mi trabajo, resolviendo en todo momento mis dudas, ayudándome no sólo en la parte científica, sino también en la personal. Durante estos años, Celia me ha enseñado que el esfuerzo constante siempre tiene su recompensa. Ella ha sido un ejemplo a seguir por su duro trabajo, constancia, perseverancia y dedicación. José Ramón me ha enseñado todo lo que sé sobre programación de software de modelización y optimización. Destaco de él su capacidad para analizar y plantear las preguntas e hipótesis adecuadas para problemas complejos, encontrando soluciones prácticas, siempre con ideas brillantes.

A ambos les agradezco que hayan confiado en mí y les doy las gracias ya que, considero que el profesional que soy hoy se lo debo a ellos.

También quiero agradecer a mis compañeros de laboratorio y amigos: Ester, Nikita y José Antonio. Junto a ellos he pasado grandes momentos que jamás olvidaré. Ellos me han enseñado todo lo que sé acerca del manejo práctico de los cromatógrafos y que lograra utilizarlos de manera totalmente autónoma. Además, también me han transmitido la forma correcta de trabajar en un laboratorio de investigación, así como llevar a cabo su gestión.

Por último, me gustaría dedicar también unas palabras a todas aquellas personas que, a pesar de no haber participado en el desarrollo de mi trabajo de investigación, han sido un pilar fundamental durante este periodo, y mi vida.

En primer lugar, a mis padres, Gerson y Conchi, porque sin su constante esfuerzo y confianza no habría sido posible disponer del tiempo y los medios necesarios para la elaboración de este trabajo de investigación. Junto a ellos, mi hermano Alejandro, su esposa Helen y desde hace 2 años mi sobrina Luna, siempre me han animado a seguir adelante, dándome su apoyo y cariño. Especial agradecimiento se merecen mis amigos Alicia e Israel. Alicia entró a mi vida el primer año de carrera, hace 10 años, mientras que Israel vino dos años más tarde cuando coincidí con él en un piso de estudiantes, haciendo que fueran los mejores años de mi vida. Ellos siempre han estado en los peores momentos cuando todo parecía imposible y gracias a ellos siempre ha ido todo bien. Finalmente, deseo agradecer a mis compañeros residentes de Bioquímica Clínica del Hospital Universitari de Bellvitge, Teresa, Dani, María, y especialmente a David y Lidia, su increíble apoyo durante la última etapa de elaboración de esta Memoria.

A todos ellos, gracias.

ABREVIATURAS

BEADS: *Baseline Estimation and Denoising using Sparsity*

Camellia sinensis: Té verde

df: *degrees of freedom*; grados de libertad

DOE: Design of experiments; diseño de experimentos

ER: Error relativo

F: Coeficiente F de Snedecor

$F(x)$: Función de una variable

FDA: *Food and Drug Administration*; Administración de Drogas y Alimentos (EE.UU.)

$h(t)$: Altura del pico cromatográfico en el tiempo t

h_0 : Altura máxima del pico

HPLC: *High performance liquid chromatography*; cromatografía líquida de alto rendimiento

$I(t)$: Función primitiva

in silico: Experimentos por ordenador mediante algoritmos de simulación y predicción computacional

\mathbf{j}_{pred} : Vector jacobiano de predicción

\mathbf{J}_{pred} : Matriz jacobiana del diseño de predicción

$\mathbf{J}_{\text{train}}$: Matriz jacobiana del diseño de entrenamiento

$\mathbf{J}_{\text{train}}^T$: Matriz jacobiana traspuesta del diseño de entrenamiento

k : Factor de retención

\bar{k}_{exp} : Factor de retención predicho promedio

$k_{\text{exp},i}$: Factor de retención experimental para el experimento i

\hat{k}_i : Factor de retención predicho para el experimento i

\hat{k}_{medio} : Factor de retención experimental promedio

LC: *Liquid chromatography*; cromatografía líquida

$\log k_w$: Factor de retención para una fase móvil constituida por agua pura

$\log k_{w,i}$: Factores de retención en fase acuosa específicos para cada soluto, relacionados con su hidrofobicidad

$(\ln k)_0$: retención en una fase móvil hipotética donde la polaridad de las fases móvil y estacionaria es la misma

LOF: *Lack of fit*; falta de ajuste

LSS: *Linear solvent strength*; Fuerza lineal del disolvente

m : Pendiente del gradiente de elución

M: Concentración molar (mol/L)

Melissa officinalis: Melisa

n_e : Número de fases móviles en el diseño experimental

n_p : Número de parámetros del modelo

n_s : Número de compuestos de referencia

n_g : Número de gradientes en el diseño experimental

NK: Neue-Kuss

NPLC: *Normal-phase liquid chromatography*; cromatografía líquida en fase normal

OMS: Organización Mundial de la Salud

$P_{o/w}$: Coeficiente de reparto octanol-agua

p_s : descriptor de la polaridad del soluto en el modelo de Bosch-Rosés

P_M^N : Descriptor de la polaridad de la fase móvil en el modelo de Bosch-Rosés

P_S^N : Descriptor de la polaridad de la fase estacionaria en el modelo de Bosch-Rosés

pK_a : Constante ácido-base de disociación

PMG1: Modelo gaussiano modificado

QbD: *Quality-by-Design*

R : Coeficiente de correlación

R_{Adj} : Coeficiente de correlación ajustado

RPLC: *Reversed-phase liquid chromatography*; cromatografía líquida en fase inversa

S : Fuerza eluyente

s_0 : Medida de la anchura de pico cromatográfico establecida sobre una base gaussiana

s_1 : Distorsión del pico cromatográfico

S_i : Fuerza eluyente para un soluto concreto

S_m : Mediana de los valores de S_i

s_{pred} : Incertidumbre de predicción

s_{x_i} : Incertidumbre de una variable

s_F^2 : Varianza de la función

s_{PE}^2 : Cuadrado del error experimental puro

$s_{x_i}^2$: Varianza de la variable i

SEP: *Standard error in predictions*; error standard de predicción

SSR: *Sum of squared residuals*; suma del cuadrado de los residuales

t : Tiempo

T : Corrección de la desviación de la linealidad de $\ln k$ vs. φ

t_0 : Tiempo muerto

t_D : *Dwell time*; tiempo de retardo

t_{ext} : Tiempo extra-columnar

t_{final} : Tiempo al finalizar el pulso

\hat{t}_g : Tiempo de retención previsto para el gradiente

t_G : Tiempo que tarda el gradiente en completarse

$t_{g,i}$: Tiempo de retención experimental en gradiente

$\hat{t}_{g,i}$: Tiempo de retención predicho en gradiente

t_{iso} : Tiempo de retención isocrático en el instante en que el soluto abandona la columna, para una fase móvil de composición φ_j

T_m : Mediana de los valores de T_i

t_p : Tiempo medido al inicio del pulso programado

t_R : Tiempo de retención

Tilia platyphyllos: Tila

UV: Ultravioleta

v/v : Volumen/volumen

w : Anchura de pico o semianchuras izquierda o derecha

x_i : Variable

φ : Concentración de modificador orgánico

φ_j : Tiempo isocrático equivalente

φ_{final} : Concentración de modificador orgánico al finalizar el gradiente

φ_{max} : Concentración de modificador orgánico máximo en el gradiente

ÍNDICE

OBJETIVOS Y DESARROLLO DE LAS INVESTIGACIONES	1
CAPÍTULO 1: Introducción	13
1.1. Cromatografía líquida de fase inversa	15
1.1.1. Ámbito de aplicación	15
1.1.2. La fase estacionaria	16
1.1.3. Factores de la fase móvil que afectan a la retención	17
1.2. Elución isocrática y en gradiente	18
1.2.1. Elución isocrática	18
1.2.2. Elución en gradiente	19
1.3. Tiempo de retraso y tiempo muerto	21
1.4. Optimización de la resolución cromatográfica	22
1.4.1. Modelización de la retención	23
1.4.2. Diseño experimental y seguimiento de picos	25
1.4.3. Modelización del perfil de los picos	27
1.4.4. Optimización interpretativa asistida por ordenador	28
1.5. Referencias	32

PARTE 1: EVALUACIÓN DEL RENDIMIENTO PREDICTIVO

DE DIVERSOS DISEÑOS EXPERIMENTALES39

CAPÍTULO 2: Beneficios del uso de pulsos de concentración de modificador orgánico en la modelización del tiempo de retención en cromatografía líquida.....41

2.1. Resumen43

2.2. Introducción44

2.3. Teoría47

2.3.1. Modelos de retención47

2.3.2. Predicción de tiempos de retención en elución en gradiente.....48

2.3.3. Corrección de las desviaciones en la retención asociadas a retrasos de tiempo cuando se utiliza integración numérica en elución en gradiente51

2.3.4. Ajuste del modelo de retención y predicciones de tiempo de retención para experimentos no isocráticos53

2.4. Parte experimental54

2.4.1. Reactivos54

2.4.2. Instrumentación y columna cromatográfica55

2.4.3. Software56

2.5. Resultados y discusión56

2.5.1. Selección del modelo de retención56

2.5.2. Efecto de los pulsos de modificador orgánico sobre la retención y la eficacia60

2.5.3. Modelización de tiempos de retención.....	66
2.5.3.1. Desviación de las predicciones respecto a los cromatogramas experimentales	66
2.5.3.2. Corrección de tiempos de retención en la integración numérica debido a la migración de los solutos en el interior de la columna	71
2.5.4. Calidad en la predicción de los tiempos de retención para diseños experimentales que involucran pulsos	79
2.5.4.1. Validación y diseños experimentales utilizados	80
2.5.4.2. Metodología	86
2.5.4.3. Comparación del rendimiento obtenido con los distintos diseños	87
2.6. Conclusiones	89
2.7. Referencias	93

CAPÍTULO 3: Diseños experimentales en cromatografía líquida:

Desarrollo y validación de un método para la inspección integral de diseños experimentales	97
3.1. Resumen	99
3.2. Introducción	100
3.3. Teoría	104
3.3.1. Predicción de tiempos de retención	104
3.3.2. Construcción de mapas de error en elución isocrática y de gradiente	107
3.3.2.1. Cálculo de incertidumbres	108

3.3.2.2. Mapas de incertidumbre relativa asociados a un diseño experimental de entrenamiento	112
3.4. Parte experimental	114
3.4.1. Reactivos	114
3.4.2. Instrumentación y columna	114
3.4.3. Software	115
3.5. Resultados y discusión	116
3.5.1. Diseños evaluados	116
3.5.2. Modelización de la retención	117
3.5.3. Mapas de incertidumbre relativa asociados a un diseño experimental de entrenamiento	123
3.5.3.1. Cálculo de las matrices Jacobianas	123
3.5.3.2. Construcción de los mapas de incertidumbre	125
3.5.4. Evaluación de los diseños	130
3.5.4.1. Comparación de las predicciones isocráticas frente a las predicciones en gradiente	131
3.5.4.2. Predicciones isocráticas	134
3.5.4.3. Predicciones en gradiente	135
3.5.5. Comparación del uso de los modelos de Neue-Kuss y de Snyder	136
3.6. Conclusiones	142
3.7. Referencias	145

CAPÍTULO 4: Diseños experimentales en cromatografía líquida: Influencia de la geometría del diseño sobre la capacidad predictiva de los modelos de retención	151
4.1. Resumen	153
4.2. Introducción	154
4.3. Parte experimental	157
4.3.1. Reactivos y experimentos isocráticos de referencia	157
4.3.2. Diseños de entrenamiento	158
4.3.3. Diseños de muestreo	166
4.4. Resultados y discusión	166
4.4.1. Gráficos de incertidumbre relativa en las predicciones	167
4.4.1.1. Diagramas de incertidumbre relativa de caja y bigotes para cada diseño	168
4.4.1.2. Diagramas triples de incertidumbre relativa que consideran cada diseño, soluto y composición	169
4.4.2. Análisis de la capacidad predictiva de los diseños de entrenamiento	175
4.4.2.1. Diseños isocráticos: ISO1–ISO5	180
4.4.2.2. Diseños isocráticos con pulsos: P1–P7	181
4.4.2.3. Diseños de gradiente con una única rampa y la misma composición inicial: SL1–SL6	184
4.4.2.4. Diseños de gradiente con una única rampa y diferente composición inicial: ML1–ML6.....	186

4.4.2.5. Diseños de gradiente con una única rampa en la que se modifican ambas composiciones inicial y final: ML7–ML13.....	188
4.4.2.6. Diseños que contienen gradientes multi-lineales: N1–N7	190
4.4.3. Comparación de los mejores diseños encontrados en cada grupo	192
4.5. Conclusiones	197
4.6. Referencias	201

PARTE 2: DESARROLLO DE MODELOS GLOBALES DE RETENCIÓN PARA SU APLICACIÓN A MUESTRAS COMPLEJAS	205
---	------------

CAPÍTULO 5: Desarrollo de metodologías para obtener modelos globales de retención que describan grupos de compuestos	207
5.1. Resumen	209
5.2. Introducción	210
5.3. Teoría	213
5.3.1. Ecuaciones para describir la retención	213
5.3.1.1. Modelo logarítmico lineal de Snyder	213
5.3.1.2. Modelo logarítmico cuadrático de Schoenmakers	214
5.3.1.3. Modelo de Bosch Rosés	214
5.3.1.4. Modelo de Neue-Kuss	215
5.3.1.5. Modelo de Jandera	216

5.3.2. Modelos globales basados en la correlación de los parámetros de los modelos para los solutos individuales	216
5.4. Parte experimental	217
5.4.1. Análisis de muestras de manzanilla	217
5.4.2. Conjunto de 15 sulfonamidas.....	220
5.4.3. Conjunto de 145 compuestos de diversa naturaleza.....	220
5.5. Resultados y discusión	221
5.5.1. Construcción de modelos de retención individuales para muestras de manzanilla	224
5.5.1.1. Selección de picos de referencia	224
5.5.1.2. Ajuste de modelos de retención individuales para los picos de referencia	226
5.5.1.3. Construcción de modelos globales basados en correlaciones entre los parámetros para todo el conjunto de picos de referencia	229
5.5.2. Desarrollo de un nuevo enfoque para construir modelos globales de retención para los picos de referencia de la manzanilla.....	235
5.5.2.1. Fijación de parámetros para desarrollar un modelo global	236
5.5.2.2. Regresión alterna basada en la ecuación de Schoenmakers	239
5.5.2.3. Regresión alterna basada en la ecuación de Snyder	241
5.5.3. Aplicación de los modelos globales a grandes conjuntos de solutos conocidos de diversa naturaleza eluidos isocráticamente	242
5.5.4. Separación de las contribuciones polares con el modelo global de Bosch-Rosés y predicciones de $\log P_{o/w}$	246

5.5.5. Simulación de huellas dactilares cromatográficas utilizando modelos globales	250
5.6. Conclusiones	255
5.7. Referencias	260

CAPÍTULO 6: Análisis de huellas dactilares cromatográficas

**de extractos de té verde, melisa y tila: Desarrollo de
modelos globales de retención sin el uso de estándares.....**

6.1. Resumen	269
6.2. Introducción	270
6.3. Teoría	273
6.3.1. Modelos individuales de retención.....	273
6.3.2. Modelos globales de retención.....	275
6.3.3. Regresión alterna para ajustar modelos globales	276
6.4. Parte experimental	278
6.5. Resultados y discusión	279
6.5.1. Plantas medicinales y gradientes de exploración	279
6.5.2. Estudio preliminar haciendo uso de un diseño experimental de gradientes no relacionados	284
6.5.3. Desarrollo de diseños de entrenamiento en forma de diamante que hacen uso de gradientes	286
6.5.3.1. Té verde	289
6.5.3.2. Melisa	290
6.5.3.3. Tila	291
6.5.4. Selección de los picos de los compuestos de referencia	292

6.5.5. Descripción del comportamiento de retención para los compuestos de referencia	295
6.5.5.1. Modelos individuales de retención	295
6.5.5.2. Modelos globales de retención.....	303
6.6. Conclusiones	311
6.7. Referencias	314

**CAPÍTULO 7: Análisis de huellas dactilares cromatográficas de extractos
de té verde, melisa y tila: Simulación de cromatogramas**

 haciendo uso de modelos globales de retención	319
7.1. Resumen	321
7.2. Introducción	322
7.3. Teoría	324
7.3.1. Modelos globales de retención para los compuestos de referencia	324
7.3.2. Extensión de los modelos globales de retención a la totalidad de los picos detectados en las plantas medicinales	326
7.3.3. Modelos globales de anchura de pico para los compuestos de referencia	329
7.4. Parte experimental	332
7.4.1. Preparación de extractos de plantas medicinales	332
7.4.2. Separación cromatográfica	333
7.4.3. Software	336
7.5. Resultados y discusión	336
7.5.1. Medida de las señales cromatográficas	336

7.5.2. Construcción de modelos globales de anchura de pico para simular cromatogramas	339
7.5.3. Algunos factores que afectan a la simulación de cromatogramas basados en modelos globales	345
7.5.3.1. Influencia de la elección del cromatograma base en las predicciones	345
7.5.3.2. Predicción de señales no asociadas a solutos retenidos	348
7.5.3.3. Picos con anchuras anormales	349
7.5.4. Validación de cromatogramas obtenidos con gradientes multi-lineales externos	350
7.6. Conclusiones	353
7.7. Referencias	355
RESUMEN Y CONCLUSIONES	361
C.1. Evaluación del rendimiento predictivo de diversos diseños experimentales	365
C.1.1. Capítulo 2: Beneficios del uso de pulsos de concentración de modificador orgánico en la modelización del tiempo de retención en cromatografía líquida.....	366
C.1.2. Capítulo 3: Diseños experimentales en cromatografía líquida: Desarrollo y validación de un método para la inspección integral de diseños experimentales.....	369

C.1.3. Capítulo 4: Diseños experimentales en cromatografía líquida: Influencia de la geometría del diseño en la capacidad predictiva de los modelos de retención.....	373
C.2. Desarrollo de modelos globales de retención para su aplicación a muestras complejas	377
C.2.1. Capítulo 5: Desarrollo de metodologías para obtener modelos globales de retención que describan grupos de compuestos	378
C.2.2. Capítulo 6: Análisis de huellas dactilares cromatográficas de extractos de té verde, melisa y tila: Desarrollo de modelos globales de retención sin el uso de estándares	382
C.2.3. Capítulo 7: Análisis de huellas dactilares cromatográficas de extractos de té verde, melisa y tila: Simulación de cromatogramas haciendo uso de modelos globales de retención	386
 CONTRIBUCIÓN DEL TRABAJO DE DOCTORADO A LAS PUBLICACIONES	 391

OBJETIVOS Y DESARROLLO DE LAS INVESTIGACIONES

Desde los años 70, el uso de la cromatografía líquida de alta resolución (HPLC, *High Performance Liquid chromatography*) ha ido creciendo en los laboratorios analíticos, siendo actualmente la técnica de separación analítica más utilizada. Desafortunadamente, la eficacia suele ser menor que la que se logra en otras técnicas de separación como la cromatografía de gases y la electroforesis capilar. Esto restringe el análisis de muestras complejas. De ahí el gran esfuerzo dedicado a aumentar la eficacia y selectividad en HPLC, con importantes mejoras desde el inicio de la técnica. A pesar de los avances en las últimas décadas, con el desarrollo de instrumentación cada vez más sofisticada, y en la tecnología de columnas (síntesis de nuevos soportes y fases estacionarias), aún existen desafíos por resolver. Además, la necesidad de proporcionar datos informativos de calidad, con un mínimo esfuerzo experimental, es imperativa en áreas científicas donde la adquisición de datos es lenta o laboriosa, como es el caso de la cromatografía líquida.

El trabajo descrito en esta Memoria de Tesis Doctoral recoge algunas propuestas de dos tipos de estrategias para mejorar el rendimiento en HPLC. A continuación se resumen brevemente los objetivos, que en definitiva son: uso de diseños experimentales óptimos y de modelos globales de retención en cromatografía líquida:

Objetivo 1. *Evaluación de la capacidad predictiva de diseños experimentales inusuales*

La obtención de modelos de retención de calidad requiere datos experimentales de calidad, para lo que resulta imprescindible una planificación y ejecución de conjuntos de experimentos perfectamente controlados. La información aportada por un diseño es máxima cuando éste incluye datos

isocráticos, distribuidos de acuerdo a los requisitos de cada soluto. Lamentablemente, las experiencias isocráticas son también las más ineficientes en cuanto a tiempo de adquisición. Por esta razón, y a pesar de ser menos informativos, los experimentos de gradiente se consideran los más prácticos. Los diseños de entrenamiento que hacen uso de gradientes proporcionan también la información necesaria para realizar el ajuste de modelos, aunque no es sencillo encontrar un conjunto de gradientes óptimo para la modelización. En esta Tesis Doctoral, se presentan varios estudios que profundizan en el análisis de diseños experimentales para mejorar la capacidad de predicción, así como el ensayo de diversos diseños experimentales inusuales.

Se han abordado los siguientes tres objetivos concretos:

Objetivo 1.1: Analizar la capacidad predictiva de diversos diseños experimentales y modelos de retención, a partir de la calidad de la información proporcionada por experimentos isocráticos, de gradiente y mixtos. Proponer el uso de experiencias isocráticas que incluyen incrementos transitorios de modificador orgánico (pulsos) con fines de modelización.

Objetivo 1.2: Desarrollar una metodología sistemática basada en la propagación de errores, para la evaluación en profundidad de la capacidad predictiva de diseños experimentales de un factor. La estrategia a desarrollar debe ser adecuada para la evaluación del rendimiento de cualquier perfil de elución, incluyendo configuraciones inusuales, no explotadas en HPLC.

Objetivo 1.3: Estudiar los efectos asociados a cambios regulares en los diseños más usuales en HPLC isocráticos y de gradiente de modificador orgánico, para la deducción de los principios generales que favorecen la calidad de la información en los diseños, y relacionarlos con las propiedades de los solutos, particularmente con su hidrofobicidad.

Objetivo 2. *Desarrollo de nuevas estrategias para la obtención de huellas dactilares cromatográficas más informativas*

La resolución de muestras complejas de productos naturales que contienen un gran número de compuestos desconocidos de distinta naturaleza, o para los que no existe disponibilidad de patrones (se trate de compuestos conocidos o desconocidos), sigue siendo un reto. En este caso, una estrategia de análisis cada vez más frecuente se basa en la obtención de las denominadas huellas dactilares cromatográficas. Por lo general, las muestras con huellas dactilares similares también poseen propiedades similares. Por lo tanto, el cromatograma obtenido tiene un potencial interés como herramienta válida en el control de calidad de este tipo de muestras. Sin embargo, para el uso exitoso de las huellas dactilares, aún se necesitan estrategias que permitan obtener cromatogramas más ricos en información.

Se han abordado los dos objetivos siguientes:

Objetivo 2.1: Obtener un conjunto de parámetros que permitan modelizar el comportamiento de un grupo de compuestos seleccionados (conocidos o desconocidos) de una huella dactilar cromatográfica, como alternativa al uso de parámetros enfocados a cada compuesto.

Objetivo 2.2: Extender el modelo de retención, obtenido a partir de compuestos seleccionados, a la predicción del cromatograma completo de la huella dactilar bajo cualquier nueva condición de elución arbitraria.

De acuerdo a los objetivos propuestos, los contenidos de esta Tesis Doctoral están divididos en dos partes:

Parte 1. *Evaluación del rendimiento predictivo de diversos diseños experimentales*

Parte 2. *Desarrollo de modelos globales de retención para su aplicación a muestras complejas*

El trabajo ha supuesto un gran esfuerzo experimental, destinado a explorar y extraer información sobre el comportamiento cromatográfico de compuestos de distinta naturaleza, así como de muestras de diferentes plantas medicinales. A lo largo de los estudios se ha utilizado un gran número de fases móviles acuo-orgánicas con acetonitrilo de diverso contenido. El trabajo ha implicado un extenso tratamiento de datos, principalmente relacionado con el ajuste de modelos que permitan predecir la retención cromatográfica, así como los perfiles de los picos en los cromatogramas. Cabe destacar también el gran esfuerzo realizado en las diferentes búsquedas bibliográficas realizadas sobre los temas investigados en este trabajo. Ello ha implicado la búsqueda, lectura y organización de una gran cantidad de información que se refleja adecuadamente en cada parte de la Memoria de Tesis Doctoral.

Supervisores

El trabajo de investigación conducente al Grado de Doctor en Química se inició en noviembre de 2019, una vez finalizado el Máster en “Técnicas Experimentales en Química”, impartido por los Departamentos de Química Analítica y Química Inorgánica de la Universitat de València. El trabajo experimental recogido en esta Memoria se ha desarrollado en el Departamento

de Química Analítica de la Universitat de València, bajo la dirección de José Ramón Torres Lapasió y María Celia García Álvarez-Coque.

Publicaciones

A continuación se muestran las publicaciones incluidas en la Memoria de Tesis Doctoral (se proporciona el factor de impacto de la revista y el ranking en la categoría de Química Analítica):

1. José Antonio Navarro Huerta, Adrián Gisbert Alonso, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
Benefits of solvent concentration pulses in retention time modelling of liquid chromatography.
Journal of Chromatography A 1597 (2019) 76–88 (Capítulo 2).
IF (2019): 4.21 (Química Analítica: Q1)
2. José Antonio Navarro Huerta, Adrián Gisbert Alonso, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
Testing experimental designs in liquid chromatography (I): Development and validation of a method for the comprehensive inspection of experimental designs.
Journal of Chromatography A 1624 (2020) 461180 (Capítulo 3).
IF (2020): 4.43 (Química Analítica: Q1)
3. Adrián Gisbert Alonso, José Antonio Navarro Huerta, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
Testing experimental designs in liquid chromatography (II): Influence of the design geometry on the prediction performance of retention models.
Journal of Chromatography A 1654 (2021) 462458 (Capítulo 4).
IF (2021): 4.17 (Química Analítica: Q1)

4. Adrián Gisbert Alonso, José Antonio Navarro Huerta, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
Global retention models and their application to the prediction of chromatographic fingerprints.
Journal of Chromatography A 1637 (2021) 461845 (Capítulo 5).
IF (2021): 4.17 (Química Analítica: Q1)
5. Adrián Gisbert Alonso, Sergio López Ureña, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
Chromatographic fingerprint-based analysis of extracts of green tea, lemon balm and linden: I. Development of global models without the use of standards.
Journal of Chromatography A 1672 (2022) 463060 (Capítulo 6).
IF (2021): 4.17 (Química Analítica: Q1)
6. Adrián Gisbert Alonso, Álvaro Navarro Martínez, José Antonio Navarro Huerta, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
Chromatographic fingerprint-based analysis of extracts of green tea, lemon balm and linden: II. Simulation of chromatograms using global models.
Journal of Chromatography A 1684 (2022) 463561 (Capítulo 7).
IF (2021): 4.17 (Química Analítica: Q1)

Comunicaciones a congresos

La investigación desarrollada también ha sido presentada en 5 congresos científicos en el periodo de junio de 2018 a noviembre de 2021 (4 comunicaciones en congresos internacionales y 2 comunicaciones en congresos nacionales, 5 fueron pósters y una se presentó como comunicación oral):

24th International Symposium on Separation Sciences (ISSS'2018)

Jasná (Eslovaquia), Junio de 2018 (internacional)

1. José Antonio Navarro Huerta, Adrián Gisbert Alonso, José Ramón Torres Lapasió, María Celia García Álvarez-Coque

Prediction of peak width and asymmetry in linear and multi-linear gradient elution: Jandera's approximation versus numerical integration

(Póster P058)

XVIII Reunión de la Sociedad Española de Cromatografía y Técnicas Afines (SECyTA'2018)

Granada (España), Octubre de 2018 (nacional)

2. José Ramón Torres Lapasió, Adrián Gisbert Alonso, José Antonio Navarro Huerta, María Celia García Álvarez-Coque

Pulses of organic solvent and their consequences to enhance isocratic and gradient predictions

(Póster P-FCH-03)

48th International Symposium on High Performance Liquid Phase Separations and Related Techniques (HPLC'2019)

Milán (Italia), Junio de 2019 (internacional)

3. José Ramón Torres Lapasió, José Antonio Navarro Huerta, Adrián Gisbert Alonso, María Celia García Álvarez-Coque
Solvent concentration pulses to improve retention modelling in liquid chromatography
(Póster P475)

27th International Symposium on Electroseparation and Liquid Phase-Separation Techniques (ITP'2020), Virtual Conference

Nanjing (China), Noviembre de 2020 (internacional)

4. José Ramón Torres Lapasió, Adrián Gisbert Alonso, José Antonio Navarro Huerta, María Celia García Álvarez-Coque
Global retention models in HPLC for prediction of fingerprints under gradient elution
(Póster)
5. José Ramón Torres Lapasió, Adrián Gisbert Alonso, José Antonio Navarro Huerta, María Celia García Álvarez-Coque
Influence of the design geometry on the predictive capability of training gradient experiments
(Póster)

XX Reunión de la Sociedad Española de Cromatografía y Técnicas Afines (SECyTA 2021), Virtual edition (España), 18–19 de Noviembre de 2021 (nacional)

6. María Celia García Álvarez-Coque, José Ramón Torres Lapasió, Adrián Gisbert Alonso, José Antonio Navarro Huerta

Development and validation of a method for the comprehensive inspection of experimental designs in HPLC

(Comunicación oral O-CF2)

Proyectos que han financiado las investigaciones

La investigación desarrollada a lo largo de la Tesis Doctoral ha sido financiada por dos Proyectos de Investigación nacionales:

1. Proyecto CTQ2016-75644-P: “*Diseño de metodologías para optimizar la calidad separativa en cromatografía líquida*”, subvencionado por el Ministerio de Economía, Industria y Competitividad. Periodo: enero de 2017 a diciembre de 2019. Investigadores responsables: María Celia García Álvarez-Coque y José Ramón Torres Lapasió.
2. Proyecto PID2019-106708GB-I00: “*Estudios fundamentales y diseño de estrategias en cromatografía mono- y bi-dimensional para potenciar la calidad de las separaciones de muestras complejas*”, subvencionado por el Ministerio de Ciencia, Innovación y Universidades. Periodo: enero de 2020 a diciembre de 2022. Investigadores responsables: María Celia García Álvarez-Coque y José Ramón Torres Lapasió.

CAPÍTULO 1

INTRODUCCIÓN

1.1. Cromatografía líquida de fase inversa

1.1.1. *Ámbito de aplicación*

En cromatografía líquida (LC, *liquid chromatography*), la retención depende de la naturaleza de la fase estacionaria, la naturaleza de los solutos, la naturaleza y composición de la fase móvil, y de otros factores como la geometría del sistema de separación, el caudal y la temperatura [1]. La cromatografía líquida de fase inversa (RPLC, *reversed-phase liquid chromatography*) es un modo de LC que emplea una fase estacionaria no polar o débilmente polar (generalmente una cadena hidrocarbonada unida químicamente a partículas de sílice porosa) y una fase móvil polar constituida por agua, a la que se añade un disolvente orgánico miscible con agua para reducir la polaridad, y que actúa como modificador [2-6]. Puesto que la polaridad de la mezcla formada en la fase móvil se parece más a la de la fase estacionaria, puede competir mejor por la desorción de los solutos apolares, lo que tiene una influencia considerable en la separación de la mezcla de solutos inyectada.

La RPLC es una técnica madura, empleada en ciencia y tecnología por químicos, bioquímicos y farmacéuticos en los campos del análisis y la purificación de sustancias. Las aplicaciones de la RPLC cubren el control ambiental, el análisis de alimentos, los análisis clínicos, el control farmacéutico e industrial, la fabricación de fármacos y de productos químicos (tanto en el campo del control de calidad analítico, como en el ámbito preparativo). Se utiliza, además, en estudios biomédicos y en la medida de propiedades físico-químicas.

Uno de los principales factores responsables del desarrollo de la RPLC fue la necesidad de separar mezclas que contenían compuestos que no eran lo

suficientemente volátiles, o poseían una estabilidad térmica insuficiente para ser analizados por cromatografía de gases, como es el caso de los compuestos iónicos orgánicos e inorgánicos, los fármacos y las biomoléculas. La retención en RPLC está relacionada principalmente con la polaridad de los solutos. Por lo tanto, el coeficiente de reparto 1-octanol/agua ($\log P_{o/w}$) de los solutos analizados se puede utilizar para predecir la retención. La RPLC es la técnica de elección para la separación de mezclas complejas de analitos dentro de una amplia gama de polaridades.

1.1.2. La fase estacionaria

En el desarrollo de métodos en LC, la selección de una columna adecuada constituye la decisión más importante. La columna puede elegirse (a menudo con éxito) entre las disponibles en el laboratorio atendiendo a la experiencia previa, a datos tomados de la bibliografía, o a las recomendaciones obtenidas de los fabricantes. En RPLC, las fases de octadecil-sílice (C18) son, con mucho, las más comunes. Las fases de octil-(C8) o butil-sílice (C4) reducen la retención, obteniéndose el efecto inverso con triacontil-sílice (C30). Las fases de alquil-fenilo, alquil-difenilo y alquilfluorofenilo ofrecen diferentes selectividades, debido a la posibilidad de interacciones dipolo inducidas, mientras que las interacciones dipolo-dipolo permanentes son apropiadas para analizar solutos polares mediante las fases alquil-ciano [7,8].

La selectividad de un sistema cromatográfico particular depende de un número importante de interacciones intermoleculares entre los solutos y las fases estacionaria y móvil [9,10]. Las principales propiedades de la fase estacionaria que determinan el comportamiento de retención de los solutos son la polaridad (que depende del área superficial y la densidad de enlace), el número de grupos silanol residuales accesibles con diferente acidez (capaces de

interactuar con los solutos neutros por puentes de hidrógeno o con la forma iónica de los solutos básicos), y la selectividad relacionada con la forma de la molécula analizada (selectividad estérica).

1.1.3. Factores de la fase móvil que afectan a la retención

Un enfoque común en el desarrollo de métodos es mantener constante la naturaleza de la fase estacionaria, procediendo así a ajustar en primer lugar la fuerza eluyente de la fase móvil y, a continuación, su selectividad. En principio, existe una amplia gama de disolventes orgánicos miscibles con agua para ser utilizados como modificadores de la fase móvil; sin embargo, sólo unos pocos son habituales en RPLC: acetonitrilo, metanol, etanol (considerado un disolvente verde), tetrahidrofurano e isopropanol [11,12]. Entre ellos, el acetonitrilo seguido del metanol es, con diferencia, el más utilizado. Las razones que hacen que el acetonitrilo sea la opción más popular como modificador son su menor viscosidad, que reduce la presión de la bomba del cromatógrafo, el valor de corte más bajo para la longitud de onda de detección UV, su fuerza eluyente suficientemente grande, su escasa reactividad y la capacidad de disolver una amplia gama de compuestos. Por su parte, el metanol es menos costoso y menos tóxico, y su mayor polaridad reduce el riesgo de precipitación del tampón.

Además del tipo y contenido de disolvente orgánico en la fase móvil, otros factores experimentales pueden afectar a la separación, tales como la presencia de otros reactivos (aditivos o tampones), el pH de la fase móvil (para muestras que contienen compuestos ionizables), el caudal de la fase móvil y la temperatura (cuando la naturaleza del empaquetado permite valores elevados), así como la fuerza iónica.

1.2. Elución isocrática y en gradiente

Las separaciones se pueden llevar a cabo en los modos de elución isocrática o en gradiente.

1.2.1. Elución isocrática

En elución isocrática, la composición de la fase móvil se mantiene constante (por ejemplo, 25 % *v/v* de metanol) durante la separación, mientras que en elución en gradiente varía (por ejemplo, cambiando progresivamente del 10 al 70 % *v/v* de metanol). Si la polaridad de los compuestos en la muestra es similar, la elución isocrática es apropiada. Si abarca un amplio intervalo, es necesario utilizar elución en gradiente.

La utilidad práctica de la RPLC se ve seriamente restringida si la polaridad de los compuestos de la muestra es similar, mientras que si el intervalo de polaridad es demasiado amplio, será difícil encontrar un conjunto de condiciones cromatográficas capaces de equilibrar un poder de separación satisfactorio para los solutos menos retenidos y una retención razonable para los más retenidos. Ello puede dar lugar a la imposibilidad práctica de resolver mezclas que contengan más de 15 a 20 compuestos mediante elución isocrática. Además, una fase móvil con composición fija a menudo no es adecuada para resolver muestras complejas con relaciones por debajo de ~20 entre los valores de factor de retención, k , para el último y el primer pico en un cromatograma:

$$\log k = \log \frac{t_R - t_0}{t_0} \quad (1.1)$$

donde k es el factor de retención (o retención relativa), en ausencia de contribuciones extra-columnares, t_R el tiempo de retención y t_0 el tiempo muerto (tiempo al que eluye un soluto no retenido). En elución isocrática, una

disminución en el porcentaje constante del modificador da como resultado un tiempo de retención mayor, con mejoras en la resolución.

La elución isocrática presenta varias ventajas, como su mayor simplicidad, menor coste, mayor robustez, el hecho de que requiere una instrumentación más simple y no es necesario reequilibrar la columna entre inyecciones consecutivas (como ocurre en elución en gradiente) [13]. Por lo tanto, muchas separaciones en RPLC se llevan a cabo en el modo isocrático. Este enfoque es adecuado cuando el cromatograma contiene el pico de un único analito, o picos de dos o más analitos, dentro de un intervalo de polaridad pequeño o moderado. En estos casos, es posible separar con cierta facilidad todos los solutos de la muestra.

1.2.2. Elución en gradiente

La elución isocrática no es recomendable si la polaridad de los solutos en la muestra se extiende en un amplio intervalo. Una solución lógica para superar esta situación es la elución en gradiente, en la que la fuerza eluyente aumenta gradualmente alterando al menos un factor experimental, a medida que avanza el análisis, de modo que se acelera la elución de los compuestos más retenidos [14-17]. Los picos se producen en el mismo orden en ambos modos de elución isocrático y de gradiente. Además, en ambos casos, el ensanchamiento de los picos aumenta con el tiempo de retención, por lo que los picos menos retenidos son los más estrechos. En ambos casos, la retención se controla variando la composición del disolvente A (el débil, agua) y del disolvente B (el fuerte, el modificador orgánico).

La solución habitual al problema general de elución en LC es la aplicación de gradientes de elución, con cambios programados de modificador orgánico. El principal objetivo de la elución en gradiente es lograr una resolución

adecuada en un tiempo razonablemente corto, realizando una única inyección. Con este propósito, la fuerza eluyente de la fase móvil debe ser inicialmente baja e incrementarse progresivamente al concentrar el modificador menos polar, a medida que avanza la separación. De esta forma, se incrementa la retención de los analitos poco retenidos y se acelera la elución de los que se hallan muy retenidos, al incrementarse el porcentaje de reparto de los solutos en la fase móvil. Por ello, se prefiere la elución en gradiente para la separación de muchas muestras, siendo ampliamente utilizada en RPLC.

Los gradientes son especialmente útiles para ralentizar la elución de pares o grupos de solutos difíciles de resolver, mientras se acelera la elución de aquellas regiones del cromatograma que contienen solutos bien resueltos o espacios entre los solutos. La elución en gradiente reduce la distancia entre los picos de solutos de polaridad progresivamente decreciente, que se espacian más regularmente en el cromatograma. Los tiempos de retención son más cortos, las anchuras de los picos son apreciablemente más pequeñas y, para picos de la misma altura, las anchuras son aproximadamente las mismas. Al final del gradiente, la fase móvil suele ser lo suficientemente fuerte como para eliminar todas las impurezas de la muestra, que de lo contrario pueden quedar retenidas interfiriendo en la interpretación de los cromatogramas de las muestras inyectadas a continuación.

Los gradientes consisten generalmente en incrementos lineales simples de un factor experimental, pero pueden añadirse varias etapas isocráticas (gradientes multi-isocráticos), o varios segmentos de diferente pendiente (gradientes multi-segmentados), que pueden ser lineales (gradientes multi-lineales), o curvilíneos [18]. Aunque, en realidad, la optimización de un gradiente lineal involucra tan sólo un factor experimental (la concentración del modificador), el gradiente estará regido por el contenido inicial de disolvente

orgánico en el gradiente y su pendiente. Las pendientes de los gradientes deben ser siempre positivas para acelerar la elución, pero ocasionalmente se pueden añadir una o más etapas isocráticas. Para los gradientes multi-lineales, que contienen dos o más segmentos lineales de distinta pendiente, se deben optimizar las coordenadas de los puntos de transición entre segmentos lineales consecutivos (o la pendiente y la longitud de cada segmento lineal) [19]. Desafortunadamente, cuando aumenta el número de nodos, la complejidad de la optimización aumenta exponencialmente. La inclusión de un nodo adicional en una búsqueda exhaustiva puede suponer un volumen de cálculo casi elevado al cuadrado. El número de programas de gradiente que se pueden implementar es virtualmente ilimitado y el éxito de la separación depende críticamente del perfil seleccionado.

1.3. Tiempo de retraso y tiempo muerto

La fase móvil que se forma en el mezclador del instrumento para formar el gradiente tiene asociado un tiempo de retraso (t_D , *dwell time*) para llegar a la entrada de la columna. Este retraso está relacionado con el volumen de permanencia, que es el volumen del sistema que debe rellenarse para que la fase móvil llegue a la entrada de la columna [20]. El volumen de permanencia está constituido por el volumen interno de la cámara de mezcla del gradiente, la tubería de conexión, las fritas y, por lo general, también el volumen interno del dispositivo de inyección (incluido el volumen del circuito de inyección). Al inicio del gradiente, el volumen de permanencia contiene la fase móvil que se forma al iniciarse el gradiente, de modo que la muestra eluye inicialmente en régimen isocrático. En LC convencional, la influencia del volumen de permanencia en la retención de los solutos es pequeña, pero no despreciable.

Por su parte, el volumen muerto en los sistemas de LC se ha definido como el volumen de fase móvil que llena el espacio entre el inyector y la celda detectora, que incluye el volumen intersticial o interpartícula accesible (el volumen entre las partículas empaquetadas) y el volumen intrapartícula (el volumen en el interior de los poros de las partículas accesibles a la fase móvil), así como el volumen de los tubos y cualquier otro componente del sistema (el denominado volumen extra-columna) [21,22].

Relacionado con este concepto está el tiempo muerto, que es el tiempo que necesita un compuesto no retenido ideal (un compuesto que no interactúa con la fase estacionaria) para cruzar la distancia entre el inyector y la celda del detector, cuando eluye a un caudal constante. La estimación del tiempo muerto en los sistemas cromatográficos es la base para el cálculo de los factores de retención (Ecuación 1.1). El cálculo del factor de retención implica trasladar el origen de los tiempos de retención al tiempo en el que aparece un compuesto no retenido, y dividir el tiempo neto resultante por el tiempo muerto. El factor de retención normaliza la retención y permite la comparación entre diferentes longitudes de columna o la elución a diferentes velocidades de flujo para la misma columna.

1.4. Optimización de la resolución cromatográfica

La LC es una técnica muy extendida en los laboratorios analíticos, pero posee la desventaja de su baja eficacia, en comparación con la cromatografía de gases. Para conseguir una resolución adecuada en tiempos de análisis lo suficientemente cortos, se han diseñado diversas estrategias de optimización. En general, encontrar las mejores condiciones de separación cromatográfica no es fácil. A pesar de ser particularmente lentas e ineficientes, las estrategias de prueba y error siguen siendo frecuentes en el desarrollo de los métodos. Sin

embargo, muchas muestras son tan complejas que el protocolo puede ser demasiado largo y a menudo se pierden las mejores (o al menos aceptables) condiciones de separación. Afortunadamente, el desarrollo de los métodos se puede acelerar notablemente con resultados más fiables mediante la aplicación de optimizaciones interpretativas (basadas en modelos) asistidas por ordenador [23-26]. También hacen posible una optimización integral, que es esencial cuando se trata de muestras complejas.

1.4.1. Modelización de la retención

Las optimizaciones interpretativas incluyen dos etapas: la modelización del sistema en base a la información proporcionada por los datos de entrenamiento y la predicción de la resolución, a través de simulaciones realizadas con un ordenador [24,26,27]. En la primera etapa, el analista desarrolla una serie de experimentos tan reducidos e informativos como sea posible, a fin de ajustar ecuaciones o entrenar algoritmos que permitan la predicción de cromatogramas bajo cualquier nueva condición, dentro del espacio experimental de interés. El objetivo es desarrollar modelos capaces de predecir la separación en cualquier nueva condición arbitraria.

En la segunda etapa de la optimización, en base a los modelos ajustados, se predice la calidad de la separación para un gran número de condiciones de separación, lo que permite encontrar el conjunto de condiciones que proporcionan la resolución máxima (o al menos, una resolución adecuada), para todos los pares de picos. En la práctica, este proceso se realiza simulando los cromatogramas para una distribución pre-establecida (regular o aleatoria) de los factores experimentales que se están optimizando. Con este propósito, se simulan cromatogramas sintéticos añadiendo las señales predichas para los compuestos objetivo existentes en la mezcla analizada.

La propiedad cromatográfica que tiene un mayor impacto en la calidad de la separación es la retención. Como se ha mencionado, en la primera etapa de la optimización se recopila información sobre el comportamiento cromatográfico de los compuestos en la muestra, que cubran una región de los factores razonablemente amplia. Esto permitirá, entre otros fines, inferir una relación que describa la retención en función de los factores experimentales, para cada soluto, que puede ser una expresión algebraica conveniente (un modelo empírico o mecanicista), o un algoritmo de caja negra, capaz de predecir tiempos de retención.

Los modelos de retención se pueden ajustar a partir de experimentos isocráticos o de gradiente [28–31], lo que significa encontrar los mejores valores de los parámetros del modelo para cada soluto, aplicando regresión por mínimos cuadrados. El proceso de regresión construye la mejor relación posible entre la respuesta y los predictores. Esto se hace minimizando la suma de los residuos al cuadrado (es decir, la diferencia al cuadrado entre las respuestas reales y las previstas, extendida a todo el conjunto de experimentos de entrenamiento).

Los modelos de retención se pueden clasificar en lineales y no lineales, para lo cual se requieren procedimientos de regresión lineal y no lineal adecuados, respectivamente. Algunos modelos no lineales pueden transformarse en lineales, modificando la variable respuesta. Una vez realizada esta transformación, se pueden aplicar procedimientos de regresión lineales (que son universales, más seguros y más simples). Se han probado exhaustivamente varios modelos que permiten predicciones exactas a concentraciones variables de disolvente orgánico [32–40].

Debe señalarse que todas las ecuaciones que describen la retención en RPLC, con k como variable respuesta, son no lineales. Cuando la variable

respuesta se transforma (por ejemplo, $\log k$), para obtener un modelo lineal más simple, la distribución de las incertidumbres asociadas a la solución se ve afectada: el conjunto de parámetros óptimos minimizará los residuos de la respuesta transformada ($\log k$), pero no los de la respuesta original (k), que es lo que realmente interesa. Este problema no existe si la regresión se realiza de forma no lineal con la variable respuesta original. Sin embargo, lamentablemente, existe una práctica generalizada de ignorar estos efectos sobre la regresión.

1.4.2. Diseño experimental y seguimiento de picos

La exactitud de las predicciones es decisiva para la fiabilidad de las optimizaciones. Para un soluto dado, la calidad de la predicción depende de la riqueza de la información proporcionada por el diseño experimental (el número y la distribución de los experimentos dentro del dominio del espacio de los factores), la ecuación seleccionada para ajustar los datos de entrenamiento y el procedimiento de ajuste [41-43]. También pueden influir en la exactitud de las predicciones los intervalos de los factores experimentales y el modo de elución (isocrático o gradiente), en los que se mide la retención, así como otras propiedades de pico para el conjunto de entrenamiento.

La fiabilidad de todo el proceso depende de la calidad del diseño experimental utilizado para construir (entrenar) los modelos de retención, que se desea den acceso a las mejores condiciones de separación con el mínimo esfuerzo y la máxima exactitud. Idealmente, el diseño de entrenamiento debería ser común para un grupo de analitos, con un número razonablemente pequeño de condiciones de elución. Las experiencias realizadas deben planificarse cuidadosamente para que sean lo más informativas posible y cubran todo el espacio de los factores. La búsqueda del mejor diseño es más compleja en el

modo gradiente que en elución isocrática, debido a la cantidad de variables y perfiles involucrados (gradientes con una o más etapas isocráticas, lineales o multi-lineales, entre otros tipos de diseños).

También es importante la reducción del esfuerzo experimental, sin deteriorar la calidad del modelo de predicción. Por lo general, cuando sólo se seleccionan como factores el contenido de modificador o la temperatura, se obtienen modelos que originan predicciones exactas, incluso utilizando diseños experimentales mínimos con solo un grado de libertad. Además, dado que la fiabilidad de los modelos ha sido ampliamente demostrada, en muchos casos prácticos se utilizan diseños experimentales sin grados de libertad. Sólo cuando se encuentran desviaciones significativas al realizar una predicción con el modelo construido, se debe repetir el ajuste del modelo incluyendo en el diseño más experimentos que los estrictamente necesarios, con el fin de reducir la incertidumbre de las predicciones.

Otro punto a considerar en la medida de los tiempos de retención es la asignación incorrecta de picos, cuando se inyectan mezclas de compuestos estándar. La asignación de cada pico a cada soluto eluido puede facilitarse mediante la inyección de estándares de los compuestos aislados, aunque se pueden inyectar disoluciones que contengan dos o más compuestos con retenciones suficientemente separadas. Si es necesario, se puede realizar un seguimiento de los picos variando la concentración de los estándares inyectados, o con la ayuda de una técnica de detección selectiva [44-46]. Por lo tanto, se pueden utilizar disoluciones multi-estándar, siempre que las diferencias en los tiempos de retención permitan identificar correctamente a los picos.

1.4.3. Modelización del perfil de los picos

Como se ha indicado, muy a menudo, la optimización de la resolución se lleva a cabo utilizando información de cromatogramas simplificados, donde sólo se consideran los tiempos de retención. Este tratamiento se halla respaldado por la idea de que la propiedad de los picos más relevante es la retención. Sin embargo, las simulaciones realistas de cromatogramas requieren tratamientos más elaborados, donde se consideren las anchuras de banda, o incluso, perfiles de pico completos. En algunos casos, también es necesario considerar las áreas de los picos individuales, para conseguir una optimización fiable (por ejemplo, para tener en cuenta un componente de menor concentración), aunque en la mayoría de las situaciones, el uso de áreas normalizadas es una opción válida. Para el cálculo de la resolución, los criterios más completos requieren una simulación total de los cromatogramas, que se construyen sumando los picos individuales.

Se han descrito en la bibliografía varias funciones matemáticas teóricas y empíricas para la descripción de los perfiles de picos cromatográficos, con resultados más o menos aceptables [47]. Los perfiles de elución de los picos cromatográficos simétricos y sin sobrecarga se describen correctamente mediante el modelo gaussiano. Sin embargo, es muy frecuente encontrar en los cromatogramas picos no ideales, mostrando colas o desviaciones frontales. Además, la descripción de los picos asimétricos debe considerar variaciones incidentales en el perfil de los picos con la composición de la fase móvil y otros factores [48,49]. Por lo tanto, se necesita un modelo de pico matemático adecuado y práctico para describir cada pico en un cromatograma. Además, el modelo debe basarse en las propiedades que habitualmente se monitorizan (tiempo de retención, eficacia, asimetría y área o altura).

Si se establece una correlación adecuada entre el tiempo de retención (que puede predecirse con gran exactitud) y los parámetros del perfil de pico, éstos últimos se podrán predecir con suficiente exactitud. Se ha encontrado que la correlación entre el tiempo de retención y las semianchuras de los picos es parabólica (aunque casi lineal) [49]. Además, todos los compuestos eluidos en una determinada columna de RPLC siguen una tendencia similar de variación de los perfiles de pico con el tiempo de retención, siempre que las interacciones establecidas sean del mismo tipo. Por lo tanto, el rendimiento de la columna se puede caracterizar utilizando parámetros de perfil de pico para unos pocos compuestos eluidos a diversos tiempos de retención.

1.4.4. Optimización interpretativa asistida por ordenador

El análisis cromatográfico de nuevas muestras implica encontrar condiciones capaces de aislar los compuestos de interés en el menor tiempo posible. Esta tarea suele ser ardua y desalentadora: se deben cumplir varios objetivos opuestos relacionados con el tiempo empleado, el nivel de resolución, el factor económico y el coste ambiental. Además, se abren demasiadas posibilidades para resolver el problema. Por lo tanto, cuando un analista se enfrenta a una muestra desconocida, no se conoce el nivel de satisfacción que se alcanzará con la solución encontrada al aplicar las decisiones adoptadas durante el proceso de optimización [23]. Siempre surge la duda de si la solución alcanzada es realmente la mejor.

Cuando el analista se enfrenta a un nuevo problema, se deben tomar en primer lugar algunas decisiones relacionadas con los factores experimentales no ajustables (la naturaleza y longitud de la columna, y la naturaleza del modificador orgánico y del tampón en la fase móvil). Estos factores generalmente se seleccionan mediante prueba y error, o sobre la base de

conocimientos previos. A continuación, se examinan los efectos de las propiedades fácilmente ajustables, como son el contenido de modificador, el pH, la temperatura o el programa de gradiente, y se varían uno o más de estos factores. Cuando los resultados no son satisfactorios, se requiere un cambio importante en el sistema cromatográfico. Este panorama sugiere que encontrar las mejores condiciones de separación cromatográfica no es sencillo.

El proceso de optimización se restringe a los factores experimentales más relevantes, que son los de mayor impacto en la resolución, o los más convenientes desde el punto de vista práctico. En el caso de la RPLC isocrática, estos factores suelen ser la concentración de uno o más modificadores, el pH y la temperatura. Cuando se trabaja en modo gradiente, el objetivo de la optimización es el patrón de variación de una o más de estas variables (es decir, el perfil del programa). Como se ha comentado, la concentración de modificador orgánico en la fase móvil es el factor más frecuentemente optimizado en RPLC. No sólo tiene un gran impacto sobre la retención y la selectividad, sino que también es una propiedad muy conveniente, debido a la flexibilidad y exactitud en la implementación de cambios; de ahí la importancia de obtener modelos de retención fiables que consideren este factor.

En LC, la fuerza eluyente es la capacidad de la fase móvil para hacer migrar los solutos retenidos por la fase estacionaria. Depende de la naturaleza de la fase estacionaria y de los solutos, así como de la composición de la fase móvil (naturaleza y concentración de los disolventes y aditivos), el pH y la temperatura de la columna. Por lo tanto, para una fase estacionaria dada, la fuerza eluyente no es exclusivamente una propiedad relacionada con el disolvente, ya que los solutos experimentan diferentes fuerzas eluyentes dependiendo de sus estructuras [12]. La fuerza eluyente de la fase móvil es un concepto práctico en LC, comúnmente utilizado para ajustar de manera óptima

la retención global de un grupo de solutos dentro de la región de retención objetivo, en el intervalo de $1 < k < 5$, o al menos $0.2 < k < 20$.

Para una fase estacionaria y un conjunto determinado de solutos, si la fuerza eluyente es demasiado alta, los tiempos de retención serán demasiado cortos y, en consecuencia, la resolución será deficiente. Por el contrario, si la fuerza eluyente es demasiado baja, los tiempos de retención serán excesivos y, por lo tanto, el tiempo de separación será demasiado largo y, debido a la dilución excesiva, la relación señal/ruido en el máximo de los picos para los analitos más retenidos se reducirá significativamente.

Una vez que se ha ajustado la fuerza eluyente, la selectividad (orden de elución y distribución de los picos) se puede optimizar sin modificar significativamente la retención global. El criterio que suele seguirse para optimizar la selectividad es resolver todos los pares de picos de las muestras objetivo, en el menor tiempo de separación total posible.

En la práctica, las optimizaciones interpretativas asistidas por ordenador se realizan simulando la separación de la muestra en un espacio factorial pre-establecido. Para encontrar las mejores condiciones, la información contenida en los cromatogramas simulados debe transformarse a un valor numérico, el cual es monitorizado a lo largo de la optimización. Idealmente, este valor debería correlacionarse con la evaluación de resolución del analista. La expresión matemática que permite evaluar la calidad de la separación de cualquier distribución de picos se denomina “función objetivo”, la cual se maximiza a lo largo del proceso de optimización.

Se han propuesto varias estrategias para optimizar las condiciones experimentales que ayuden a resolver problemas de elución complejos [25]. Como se ha comentado, para llevar a cabo una optimización rigurosa, es conveniente modelizar no sólo la retención, sino también el perfil de los picos

definido por sus semianchuras izquierda y derecha (y eventualmente, su altura). El proceso de optimización se inicia imitando un diseño experimental sintético exhaustivo. Con este propósito, primero se define un conjunto artificial de condiciones de separación, ya sea isocráticas o de gradiente. Para cada una de esas condiciones sintéticas, se simula el cromatograma correspondiente, o una simplificación del mismo, y se evalúa la calidad de la distribución de picos predicha. Este proceso se extiende a todas las condiciones consideradas en el diseño sintético, para averiguar qué condición se espera produzca la máxima calidad de separación.

La fiabilidad de las condiciones optimizadas depende, por un lado, de la exactitud alcanzada en la simulación de cromatogramas y, por otro, de la exhaustividad del diseño experimental sintético. El primer requisito implica la selección de un sistema de predicción que garantice el máximo nivel de fiabilidad y realismo. Esto es particularmente importante para las separaciones críticas, o cuando se está considerando el tamaño de la señal. Se puede considerar que este objetivo se halla satisfactoriamente resuelto con los desarrollos actuales.

El segundo requisito significa que el diseño experimental sintético debe ofrecer un nivel de precisión similar al alcanzable en la traducción práctica de las condiciones óptimas al cromatógrafo. La satisfacción de este objetivo puede implicar tiempos de cálculo excesivamente elevados, pero el desarrollo de nuevas herramientas de cálculo y procesadores más rápidos están permitiendo que sean accesibles problemas que en el pasado eran inviables, como los considerados en esta Memoria de Tesis.

1.5. Referencias

- [1] L.R. Snyder, J.J. Kirkland, *Introduction to Modern Liquid Chromatography*, 2ª edición, John Wiley & Sons, Nueva York, EE.UU., 1979.
- [2] W.R. Melander, C. Horváth, Reversed-phase chromatography, en *High-Performance Liquid Chromatography*, Vol. 2, Academic Press, Nueva York, EE.UU., 1980, págs. 113–319.
- [3] W.J. Lough, Reversed phase liquid chromatography, en *Encyclopedia of Analytical Chemistry* (editado por R.A. Meyers), John Wiley & Sons, Inc., Nueva York, EE.UU., 2006.
- [4] J.J. Pesek, M.T. Matyska, Reversed-phase chromatography: Description and applications, en *Encyclopedia of Chromatography* (editado por J. Cazes), Taylor & Francis, Nueva York, EE.UU., 2006, págs. 719–722.
- [5] A. Soliven, S. Kayillo, R.A. Shalliker, Reversed phase liquid chromatography, en *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering* (editado por J. Reedijk), Elsevier, Amsterdam, Holanda, 2013.
- [6] M.C. García Álvarez-Coque, J.J. Baeza Baeza, G. Ramis Ramos, Reversed phase liquid chromatography, en *Analytical Separation Science Series* (editado por J.L. Anderson, A. Stalcup, A. Berthod, V. Pino), Vol. 1 (Liquid Chromatography), Capítulo 8, Wiley-VCH, Nueva York, EE.UU., 2015, págs. 159–197.
- [7] J.J. Pesek, M.T. Matyska, Reversed-phase stationary phases, en *Encyclopedia of Chromatography* (editado por J. Cazes), Taylor & Francis, Nueva York, EE.UU., 2006, págs. 723–726.
- [8] U.D. Neue, Stationary phase characterization and method development, *J. Sep. Sci.* 30 (2007) 1611–1627.

-
- [9] G.J. Eppert, P. Heitmann, Selectivity optimization of stationary phases, *LC-GC Eur.* 16 (2003) 698–705.
- [10] J.W. Dolan, Selectivity in reversed-phase LC separations, Part I: Solvent-type selectivity, *LCGC North America* 28 (2010) 1022–1027.
- [11] V.J. Barwick, Strategies for solvent selection, *Trends Anal. Chem.* 16 (1997) 293–309.
- [12] G. Ramis Ramos, M.C. García Álvarez-Coque, Solvent selection in liquid chromatography, en *Liquid Chromatography: Fundamentals and Instrumentation* (editado por S. Fanali, P. Haddad, C.F. Poole, P.J. Schoenmakers, D. Lloyd), Elsevier, Amsterdam, Holanda, 2013, págs. 223–249.
- [13] J.W. Dolan, LC method scaling, Part I. Isocratic separations, *LCGC North Am.* (2014) 32, 98, 100, 102.
- [14] P. Jandera, Can the theory of gradient liquid chromatography be useful in solving practical problems?, *J. Chromatogr. A* 1126 (2006) 195–218.
- [15] J.E. Haky, D.A. Teifer, Gradient elution, en *Encyclopedia of Chromatography* (editado por J. Cazes), Taylor & Francis, Nueva York, EE.UU., 2006, págs. 393–396.
- [16] L.R. Snyder, J.W. Dolan, *High-Performance Gradient Elution*, John Wiley & Sons, Inc., Hoboken, NJ., EE.UU., 2007.
- [17] J.W. Dolan, L.R. Snyder, Gradient elution chromatography, en *Encyclopedia of Analytical Chemistry* (editado por R.A. Meyers), John Wiley & Sons, Nueva York, EE.UU., 2012.
- [18] J.W. Dolan, L.R. Snyder, Theory and practice of gradient elution, en *Liquid Chromatography: Fundamentals and Instrumentation* (editado por S. Fanali, P. Haddad, C.F. Poole, P.J. Schoenmakers, D. Lloyd), Elsevier, Amsterdam, Holanda, 2013, págs. 269–282.
-

- [19] P. Nikitas, A. Pappa-Louisi, A. Papageorgiou, Simple algorithms for fitting and optimisation for multi-linear gradient elution in reversed-phase liquid chromatography, *J. Chromatogr A* 1157 (2007) 178–186.
- [20] V. Concha Herrera, G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Limits of multi-linear gradient optimisation in reversed-phase liquid chromatography, *J. Chromatogr. A* 1063 (2005) 79–88.
- [21] S. Pous Torres, J.R. Torres Lapasió, M.C. García Álvarez-Coque; Performance of markers and the homologous series method for dead time estimation in reversed-phase liquid chromatography, *J. Liq. Chromatogr. Rel. Technol.* 32 (2009) 1065–1083.
- [22] S. Pous Torres, J.R. Torres Lapasió, J.J. Baeza Baeza, M.C. García Álvarez-Coque, Alternating iterative regression method for dead time estimation from experimental designs, *Anal. Bioanal. Chem.* 394 (2009) 625–636.
- [23] J.R. Torres Lapasió, M.C. García Álvarez-Coque, Levels in the interpretive optimisation of selectivity in high-performance liquid chromatography: A magical mystery tour, *J. Chromatogr. A* 1120 (2006) 308–321.
- [24] M.C. García Álvarez-Coque, J.R. Torres Lapasió, J.J. Baeza Baeza, Models and objective functions for the optimisation of selectivity in reversed-phase liquid chromatography, *Anal. Chim. Acta* 579 (2006) 125–145.
- [25] J.R. Torres Lapasió, M.C. García Álvarez-Coque, Liquid chromatography: Strategies for optimization, *Encyclopedia of Analytical Science* (editada por P.J. Worsfold, A. Townshend, C. Poole, M. Miró), Vol. 6, 3ª edición, Elsevier, Waltham, MA, EE.UU., 2019, págs. 252–260.

-
- [26] J.R. Torres Lapasió, M.C. García Álvarez-Coque, J.J. Baeza Baeza, Modelling and optimisation strategies in reversed-phase liquid chromatography, en *Recent Research Developments in Pure & Applied Analytical Chemistry* 3 (2001) 169–184, Transworld Research Network, Trivandrum (India).
- [27] M.C. García Álvarez-Coque, G. Ramis Ramos, J.R. Torres Lapasió, C. Ortiz Bolsico, Modeling of retention in reversed phase liquid chromatography, en *Analytical Separation Science Series* (editado por J.L. Anderson, A. Stalcup, A. Berthod, V. Pino), Vol. 1 (Liquid Chromatography), Capítulo 9, Wiley, Nueva York, EE.UU., 2015, págs. 199 226.
- [28] P.J. Schoenmakers, H.A.H. Billiet, L. De Galan, Systematic study of ternary solvent behaviour in reversed-phase liquid chromatography, *J. Chromatogr. A* 218 (1981) 261 284.
- [29] M.A. Quarry, R.L. Grob, L.R. Snyder, Prediction of precise isocratic retention data from two or more gradient elution runs: Analysis of some associated errors, *Anal. Chem.* 58 (1986) 907 917.
- [30] G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Error analysis and performance of different retention models in the transference of data from/to isocratic/gradient elution, *J. Chromatogr. A* 1018 (2003) 169 181.
- [31] P. Nikitas, A. Pappa-Louisi, Retention models for isocratic and gradient elution in reversed-phase liquid chromatography, *J. Chromatogr. A* 1216 (2009) 1737 1755.
- [32] P.J. Schoenmakers, H.A.H. Billiet, R. Tijssen, L. de Galan, Gradient selection in reversed-phase liquid chromatography, *J. Chromatogr.* 149 (1978) 519 537.
-

- [33] K. Valko, L.R. Snyder, J.L. Glajch, Retention in reversed-phase liquid chromatography as a function of mobile-phase composition, *J. Chromatogr. A* 656 (1993) 501–520.
- [34] W. Prus, Y. Vander Heyden, P. Kus, D.L. Massart, T. Kowalska, A statistical evaluation of selected RPHPLC retention models, *Acta Chromatogr.* 7 (1997) 193–209.
- [35] T. Baczek, M. Markuszewski, R. Kaliszan, M.A. van Straten, H.A. Claessens, Linear and quadratic relationships between retention and organic modifier content in eluent in reversed phase high-performance liquid chromatography: A systematic comparative statistical study, *J. High Resolut. Chromatogr.* 23 (2000) 667–676.
- [36] J. Ko, J.C. Ford, Comparison of selected retention models in reversed-phase liquid chromatography, *J. Chromatogr. A* 913 (2001) 3–13.
- [37] W. Zapala, K. Kaczmarski, T. Kowalska, Comparison of different retention models in normal and reversed-phase liquid chromatography with binary mobile phases, *J. Chromatogr. Sci.* 40 (2002) 575–580.
- [38] P. Nikitas, A. Pappa-Louisi, P. Agrafiotou, Effect of the organic modifier concentration on the retention in reversed-phase liquid chromatography II. Tests using various simplified models, *J. Chromatogr. A* 946 (2002) 33–45.
- [39] A. Pappa-Louisi, P. Nikitas, P. Balkatzopoulou, C. Malliakas, Two- and three-parameter equations for representation of retention data in reversed-phase liquid chromatography, *J. Chromatogr. A* 1033 (2004) 29–41.
- [40] Ł. Komsta, A new general equation for retention modeling from the organic modifier content of the mobile phase, *Acta Chromatographica* 22 (2010) 267–279.

-
- [41] S.N. Deming, S.L. Morgan, Experimental design: A chemometric approach, 2ª edición, en *Data Handling in Science and Technology*, Vol. 11, Elsevier Science, Amsterdam, Holanda, 1993.
- [42] D.B. Hibbert, Experimental design in chromatography: A tutorial review, *J. Chromatogr. B* 910 (2012) 2–13.
- [43] E. Rozet, P. Lebrun, P. Hubert, Design spaces for analytical methods, *Trends Anal. Chem.* 42 (2013) 157–167.
- [44] A.C.J.H. Drouen, H.A.H. Billiet, L. De Galan, Multiwavelength absorbance detection for solute recognition in liquid chromatography, *Anal. Chem.* 57 (1985) 962–968.
- [45] G. Stoev, A. Mihailova, Increasing the reliability of the identification by high-performance liquid chromatography by means of selective and/or sensitive detection, *J. Chromatogr. A* 869 (2000) 275–284.
- [46] P.V. van Zomeren, A. Hoogvorst, P.M.J. Coenegracht, G.J. de Jong, Optimisation of high-performance liquid chromatography with diode array detection using an automatic peak tracking procedure based on augmented iterative target transformation factor analysis, *Analyst* 129 (2004) 241–248.
- [47] J.R. Torres Lapasió, J.J. Baeza Baeza, M.C. García Álvarez-Coque, Modeling of peak shape and asymmetry, en *Chemometrics in Chromatography* (editado por Ł. Komsta, Y. Vander Heyden, J. Sherma), Chromatographic Science Series, Vol. 111, Capítulo 12, CRC Press, Taylor and Francis Group, Boca Raton, FL., EE.UU., 2018, págs. 217–238.
- [48] M.J. Ruiz Ángel, S. Carda Broch, M.C. García Álvarez-Coque, Peak half-width plots to study the effect of organic solvents on the peak

- performance of basic drugs in micellar liquid chromatography, *J. Chromatogr. A* 1217 (2010) 1786–1798.
- [49] J.J. Baeza Baeza, M.J. Ruiz Ángel, M.C. García Álvarez-Coque, S. Carda Broch, Half-width plots, a simple tool to predict peak shape, reveal column kinetics and characterise chromatographic columns in liquid chromatography: State of the art and new results, *J. Chromatogr. A* 1314 (2013) 142–153.

Parte 1

EVALUACIÓN DEL RENDIMIENTO PREDICTIVO DE DIVERSOS DISEÑOS EXPERIMENTALES

CAPÍTULO 2

BENEFICIOS DEL USO DE PULSOS DE CONCENTRACIÓN DE MODIFICADOR ORGÁNICO EN LA MODELIZACIÓN DEL TIEMPO DE RETENCIÓN EN CROMATOGRAFÍA LÍQUIDA

2.1. Resumen

Con fines de modelización, se exploran las ventajas y desventajas del uso de diseños experimentales isocráticos que incluyen incrementos transitorios de modificador orgánico (pulsos), en las fases móviles de menor fuerza eluyente. Para los solutos retenidos, este tipo de diseño mixto ofrece una capacidad predictiva similar o mejor que los diseños de gradiente, un tiempo de medida más corto que los diseños puramente isocráticos y parámetros del modelo de retención que concuerdan con los derivados de experimentos isocráticos puros, con incertidumbres similares. Los tiempos de retención predichos son comparables a los ofrecidos por modelos ajustados a partir de diseños isocráticos puros, siendo el consumo de disolvente orgánico apreciablemente menor. Desde un punto de vista práctico, los diseños mixtos que incluyen pulsos se pueden construir fácilmente insertando en las fases móviles isocráticas más lentas un pulso de corta duración a un tiempo intermedio. Esto permite que los solutos más rápidos mantengan una retención apreciable en el sector inicial del programa de elución, previo al pulso, y que los solutos lentos eluyan tras el pulso, también en tiempos aceptables. El ajuste de los datos de retención obtenidos con fases móviles a las que se han añadido pulsos es más simple, en comparación con la elución en gradiente, ya que supone una resolución simplificada de la ecuación integral de elución en gradiente, debido a la presencia de sectores isocráticos. Sin embargo, los experimentos que involucran pulsos revelan la existencia de discrepancias en las predicciones ofrecidas por la ecuación fundamental de elución en gradiente, cuando se resuelve mediante integración numérica, para los solutos que eluyen en las proximidades del pulso. La corrección de tales discrepancias implica la inclusión de retrasos intra-columnares en los cambios de concentración de modificador orgánico, dictados por el programa de gradiente, en la posición instantánea del soluto a lo largo de toda la migración.

2.2. Introducción

Desde los años 70, el uso de la cromatografía líquida (LC, *Liquid chromatography*) ha ido creciendo en los laboratorios analíticos, debido a su sensibilidad, robustez, facilidad de uso y aplicabilidad a múltiples problemas en diversos campos (análisis ambiental, farmacéutico, clínico y de alimentos) [1,2]. La cromatografía líquida en fase inversa (RPLC, *Reversed-phase liquid chromatography*) es el modo de LC más habitual para el análisis de compuestos no volátiles, desde pequeñas moléculas hasta grandes macromoléculas biológicas, abarcando una amplia gama de polaridades [3]. La etapa más importante en el desarrollo de un método cromatográfico sigue siendo la elección de la columna. En la actualidad, se comercializan cientos de columnas para RPLC, con rendimientos muy diferentes. En cambio, la elección del modificador orgánico es particularmente limitada, reduciéndose casi exclusivamente a mezclas de acetonitrilo o metanol con un tampón acuoso. No obstante, el comportamiento cromatográfico puede modularse ampliamente variando la concentración del modificador en la fase móvil. Dado que las condiciones inicialmente seleccionadas por el analista rara vez proporcionan una resolución suficientemente satisfactoria (excepto para muestras muy simples), se debe aplicar un protocolo de optimización para encontrar la composición de fase móvil isocrática o el programa de gradiente más apropiados [4,5].

La elución isocrática es adecuada para muestras que contienen un pequeño grupo de analitos, dentro de un intervalo de polaridades pequeño o moderado. En este caso, todos los solutos se resolverán en tiempos razonables con la fase móvil adecuada. Por el contrario, este modo de elución no es recomendable cuando los solutos cubren un amplio intervalo de polaridades. En este caso, son posibles dos escenarios. El primero conduce a un cromatograma en el que los

solutos más retenidos eluyen en tiempos adecuados a porcentajes de modificador elevados, pero los picos iniciales presentan una resolución deficiente, o incluso eluyen junto con el frente del disolvente. El segundo escenario es el opuesto, dando lugar a un cromatograma donde los solutos menos retenidos se hallan bien resueltos a porcentajes bajos de modificador, pero los solutos más retenidos eluyen en tiempos excesivos, con picos ensanchados y problemas de sensibilidad. De este modo, algunos solutos altamente retenidos pueden pasar desapercibidos, sin poderse distinguir de la línea base, o incluso aparecerán superpuestos con el cromatograma de la siguiente muestra inyectada.

Por lo tanto, utilizando elución isocrática, no es posible mejorar ambos extremos del cromatograma a la vez. Esta incompatibilidad se conoce como el "problema general de elución cromatográfica" [2]. La solución habitual es la aplicación de un gradiente de modificador orgánico, en el que su concentración se modifica gradualmente de acuerdo a un programa [6-9]. El objetivo principal es obtener una resolución adecuada para todos los componentes de la muestra, aumentando la retención de los solutos poco retenidos y reduciéndola para los fuertemente retenidos. Con este propósito, la fuerza eluyente de la fase móvil debe ser inicialmente baja e incrementarse a medida que avanza la separación (por ejemplo, aumentando el porcentaje de modificador orgánico).

La optimización interpretativa de la resolución (basada en modelos) permite encontrar las condiciones que separan simultáneamente todos los compuestos de una muestra, o al menos, los compuestos objetivo [4,5,10]. La primera etapa en estas optimizaciones consiste en la recolección de información sobre el comportamiento cromatográfico de los solutos de la muestra, enfocándose principalmente en la retención y cubriendo amplias regiones de los factores experimentales involucrados en la modulación del comportamiento

cromatográfico. Para ello, los datos se recogen en condiciones de elución controladas, según un diseño experimental preestablecido. Para cada soluto, se ajusta un modelo matemático que describe adecuadamente el comportamiento cromatográfico en función de los factores experimentales. Los modelos permiten predecir los tiempos de retención y otras propiedades de los picos, para solutos particulares y en diferentes condiciones, en ambos modos isocrático o de gradiente [4].

En este capítulo, se estudia el uso de aumentos transitorios (pulsos) en la concentración del modificador orgánico en LC. Se examinan los beneficios e inconvenientes del uso de diseños mixtos, que incluyen experimentos isocráticos y/o de gradiente con incrementos transitorios de la concentración de modificador orgánico, con el objetivo principal de obtener diseños experimentales más informativos. Los pulsos permiten obtener información sobre la retención para solutos altamente hidrofóbicos, manteniendo una baja concentración de modificador durante la mayor parte de la elución. Este tipo de programa de elución reduce el tiempo de retención de los compuestos más retenidos a valores razonables. Además, constituye una posibilidad interesante en la construcción de diseños experimentales isocráticos, ya que permite tener acceso a medidas a baja fuerza eluyente, lo que garantiza la obtención de información enriquecida sobre el comportamiento de retención de los solutos de elución lenta. A diferencia de los diseños isocráticos con pulsos, en los diseños constituidos por rampas consecutivas de gradiente, las concentraciones bajas de modificador orgánico tienen una participación marginal. Por su parte, la presencia transitoria de una alta concentración de modificador, durante un corto tiempo, da lugar a tiempos de retención medibles, donde las concentraciones más bajas de modificador en el pulso tienen un peso significativo.

2.3. Teoría

2.3.1. Modelos de retención

En RPLC, el factor experimental generalmente optimizado es el contenido de modificador orgánico en la fase móvil, debido a su gran impacto sobre la fuerza eluyente y la selectividad. Además, también puede ser fácilmente alterado en amplios intervalos para modular la retención de una gran variedad de compuestos. La bibliografía proporciona una amplia variedad de modelos útiles para describir el comportamiento de retención [4,11-14], que permiten predecir el factor de retención (k), en función de la fracción volumétrica del modificador, ϕ . En este trabajo, se han considerado los modelos indicados a continuación, donde los parámetros de ajuste adoptan valores particulares para cada soluto, columna y modificador.

(i) Modelo logarítmico-lineal

Este modelo fue propuesto por Snyder *et al.* [2], y puede expresarse como:

$$\ln k = c_0 + c_1 \phi = \ln k_w - S \phi \quad (2.1)$$

donde

$$k = \frac{t_R - t_0}{t_0 - t_{\text{ext}}} \quad (2.2)$$

siendo t_{ext} el tiempo extra-columnar, que muy a menudo se desprecia en el cálculo de k . En la Ecuación (2.1), la intersección de la línea recta ajustada, $\ln k_w$, se refiere al valor de k extrapolado para una fase móvil compuesta de agua pura. La pendiente S indica la sensibilidad de la retención a cambios en el contenido de modificador orgánico, siendo una medida de la fuerza eluyente de la fase móvil.

(ii) *Modelo logarítmico-cuadrático*

La Ecuación (2.1) describe con exactitud la retención en RPLC sólo en intervalos de modificador orgánico estrechos o moderados. Para concentraciones de modificador extremadamente altas o bajas, se encuentran importantes desviaciones de la linealidad. En este caso, se puede utilizar el modelo logarítmico-cuadrático propuesto por Schoenmakers *et al.* [15], que proporciona ajustes de mayor exactitud:

$$\ln k = c_0 + c_1 \varphi + c_2 \varphi^2 = \ln k_w - S \varphi + T \varphi^2 \quad (2.3)$$

(iii) *Modelo de Neue-Kuss (NK)*

El modelo propuesto por Neue y Kuss [16] ha demostrado un excelente rendimiento en elución isocrática, en amplios dominios de modificador orgánico, siendo actualmente uno de los modelos más exactos para predecir la retención en RPLC:

$$k = k_w (1 + c \varphi)^2 e^{-\frac{B \varphi}{1 + c \varphi}} \quad (2.4)$$

donde k_w es, como en la Ecuación (2.1), el factor de retención extrapolado para una fase móvil constituida únicamente por agua, c es un parámetro de curvatura y B una medida de la fuerza eluyente.

2.3.2. *Predicción de tiempos de retención en elución en gradiente*

Los tiempos de retención se pueden obtener resolviendo la expresión conocida como ecuación fundamental de elución en gradiente, que se indica a continuación [8,17 19]:

$$t_0 - t_{\text{ext}} = \int_0^{t_D} \frac{dt}{k_0} + \int_{t_D}^{t_g - t_0} \frac{dt}{k(\varphi(t))} \quad (2.5)$$

donde t_0 es el tiempo muerto, k_0 el factor de retención al inicio del gradiente, t_g el tiempo de retención de un soluto bajo las condiciones de gradiente ensayadas, y $k(\varphi(t))$ es una expresión que describe la retención en función del programa de gradiente. El uso de conexiones entre el mezclador y la columna, con un volumen apreciable, implica la introducción de un cierto retraso en la llegada de los cambios de composición dictados por el programa de gradiente, a la entrada de la columna. Este retraso es cuantificado mediante el parámetro t_D (tiempo de permanencia o *dwelt time*). Por lo tanto, el perfil de gradiente real debe obtenerse sumando t_D a los valores de tiempo programados.

La función $k(\varphi(t))$ implica dos ecuaciones anidadas: la dependencia de la concentración del modificador con el tiempo (el programa de gradiente) y el factor de retención en función de la concentración de modificador (el modelo de retención). El tiempo de retención puede calcularse para cualquier gradiente, siempre que se conozca la función $k(\varphi(t))$, pero la solución analítica sólo puede obtenerse en casos limitados. Si el gradiente implica únicamente cambios lineales entre $\ln k$ y φ (modelo de Snyder, Ecuación (2.1)), y entre la concentración de modificador y el tiempo t (cuando el gradiente es lineal), la Ecuación (2.5) tendrá la siguiente solución analítica:

$$t_g = t_0 + t_D + \frac{1}{S m} \ln \left[1 + S m (t_0 k_w e^{-S \varphi_0} - t_D) \right] \quad (2.6)$$

Por conveniencia, en el desarrollo de la Ecuación (2.6), el programa de gradiente se ha desplazado para compensar el valor de t_D :

$$= t_0 + m (t - t_D) \quad (2.7)$$

donde m es la pendiente del gradiente, c_0 la concentración inicial del gradiente, y S y k_w son los parámetros del modelo expresado en la Ecuación (2.1), para cada soluto.

Por su parte, la integración analítica del modelo NK (Ecuación (2.4)) da lugar a:

$$t_g = \frac{\ln H [1 + c(a - mt_0)] - B(a - mt_0)}{m(B - c \ln H)} \quad (2.8)$$

donde

$$H = mB \ln k_w \left(t_0 - \frac{t_D}{k_w} \right) + e^{\frac{B(a + mt_D)}{1 + c(a + mt_D)}} \quad (2.9)$$

siendo k_w , c y B los parámetros del modelo para cada soluto, m de nuevo la pendiente del gradiente, y a el contenido de modificador orgánico al inicio de la rampa del gradiente ($t = t_D$). Las Ecuaciones (2.6), (2.8) y (2.9) son válidas para solutos que eluyen a lo largo de una única rampa de gradiente.

La combinación de la función del programa de gradiente y el modelo de retención a menudo conduce a expresiones, para la Ecuación (2.5), que carecen de una solución analítica simple. En tales casos, el tiempo de retención en gradiente, t_g , se puede obtener mediante integración numérica, dividiendo la integral en etapas infinitesimales [19]. El límite superior de la integral, cuya suma es igual o superior a t_0 proporciona la solución:

$$t_0 - t_{\text{ext}} = \int_0^{t_D} \frac{dt}{k_0} + \int_{t_D}^{t_g - t_0} \frac{dt}{k(\varphi(t))} = \frac{t_D}{k_0} + \int_{t_D}^{t_1} \frac{dt}{k(\varphi(t))} \quad (2.10)$$

$$+ \int_{t_1}^{t_2} \frac{dt}{k(\varphi(t))} + \dots + \int_{t_{i-1}}^{t_i} \frac{dt}{k(\varphi(t))} + \int_{t_i}^{t_{i+1} - t_0} \frac{dt}{k(\varphi(t))}$$

Si las etapas en las que se divide la integral son suficientemente pequeñas, se puede asumir que $k(t)$ será constante en cada etapa infinitesimal:

$$t_0 - t_{\text{ext}} \approx \frac{t_D}{k_0} + \frac{t_1 - t_D}{k(\varphi(t_1))} + \frac{t_2 - t_1}{k(\varphi(t_2))} + \dots + \frac{t_i - t_{i-1}}{k(\varphi(t_i))} + \frac{t_{i+1} - t_i}{k(\varphi(t_{i+1}))} \quad (2.11)$$

donde $k \approx k_i \approx k_{i+1}$ si el paso de integración ($t_{i+1} - t_i$) es suficientemente pequeño.

Dependiendo de la disponibilidad de la función primitiva en la integral y el nivel de exactitud requerido para la solución, el analista debe tomar una decisión sobre el empleo de integración analítica o numérica. Para este estudio se ha utilizado integración numérica, que resulta competitiva siempre que la precisión requerida no sea muy inferior a 0.0001 min. En la optimización de gradientes, o en el ajuste de los datos de gradiente con fines de modelización, la integración numérica es una opción válida, a pesar de la baja velocidad de cálculo.

2.3.3. Corrección de las desviaciones en la retención asociadas a retrasos de tiempo cuando se utiliza integración numérica en elución en gradiente

En elución en gradiente, los cambios programados en la composición del eluyente alcanzan la posición del soluto con un retraso cada vez mayor, a lo largo de su migración, pudiéndose distinguir varias contribuciones independientes. La primera contribución se corresponde con el tiempo de permanencia (t_D), asociado al volumen muerto correspondiente a las conexiones entre el mezclador y la entrada de la columna. Si se conectaran varias columnas en tándem, los volúmenes de las conexiones entre las columnas insertadas antes de la columna en la que migra el soluto introducirían un segundo retraso. Un tercer retraso está asociado al tiempo que necesita el frente del eluyente para

alcanzar la posición del soluto desde la entrada de la columna (retraso intra-columnar) [18-20].

El retraso intra-columnar es difícil de implementar cuando se lleva a cabo la integración numérica, en contraste con otros tipos de retrasos que pueden incorporarse fácilmente en la solución de la ecuación fundamental de la elución en gradiente (Ecuaciones (2.10) y (2.11)). La dificultad radica en que se necesita monitorizar la posición real del soluto dentro de la columna a lo largo del programa de gradiente, lo que implica evaluar la magnitud de la integral a lo largo de las múltiples etapas infinitesimales sucesivas. Esto se puede corregir modificando la Ecuación (2.11) de la siguiente manera:

$$t_0 - t_{\text{ext}} = \frac{t_D}{k_0} + \frac{t_1 - t_D}{k(\varphi(t_1 - \tau_1))} + \frac{t_2 - t_1}{k(\varphi(t_2 - \tau_2))} + \dots$$

$$+ \frac{t_i - t_{i-1}}{k(\varphi(t_i - \tau_i))} + \frac{t_{i+1} - t_i}{k(\varphi(t_{i+1} - \tau_{i+1}))}$$
(2.12)

donde cada uno de los términos τ_i es la suma acumulada hasta el término i :

$$\tau_i = \frac{t_D}{k_0} + \sum_{j=1}^i \frac{t_{j+1} - t_j}{k(\varphi(t_j - \tau_j))}$$
(2.13)

Así, el tiempo de retención en gradiente (t_g) vendrá dado por la suma $t_i + t_0$, siendo t_i el tiempo transcurrido a lo largo del gradiente en la Ecuación (2.12), en el que la suma total coincide con $t_0 - t_{\text{ext}}$. La corrección intra-columnar depende del soluto y debe implementarse mediante iteraciones en la integración numérica. La razón de ello es que la posición del soluto, calculada a lo largo del programa de gradiente, ocurre antes de lo que se calcula sin considerar el retraso. Esto implica que debe calcularse un nuevo retraso con la posición corregida. Las sucesivas correcciones tienden a cero, convergiendo el proceso rápidamente.

En caso de llevar a cabo la integración analítica, el retraso intra-columnar se considera intrínsecamente a través de los límites de la integral, siempre y cuando el gradiente posea una única rampa lineal. Sin embargo, cuando se utilizan gradientes multi-lineales, los límites inferiores de los términos de la integral, asociados a cada segmento lineal deben corregirse considerando la fracción de columna migrada al inicio de cada segmento.

2.3.4. Ajuste del modelo de retención y predicciones de tiempo de retención para experimentos no isocráticos

Cuando el conjunto de entrenamiento incluye exclusivamente datos isocráticos, el modelo de retención puede ajustarse directamente utilizando procedimientos lineales o no lineales, dependiendo del modelo de retención seleccionado (véanse las Ecuaciones (2.1), (2.3) y (2.4)). Cuando los datos de origen que se ajustan son tiempos de retención en gradiente, el cálculo de los parámetros del modelo para cada soluto es más complejo. La situación más simple corresponde a los modelos de retención para los cuales la ecuación fundamental (Ecuación (2.5)) posee solución analítica, y t_g se puede despejar, como ocurre con las Ecuaciones (2.1) y (2.4), siempre que el gradiente sea lineal (Ecuaciones (2.6), (2.8) y (2.9)). En tales casos, la solución incluye los parámetros que definen cada gradiente del conjunto de entrenamiento (la pendiente y la intersección, m y t_0 , respectivamente), el tiempo de permanencia del sistema, los tiempos muerto y extra-columnar, y los parámetros del modelo que describe el comportamiento del soluto (por ejemplo, k_w y S en la Ecuación (2.1)). Para obtener los parámetros del modelo, se requiere un ajuste no lineal, donde se monitoriza la concordancia entre los valores de t_g predichos y experimentales, a medida que se alteran los parámetros del modelo utilizando un algoritmo que busca la máxima similitud.

En este estudio, el proceso de ajuste de los parámetros del modelo se ha controlado mediante el algoritmo de Powell [21]. Para diseños experimentales constituidos por cinco programas de elución, una evaluación de alta precisión (hasta alcanzar la precisión interna de la máquina) necesitó típicamente de 50 iteraciones.

2.4. Parte experimental

2.4.1. Reactivos

Para este estudio, se consideró la separación de 14 sulfonamidas (Sigma, Roedermark, Alemania): (1) sulfaguanidina, (2) sulfanilamida, (3) sulfadiazina, (4) sulfatiazol, (5) sulfapiridina, (6) sulfamerazina, (7) sulfametazina, (8) sulfametizol, (9) sulfamonometoxina, (10) sulfacloropiridazina, (11) sulfametoxazol, (12) sulfisoxazol, (13) sulfadimetoxina, y (14) sulfaquinoxalina. Para cada sulfonamida se preparó una disolución madre de $100 \mu\text{g}\cdot\text{mL}^{-1}$, utilizando agua nanopura (obtenida con un sistema de purificación de Adrona B30 Trace, Burladingen, Alemania), para lo que se utilizó un baño de ultrasonidos (Elmasonic, Singen, Alemania). A partir de ella, se hicieron las diluciones oportunas para obtener áreas de pico similares para todas las sulfonamidas.

La separación cromatográfica se llevó a cabo, tanto en modo isocrático como en modo gradiente, utilizando fases móviles preparadas con acetonitrilo grado HPLC (Scharlau, Barcelona) y una disolución acuosa de dihidrógenofosfato de sodio anhidro (Fluka, Alemania) 0.01 M preparado con agua nanopura. El pH se fijó a 3.0 mediante la adición de cantidades adecuadas de HCl y NaOH 0.1 M (Scharlau). El conjunto de entrenamiento consistió en cinco experimentos isocráticos realizados con fases móviles que contenían

acetonitrilo al 10, 13, 16, 20 y 25 % (v/v). Se llevaron a cabo inyecciones duplicadas, con el fin de comprobar la reproducibilidad de las medidas.

Antes de su inyección en el sistema cromatográfico, todas las disoluciones se filtraron a través de membranas de Nylon de 0.45 μm (Micron Separations, Westboro, MA, EE.UU.).

2.4.2. Instrumentación y columna cromatográfica

El análisis se realizó con un equipo cromatográfico modelo HP1100 (Agilent, Waldbronn, Alemania), compuesto por los siguientes módulos: bomba cuaternaria, autoinyector equipado de viales de 2 mL, compartimento de columna termostatzado y detector UV-Vis ajustado a una longitud de onda de medida de 254 nm. El volumen de inyección fue de 20 μL y el flujo de fase móvil se mantuvo constante a 1.0 mL \cdot min⁻¹. Para realizar los análisis, se utilizó una columna Zorbax Eclipse XDB-C18 (150 \times 4.6 mm), con un tamaño de partícula de 5 μm (Agilent). Todas las inyecciones se llevaron a cabo en condiciones de temperatura controlada, a 25 °C. Se determinó el tiempo muerto del sistema para diferentes composiciones de la fase móvil mediante la inyección de KBr (Acros Organics, Fair Lawn, NJ, EE.UU.). El tiempo de permanencia (1.16 min) se midió haciendo uso de un gradiente de acetona. El tiempo extra-columnar fue de 0.12 min.

La medida de pH de las fases móviles se realizó haciendo uso de un pH-metro (modelo MicropH 2002, Crison, Barcelona), que utilizaba un electrodo de membrana de vidrio (modelo 8102, Orion, Barcelona), un electrodo de referencia de Ag/AgCl y una disolución de KCl 3.0 M en el puente salino.

2.4.3. Software

Para la adquisición de las señales, se utilizó una estación de trabajo OpenLAB CDS LC (Agilent, revisión B.04.03). Para la medida de las propiedades de pico (tiempo de retención y semianchuras, entre otras) se hizo uso del software MICHROM [22]. El tratamiento de datos se llevó a cabo con funciones desarrolladas por el grupo de investigación en Matlab 2016b (The MathWorks Inc., Natick, MA, EE.UU.).

2.5. Resultados y discusión

2.5.1. Selección del modelo de retención

Para evaluar la calidad del ajuste proporcionado por las Ecuaciones (2.1), (2.3) y (2.4), se calcularon los siguientes estadísticos [23,24]:

(i) *Error relativo medio*

$$ER = \frac{\sum_{i=1}^{ne} |\hat{k}_i - k_{\text{exp},i}|}{\sum_{i=1}^{ne} \hat{k}_{\text{exp},i}} \times 100 \quad (2.14)$$

(ii) *Coficiente de correlación convencional*

$$R = \sqrt{1 - \frac{\sum_{i=1}^{ne} (\hat{k}_i - k_{\text{exp},i})^2}{\sum_{i=1}^{ne} (k_{\text{exp},i} - \bar{k}_{\text{exp},i})^2}} \quad (2.15)$$

(iii) *Coefficiente de correlación ajustado*

$$R_{aj} = \sqrt{1 - (1 - R^2) \times \frac{ne - 1}{ne - np - 1}} \quad (2.16)$$

(iv) *Estadístico F de Snedecor*

$$F = \frac{\sum_{i=1}^{ne} (\hat{k}_i - \hat{k}_{medio})^2}{\sum_{i=1}^{ne} (k_{exp,i} - \hat{k}_i)^2} \quad (2.17)$$

$$ne - np$$

En las Ecuaciones (2.14) a (2.17), ne es el número de experimentos (fases móviles) en el diseño experimental, np es el número de parámetros del modelo, $k_{exp,i}$ y \hat{k}_i son los factores de retención experimental y predicho para el experimento i , y \hat{k}_{medio} y \bar{k}_{exp} son los factores de retención experimental y predicho promedio, respectivamente. Cabe señalar que, a pesar de su uso frecuente, el coeficiente de correlación no es apropiado para comparar modelos que involucran un número diferente de parámetros.

La Tabla 2.1 muestra el rendimiento de los modelos de retención considerados, en base a los criterios estadísticos antes mencionados. Entre ellos, el modelo de Snyder (Ecuación (2.1), con dos parámetros, ofreció las predicciones más deficientes, con errores relativos del orden del 4-5 %. Con los modelos de Schoenmakers (Ecuación (2.3)) y NK (Ecuación (2.4)), los dos con tres parámetros, el rendimiento fue excelente, observándose errores de predicción generalmente comprendidos en el intervalo de 0.3 a 0.5 %, ambos con una capacidad de predicción casi idéntica.

Tabla 2.1. Estadísticos que miden la calidad de ajuste correspondientes a tres modelos de retención para el conjunto de 14 sulfonamidas, considerando un diseño experimental con cinco fases móviles isocráticas (10, 13, 16, 20 y 25 % de acetonitrilo (v/v)). Para la sulfadimetoxina y sulfaquinoxalina, cuya retención con una fase móvil conteniendo 10 % de acetonitrilo excedió las 2.5 horas, el diseño incluyó sólo las otras cuatro fases móviles.

Soluta	Estadísticos	Modelo de retención		
		Ec. (2.1)	Ec. (2.3)	Ec. (2.4)
Sulfaguanidina	<i>ER</i>	5.15	0.24	0.67
	<i>R</i>	0.99719	0.99999	0.99999
	<i>R_{aj}</i>	0.99437	0.99999	0.99994
	<i>F Snedecor</i>	78.2	16953.3	2345.1
Sulfanilamida	<i>ER</i>	3.96	0.45	0.14
	<i>R</i>	0.99821	0.99999	0.99999
	<i>R_{aj}</i>	0.99641	0.99998	1.00000
	<i>F Snedecor</i>	85.4	3976.7	38684.4
Sulfadiazina	<i>ER</i>	4.89	0.46	0.24
	<i>R</i>	0.99769	0.99999	0.99999
	<i>R_{aj}</i>	0.99538	0.99998	0.99999
	<i>F Snedecor</i>	204.0	13348.3	43689.3
Sulfatiazol	<i>ER</i>	5.87	0.46	0.22
	<i>R</i>	0.99745	0.99999	0.99999
	<i>R_{aj}</i>	0.99489	0.99998	0.99999
	<i>F Snedecor</i>	264.5	21293.8	89193.9
Sulfapiridina	<i>ER</i>	5.34	0.47	0.56
	<i>R</i>	0.99769	0.99999	0.99999
	<i>R_{aj}</i>	0.99538	0.99997	0.99996
	<i>F Snedecor</i>	266.9	17500.4	11159.9
Sulfamerazina	<i>ER</i>	5.15	0.61	0.67
	<i>R</i>	0.99776	0.99999	0.99998
	<i>R_{aj}</i>	0.99553	0.99996	0.99994
	<i>F Snedecor</i>	260.7	9923.0	7339.8
Sulfametazina	<i>ER</i>	5.48	0.58	0.69
	<i>R</i>	0.99782	0.99999	0.99998
	<i>R_{aj}</i>	0.99564	0.99996	0.99994
	<i>F Snedecor</i>	318.3	14451.6	9226.3

Tabla 2.1 (continuación).

Solutos	Estadísticos	Modelo de retención		
		Ec. (2.1)	Ec. (2.3)	Ec. (2.4)
Sulfametizol	<i>ER</i>	5.43	0.38	0.31
	<i>R</i>	0.99808	0.99999	0.99999
	<i>R_{aj}</i>	0.99615	0.99998	0.99999
	<i>F</i> Snedecor	403.9	38320.1	57137.0
Sulfamonometoxina	<i>ER</i>	5.07	0.58	0.68
	<i>R</i>	0.99843	0.99999	0.99999
	<i>R_{aj}</i>	0.99686	0.99997	0.99995
	<i>F</i> Snedecor	518.9	20065.7	13698.5
Sulfacloropiridazina	<i>ER</i>	4.66	0.37	0.30
	<i>R</i>	0.99852	0.99999	0.99999
	<i>R_{aj}</i>	0.99704	0.99998	0.99999
	<i>F</i> Snedecor	486.5	33863.1	54464.7
Sulfametoxazol	<i>ER</i>	4.38	0.49	0.48
	<i>R</i>	0.99873	0.99999	0.99999
	<i>R_{aj}</i>	0.99746	0.99997	0.99997
	<i>F</i> Snedecor	576.8	21244.6	20922.2
Sulfisoxazol	<i>ER</i>	4.29	0.41	0.45
	<i>R</i>	0.99887	0.99999	0.99999
	<i>R_{aj}</i>	0.99774	0.99998	0.99998
	<i>F</i> Snedecor	709.1	39899.2	30554.2
Sulfadimetoxina	<i>ER</i>	4.00	0.92	0.64
	<i>R</i>	0.99927	0.99999	0.99999
	<i>R_{aj}</i>	0.99781	a	a
	<i>F</i> Snedecor	411.6	3800.4	7706.0
Sulfaquinoxalina	<i>ER</i>	3.51	0.64	0.42
	<i>R</i>	0.99946	0.99999	0.99999
	<i>R_{aj}</i>	0.99838	a	a
	<i>F</i> Snedecor	595.8	8723.1	19996.4

^a Dato no computable debido a la falta de grados de libertad: $n - p - 1 = 0$ para la Ecuación (2.16).

Finalmente, se seleccionó el modelo NK por su buen comportamiento en la realización de extrapolaciones, la baja incertidumbre en la estimación de los parámetros del modelo, y la posibilidad de obtener una solución analítica relativamente simple de la ecuación fundamental (Ecuaciones (2.8) y (2.9)). Cabe resaltar que la simulación de cromatogramas se llevó a cabo considerando también la predicción de los perfiles de pico, de acuerdo con los desarrollos descritos en la Referencia [25].

2.5.2. Efecto de los pulsos de modificador orgánico sobre la retención y la eficacia

Con el fin de comprender el mecanismo de elución en LC utilizando gradientes, Dolan sugirió la siguiente analogía [26]: “un soluto se asienta en la cabeza de la columna hasta que aparece un disolvente lo suficientemente fuerte para empujarlo a través de ella, dejando atrás a los otros solutos; a partir de ese momento el soluto viaja hacia la salida de la columna con bastante rapidez”. Esta analogía ayudará a comprender las explicaciones siguientes.

En elución en gradiente, la afinidad de los solutos por la fase móvil se ve favorecida a medida que aumenta la fuerza eluyente, requiriéndose en ocasiones un programa complejo que se adapte a los requisitos de los solutos [27]. En la práctica cromatográfica, en ocasiones, dichos gradientes contienen segmentos con pequeña pendiente, seguidos de fuertes incrementos en el contenido de modificador orgánico. En este caso, los solutos sufren una fuerte aceleración, lo que beneficia la reducción de los tiempos de retención de los solutos ya separados. Consideremos lo que sucedería si esta alta concentración de modificador orgánico se mantiene constante durante un tiempo determinado, para recuperar a continuación la concentración anterior al súbito aumento, dando lugar a un aumento transitorio de la concentración de modificador

orgánico (véase, por ejemplo, los programas de elución representados en los cromatogramas de las Figuras 2.1b a 2.1d). En este caso particular, los solutos de alta hidrofobicidad, que eluyen posteriormente al aumento transitorio, pasarían de moverse muy rápido a hacerlo de nuevo lentamente. El alcance de la magnitud del efecto dependerá de la hidrofobicidad del soluto. Para mayor brevedad, en adelante los aumentos transitorios de modificador orgánico serán referidos como "pulsos".

Debe indicarse que los efectos descritos a continuación serían indeseables con fines de optimización, debido a que se produce una mayor superposición de los picos cromatográficos, especialmente en la región del pulso (o en la inmediatamente posterior), y a que los solutos sufren reducciones importantes de la eficacia en la región inmediata al pulso, como consecuencia de la disminución en la velocidad de migración. Sin embargo, como se mostrará más adelante, el uso de pulsos de modificador orgánico es beneficioso en la modelización de la retención. El efecto obtenido es similar a trasladar en bloque, a tiempos de retención más bajos, compuestos que eluyen tardíamente. Es decir, es como si se aplicara al soluto un impulso en una región intermedia, haciendo uso de un tipo de gradiente escalonado con cambios de pendiente repentinos, positivos y negativos. Los pulsos pueden considerarse como un caso especial de elución multi-isocrática.

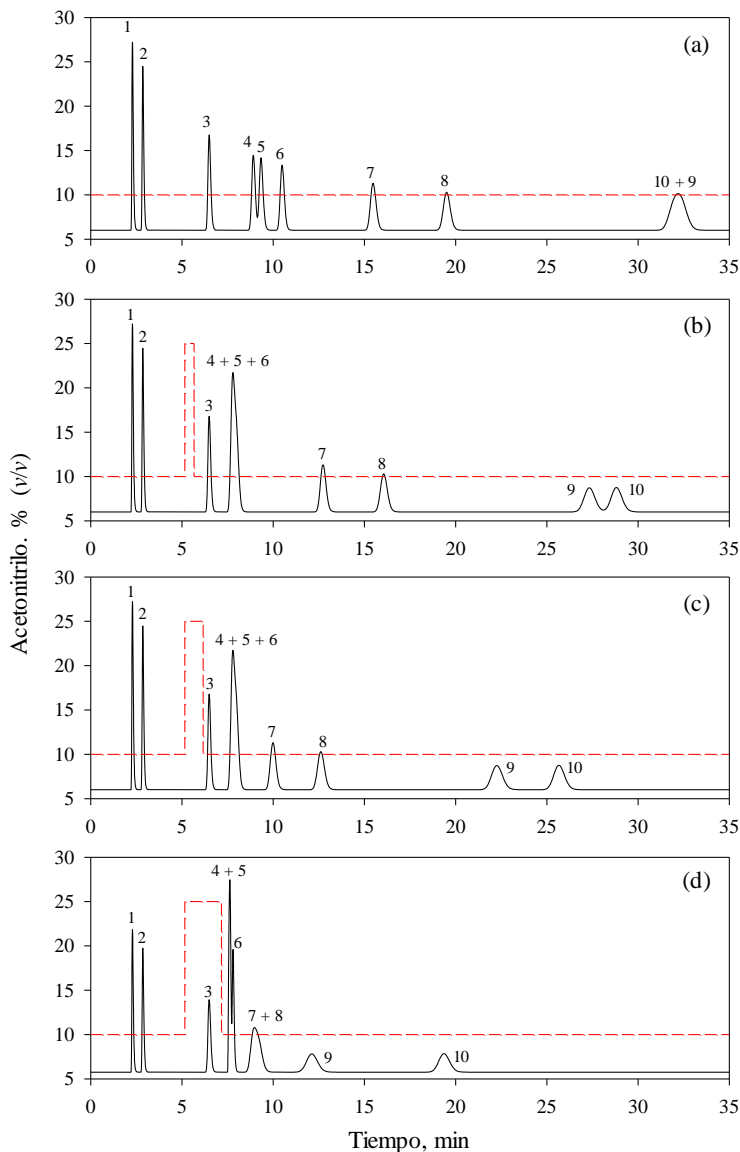


Figura 2.1. Efecto de la aplicación de un pulso de duración creciente en la separación de una mezcla de 10 sulfonamidas: (a) elución isocrática con un 10 % de acetonitrilo, y pulsos donde la concentración de acetonitrilo se incrementa al 25 %, con una duración de: (b) 0.5 min, (c) 1.0 min, y (d) 2.0 min. Todos los pulsos se aplicaron a los 5.16 min (4.0 min + 1.16 min de tiempo de permanencia). La identidad de las señales y otros detalles se encuentran en la Sección 2.4.1.

La realización de separaciones con ayuda de pulsos constituye una interesante posibilidad en la obtención de información cromatográfica, para solutos altamente hidrofóbicos eluidos a bajos contenidos de modificador orgánico (el contenido de modificador previo y posterior al pulso), debido a la reducción significativa de sus tiempos de retención. Cabe tener en cuenta que, a lo largo de un gradiente convencional, las concentraciones de modificador más altas que experimenta un soluto son las únicas que contribuyen significativamente a su migración, mientras que el efecto de las concentraciones más bajas es depreciable, excepto para los solutos que eluyen a mayor velocidad. En otras palabras, el uso de pulsos proporciona información sobre el comportamiento de los solutos en las fases móviles más lentas del diseño experimental, que de otro modo darían lugar a tiempos de retención excesivos para los solutos más hidrofóbicos. No se puede prescindir de esta información, que puede ser decisiva en la separación de algunos solutos. Por otro lado, debe indicarse que diseños experimentales con intervalos estrechos de modificador orgánico dan lugar a la obtención de parámetros del modelo de retención afectados de una mayor incertidumbre.

Con el fin de explicar el efecto del uso de pulsos de modificador orgánico sobre la elución de una mezcla compuesta por un conjunto de 10 sulfonamidas, en las Figuras 2.1 y 2.2 se muestran dos secuencias de cromatogramas simulados. En cada cromatograma, la concentración de modificador se mantiene al 10 %, antes y después del pulso en el que se incrementa bruscamente al 25 %. En la primera secuencia (Figura 2.1), el pulso se aplica a un tiempo $t = 4.0$ min y la duración del pulso se incrementa a lo largo de los experimentos posteriores, desde 0.5 a 2.0 min.

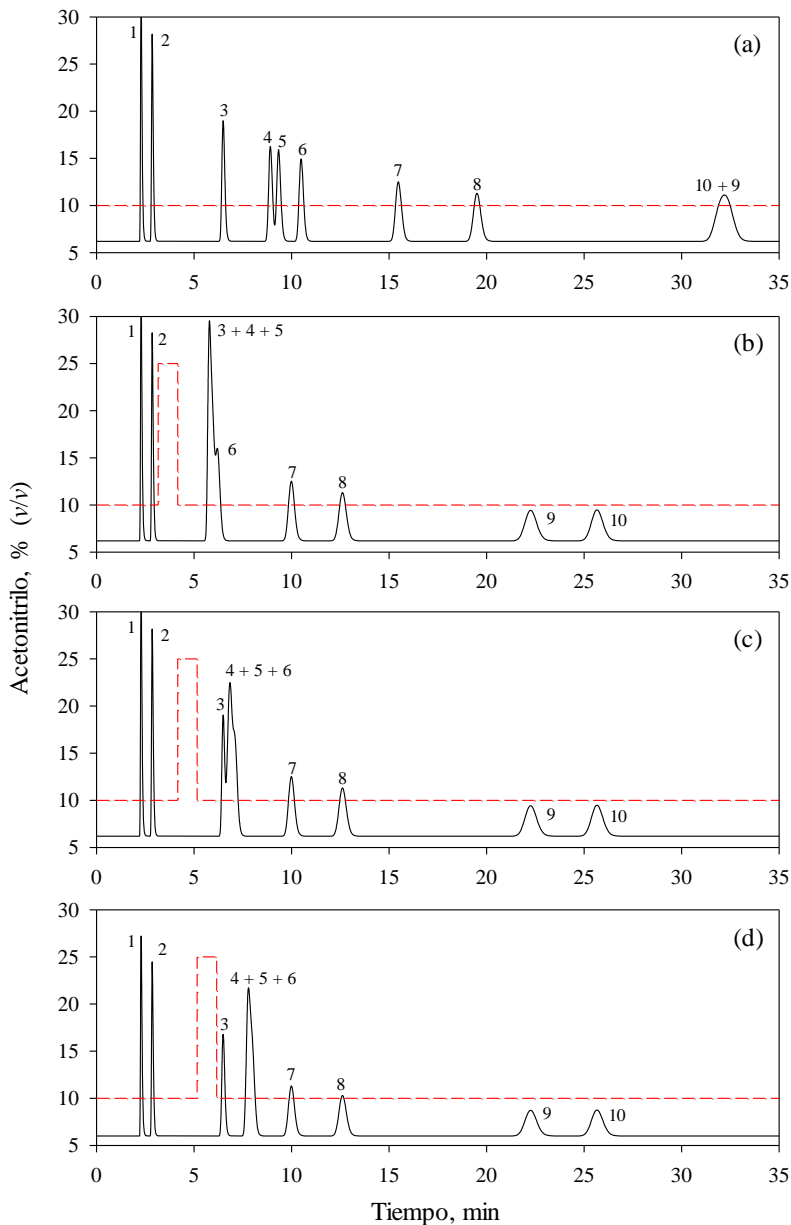


Figura 2.2. Efecto del desplazamiento de un pulso de 1 min de duración en la separación de la mezcla de 10 sulfonamidas. El pulso comienza a los: (b) 2 min, (c) 3 min y (d) 4 min (a lo que debe añadirse el tiempo de permanencia, 1.16 min). En (a) se muestra, con fines comparativos, la elución isocrática para 10 % de acetonitrilo. Otros detalles se encuentran en la Sección 2.4.1.

En la segunda secuencia (Figura 2.2), un pulso de 1.0 min de duración se desplaza gradualmente desde 2.0 a 4.0 min. Cabe señalar que los programas de elución se trazan teniendo en cuenta el retraso debido al volumen de permanencia (es decir, en el cromatograma se ha representado la ubicación real del pulso a la entrada de la columna). Con fines comparativos, las Figuras 2.1a y 2.2a muestran el resultado de la elución en ausencia de un pulso, utilizando una fase móvil isocrática con un contenido de acetonitrilo del 10 %, es decir, la concentración que se mantiene en los demás cromatogramas antes y después del pulso. En todos los casos, como era de esperar, los dos primeros solutos (solutos 1 y 2), que eluyen antes del pulso, no se ven afectados por éste. Los detalles relativos a la simulación de cromatogramas se dan en las Secciones 2.3.1 a 2.3.3.

La Figura 2.1 muestra que los solutos que eluyen tras el pulso acortan su tiempo de elución respecto al obtenido en ausencia de un pulso. Esta reducción en la retención es considerablemente mayor para los compuestos más lentos (solutos 7 a 10), los cuales se ven afectados de forma más significativa al aumentar la duración del pulso (cuanto más prolongada es la retención, mayor es la reducción). Aunque el trabajo expuesto en este capítulo se centra únicamente en la modelización de la retención, se puede observar que los pulsos también tienen efectos sobre la selectividad. Así, los solutos 9 y 10, que co-eluyen en ausencia de pulso, se resuelven aceptablemente cuando se aplica un pulso. Además, se observa una modificación en la resolución de los solutos 4 a 6. Téngase en cuenta que el pulso afecta a los solutos después de un retraso adicional debido al tiempo muerto. Por lo tanto, en los cromatogramas, los efectos de un pulso serán perceptibles tras un tiempo $t_p + t_D + t_0$, donde t_p es el tiempo medido al inicio del pulso programado.

Cuando un pulso de duración fija se desplaza a tiempos más largos (Figuras 2.2b a 2.2d), los solutos más retenidos no se ven afectados. Tampoco se ve afectado el tiempo de retención de un soluto que eluye muy cerca del final del pulso (véase el soluto 3 en las Figuras 2.2 c y 2.2d). Por el contrario, se puede comprobar que los cuatro solutos intermedios experimentan importantes variaciones en su elución.

2.5.3. Modelización de tiempos de retención

2.5.3.1. Desviación de las predicciones respecto a los cromatogramas experimentales

Con el fin de inspeccionar con más detalle el efecto de los pulsos de modificador orgánico en cromatogramas reales, se diseñaron cuatro experimentos con el conjunto de 10 sulfonamidas (Figura 2.3). Se ejecutaron pulsos de 1 min de duración, desplazados gradualmente a tiempos más elevados. En cada análisis, tras el pulso, la elución finalizó con un gradiente de acetonitrilo del 10 al 25 % en el intervalo de 10 a 12 min (a lo que se sumaría el tiempo de permanencia). Si la elución isocrática se hubiera mantenido tras el pulso (sin aplicar un gradiente a continuación), la elución de los solutos 8-10 habría alcanzado un tiempo de aproximadamente 25 min (como se observa en la Figura 2.2b). Cabe señalar que la señal correspondiente a la sulfonamida menos retenida, la sulfaguanidina, aparece dividida en dos picos debido a la degradación del compuesto.

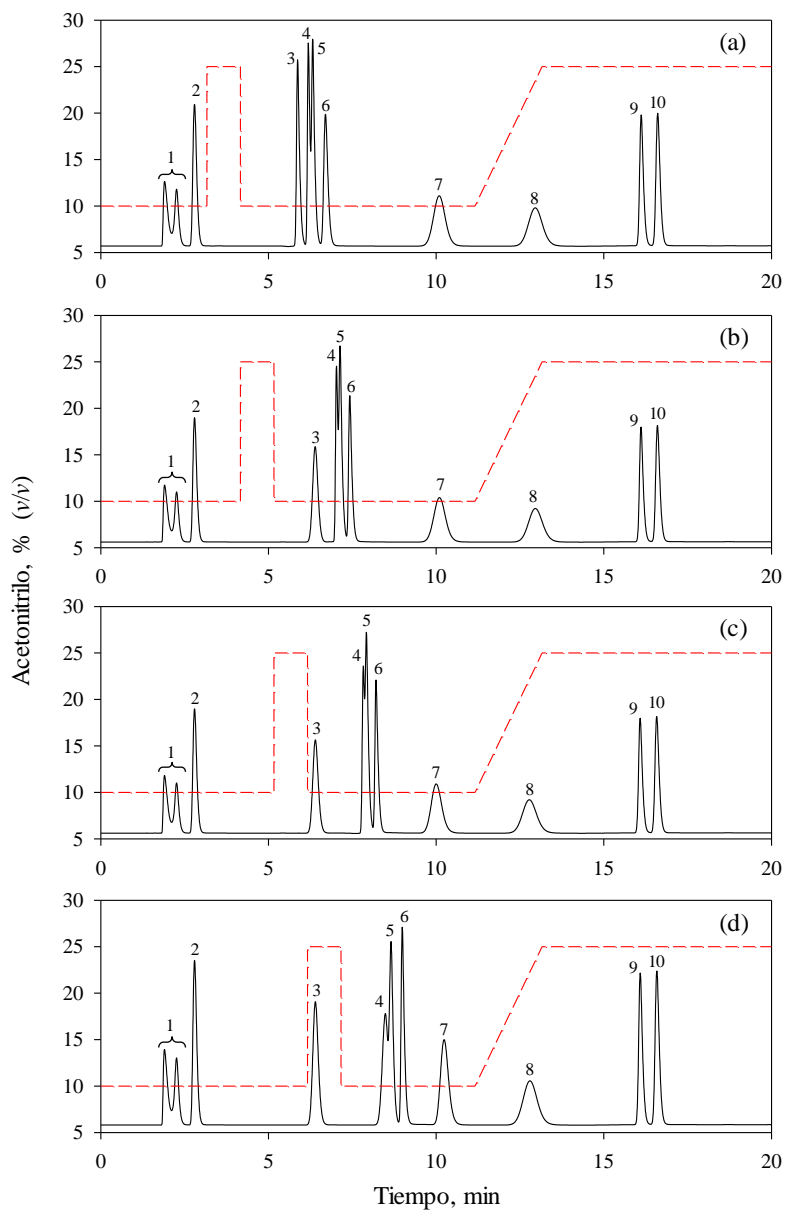


Figura 2.3. Cromatogramas experimentales mostrando el efecto del desplazamiento de un pulso de 1 min de duración, en la separación de una mezcla de 10 sulfonamidas. El pulso comienza a los: (a) 2 min, (b) 3 min, (c) 4 min y (d) 5 min (a lo que debe añadirse el t_D de 1.16 min). Entre los 10 y 12 min se aplica un gradiente rápido. Véase la Sección 2.4.1 para conocer la identidad de los solutos.

Además, debe tenerse en cuenta que no hay ninguna consecuencia en la línea base asociada a la aplicación, tanto del pulso como del gradiente rápido, debido al mismo nivel de concentración del tampón preparado con el sistema fosfato 0.01 M, utilizado en las dos disoluciones que se mezclan (acetonitrilo al 10 y al 25 %) para generar los pulsos y los gradientes. Se comprobó que, en general, la fluctuación de la presión de la columna debido al cambio transitorio de modificador orgánico en la fase móvil variaba entre 119 y 136 bares (véase la Figura 2.4). Como se observa en la secuencia de cromatogramas simulados obtenidos para la elución isocrática (Figura 2.2), la posición donde se aplica el pulso no afecta a los picos que eluyen suficientemente alejados del aumento transitorio. Sólo se ven afectados los picos que eluyen en su vecindad, ya sea dentro o tras el pulso, dependiendo de su duración.

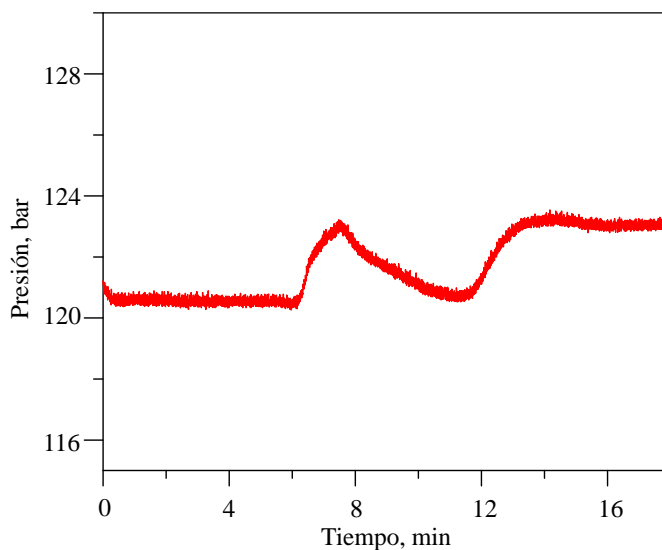


Figura 2.4. Perfil de la presión ejercida sobre la columna a lo largo de la aplicación del pulso mostrado en la Figura 2.5c.

Las Figuras 2.5a y 2.5b muestran los cromatogramas predichos obtenidos al aplicar un pulso de modificador orgánico a los 5.16 min (4 min + tiempo de permanencia). La Figura 2.5c corresponde al cromatograma experimental obtenido en el laboratorio para esta misma situación. Las predicciones se llevaron a cabo de acuerdo con el modelo NK, utilizando dos diseños experimentales que contenían como conjunto de entrenamiento: (i) los datos obtenidos con las cinco fases móviles isocráticas del diseño de entrenamiento para construir los modelos (Figura 2.5a), y (ii) un conjunto mixto en el que los datos obtenidos con los cuatro experimentos que involucraban un pulso (mostrados en la Figura 2.3) se procesaron conjuntamente con los cinco experimentos isocráticos (Figura 2.5b).

En principio, y tal y como se muestra en la Tabla 2.1, el modelo NK garantiza un error muy bajo en la predicción de los tiempos de retención. Teniendo esto en cuenta, las deficientes predicciones de los tiempos de retención, en presencia de cambios bruscos en la concentración de modificador orgánico (Figura 2.5), llaman fuertemente la atención. Podría pensarse que el conjunto de entrenamiento formado por fases móviles isocráticas no proporciona suficiente información para este tipo de predicciones. Sin embargo, la incorporación, al diseño experimental, de la información de los cromatogramas obtenidos aplicando pulsos, es decir, el uso del diseño mixto, apenas mejoró la situación (Figura 2.5b). Esto sugiere que es necesario considerar otro tipo de corrección en la integración numérica, con el fin de mejorar las predicciones de retención de solutos que eluyen en las proximidades de los pulsos.

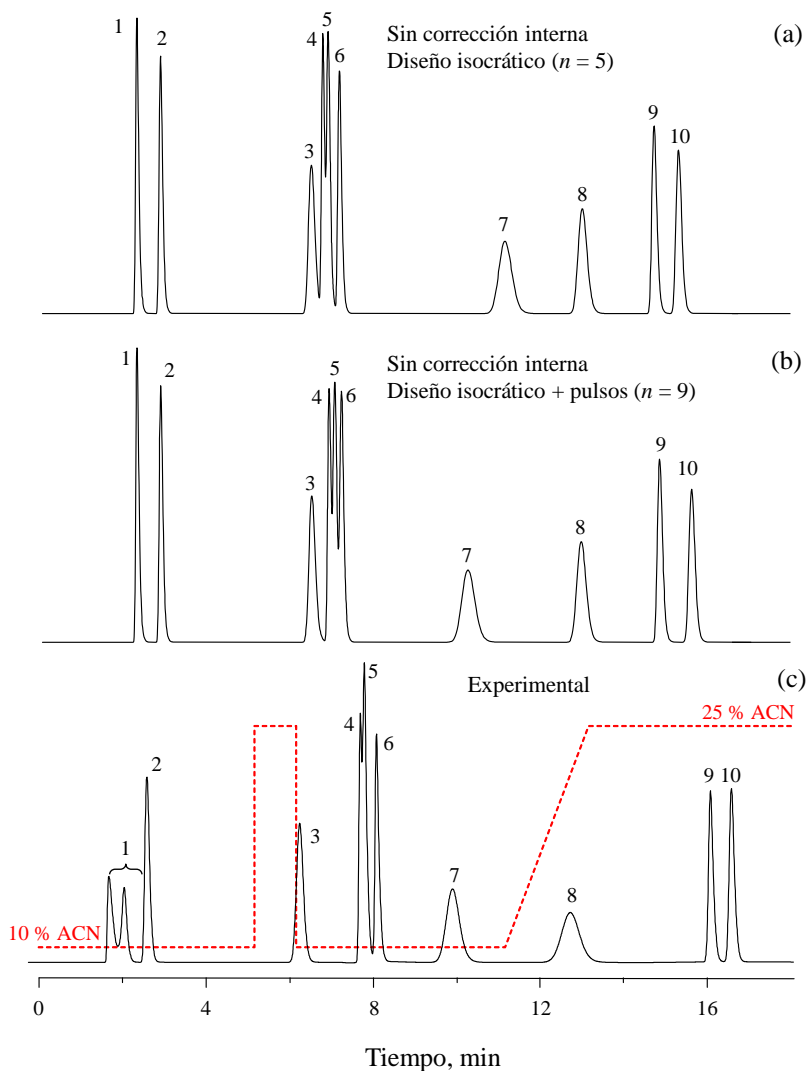


Figura 2.5. Cromatogramas predichos (a,b) y experimental (c), utilizando el programa de elución representado en (c) para la separación de una mezcla de 10 sulfonamidas. Las predicciones se realizaron con conjuntos de entrenamiento constituidos por: (a) los cinco experimentos isocráticos indicados en la Sección 2.4.1, y (b) esos cinco experimentos junto con los cuatro pulsos mostrados en la Figura 2.3. El programa de elución consistió en un pulso de 1 min de duración aplicado a los 4 min, seguido de un gradiente rápido entre los 10 y 12 min (al que debe añadirse el tiempo de permanencia de 1.16 min). Otros detalles se detallan en la Sección 2.4.1.

2.5.3.2. *Corrección de tiempos de retención en la integración numérica debido a la migración de los solutos en el interior de la columna*

Uno de los objetivos de esta investigación es la mejora de las predicciones en condiciones críticas de elución, tales como gradientes que incluyen cambios bruscos en la concentración de modificador, o situaciones en las que participan composiciones extremas, como es el caso de los pulsos. Debe tenerse en cuenta que, cuando un cambio en la composición de la fase móvil (es decir, un cambio en el programa de gradiente o un aumento transitorio) llega a la entrada de la columna, sus efectos requieren un pequeño tiempo adicional para alcanzar la vecindad del soluto, debido a la distancia recorrida por éste desde la entrada. Así, como se comentó en la Sección 2.3.3, la composición instantánea que experimenta el soluto incluye dos retrasos: uno de ellos es el tiempo de permanencia (asociado a la distancia recorrida por la fase móvil desde el mezclador hasta la entrada de la columna); el segundo es un retraso intra-columnar que tiene en cuenta la discrepancia gradual entre el programa de gradiente y la composición instantánea en la posición a la que se encuentra el soluto. Naturalmente, la magnitud de dicho retraso intra-columnar aumenta a lo largo de la migración del soluto; de ahí que se incremente la necesidad de realizar una corrección (el retraso máximo es t_0 y se encuentra a la salida de la columna).

Como se ha comentado, el retraso intra-columnar se tiene en cuenta en la integración analítica para gradientes con una sola rampa, puesto que ésta se basa en la aplicación de la regla de Barrow. Esta regla establece que la integral se calcula en dos momentos de la elución, uno de ellos donde el retraso intra-columnar es t_D (el límite inferior de la integral), y el otro a la salida de la columna, donde dicho retraso es $t_g - t_0$ (el límite superior de la integral en la Ecuación (2.5)). Por el contrario, la integración numérica requiere corregir cada

término infinitesimal hasta el momento en el que el soluto abandona la columna. En este trabajo, el retraso intra-columnar hasta una determinada ubicación en la columna se ha incorporado en la predicción de los tiempos de retención, a través de las Ecuaciones (2.12) y (2.13). Cabe señalar que la verdadera variable en la función $k(\varphi(t))$ es la concentración del modificador orgánico y sólo cuando el retraso implica una variación en su contenido, la corrección es significativa [20].

Con el fin de evaluar las mejoras obtenidas en las predicciones de la retención, tras introducir el retraso intra-columnar, se plantearon dos escenarios diferentes: cromatogramas obtenidos mediante un gradiente lineal o multi-lineal (Figura 2.6) y mediante la aplicación de un pulso (Figura 2.7). En cada caso, se comparan las predicciones sin tener en cuenta la corrección intra-columnar en la integración numérica (Figuras 2.6a y 2.6d), y considerando dicha corrección (Figuras 2.6b y 2.6e, y 2.7a y 2.7b), respecto a los resultados obtenidos en el cromatograma experimental (Figuras 2.6c y 2.6f, y 2.7c).

En el primer escenario (Figura 2.6), donde los cambios de composición de la fase móvil son graduales y crecientes, los efectos del retraso intra-columnar no fueron significativos, ya que debido a que los cambios no son bruscos, la diferencia del contenido de modificador orgánico a la entrada de la columna no difiere apreciablemente del existente en la vecindad del soluto. Por lo tanto, la corrección intra-columnar no se tradujo en grandes diferencias en la predicción de los tiempos de retención. Este hecho explica por qué este tipo de corrección no se considera en la integración numérica para gradientes.

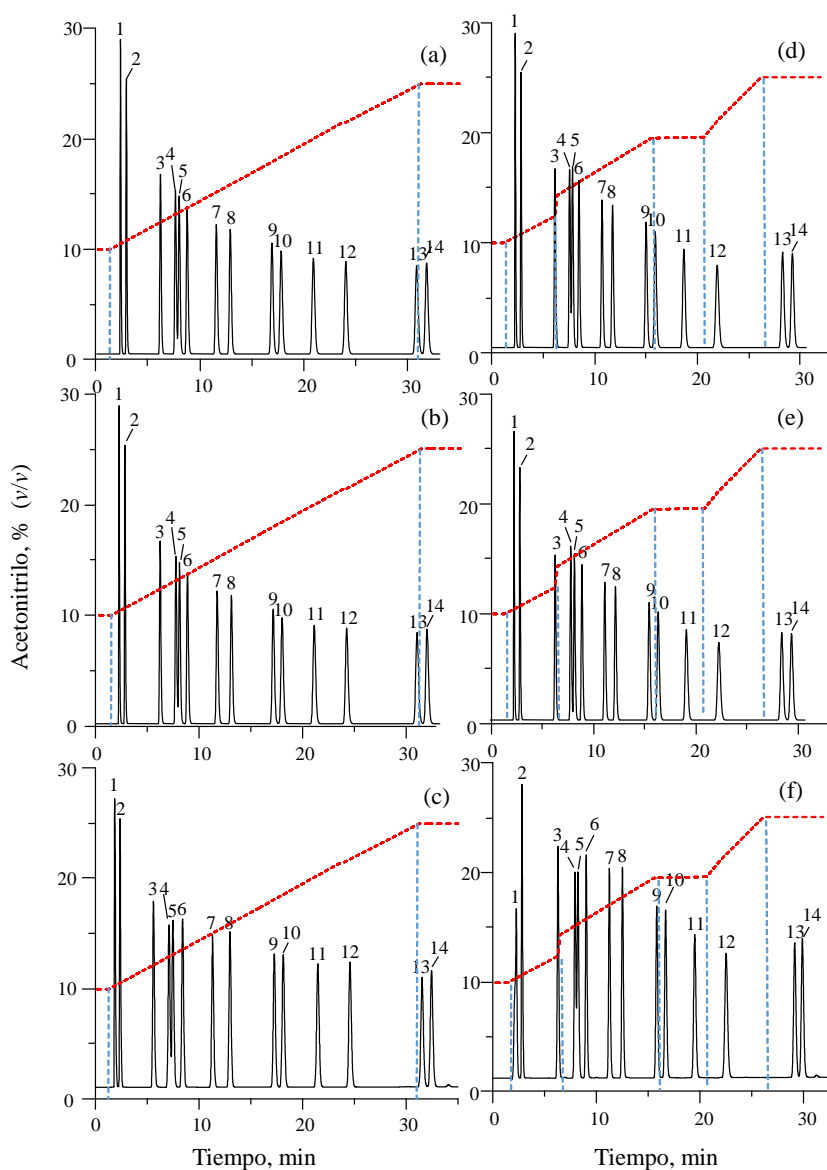


Figura 2.6. Cromatogramas predichos (a,b,d,e) y experimentales (c,f), correspondientes a la separación de la mezcla de 10 sulfonamidas, haciendo uso de un gradiente lineal (a,b,c) y multi-lineal (d,e,f). Las predicciones se realizaron mediante integración numérica, (a) sin aplicar y (b) aplicando la corrección intra-columnar. El programa de gradiente utilizado se ha superpuesto en los cromatogramas.

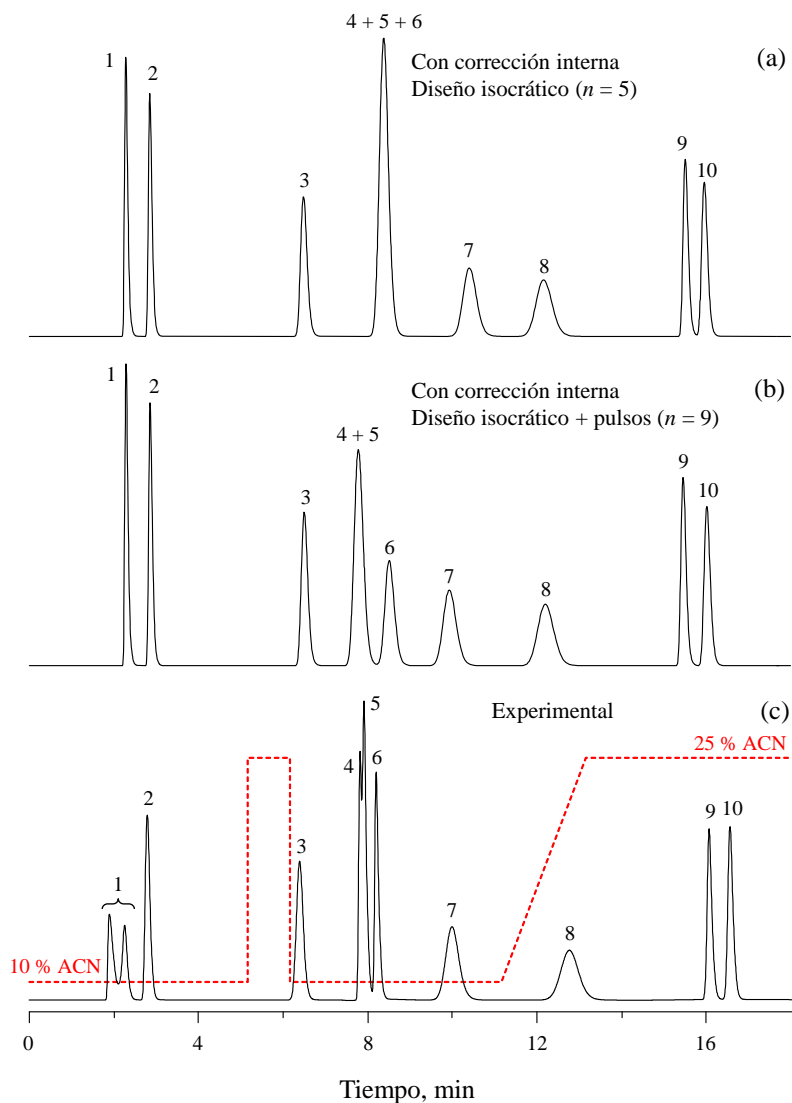


Figura 2.7. Cromatogramas predichos (a,b) y experimental (c) para la separación de una mezcla de 10 sulfonamidas. Las predicciones se realizaron mediante integración numérica, utilizando conjuntos de entrenamiento constituidos por: (a) los cinco experimentos isocráticos indicados en la Sección 2.4.1, y (b) los cinco experimentos isocráticos conjuntamente con los cuatro pulsos mostrados en la Figura 2.3. En este caso, y en contraposición a los cromatogramas mostrados en la Figura 2.5, las predicciones se llevaron a cabo implementando la corrección intra-columnar. Compárense los resultados obtenidos con los mostrados en la Figura 2.5.

A pesar de esto, si los cromatogramas de la Figura 2.6 se inspeccionan en detalle, las discrepancias y el rendimiento de las correcciones son perceptibles (véase, por ejemplo, los picos de los solutos 11 y 12 en las Figuras 2.6d, 2.6e y 2.6f; la representación de líneas verticales discontinuas delimitando las transiciones entre los segmentos lineales del gradiente sirven de referencia para evaluar las diferencias en los tiempos de retención).

En el segundo escenario, en el que se aplica un pulso en un cromatograma que termina con un gradiente lineal, las mejoras en los cromatogramas predichos, tras la aplicación de la corrección intra-columnar (Figura 2.7), son apreciables respecto a las predicciones realizadas sin aplicar la corrección (Figura 2.5). Esto puede explicarse teniendo en cuenta que, cuando se producen cambios súbitos en el contenido de modificador orgánico en la fase móvil, la velocidad del soluto dentro de la columna puede variar drásticamente entre las dos composiciones extremas que afectan a la vecindad del soluto: la velocidad puede ser muy lenta cuando el soluto migra experimentando la composición de modificador más baja en el pulso, y muy rápida cuando lo hace a la composición más elevada. La diferencia de velocidad en la proximidad del cambio en la concentración de modificador orgánico, al final del pulso, puede ser crítica para solutos de polaridad próxima, conduciendo a importantes errores de predicción, si se ignora la corrección intra-columnar.

La Figura 2.7 muestra las consecuencias de incluir la corrección intra-columnar en el segundo escenario, tanto en las predicciones obtenidas con el conjunto de entrenamiento constituido por cinco experimentos isocráticos, como con el diseño extendido que incluye, además de los experimentos isocráticos, los cuatro experimentos con pulsos (mostrados en las Figuras 2.3a a 2.3d). Nótese que la corrección intra-columnar no necesita incluir en el diseño

experimental la información obtenida con las experiencias con pulsos, con el fin de originar buenas predicciones para la mayoría de los picos. La concordancia entre los cromatogramas experimental y predicho es muy satisfactoria, aunque no perfecta. Problemas como los asociados con deformaciones en el perfil del pulso, efectos de difusión, o de re-equilibrio de la fase estacionaria, pueden contribuir a estas pequeñas variaciones.

La Figura 2.8 complementa la información de la Figura 2.7, mostrando los cromatogramas simulados considerando los retrasos intra-columnares, que deben compararse con los cromatogramas experimentales análogos de la Figura 2.3. Debe mencionarse que los cromatogramas predichos de la Figura 2.8 se obtuvieron haciendo uso del conjunto de entrenamiento de cinco experimentos isocráticos. Además, debe indicarse que los perfiles de pico se predijeron utilizando la aproximación de Jandera [28], que tal como se observa, tiende a sobreestimar la anchura de pico para los solutos que eluyen cerca del pulso, siendo la predicción para los otros picos satisfactoria.

Finalmente, debe resaltarse la importancia de utilizar un paso suficientemente pequeño en la integración numérica, para predecir cromatogramas que incluyen pulsos. En este trabajo, se ha utilizado un paso de integración de 0.0001 min. La Figura 2.9 muestra las consecuencias en las predicciones al utilizar pasos de integración variables, en el intervalo comprendido entre 10^{-5} y 0.1 min. Un valor de 0.01 min suele ser una buena opción para gradientes lineales y multi-lineales convencionales. En presencia de fuertes transiciones, especialmente a concentraciones decrecientes, este tamaño de paso no es lo suficientemente pequeño, dando lugar a predicciones erróneas para picos que eluyen críticamente cerca del final del pulso aplicado.

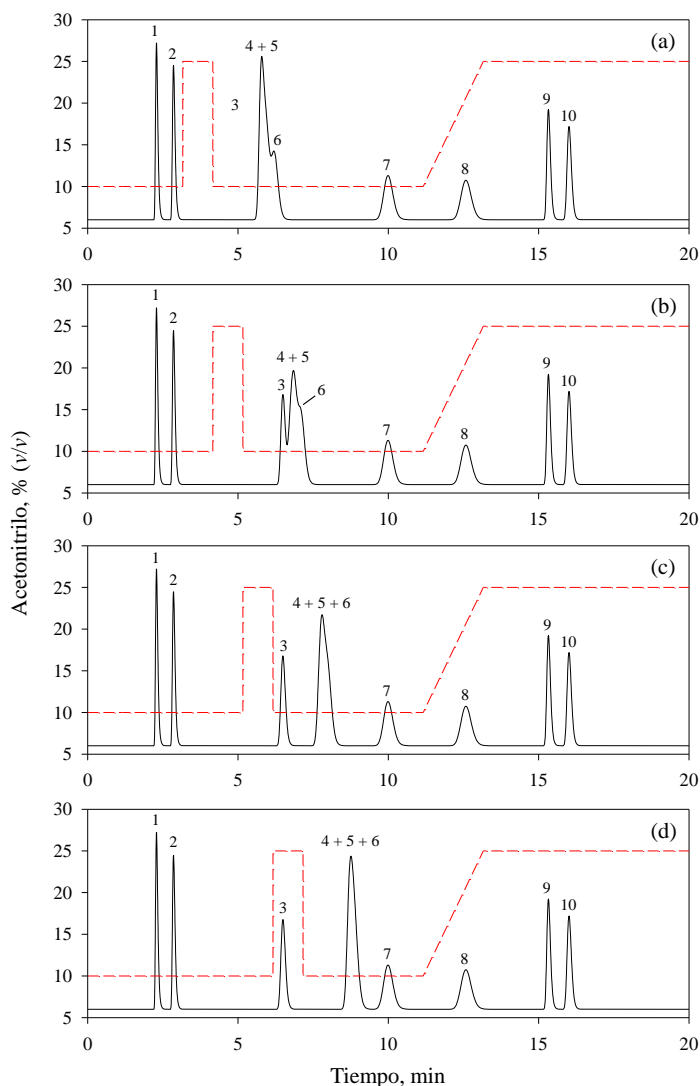


Figura 2.8. Cromatogramas predichos que muestran el efecto del desplazamiento de un pulso de 1 min de duración, en la separación de una mezcla de 10 sulfonamidas, iniciándose a los: (a) 2 min, (b) 3 min, (c) 4 min y (d) 5 min. A estos tiempos ha de añadirse el tiempo de permanencia (1.16 min). Además del pulso, se aplicó un gradiente rápido en el periodo entre 10 y 12 min. Los cromatogramas se predijeron utilizando integración numérica y el conjunto de entrenamiento compuesto por los cinco experimentos isocráticos, introduciendo además la corrección intra-columnar. Compárese con los cromatogramas experimentales correspondientes mostrados en la Figura 2.3.

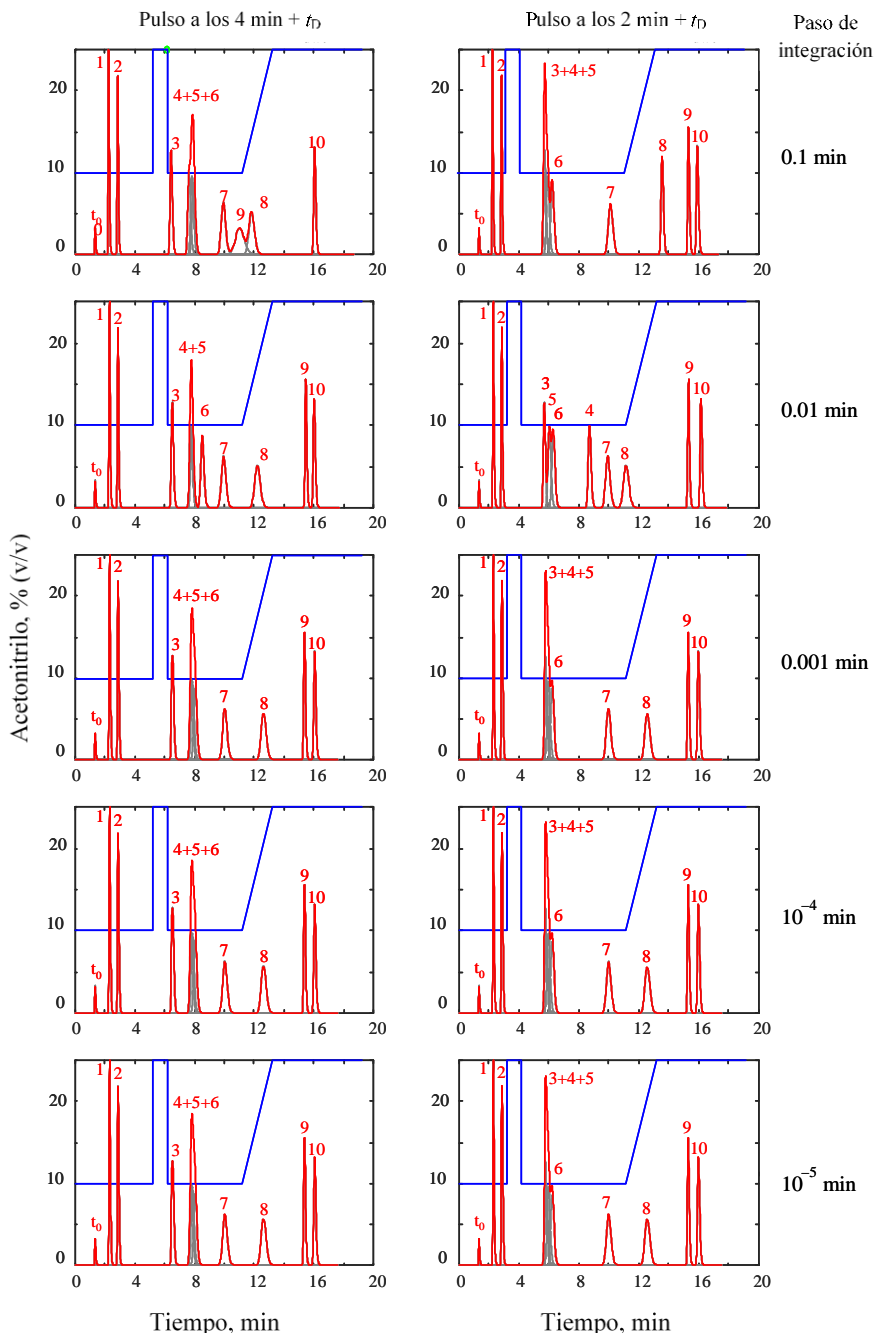


Figura 2.9. Efecto de la magnitud del paso de integración sobre la calidad de las predicciones.

2.5.4. Calidad en la predicción de los tiempos de retención para diseños experimentales que involucran pulsos

Como se comentó anteriormente, el uso de pulsos da lugar a una reducción significativa de los tiempos de retención para los solutos más hidrofóbicos, cuando éstos eluyen en condiciones puramente isocráticas (Figuras 2.1 y 2.2). En esta sección, se evalúa el rendimiento de diseños experimentales mixtos constituidos por fases isocráticas puras, a los que se añaden pulsos, lo que se compara con el ofrecido por diseños que incluyen gradientes lineales convencionales, cuyo propósito es también la reducción de los tiempos de análisis. Debe mencionarse que varios paquetes de software disponibles comercialmente implementan tratamientos orientados a conjuntos de entrenamiento constituidos por gradientes lineales simples, con fines de modelización, generalmente utilizando el modelo de retención logarítmico lineal (Ecuación (2.1)). En este capítulo, los datos obtenidos, utilizando diversos conjuntos de entrenamiento, se han ajustado de acuerdo al modelo NK (Ecuación (2.4)), que ofrece un error bajo en amplios intervalos de concentración de modificador orgánico en la fase móvil, tal y como se comentó en la Sección 2.5.1. El software aplicado fue desarrollado por el grupo de investigación, lo que permitió una mayor flexibilidad en el diseño de experimentos y realización de cálculos, respecto al uso de software comercial, así como una comparación más justa, ya que no es posible conocer todos los detalles de la arquitectura de un software comercial, lo que puede conducir a conclusiones erróneas.

Cuando se lleva a cabo un ajuste no lineal, es importante elegir valores iniciales apropiados para los parámetros del modelo, para evitar soluciones colapsadas en un óptimo local. Inicialmente, los ajustes se realizaron asignando valores positivos arbitrarios, idénticos para todos los solutos. Los parámetros

encontrados para aquellos solutos en los que la regresión tuvo éxito se utilizaron, en un siguiente análisis, como valores iniciales para los solutos que fallaron. Por lo general, se utilizaron como valores predeterminados $k_w = 800$, $c = 4$, y $B = 50$, para modelizar la retención de solutos típicos, utilizando diseños favorables y ajustando los datos con el modelo NK (Ecuación (2.4)). Para los solutos más hidrofóbicos, utilizando diseños menos informativos, los ajustes requirieron valores de k_w más elevados (por ejemplo, $k_w = 5000$).

2.5.4.1. Validación y diseños experimentales utilizados

La evaluación del rendimiento de diseños experimentales utilizados en la modelización de la retención, llevada a cabo en este trabajo, considera varios aspectos:

- (i) Verificar la calidad de los parámetros de los modelos de retención ajustados utilizando diseños experimentales que incluyen pulsos, respecto a los obtenidos a partir de diseños puramente isocráticos, que son los más informativos.
- (ii) Evaluar la capacidad predictiva de los modelos obtenidos a partir de diferentes diseños experimentales que incluyen pulsos o gradientes. La evaluación implica tanto la predicción de los datos de entrenamiento (los utilizados en el ajuste de los modelos de retención), como predicciones externas con fases móviles isocráticas de baja fuerza eluyente. Dichas composiciones participan, en menor medida, al comienzo de los gradientes, traducándose en errores más importantes.
- (iii) Evaluar si los diseños que incluyen pulsos son competitivos en cuanto al consumo de disolvente orgánico y al tiempo de análisis, en comparación con los diseños puramente isocráticos y aquellos que contienen exclusivamente gradientes.

Para realizar este estudio, se inspeccionaron los dos solutos más lentos del conjunto de 14 sulfonamidas (sulfadimetoxina y sulfaquinoxalina, solutos 13 y 14, véanse las Tablas 2.2 y 2.3), que son los más afectados por el uso de pulsos. Para los solutos 7 a 12, los beneficios son menores. La medida de los tiempos de retención para los solutos 13 y 14, analizados mediante elución isocrática, resultó problemática, ya que a la concentración más baja del diseño experimental (acetonitrilo al 10 %), ascendieron hasta alrededor de 2.5 y 3 horas, respectivamente.

El estudio tuvo en cuenta, por un lado, el diseño isocrático utilizado en el laboratorio, que incluía cinco fases móviles (acetonitrilo al 10, 13, 16, 20 y 25 %, Sección 2.4.1), y por otro, los siguientes cuatro diseños (Figura 2.10):

- (i) El primer diseño es el más simple (Figura 2.10a), ya que sólo inserta en el experimento más lento del diseño isocrático base, un pulso de 10 min de duración, comenzando a los 20 min (a lo que hay que añadir el tiempo de permanencia). A lo largo de estos primeros 20 min, los solutos más rápidos eluyen isocráticamente al 10 % de acetonitrilo, proporcionando medidas de calidad para estos solutos. Los solutos más lentos también se benefician, ya que el pulso reduce significativamente sus tiempos de elución al 10 % de acetonitrilo.
- (ii) El segundo diseño contiene tres experimentos isocráticos y dos pulsos de 12 y 15 min de duración (Figura 2.10b) que, como en el caso anterior, comienzan a los 20 min. La duración del pulso se estableció de modo que el tiempo de retención del soluto más lento en ambos experimentos estuviera próximo a una hora. Esto implica una reducción a un tercio del tiempo que se obtendría si la elución fuera puramente isocrática al 10 % de acetonitrilo.

- (iii) El tercer diseño incluye cinco gradientes lineales, donde cuatro de ellos alcanzan la concentración máxima de modificador orgánico, ϕ_{final} (25, 20, 15 y 13 % v/v) en 60 min, mientras que para el quinto se alcanza el 25 % (v/v) en 40 min. El gradiente más lento es prácticamente isocrático, por lo que es uno de los diseños de gradiente más favorables en términos informativos.
- (iv) El cuarto diseño consta de cinco gradientes en el intervalo de acetonitrilo 10-25 %, con un tiempo de gradiente (t_G) variable. Con ello se consigue una mayor reducción del tiempo de análisis respecto al diseño (iii).

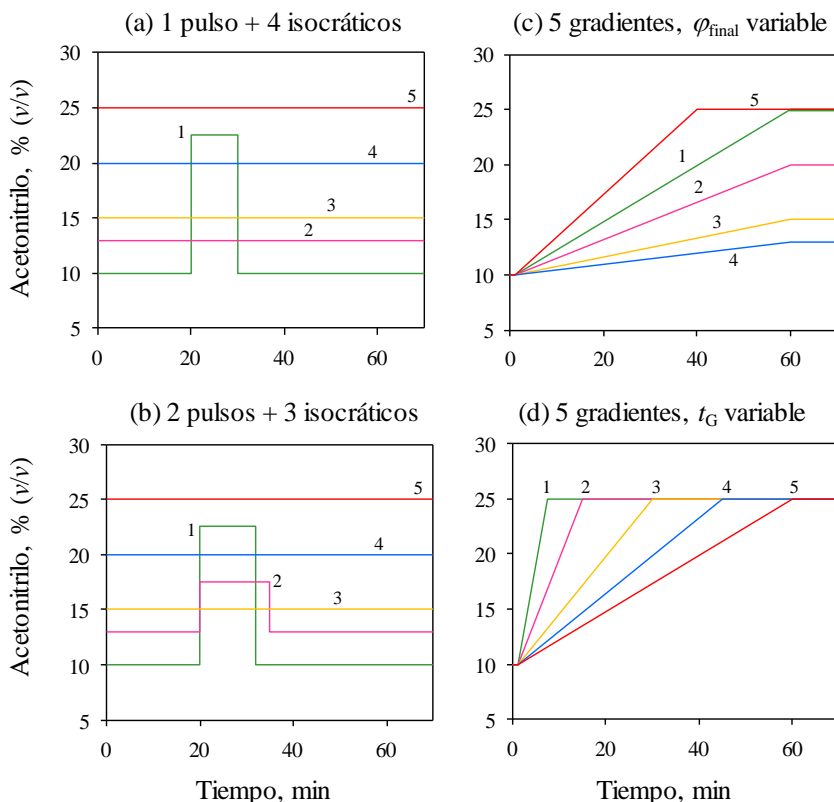


Figura 2.10. Diseños experimentales utilizados para obtener los parámetros del modelo de retención para los solutos 13 y 14. Para una descripción detallada de los diseños, consúltese la Sección 2.5.4.1.

Tabla 2.2. Rendimiento predictivo de las dos sulfonamidas más hidrofóbicas (ver Sección 2.5.4) y consumo de acetonitrilo para los experimentos de los conjuntos de entrenamiento.^{a,b}

Diseño	Experimento	Sulfadimetoxina <i>f_{R,objetivo}</i>	Sulfaquinoxalina <i>f_{R,objetivo}</i>	Residuo Acetonitrilo, mL
5 isocráticos	1	154.91	180.32	18.03
	2	76.17	88.82	11.55
	3	51.06	58.74	8.81
	4	22.83	24.70	4.94
	5	12.57	12.58	3.15
Tiempo de análisis acumulado / 5 experimentos, min				365.16
Residuo total, mL				46.48
1 pulso + 4 isocráticos	1	69.33	83.70	9.62
	2	76.17	88.82	11.55
	3	51.06	58.74	8.81
	4	22.83	24.70	4.94
	5	12.57	12.57	3.15
Tiempo de análisis acumulado / 5 experimentos, min				268.54
Residuo total, mL				38.07
2 pulsos + 3 isocráticos	1	52.23	64.38	7.94
	2	56.44	67.72	9.48
	3	51.06	58.74	8.81
	4	22.83	24.70	4.94
	5	12.57	12.58	3.15
Tiempo de análisis acumulado / 5 experimentos, min				228.13
Residuo total, mL				34.33

Tabla 2.2. (continuación)

Diseño	Experimento	Sulfadimetoxina		Sulfaquinoxalina		Residuo	
		$t_{R,objetivo}$	$t_{R,encentrado}$	$t_{R,objetivo}$	$t_{R,encentrado}$	Acetonitrilo, mL	
5 gradientes, variable	1	43.81	43.81	46.05	46.05	7.18	
	2	53.48	53.49	56.89	56.90	8.33	
	3	74.10	74.12	82.15	82.17	10.79	
	4	92.97	92.96	105.67	105.66	12.82	
	5	35.33	35.31	36.68	36.67	6.11	
		Tiempo de análisis acumulado / 5 experimentos, min					
		Residuo total, mL					
5 gradientes, t_G variable	1	17.27	17.27	17.48	17.47	3.72	
	2	21.57	21.58	22.00	22.00	4.29	
	3	30.13	30.15	31.00	31.02	5.42	
	4	37.65	37.66	39.23	39.25	6.41	
	5	43.81	43.78	46.05	46.04	7.18	
		Tiempo de análisis acumulado / 5 experimentos, min					
		Residuo total, mL					
		55.77					
		27.02					

^a Para el diseño isocrático, la concentración de acetonitrilo fue de 10, 13, 16, 20 y 25 % (v/v) para los experimentos 1 – 5, respectivamente. Para otros diseños véase la Figura 2.10.

^b Véase la Sección 2.5.4 para conocer el significado de $t_{R,objetivo}$ y $t_{R,encentrado}$

Tabla 2.3. Parámetros del modelo de retención de Neue-Kuss (Ecuación (2.4)), obtenidos a partir de los diseños experimentales mostrados en la Figura 2.10 y tiempos de retención predichos haciendo uso de los parámetros del modelo, para los dos solutos más hidrofóbicos eluidos en las condiciones de elución isocráticas más lentas (10 y 13 % (v/v) de acetonitrilo).

Diseños	Parámetros de Neue-Kuss			t_R isocrático, min	
	k_w	c	B	10% acetonitrilo	13% acetonitrilo
Sulfadimetoxina					
5 experimentos isocráticos	6617 ±179	4.52 ±0.06	68.2 ±0.79	154.9	76.2
1 pulso + 4 isocráticos	6673 ±321	4.53 ±0.7	68.4 ±1.1	155.0	76.2
2 pulsos + 3 isocráticos	6590 ±637	4.51 ±0.13	68.1 ±2.1	154.8	76.2
5 gradientes, φ variable	6106 ±816	4.37 ±0.25	66.1 ±3.5	154.4	76.2
5 gradientes, t_G variable	4172 ±3210	3.88 ±1.08	58.3 ±16.2	147.4	75.0
Sulfaquinoxalina					
5 experimentos isocráticos	4816 ±88.6	3.30 ±0.04	53.9 ±0.50	180.3	88.9
1 pulso + 4 isocráticos	4817 ±159	3.30 ±0.04	53.9 ±0.67	180.4	88.8
2 pulsos + 3 isocráticos	4752 ±300	3.29 ±0.08	53.7 ±1.22	180.1	88.8
5 gradientes, φ variable	4618 ±399	3.23 ±0.16	52.9 ±2.1	179.9	88.8
5 gradientes, t_G variable	2930 ±2233	2.65 ±1.02	44.5 ±13.9	169.0	86.9

2.5.4.2. Metodología

Con el fin de obtener resultados comparables con los diferentes diseños presentados en la Figura 2.10, se aplicó la siguiente metodología:

En una primera etapa, se obtuvieron para cada soluto los parámetros k_w , c y B del modelo NK (Ecuación (2.4)), mediante regresión no lineal. En estos ajustes, se utilizaron los tiempos de retención medidos con los cinco experimentos isocráticos empleados como conjunto de entrenamiento (véase la parte superior de la Tabla 2.2). Como se comentó anteriormente, los estadísticos de regresión se muestran en la Tabla 2.1. Por su parte, la Tabla 2.3 muestra los parámetros del modelo NK para sulfadimetoxina y sulfaquinoxalina, obtenidos a partir de los datos experimentales isocráticos. Estos parámetros son los más exactos que podrían obtenerse y se utilizarán para evaluar los resultados ofrecidos por los otros diseños.

Los modelos obtenidos con los datos isocráticos se utilizaron para predecir los tiempos de retención de las dos sulfonamidas, para cada uno de los diseños de la Figura 2.10 ($t_{R,objetivo}$ en la Tabla 2.2). Para obtener estos valores se aplicó integración numérica, tal y como se explica en las Secciones 2.3.2 y 2.3.3, utilizando un paso de integración de 10^{-5} min, muy por debajo del error experimental. De ahí que las incertidumbres asociadas a la integración fueran despreciables. Ello permitió utilizar los valores de $t_{R,objetivo}$ en cálculos posteriores, como si éstos fueran datos experimentales. También permitió realizar una comparación adecuada del rendimiento de los diferentes diseños. A continuación, se utilizaron los valores de $t_{R,objetivo}$ obtenidos para cada diseño como datos de entrenamiento, con el fin de estimar los parámetros del modelo NK correspondientes, y con ellos predecir los tiempos de retención de las dos sulfonamidas más hidrofóbicas, eluidas de acuerdo a los experimentos de cada diseño. Los cálculos se llevaron a cabo de acuerdo a la metodología explicada

en las Secciones 2.3.2 a 2.3.4. En la Tabla 2.2 se indica el tiempo utilizado en términos de tiempo de análisis. Naturalmente, este tiempo dependerá de la forma en la que se inyecten los solutos (por ejemplo, individualmente, o por grupos compatibles de solutos en cada inyección). El tiempo de computación para el ajuste fue de aproximadamente 2 min por soluto.

2.5.4.3. Comparación del rendimiento obtenido con los distintos diseños

En la Tabla 2.3, se comparan los parámetros del modelo NK obtenidos ajustando los valores de $t_{R,objetivo}$ para los cuatro diseños de la Figura 2.10, con los parámetros evaluados a partir de los datos isocráticos experimentales. Como se puede observar, los diseños con pulsos proporcionaron parámetros del modelo prácticamente coincidentes con los obtenidos a partir de los experimentos isocráticos, siendo el diseño que contenía un único pulso el que ofreció los resultados más exactos (Figura 2.10a). Los diseños con gradientes ofrecieron parámetros más discrepantes, particularmente el diseño que incluía los gradientes más rápidos (Figura 2.10d). La Tabla 2.3 también proporciona los errores estándar de los parámetros estimados para cada diseño. A partir de estos valores, se puede concluir que no sólo los parámetros ofrecidos por los diseños que incluyen experimentos con pulsos son más coincidentes con los obtenidos a partir de datos puramente isocráticos (cuando se comparan con los obtenidos con datos de gradiente), sino que también los intervalos de confianza son más estrechos, denotando en conjunto resultados de mejor calidad.

Como puede observarse en la Tabla 2.2, los valores de $t_{R,objetivo}$ se reprodujeron perfectamente a partir de los parámetros del modelo ajustados para cada diseño, con discrepancias entre los valores objetivo y los encontrados ($t_{R,objetivo}$ y $t_{R,encontrado}$), generalmente próximas a 0.01 min, e incluso por debajo de este valor. Esto significa que los modelos ajustados con los datos de cada

diseño predican sus propios datos de entrenamiento con excelente exactitud. Sin embargo, ello no implica una capacidad predictiva similar para experimentos fuera del dominio (especialmente para eluyentes muy lentos), u obtenidos con otros diseños. La Tabla 2.3 indica el tiempo de retención esperado para los eluyentes más lentos en el diseño isocrático (para acetonitrilo al 10 y 13 %), utilizando los parámetros del modelo indicados en la tabla. Se puede observar que los diseños con pulsos y el diseño de gradiente de la Figura 2.10c conducen a predicciones extrapoladas con una calidad prácticamente igual a la del diseño isocrático. El diseño de gradiente de la Figura 2.10d proporciona las predicciones extrapoladas más deficientes.

La Tabla 2.2 también incluye los tiempos de análisis asociados a los cinco experimentos, para cada uno de los diseños, así como el consumo de acetonitrilo, utilizando el tiempo de retención del soluto más tardío, la sulfaquinoxalina. Puede observarse que los diseños con pulsos implican tiempos de análisis y consumos de disolvente orgánico intermedios, entre los obtenidos con los dos diseños de gradiente estudiados. Sin embargo, debe tenerse en cuenta que el diseño de gradiente con menor consumo y tiempo de análisis (el diseño de gradiente con t_G variable representado en la Figura 2.10d) fue también el que produjo mayores errores en la estimación de los parámetros de los modelos de retención, y en las predicciones de retención a bajas concentraciones de modificador orgánico (Tabla 2.3). La razón del peor rendimiento de dicho diseño puede atribuirse a la escasa información que proporciona sobre el comportamiento global de la retención, para composiciones de baja fuerza eluyente, que son importantes para la separación. El diseño de la Figura 2.10c (con t_{final} variable) mejora esta insuficiencia, pero los diseños mixtos que incluyen pulsos y experiencias isocráticas mejoran aún más los resultados, especialmente el diseño de la Figura 2.10a.

Cabe señalar que los diseños que contienen pulsos implican reducciones considerables en el tiempo de análisis y un consumo de modificador orgánico algo más bajo, respecto a la elución puramente isocrática. Por lo tanto, considerando los resultados obtenidos a nivel global, es posible concluir que los diseños con pulsos son muy competitivos, respecto a los diseños de gradiente, para la obtención de modelos de retención de gran calidad.

2.6. Conclusiones

En LC, es crucial la obtención de modelos de gran exactitud, con el fin de predecir la retención en el desarrollo de métodos de análisis. Con este propósito, se requieren diseños experimentales elaborados. Entre las diferentes opciones posibles en RPLC, los diseños experimentales isocráticos son los que proporcionan la información más rica sobre el comportamiento de retención de los solutos. Sin embargo, su uso se ve obstaculizado por los largos tiempos de retención que experimentan los compuestos más hidrofóbicos, cuando son eluidos a bajas concentraciones de modificador orgánico. Los diseños experimentales de gradiente son capaces de reducir sustancialmente los largos tiempos de adquisición, pero proporcionan información más pobre sobre el comportamiento de retención cuando se utilizan composiciones de baja fuerza eluyente, que pueden ser decisivas para la separación.

En este capítulo, se exploran las ventajas y desventajas del uso de diseños isocráticos en los que se inserta un incremento transitorio de modificador orgánico (un pulso) en una posición intermedia de los experimentos isocráticos que muestran una menor fuerza eluyente. La aplicación de pulsos en la modelización de la retención tiene como objetivo reducir el tiempo de elución de los solutos lentos, cuando se recopilan los datos de entrenamiento para el ajuste de modelos de retención.

Desde un punto de vista práctico, los diseños mixtos con pulsos se construyen fácilmente reemplazando las fases móviles isocráticas más lentas por experimentos que contienen un pulso insertado a un tiempo intermedio. Esto permite la elución de los solutos más rápidos en el sector previo al pulso, mientras que los solutos más retenidos eluyen tras el pulso, en tiempos aceptables. Si el diseño se lleva a cabo ejecutando los experimentos de mayor a menor fuerza eluyente, el ajuste de los datos en los ensayos isocráticos ya medidos permite seleccionar la duración y posición más adecuadas para el pulso en los experimentos con menor fuerza eluyente. Por otro lado, el ajuste de los modelos de retención utilizando datos obtenidos en experimentos que involucran diseños mixtos con experiencias isocráticas y pulsos es más simple que cuando se utilizan datos de experimentos de gradiente, ya que permite resolver una versión simplificada de la ecuación fundamental de elución de gradiente, debido a la presencia de sectores isocráticos.

El estudio ha revelado que los diseños isocráticos mixtos que incluyen pulsos ofrecen una mejor capacidad predictiva, frente a extrapolaciones, que los diseños experimentales de gradiente, y un tiempo de medida más corto que los diseños puramente isocráticos. Los modelos de retención ajustados, a partir de estos diseños mixtos, prácticamente coinciden con los obtenidos de diseños experimentales isocráticos convencionales. La exactitud en la predicción de los tiempos de retención también es comparable, pero el tiempo de análisis para los compuestos más hidrofóbicos, haciendo uso de los eluyentes más lentos, se reduce considerablemente. Además, este tipo de diseño mixto también es competitivo en términos de consumo de disolvente orgánico.

Los experimentos que implican pulsos mostraron, sin embargo, la existencia de discrepancias en las predicciones, al aplicar integración numérica, para los solutos que eluyen en regiones cercanas al pulso, lo que puede atribuirse

erróneamente a un re-equilibrado insuficiente de la columna. Estas desviaciones, que son insignificantes para los experimentos de gradientes convencionales, se evidencian claramente en experimentos que involucran fuertes alteraciones en el contenido de modificador orgánico, a lo largo del programa de elución (como los que incluyen pulsos). La corrección de tales desviaciones implica la inclusión de retrasos en la llegada de los cambios de composición del gradiente a la posición instantánea del soluto, a lo largo de toda la migración. Una vez que se incorporan tales correcciones intra-columnares en la integración numérica, se obtienen predicciones de retención exactas para los experimentos que involucran pulsos. Sin embargo, para la predicción del perfil de pico haciendo uso de la aproximación de Jandera, se obtuvo una sobreestimación de la anchura para los solutos que eluían cercanos al pulso. Debe indicarse que, en las experiencias con pulsos, la línea base no mostró alteraciones perceptibles tras la introducción de variaciones repentinas en la composición de la fase móvil.

Las predicciones realizadas al ignorar la corrección del desplazamiento intra-columnar, utilizando integración numérica, serán exactas siempre que los solutos no eluyan en las cercanías del pulso. Cuando el objetivo es desarrollar un diseño experimental que incluya pulsos, la ubicación del pulso puede establecerse arbitrariamente, siendo la mejor opción situarlo en una región intermedia del cromatograma que no contenga picos (teniendo en cuenta que sus efectos tardarán un tiempo en desaparecer tras finalizar el pulso: $t_{\text{final}} + t_0$). Para solutos que eluyen demasiado cerca del pulso, puede resultar más práctico desplazar la ubicación del pulso, para evitar la introducción de correcciones intra-columnares. Para los solutos intermedios (como los solutos 7 y 8 de este estudio), que eluyen suficientemente alejados del pulso, los beneficios fueron

similares aunque inferiores a los obtenidos para los solutos más lentos (como los solutos 9 y 10).

Finalmente, vale la pena comentar que, para acortar el tiempo de retención de los solutos lentos, existen enfoques alternativos:

- (i) La recopilación de datos de entrenamiento isocráticos a niveles de concentración más altos (por ejemplo, 15, 20, 25 y 30 % (v/v) frente a 10, 13, 16, 20 y 25 % (v/v), utilizados en este trabajo). Para evaluar esta posibilidad, debe tenerse en cuenta que un diseño debe satisfacer simultáneamente los requisitos de información de solutos de polaridad muy diversa. El desplazamiento del diseño a concentraciones más elevadas será perjudicial para los solutos más rápidos, ya que su elución se aceleraría excesivamente, pudiéndose dar el caso de que co-eluyan en la región del tiempo muerto. Además, la eliminación de los datos para acetonitrilo al 10 y 13 % significa que los gradientes cubiertos por el diseño se iniciarían en una situación desfavorable para los solutos más rápidos, que eluirían en una ventana de tiempo demasiado estrecha.
- (ii) La obtención de los datos de entrenamiento utilizando un flujo de fase móvil más elevado (así, por ejemplo, los tiempos de retención se reducen en un factor de dos si se duplica el flujo). Ésta puede ser una estrategia válida, siempre que los efectos de la difusión térmica sean despreciables y la columna permita una presión mayor. Naturalmente, aumentar el flujo no es mutuamente excluyente con el uso de experimentos con pulsos, que pueden aplicarse a flujos más elevados.
- (iii) La recopilación de datos de entrenamiento haciendo uso de una columna con la misma naturaleza química y diferente diámetro y tamaño de partícula. Por ejemplo, haciendo uso de una columna de 50×2.1 mm con un tamaño de partícula de $1.7 \mu\text{m}$ y un flujo de $0.2 \text{ mL} \cdot \text{min}^{-1}$, el tiempo

de retención para cada soluto se reduce en un factor de aproximadamente 2.88, respecto a la columna original de 150×4.6 mm con un tamaño de partícula de $5 \mu\text{m}$ y un flujo de $1.0 \text{ mL}\cdot\text{min}^{-1}$. Ésta es otra opción válida, siendo el único problema poder disponer en el laboratorio de una columna con la geometría reducida adecuada (y también la bomba apropiada).

2.7. Referencias

- [1] G. Guiochon, The limits of the separation power of unidimensional column liquid chromatography, *J. Chromatogr. A* 1126 (2006) 6–49.
- [2] L.R. Snyder, J.J. Kirkland, J.W. Dolan, *Introduction to Modern Liquid Chromatography*, 2^a ed., John Wiley & Sons, Nueva York, 2011.
- [3] M.C. García Álvarez-Coque, J.J. Baeza Baeza, G. Ramis Ramos, Reversed phase liquid chromatography, en *Analytical Separation Science Series*, Vol. 1 (editado por J. Anderson, A. Berthod, V. Pino, A.M. Stalcup), Wiley, Nueva York, 2015.
- [4] M.C. García Álvarez-Coque, J.R. Torres Lapasió, J.J. Baeza Baeza, Models and objective functions for the optimisation of selectivity in reversed-phase liquid chromatography, *Anal. Chim. Acta* 579 (2006) 125–145.
- [5] R. Cela, E.Y. Ordoñez, J.B. Quintana, R. Rodil, Chemometric-assisted method development in RPLC, *J. Chromatogr. A* 1287 (2013) 2–22.
- [6] P. Jandera, Can the theory of gradient liquid chromatography be useful in solving practical problems?, *J. Chromatogr. A* 1126 (2006) 195–218.
- [7] J.E. Haky, D.A. Teifer, Gradient elution, en *Encyclopedia of Chromatography*, Taylor and Francis, Nueva York, 2006.

- [8] L.R. Snyder, J.W. Dolan, *High-Performance Gradient Elution*, John Wiley & Sons, Hoboken, NJ, 2007.
- [9] J.W. Dolan, L.R. Snyder, Gradient elution chromatography, en *Encyclopedia of Analytical Chemistry* (editado por R.A. Meyers), John Wiley & Sons, Nueva York, 2012.
- [10] J.R. Torres Lapasió, M.C. García Álvarez-Coque, Levels in the interpretive optimisation of selectivity in high-performance liquid chromatography: A magical mystery tour, *J. Chromatogr. A* 1120 (2006) 308–321.
- [11] A. Pappa-Louisi, P. Nikitas, Non-linear least-squares fitting with Microsoft Excel Solver and related routines in HPLC modelling of retention: I. Considerations of the problems of the method, *Chromatographia* 53 (2000) 477–486.
- [12] A. Pappa-Louisi, P. Nikitas, Non-linear least-squares fitting with Excel Solver and related routines in HPLC modelling of retention: II. Considerations of selection of optimal fit, *Chromatographia* 53 (2000) 487–493.
- [13] A. Pappa-Louisi, P. Nikitas, P. Balkatzopoulou, C. Malliakas, Two- and three-parameter equations for representation of retention data in reversed-phase liquid chromatography, *J. Chromatogr. A* 1033 (2004) 29–41.
- [14] P. Nikitas, A. Pappa-Louisi, Retention models for isocratic and gradient elution in reversed-phase liquid chromatography, *J. Chromatogr. A* 1216 (2009) 1737–1755.
- [15] P.J. Schoenmakers, H.A.H. Billiet, L. de Galan, Description of solute retention over the full range of mobile phase compositions in reversed-phase liquid chromatography, *J. Chromatogr.* 282 (1983) 107–121.

-
- [16] U.D. Neue, H.J. Kuss, Improved reversed-phase gradient retention modeling, *J. Chromatogr. A* 1217 (2010) 3794–3803.
- [17] P. Nikitas, A. Pappa-Louisi, Expressions of the fundamental equation of gradient elution and a numerical solution of these equations under any gradient profile, *Anal. Chem.* 77 (2005) 5670–5677.
- [18] J.J. Baeza Baeza, M.C. García Álvarez-Coque, Some insights on the description of gradient elution in reversed phase-liquid chromatography, *J. Sep. Sci.* 37 (2014) 2269–2277.
- [19] C. Ortiz Bolsico, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Optimization of gradient elution with serially-coupled columns. Part I: Single linear gradients, *J. Chromatogr. A* 1350 (2014) 51–60.
- [20] C. Ortiz Bolsico, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Optimisation of gradient elution with serially-coupled columns. Part II: Multi-linear gradients, *J. Chromatogr. A* 1373 (2014) 51–60.
- [21] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3^a ed., Cambridge University Press, Nueva York, 2007.
- [22] J.R. Torres Lapasió, *MICHRUM Software*, Marcel Dekker, Nueva York, 2000.
- [23] I.E. Frank, R. Todeschini, *The Data Analysis Handbook: Data Handling in Science and Technology*, Vol. 14, Elsevier Science, Amsterdam, 1994.
- [24] S.N. Deming, Y. Michotte, D.L. Massart, L. Kaufman, B.G.M. Vandeginste, *Handbook of Chemometrics and Qualimetrics*, Part B, Elsevier Science, Amsterdam, 1998.
- [25] J.A. Navarro Huerta, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Estimation of peak capacity based on peak simulation, *J. Chromatogr. A* 1574 (2018) 101–113.
-

- [26] J.W. Dolan, The hazards of adjusting gradients, *LCGC North Am.* 20 (2002) 940–946.
- [27] V. Concha Herrera, G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Limits of multi-linear gradient optimisation in reversed-phase liquid chromatography, *J. Chromatogr. A* 1063 (2005) 79–88.
- [28] P. Jandera, Predictive calculation methods for optimization of gradient elution using binary and ternary solvent gradients, *J. Chromatogr.* 485 (1989) 113–141.

CAPÍTULO 3

DISEÑOS EXPERIMENTALES EN CROMATOGRAFÍA LÍQUIDA: DESARROLLO Y VALIDACIÓN DE UN MÉTODO PARA LA INSPECCIÓN INTEGRAL DE DISEÑOS EXPERIMENTALES

3.1. Resumen

El propósito de la optimización interpretativa, en cromatografía líquida, es la predicción de la resolución mediante el uso de modelos apropiados que describan la retención de los solutos eluidos. La fiabilidad del proceso depende críticamente de la calidad del diseño experimental utilizado. En este capítulo, se desarrolla, valida y aplica una metodología general destinada a evaluar la calidad de cualquier diseño experimental de entrenamiento utilizado para ajustar modelos de retención. La metodología propuesta se basa en la evaluación sistemática de las incertidumbres asociadas a la predicción de los tiempos de retención, en exploraciones exhaustivas de las condiciones experimentales, tanto en el modo isocrático como de gradiente, siendo capaz de evaluar diseños experimentales de complejidad arbitraria. Para ello, se ha hecho uso de cinco diseños experimentales ampliamente utilizados para modelizar la retención, de acuerdo al modelo lineal de fuerza del disolvente (LSS, *Linear Solvent Strength*) y al modelo de Neue-Kuss (NK), utilizando un conjunto de 14 sulfonamidas de diversa polaridad como compuestos de prueba. Los resultados se presentan en términos de incertidumbres relativas en las predicciones, que proporcionan resultados significativos e interpretables. La magnitud de las incertidumbres, junto con los cambios sistemáticos, coherentes y lógicos, observados al disminuir la polaridad de los solutos, corroboran los resultados. El modelo NK ofreció menores incertidumbres y predicciones exentas de sesgo, mientras que el LSS dio lugar a falta de ajuste. Se confirma que los diseños de entrenamiento isocráticos, ampliamente aceptados como los más informativos, son los mejores. La conclusión general del estudio es que los gradientes se predicen con incertidumbres más bajas, independientemente del diseño experimental de entrenamiento. Además, la predicción de gradientes es más insensible al tipo de diseño de entrenamiento utilizado, respecto a la predicción de experiencias isocráticas. Por otro lado, la predicción de estas últimas se deteriora rápidamente al incrementarse el contenido de modificador orgánico en la fase móvil. Ello explica la mayor calidad de la predicción de gradientes, incluso cuando se utilizan modelos sesgados.

3.2. Introducción

La necesidad de proporcionar datos informativos de calidad, con un mínimo esfuerzo experimental, es imperativa en áreas científicas donde la adquisición de datos es lenta o laboriosa, como es el caso de la cromatografía líquida (LC, *liquid chromatography*). Entre las posibles formas de obtener información de calidad, la alternativa más racional es el uso de metodologías de diseño de experimentos (DOE, *design of experiments*) [1,2]. Hoy en día, el diseño de experimentos es, junto con la tecnología analítica de procesos, la principal herramienta utilizada en el diseño de procedimientos analíticos para la industria farmacéutica (QbD, *Quality-by-Design*) [3]. En el desarrollo de métodos en LC, el diseño de experimentos se ha utilizado tradicionalmente para: (i) clasificar por orden de importancia los factores experimentales que afectan a los procedimientos y establecer condiciones preliminares (por ejemplo, mediante la exploración de columnas cromatográficas o disolventes), (ii) ajustar los factores más influyentes a través de la optimización de los métodos, y (iii) evaluar o pronosticar la robustez de las condiciones óptimas seleccionadas [2]. Este capítulo se centra en el segundo tipo de estudios, concretamente en la evaluación de la calidad de los diseños experimentales en el campo de la cromatografía líquida de fase inversa convencional (RPLC, *reversed-phase liquid chromatography*), con el fin de obtener las mejores condiciones experimentales en términos de resolución.

El desarrollo sistemático de métodos para LC se basa en la mejora de la resolución, generalmente con la ayuda de modelos que relacionan el tiempo de retención de los solutos con factores experimentales, como son el contenido de modificador orgánico, el pH o la temperatura. Los modelos de retención se construyen con la información obtenida a partir de estándares, ajustando los datos adquiridos a partir de un pequeño número de experimentos

cuidadosamente planificados, que siguen una cierta distribución, a lo que se denomina diseño experimental [4–6]. La fiabilidad de las expectativas de resolución y, por lo tanto, el éxito de la optimización cromatográfica depende de forma crítica de la calidad de los modelos de retención obtenidos. De ahí la importancia de los diseños experimentales, que determinan la calidad de la información disponible para realizar los ajustes.

Con frecuencia, el desarrollo de un método analítico se inicia llevando a cabo un gradiente de exploración, que ayuda a establecer el modo de elución y el intervalo de concentraciones de modificador orgánico más adecuados para la muestra que debe ser analizada. Tras ello, los datos de retención se adquieren a partir de un pequeño número de experiencias isocráticas o de gradiente, a lo que se denomina diseño experimental de entrenamiento, que se establece de acuerdo a la distribución de polaridades de los componentes de la muestra. A partir de este diseño de entrenamiento, se obtienen modelos de retención adecuados.

Los diseños experimentales isocráticos de entrenamiento son los más informativos, siendo su tratamiento bastante simple. No obstante, poseen el importante inconveniente de necesitar largos tiempos de adquisición de datos [7]. Por esta razón, y a pesar de ser menos informativos, los experimentos de gradiente se consideran los más prácticos [8]. Los diseños de entrenamiento que hacen uso de gradientes proporcionan también la información necesaria para realizar el ajuste de modelos, aunque no es sencillo encontrar un conjunto de gradientes óptimo para la modelización. Además, la consideración de la complejidad del gradiente queda fuera de las posibilidades de las mejores estrategias de diseño de experimentos basadas en el uso de derivadas, ya que el cálculo del tiempo de retención en gradiente implica la resolución de una ecuación integral, lo que sólo es posible en algunas circunstancias. Por otro

lado, sólo se pueden aplicar estrategias de diseño de experimentos geométricas, basadas en propiedades tales como la ortogonalidad, la capacidad de rotación o la uniformidad [9].

De acuerdo a su objetivo, los diseños utilizados en cromatografía líquida pueden dividirse, considerando en primer lugar el número de variables a estudiar en diseños de cribado y de optimización [10]. Cuando se dispone de cinco o más variables, es necesario realizar previamente diseños de cribado (diseños factoriales fraccionados [11] o de Plackett-Burman [12]) para clasificar y reducir las variables. Una vez halladas las variables más importantes (cuatro o menos), se aplican diseños centrales compuestos [13], de Box-Behnken [14] o de Taguchi [15], normalmente en combinación con modelos de retención polinomiales. Con modelos específicos y el apoyo de sus propiedades matemáticas, son posibles otras propuestas más elaboradas y completas. Los diseños D- o G-óptimos pertenecen a esta categoría, y su propósito es analizar las propiedades de la denominada “matriz de diseños” [16,17], que relaciona la distribución de los experimentos con el modelo de retención.

El grupo investigador ha aplicado diseños experimentales G-óptimos a la separación de solutos ionizables, en dominios isocráticos a temperatura constante, y pH y contenido de modificador orgánico variables [18]. La estrategia utilizada en ese estudio se basó en la adición de un punto experimental en la región del diseño con el mayor error de predicción esperado. Este proceso se repitió secuencialmente. Las predicciones cruzadas entre datos experimentales isocráticos y de gradiente (en un sentido u otro) también han sido estudiadas bajo la perspectiva de los errores asociados a la transferencia entre modos de elución [19,20]. Sin embargo, se encontró que las predicciones se hallaban fuertemente restringidas por problemas de cálculo y eran inviables

en muchas situaciones de gran interés práctico. Además, la lenta velocidad de cálculo impedía el cálculo masivo de las derivadas parciales del modelo de retención, necesarias para la evaluación de la calidad del diseño. Este tipo de cálculos es necesario siempre que la combinación del modelo de retención y el programa de gradiente no dé lugar a una expresión algebraica, cuando se resuelve la ecuación fundamental del gradiente de elución. Recientemente se ha encontrado una excelente capacidad predictiva en diseños no convencionales, donde los experimentos isocráticos se combinan con otros que involucran aumentos transitorios (pulsos) de modificador orgánico [21] (descritos en el Capítulo 2). Se ha encontrado que estos diseños inusuales, cuyo tiempo de adquisición es competitivo respecto al uso de gradientes, da lugar a buenas predicciones para solutos lentos eluidos en condiciones de baja fuerza eluyente.

Por lo tanto, aún era necesario desarrollar una metodología capaz de clasificar, en términos de calidad predictiva, cualquier conjunto de diseños experimentales para realizar tanto predicciones isocráticas como de gradiente, a fin de establecer el mejor diseño experimental de entrenamiento. En este capítulo, se propone, desarrolla, valida y aplica una metodología universal para evaluar la calidad de diseños experimentales de entrenamiento, que puede ser utilizada con configuraciones arbitrarias (implicando gradientes multi-lineales, experimentos multi-isocráticos, pulsos, etc.), capaz de revelar el mejor diseño. Una de las ventajas del método es que no se ve limitado por la disponibilidad de expresiones algebraicas para la predicción de tiempos de retención en elución en gradiente.

Mediante el uso de parámetros de modelos de retención de gran exactitud, ajustados a partir de datos isocráticos, se evalúa la capacidad predictiva cuando el diseño experimental de entrenamiento se aplica de forma exhaustiva a predicciones tanto isocráticas como de gradiente. Como solutos de prueba, se

utiliza un conjunto de compuestos de diversa polaridad, que se hallan apropiadamente o, al contrario, deficientemente cubiertos por los diseños de entrenamiento. La metodología descrita a continuación se aplica al análisis del rendimiento predictivo y las propiedades de cinco de los diseños experimentales de entrenamiento más habituales en RPLC, utilizando el modelo de retención de fuerza eluyente propuesto por Snyder (LSS) y el modelo NK. El objetivo final es obtener conclusiones sobre el rendimiento de los diseños de entrenamiento, y eventualmente encontrar los diseños óptimos (véase también el Capítulo 4), que permitan lograr los mejores ajustes y predicciones utilizando el modelo de retención seleccionado.

3.3. Teoría

3.3.1. Predicción de tiempos de retención

Para este estudio, se han considerado dos modelos de retención ampliamente utilizados. Uno de ellos fue propuesto por Snyder [8] y en él se basa la teoría de la fuerza eluyente del disolvente, ampliamente utilizada en los laboratorios de rutina y en software de optimización. El modelo se basa en una relación lineal entre el logaritmo del factor de retención (k) y la fracción volumétrica de modificador orgánico en la fase móvil (φ):

$$\log k = \log \frac{t_R - t_0}{t_0 - t_{\text{ext}}} = \log k_w - S \varphi \quad (3.1)$$

donde t_R , t_0 y t_{ext} son los tiempos de retención isocrático, tiempo muerto y tiempo extra-columnar, respectivamente. El modelo de Snyder incluye dos parámetros ajustables: la fuerza eluyente (S) y el logaritmo del factor de retención para una fase móvil constituida por agua pura ($\log k_w$). La Ecuación (3.1) ofrece buenas predicciones sólo en intervalos estrechos y moderados de

concentración de modificador orgánico. Es aplicable a elución en gradiente, pero su integración es únicamente posible en situaciones concretas, como cuando el programa de gradiente consta de una única rampa lineal. Sin embargo, es posible utilizar soluciones combinadas para programas con rampas multi-lineales.

En 2010, Neue y Kuss [22] propusieron un modelo que ofrece un excelente rendimiento en amplios intervalos de modificador orgánico:

$$k = k_w (1 + c \varphi)^2 e^{-\frac{B \varphi}{1 + c \varphi}} \quad (3.2)$$

donde c es un parámetro que describe la curvatura del modelo, y B un parámetro relacionado con la distribución del soluto entre ambas fases, debido a sus interacciones hidrofóbicas con ambas. Cuando la relación es lineal ($c = 0$), el parámetro B coincide con el parámetro S de la Ecuación (3.1). Este modelo no sólo da lugar a excelentes ajustes en elución isocrática, sino que también posee función primitiva (antiderivada), en el caso de utilizar gradientes lineales.

Los gradientes lineales simples que comienzan tras un cierto retraso en el tiempo (el denominado tiempo de permanencia, t_D), con pendiente m y ordenada en el origen a , de modo que $\varphi = \varphi_0$ cuando $t = t_D$, se pueden expresar como:

$$\varphi = a + mt = \varphi_0 + m(t - t_D) \quad (3.3)$$

Para los gradientes lineales, las funciones primitivas ($I(t)$) de los modelos LSS y NK son, respectivamente:

$$I(t) = \frac{e^{S(a+mt)}}{k_w S m} \quad (3.4)$$

$$I(t) = \frac{e^{\frac{B(a+mt)}{1+c(a+mt)}}}{k_w m B} \quad (3.5)$$

El tiempo de retención en elución en gradiente se obtiene despejando el límite superior de la ecuación fundamental [23]:

$$t_0 - t_{\text{ext}} = \int_0^{t_D} \frac{dt}{k(\varphi_0)} + \int_{t_D}^{t_g - t_0} \frac{dt}{k(\varphi(t))} = \frac{t_D}{k_0} + I(t_g - t_0) - I(t_D) \quad (3.6)$$

Para el modelo LSS, y tras operar la Ecuación (3.6), se obtiene la siguiente expresión:

$$t_g = t_0 + t_D + \frac{1}{S m} \ln \left[1 + S m (t_0 k_w e^{-S\varphi_0} - t_D) \right] \quad (3.7)$$

donde, por conveniencia, el programa de gradiente se ha desplazado para compensar t_D , de modo que la rampa lineal comience en $t = 0$.

Para el modelo NK se encuentra la siguiente solución analítica:

$$t_g = \frac{\ln H [1 + c(a - mt_0)] - B(a - mt_0)}{m(B - c \ln H)} \quad (3.8)$$

donde H viene dado por:

$$H = m B \ln k_w \left(t_0 - \frac{t_D}{k_w} \right) + e^{\frac{B(a + mt_D)}{1 + c(a + mt_D)}} \quad (3.9)$$

Las Ecuaciones (3.7–3.9) son válidas para solutos que eluyen a lo largo de una única rampa de gradiente. Con gradientes multi-lineales, la solución requiere el cálculo de cada segmento lineal por separado. En este caso, la Ecuación (3.6) debe adaptarse para incluir los diferentes segmentos de gradiente. Se deben sumar las contribuciones (Ecuaciones (3.4) o (3.5)) de

aquellos segmentos donde el soluto aún permanece dentro de la columna, tras completar el o los segmento/s del programa de gradiente:

$$\begin{aligned}
 t_0 - t_{\text{ext}} &= \int_0^{t_D} \frac{dt}{k(\varphi_0)} + \int_{t_D}^{t_1} \frac{dt}{k(\varphi_1(t))} + \dots + \int_{t_{n-1}}^{t_g - t_0} \frac{dt}{k(\varphi_n(t))} = \\
 &= \frac{t_D}{k_0} + \sum_{i=1}^{n-1} (I(t_{i+1}) - I(t_i)) + I(t_g - t_0) - I(t_{n-1})
 \end{aligned}
 \tag{3.10}$$

Finalmente, se despeja el tiempo de retención para elución en gradiente (t_g) del último término de la integral. Una solución más práctica es encontrar numéricamente t_g a partir de la Ecuación (3.6), aplicando métodos de búsqueda de raíces [24].

La siguiente sección explica las bases del cálculo sistemático de las incertidumbres en la predicción de tiempos de retención, cuando se trabaja en los modos isocrático y de gradiente, asociados a cualquier diseño experimental de configuración y complejidad arbitrarias.

3.3.2. Construcción de mapas de error en elución isocrática y de gradiente

La teoría de propagación de errores permite el cálculo de la incertidumbre asociada a la predicción de cualquier expresión compleja $F(x_1, x_2, \dots)$, en función de las incertidumbres de las variables (s_{x1}, s_{x2}, \dots), a través de la siguiente ecuación general [16]:

$$s_F^2 = s_{x_1}^2 \left(\frac{\partial F}{\partial x_1} \right)^2 + s_{x_2}^2 \left(\frac{\partial F}{\partial x_2} \right)^2 + \dots
 \tag{3.11}$$

A continuación, se detalla la aplicación de la Ecuación (3.11) al cálculo de las incertidumbres en la predicción de tiempos de retención, utilizando el modo gradiente, mediante la propagación de las incertidumbres en los parámetros de

regresión obtenidos a partir de un conjunto de experimentos isocráticos o de gradiente.

3.3.2.1. Cálculo de incertidumbres

El cálculo de las incertidumbres se basa en la combinación de matrices jacobianas, que recogen las derivadas parciales del modelo de retención respecto a sus n_{par} parámetros ($c_1, c_2 \dots$; por ejemplo, $c_1 = k_w$ y $c_2 = S$ para la Ecuación (3.1)), correspondientes a cada condición experimental i . En nuestro estudio, se utilizan dos matrices jacobianas:

- (i) $\mathbf{J}_{\text{train}}$, que contiene las derivadas parciales del modelo de retención, para los experimentos utilizados en el ajuste del modelo (es decir, el diseño experimental de entrenamiento),
- (ii) \mathbf{J}_{pred} , que contiene las derivadas parciales del modelo de retención, para el experimento (o experimentos), cuya incertidumbre (y eventualmente, retención) se está pronosticando.

Como puede verse, $\mathbf{J}_{\text{train}}$ y \mathbf{J}_{pred} incluyen las contribuciones del modelo para cada soluto eluido y geometría de los diseños estudiados. Para $\mathbf{J}_{\text{train}}$, las contribuciones participan a través de los parámetros c_i en los respectivos modelos, y para \mathbf{J}_{pred} , a través de los valores de los parámetros que describen los experimentos. Téngase en cuenta que cada experimento de gradiente i (de entrenamiento o predicho) viene definido por un conjunto específico de parámetros en la Ecuación (3.3): a (o ϕ_0) y m .

De forma genérica, ambas matrices jacobianas pueden escribirse de la siguiente forma:

$$\mathbf{J} = \begin{pmatrix} \left[\frac{\partial t_g}{\partial c_1} \right]_{i=1} & \left[\frac{\partial t_g}{\partial c_2} \right]_{i=1} & \dots & \left[\frac{\partial t_g}{\partial c_{n_{\text{par}}}} \right]_{i=1} \\ \left[\frac{\partial t_g}{\partial c_1} \right]_{i=2} & \left[\frac{\partial t_g}{\partial c_2} \right]_{i=2} & \dots & \left[\frac{\partial t_g}{\partial c_{n_{\text{par}}}} \right]_{i=2} \\ \vdots & \vdots & \ddots & \vdots \\ \left[\frac{\partial t_g}{\partial c_1} \right]_{i=n_{\text{exp}}} & \left[\frac{\partial t_g}{\partial c_2} \right]_{i=n_{\text{exp}}} & \dots & \left[\frac{\partial t_g}{\partial c_{n_{\text{par}}}} \right]_{i=n_{\text{exp}}} \end{pmatrix} \quad (3.12)$$

En este trabajo, t_g se referirá al tiempo de retención, independientemente del modo de elución (los experimentos isocráticos pueden considerarse como un caso particular de los experimentos de gradiente con $m = 0$). Por lo tanto, los valores isocráticos y de gradiente pueden analizarse con la misma metodología. Las dos contribuciones mencionadas anteriormente (modelo y experimento) se reflejan en la estructura de las matrices jacobianas. Así, $\mathbf{J}_{\text{train}}$ (la matriz jacobiana asociada al diseño de entrenamiento) incluye tantas columnas como parámetros ajustados (n_{par}) en el modelo de un soluto, y tantas filas como experimentos en el diseño de entrenamiento (n_{exp}). Se debe tener en cuenta que, en $\mathbf{J}_{\text{train}}$, deben incluirse necesariamente varias filas porque la obtención de los modelos requiere un número de experimentos igual o mayor al de los parámetros del modelo de retención (los grados de libertad deben ser cero o un número positivo). Mientras tanto, \mathbf{J}_{pred} puede tener varias filas o sólo una. En caso de poseer sólo una fila, y siguiendo las pautas de nomenclatura en Quimiometría, la matriz jacobiana será un vector fila escrito en minúsculas y negrita: \mathbf{j}_{pred} , correspondiendo al experimento cuya incertidumbre se está pronosticando (con $n_{\text{exp}} = 1$).

Para el modelo LSS, las derivadas parciales necesarias para construir $\mathbf{J}_{\text{train}}$ y \mathbf{j}_{pred} se calculan de la siguiente forma:

$$\frac{\partial t_g}{\partial c_1} = \frac{\partial t_g}{\partial k_w} = \frac{t_0 e^{-S\varphi_0}}{1 + \alpha - S m t_D} \quad (3.13)$$

$$\frac{\partial t_g}{\partial c_2} = \frac{\partial t_g}{\partial S} = \frac{1}{S^2 m} \left[\frac{\alpha - S m t_D - S \varphi_0 \alpha}{1 + \alpha - S m t_D} - \ln(1 + \alpha - S m t_D) \right] \quad (3.14)$$

donde:

$$\alpha = S m t_0 k_w e^{-S\varphi_0} \quad (3.15)$$

Para el modelo NK, las derivadas parciales son demasiado complejas, por lo que es preferible utilizar aproximaciones numéricas de las derivadas. Éstas se calculan alejando ligeramente los parámetros de los valores nominales, e inspeccionando el efecto sobre el tiempo de retención del gradiente previsto (\hat{t}_g). Con el fin de ahorrar operaciones, esta inspección se puede realizar examinando las variaciones en \hat{t}_g de forma unilateral (ya sea aumentando o disminuyendo el parámetro examinado):

$$\frac{\partial [\hat{t}_g(c_1, \dots, c_i, \dots)]}{\partial c_i} = \lim_{h \rightarrow 0} \frac{\hat{t}_g(c_1, \dots, c_i + h, \dots) - \hat{t}_g(c_1, \dots, c_i, \dots)}{h} \quad (3.16)$$

La Ecuación (3.16) incluye dos fuentes de error: el error de truncamiento y el de redondeo [25]. El primero proviene de la eliminación de los términos de mayor valor en la expansión de la serie de Taylor, en la que se basa la Ecuación (3.16):

$$\frac{\hat{t}_g(c_1, \dots, c_i + h, \dots) - \hat{t}_g(c_1, \dots, c_i, \dots)}{h} = \frac{\partial [\hat{t}_g(c_1, \dots, c_i, \dots)]}{\partial c_i} + \frac{1}{2} h \frac{\partial^2 [\hat{t}_g(c_1, \dots, c_i, \dots)]}{\partial c_i^2} + \dots \quad (3.17)$$

Por otro lado, el error de redondeo posee varias contribuciones, siendo la más importante la magnitud de h a lo largo de la aproximación a cero. Los errores de truncamiento y redondeo pueden ser menos significativos que las incertidumbres en el cálculo de \hat{t}_g , resolviendo la Ecuación (3.6) con métodos de búsqueda de raíces. De hecho, los tiempos de retención en gradiente predichos se ven afectados por la incertidumbre asociada a la resolución de la ecuación fundamental de elución en gradiente, y esta incertidumbre puede ser más grave que los errores de truncamiento y redondeo.

En este capítulo, se ha aplicado una variación del método de Ridders [25,26] para el cálculo de las derivadas numéricas en las matrices jacobianas. Este método se basa en la extrapolación a cero de la diferencia central (compárese con la Ecuación (3.16)):

$$\frac{\partial[\hat{t}_g(c_1, \dots, c_i \dots)]}{\partial c_i} = \lim_{h \rightarrow 0} \frac{\hat{t}_g(c_1, \dots, c_i + h \dots) - \hat{t}_g(c_1, \dots, c_i - h \dots)}{2h} \quad (3.18)$$

A lo largo del cálculo, se han monitorizado varias medidas de calidad relacionadas con el nivel de ruido de la derivada, cuando h se halla demasiado cercana a cero (ver Sección 3.5.3.1). Una vez obtenidas las matrices jacobianas, se calcula la varianza de la siguiente forma, al tiempo de retención predicho para el soluto en estudio [16]:

$$s_{\text{pred}}^2 = s_{\text{PE}}^2 \left(\mathbf{j}_{\text{pred}} (\mathbf{J}_{\text{train}}^T \mathbf{J}_{\text{train}})^{-1} \mathbf{j}_{\text{pred}}^T \right) \quad (3.19)$$

Debe tenerse en cuenta que \mathbf{j}_{pred} está escrito en minúsculas y negrita, ya que se refiere a una sola condición experimental (un vector fila conteniendo las derivadas parciales). En la Ecuación (3.19), s_{PE}^2 es el cuadrado del denominado error experimental puro, que se define como:

$$s_{PE} = \sqrt{\frac{\sum_{i=1}^{n_{\text{exp}}} (t_{g,i} - \bar{t}_g)^2}{n-1}} \quad (3.20)$$

mientras que s_{pred}^2 es la desviación estándar de los tiempos de retención correspondiente a experimentos replicados en el centro del diseño. En problemas de regresión, s_{PE} se reemplaza a menudo por el error estándar en las predicciones (*SEP, standard error in predictions*), que se mide a partir de la dispersión de los datos experimentales alrededor de la curva de regresión del modelo de retención:

$$SEP = \sqrt{\frac{\sum_{i=1}^{n_{\text{exp}}} (\hat{t}_{g,i} - t_{g,i})^2}{n_{\text{exp}} - n_{\text{par}}}} \quad (3.21)$$

En la ecuación anterior, $\hat{t}_{g,i}$ y $t_{g,i}$ son los tiempos de retención predicho y experimental en la condición i del diseño experimental de entrenamiento. La validez de la Ecuación (3.21) está sujeta a la ausencia de correlación entre los errores y los tiempos de retención (ausencia de sesgo).

3.3.2.2. Mapas de incertidumbre relativa asociados a un diseño experimental de entrenamiento

En la Ecuación (3.19), $\mathbf{J}_{\text{train}}$ cuantifica la calidad informativa proporcionada por los datos utilizados para construir los modelos de retención, mientras que \mathbf{j}_{pred} evalúa el nivel de dificultad de la condición experimental a predecir, por lo que una predicción demasiado propensa a error tendrá un alto valor de \mathbf{j}_{pred} . La predicción de las incertidumbres asociadas a un determinado diseño experimental se puede inspeccionar exhaustivamente a partir del valor de una

matriz \mathbf{J}_{pred} (en lugar de un vector \mathbf{j}_{pred}), modificando gradualmente las condiciones experimentales predichas y manteniendo constante $\mathbf{J}_{\text{train}}$.

Los valores de la incertidumbre (s_{pred}) se pueden representar en función de la variable que se modifica sistemáticamente, a lo largo de la exploración (la composición de la fase móvil en elución isocrática y la pendiente del programa en elución en gradiente), lo que da lugar a “mapas de incertidumbre de predicción”. Se considerarán dos tipos de exploración de las condiciones y, por lo tanto, dos tipos de representaciones:

- (i) mapas isocráticos ($s_{\text{pred}} = F(\varphi)$), en los que se realiza una exploración exhaustiva de la composición de las fases móviles, al incrementar la concentración de modificador (φ), y
- (ii) mapas de gradiente ($s_{\text{pred}} = F(m)$), donde se examina el efecto de aumentar sistemáticamente la pendiente del gradiente (m), hasta alcanzar un valor máximo φ_{max} (fijado como objetivo a alcanzar en $t = t_G$).

Ambos mapas permiten una inspección exhaustiva de la capacidad predictiva de cualquier diseño experimental, bajo una perspectiva estadística. Para un análisis más sencillo de los resultados, se han utilizado incertidumbres relativas en las predicciones ($s_{\text{R,pred}}$):

$$s_{\text{R,pred}} = 100 \frac{s_{\text{pred}}}{\hat{t}_g} \quad (3.22)$$

que se representan en función de φ o m (la pendiente del gradiente), en lugar de las incertidumbres absolutas (s_{pred} , Ecuación (3.19)).

3.4. Parte experimental

3.4.1. Reactivos

En el estudio realizado, se consideró un conjunto de 14 sulfonamidas: (1) sulfaguanidina, (2) sulfanilamida, (3) sulfadiazina, (4) sulfatiazol, (5) sulfapiridina, (6) sulfamerazina, (7) sulfametazina, (8) sulfametizol, (9) sulfamonometoxina, (10) sulfacloropiridazina, (11) sulfametoxazol, (12) sulfisoxazol, (13) sulfadimetoxina, y (14) sulfaquinoxalina (Sigma, Rödermark, Alemania). Para cada sulfonamida se preparó una disolución madre de $100 \mu\text{g}\cdot\text{mL}^{-1}$, utilizando agua nanopura (obtenida con un sistema de purificación de Adrona B30 Trace, Burladingen, Alemania), ayudándose de un baño de ultrasonidos (Elmasonic, Singen, Alemania).

La separación cromatográfica se realizó tanto en modo isocrático como en gradiente, utilizando mezclas preparadas con acetonitrilo grado HPLC (Scharlau, Barcelona) y una disolución acuosa de dihidrógenofosfato de sodio anhidro (Fluka, Alemania) en concentración 0.01 M, preparada con agua nanopura. El pH se fijó a 3.0 mediante la adición de HCl y NaOH (Scharlau), ambos en concentración 0.01 M. Las inyecciones se realizaron por duplicado.

Antes de su inyección en el sistema cromatográfico, todas las disoluciones se filtraron a través de membranas de Nylon de $0.45 \mu\text{m}$ (Micron Separations, Westboro, MA, EE.UU.).

3.4.2. Instrumentación y columna

El análisis se realizó con un equipo cromatográfico modelo HP1100 (Agilent, Waldbronn, Alemania), compuesto por los siguientes módulos: bomba cuaternaria, autoinyector equipado de viales de 2 mL, compartimento de columna termostaticado y detector UV-Vis ajustado a una longitud de onda de

medida de 254 nm. El volumen de inyección fue de 20 μL y el flujo de fase móvil se mantuvo constante a $1.0 \text{ mL}\cdot\text{min}^{-1}$.

Para realizar los análisis, se utilizó una columna Zorbax Eclipse XDB-C18 (150 mm \times 4.6 mm), con un tamaño de partícula de 5 μm (Agilent). Todas las inyecciones se llevaron a cabo en condiciones de temperatura controlada, a 25 $^{\circ}\text{C}$. Se determinó el tiempo muerto del sistema para diferentes composiciones de la fase móvil, mediante la inyección de KBr (Acros Organics, Fair Lawn, NJ, EE.UU.). El tiempo de permanencia (1.16 min) se midió haciendo uso de un gradiente de acetona. El tiempo extra-columnar fue de 0.12 min.

La medida del pH de las fases móviles se realizó haciendo uso de un pH-metro (Crison, modelo MicropH 2002, Barcelona), que utilizaba un electrodo de membrana de vidrio (modelo 8102, Orion, Barcelona), un electrodo de referencia de Ag/AgCl y una disolución de KCl 3.0 M en el puente salino.

3.4.3. Software

Para la adquisición de las señales, se utilizó una estación de trabajo OpenLAB CDS LC (Agilent, revisión B.04.03). Para la medida de las propiedades de pico (tiempo de retención y semianchuras, entre otras) se hizo uso del software MICHROM [27]. El tratamiento de datos se llevó a cabo con funciones desarrolladas por el grupo de investigación en Matlab 2016b (The MathWorks Inc., Natick, MA, EE.UU.).

3.5. Resultados y discusión

Se desarrolla y valida la metodología descrita en la Sección 3.3, para evaluar la calidad de diseños experimentales con diversos perfiles. La calidad de los resultados se ilustra investigando exhaustivamente las propiedades de cinco diseños experimentales de entrenamiento ampliamente utilizados.

3.5.1. Diseños evaluados

Las Figuras 3.1 y 3.2 muestran los cinco diseños de entrenamiento considerados en este trabajo (todos conteniendo, a su vez, cinco experimentos):

- (i) Diseño ISO1: conjunto de experimentos isocráticos que concentran gradualmente el contenido de modificador orgánico a valores bajos (10, 13, 16, 20 y 25 % de acetonitrilo), con el fin de muestrear mejor el intervalo de modificador que proporciona una retención elevada.
- (ii) Diseño ISO2: conjunto de experimentos isocráticos fuertemente enfocados en el dominio de baja fuerza eluyente (10–14 % de acetonitrilo).
- (iii) Diseño G1: conjunto de cinco gradientes que proporcionan un tiempo de gradiente, t_G , constante ($60 \text{ min} + t_D$).
- (iv) Diseño G2: conjunto de gradientes que alcanzan la misma concentración final, φ_F (25 % de acetonitrilo). Este diseño, junto con el diseño G1, son representativos del tipo de diseños de gradiente frecuentemente utilizados para modelizar la retención.
- (v) Diseño G3: conjunto de gradientes donde el inicio de las rampas se dispone en tres niveles, y tanto t_G como φ_F son variables. Este diseño pretende combinar, en cierta medida, las ventajas de los cuatro diseños anteriores.

Para interpretar los resultados, en las Figuras 3.1 y 3.2 se han superpuesto los tiempos de retención de los 14 compuestos de prueba, para todos los experimentos de cada diseño. Algunos diseños no fueron capaces de proporcionar la información requerida para modelizar algunos compuestos, debido al insuficiente número de grados de libertad ($df = 0$ o < 0). En este trabajo, se consideraron iguales dos experimentos si los valores de t_g diferían en menos de 0.1 min.

3.5.2. Modelización de la retención

En la práctica, el uso de diseños experimentales isocráticos se ve limitado por los tiempos de retención excesivos, obtenidos para los solutos más hidrofóbicos cuando se eluyen con fases móviles con un bajo contenido de modificador. Sin embargo, existe un acuerdo general sobre el hecho de que este tipo de diseño proporciona la información más rica sobre la retención de los solutos eluidos. Además, los parámetros de los modelos de retención estimados con experiencias isocráticas se consideran los más fiables. En consecuencia, para el desarrollo de este estudio, los parámetros del modelo utilizado para describir cada soluto se obtuvieron ajustando los datos experimentales obtenidos, con las cinco fases móviles del diseño ISO1, a los modelos LSS y NK (Ecuaciones (3.1) y (3.2), respectivamente).

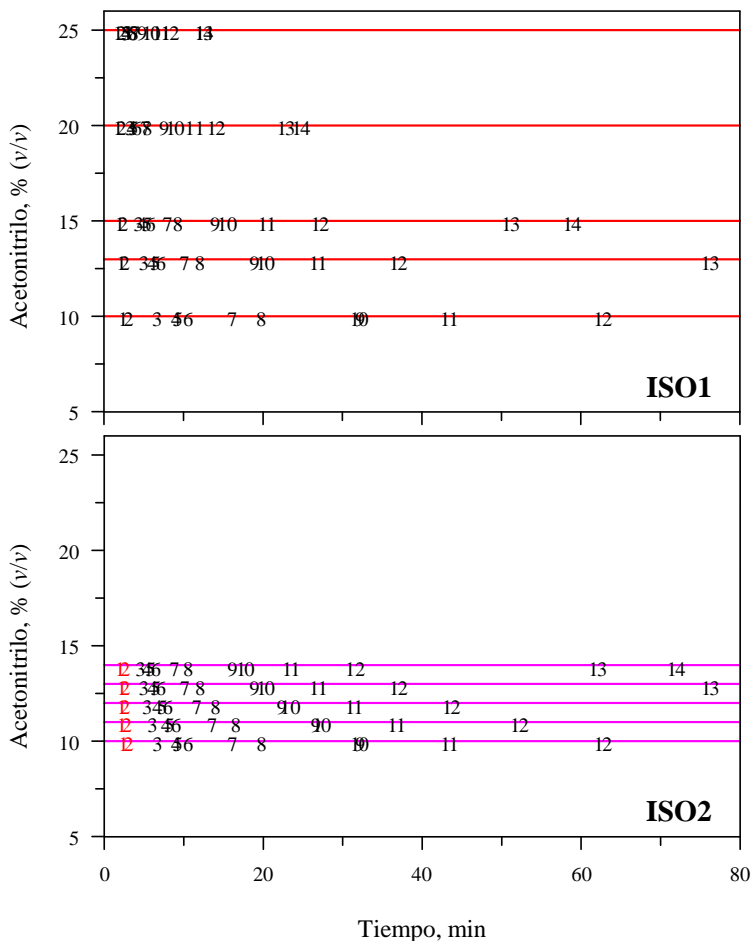


Figura 3.1. Diseños experimentales de entrenamiento bajo evaluación, formados con experiencias isocráticas. Se indican (superpuestos) los tiempos de retención de los solutos para cada experimento. Los solutos problemáticos para el diseño ISO2 (véase el texto) fueron: (1) sulfaguanidina y (2) sulfanilamida (el diseño ISO1 no presentó ningún problema). Algunos solutos hidrofóbicos que eluyen en ciertas condiciones, más allá de los 80 min, no se han representado. Consúltese la Sección 3.4.1 para conocer los códigos de identificación de los solutos y la Sección 3.5.1 para obtener más detalles.

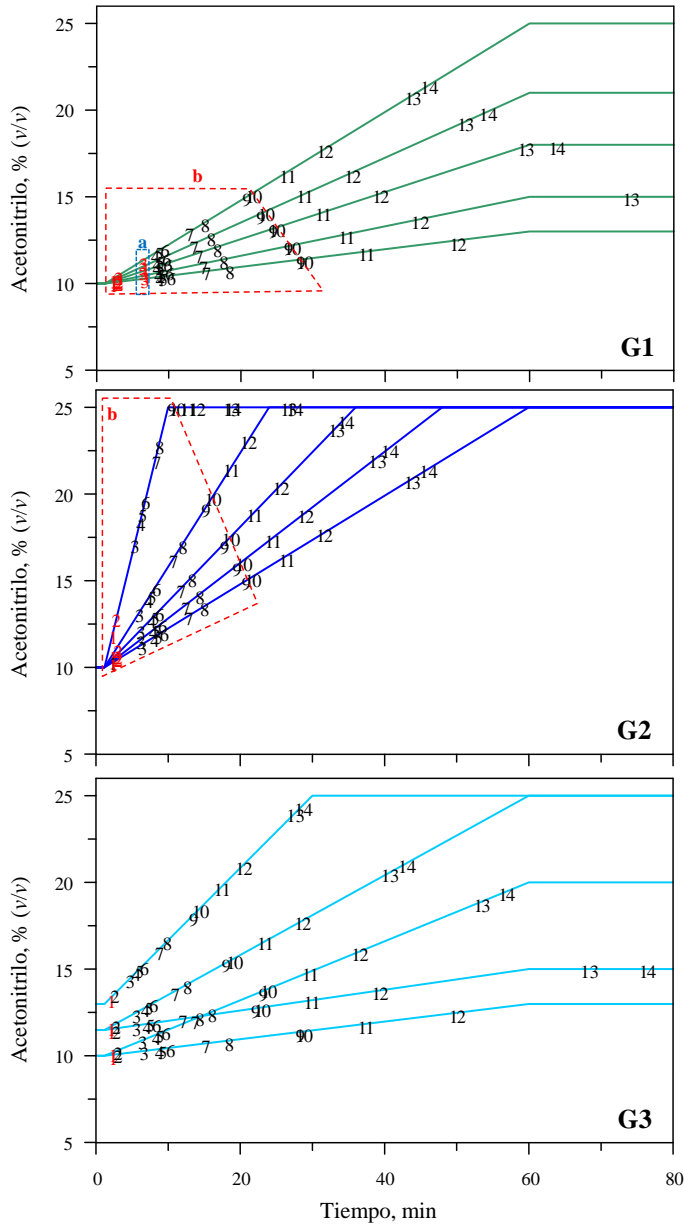


Figura 3.2. Diseños experimentales de entrenamiento bajo evaluación, formados con gradientes. Los solutos problemáticos fueron: (1) sulfaguanidina, (2) sulfanilamida y (3) sulfadiazina. Específicamente, la cobertura fue crítica para los solutos 1 a 3 en G1, 1 y 2 en G2, y 1 en G3. Otros detalles se pueden consultar en la Figura 3.1.

Tabla 3.1. Parámetros de regresión y errores estándar estimados para el modelo LSS (Ecuación (3.1)), utilizando los datos experimentales adquiridos con el Diseño ISO 1. Se dan los siguientes estadísticos de ajuste^{ab}: coeficiente de correlación ajustado (R^2_{adj}), error relativo medio en las predicciones (ER) referido a la retención media, F-Snedecor y error estándar en la predicción (SEP), que aproxima el error experimental puro (s_{PE}) en ausencia de falta de ajuste. Los coeficientes corresponden a ajustes realizados con ϕ expresados en tanto por uno ($v/\%$) (también en la Tabla 3.2).

Solutos	Parámetros modelo LSS					
	k_w	S	R^2_{adj}	$ER, \%$	$F\text{-Snedecor}$	SEP, min
Sulfaganidina	0.28 ± 0.10	-5.5 ± 0.7	0.98876	5.149	78.1	0.0491
Sulfamilamida	0.64 ± 0.08	-4.4 ± 0.5	0.99283	3.957	85.4	0.0652
Sulfadiazina	2.38 ± 0.11	-9.4 ± 0.8	0.99078	4.884	204.1	0.2281
Sulfatiazol	3.11 ± 0.03	-13.0 ± 0.3	0.98980	5.864	264.5	0.3274
Sulfapiridina	3.04 ± 0.12	-11.9 ± 0.9	0.99078	5.341	266.9	0.3324
Sulfamerazina	3.11 ± 0.11	-11.3 ± 0.9	0.99107	5.155	260.7	0.3750
Sulfametazina	3.81 ± 0.12	-13.4 ± 1.0	0.99129	5.486	318.3	0.5827
Sulfametizol	4.22 ± 0.12	-15.3 ± 1.0	0.99232	4.537	403.9	0.6797
Sulfamonometoxina	4.85 ± 0.12	-16.3 ± 1.0	0.99372	5.073	518.9	1.0262
Sulfaclopiridazina	4.62 ± 0.10	-14.0 ± 0.8	0.99409	4.656	486.5	1.0158
Sulfametoxazol	4.96 ± 0.03	-14.3 ± 0.3	0.99493	4.378	576.8	1.2840
Sulfisoxazol	5.51 ± 0.02	-16.0 ± 0.2	0.99548	4.288	709.1	1.7471
Sulfadimetoxina	6.9 ± 0.2	-21.2 ± 1.4	0.99286	5.636	541.1	5.3630
Sulfaguinoxalina	7.15 ± 0.02	-21.6 ± 0.2	0.99558	4.481	881.5	4.9273

^a Las expresiones que permiten el cálculo de los estadísticos de ajuste R^2_{adj} , ER y $F\text{-Snedecor}$ vienen dadas en la Sección 2.5.1 del Capítulo 2

^b Error estándar en la predicción (SEP) calculado según la Ecuación (3.21)

Tabla 3.2. Parámetros de regresión y errores estándar estimados para el modelo NK (Ecuación (3.2)), usando los datos experimentales adquiridos con el Diseño ISO1. Se dan los siguientes estadísticos de ajuste: coeficiente de determinación ajustado (R^2_{adj}), error relativo medio en las predicciones (ER) referido a la retención media, F-Snedecor y error estándar en la predicción (SEP), que aproxima el error experimental puro (δ_{PE}) en ausencia de falta de ajuste. Para más detalles consúltese la Tabla 3.1.

Solutos	Parámetros modelo NK				R^2_{adj}	ER, %	F-Snedecor	SEP, min
	k_w	c	B					
Sulfaguamidina	4.40 ± 0.10	5.635 ± 0.010	40.8 ± 0.3		0.99989	0.676	2332	0.0072
Sulfamilamida	4.62 ± 0.03	3.26 ± 0.02	43.84 ± 0.12		1.00000	0.136	38695	0.0025
Sulfadiazina	31.5 ± 0.2	4.046 ± 0.008	37.30 ± 0.06		0.99999	0.235	43960	0.0127
Sulfatiazol	87 ± 4	4.29 ± 0.12	47.8 ± 1.3		0.99999	0.218	88864	0.0144
Sulfapiridina	67.0 ± 1.3	3.90 ± 0.05	41.4 ± 0.3		0.99992	0.562	11168	0.0418
Sulfamerazina	67 ± 2	3.79 ± 0.02	39.1 ± 0.4		0.99989	0.672	7343	0.0575
Sulfametazina	149 ± 2	3.79 ± 0.02	43.55 ± 0.12		0.99988	0.692	9228	0.0879
Sulfametizol	237 ± 2	3.68 ± 0.02	46.21 ± 0.11		0.99998	0.313	57212	0.0463
Sulfamonometoxina	393 ± 4	3.26 ± 0.02	43.84 ± 0.12		0.99990	0.677	13703	0.1625
Sulfacoloropiridazina	261 ± 7	3.000 ± 0.005	37.1 ± 0.2		0.99998	0.297	54450	0.0779
Sulfametoxazol	335 ± 8	2.710 ± 0.015	34.9 ± 0.2		0.99994	0.482	20915	0.1735
Sulfisoxazol	596 ± 12	2.634 ± 0.006	37.1 ± 0.2		0.99995	0.448	30550	0.2172
Sulfadimetoxina	6441 ± 165	4.456 ± 0.009	67.4 ± 0.4		0.99997	0.332	70975	0.3690
Sulfaquinoxalina	4808 ± 59	3.301 ± 0.002	53.91 ± 0.15		0.99999	0.205	214489	0.2518

Las Tablas 3.1 y 3.2 proporcionan los parámetros de regresión, las incertidumbres en su determinación y otras estadísticas de ajuste para los dos modelos de retención (LSS y NK, respectivamente). Para los ajustes y cálculos derivados a lo largo de este trabajo, la concentración de modificador orgánico se expresó como fracción volumétrica (v/v), aunque en las gráficas se exprese como porcentaje.

Los modelos de retención ensayados muestran fuertes diferencias:

- (i) El modelo LSS ofrece un rendimiento bastante modesto (Tabla 3.1), con $R^2_{aj} > 0.99$ y errores relativos de predicción de los tiempos de retención de alrededor del 5 %.
- (ii) Contrasta notablemente el excelente rendimiento del modelo NK (Tabla 3.2), con $R^2_{aj} > 0.9999$ y errores relativos de predicción del orden del 0.7 % o inferiores.

Se seleccionó el intervalo de concentración del modificador (10 a 25 % de acetonitrilo), atendiendo a los tiempos de retención en elución isocrática del conjunto de sulfonamidas, procurando que fueran inferiores a una hora. El pequeño intervalo de modificador utilizado debería favorecer al modelo LSS. Sin embargo, el estudio estadístico indicó que esta ecuación no describe fielmente el comportamiento de retención experimental. A pesar de ello, se ha incluido en este trabajo debido a su importancia y amplio uso, pero es de esperar que produzca sesgos, además de predicciones inciertas y problemas potenciales en los ajustes y las predicciones.

Para un determinado soluto, no es factible la selección de una condición representativa de la cual se puedan obtener réplicas, por lo que no es posible medir s_{PE} experimentalmente con la Ecuación (3.20), a fin de calcular las incertidumbres relativas en las predicciones ($s_{R,pred}$ en la Ecuación (3.19)). Esto se puede explicar por las variaciones extremas que se producen en el tiempo de

retención para los distintos solutos (y distintas fases móviles para el mismo soluto). Por esta razón, la magnitud de s_{PE} se estableció a partir de la dispersión alrededor del modelo de retención, utilizando la Ecuación (3.21) y todos los experimentos disponibles. La última columna de las Tablas 3.1 y 3.2 proporciona el error estándar en las predicciones (SEP), que como se ha comentado, se aproxima al error experimental puro que aparece en la Ecuación (3.19).

3.5.3. Mapas de incertidumbre relativa asociados a un diseño experimental de entrenamiento

La Ecuación (3.19) permite predecir el error estándar asociado a cualquier experimento cromatográfico, utilizando la información proporcionada por cualquier conjunto de entrenamiento. Esta Sección ilustra el cálculo de los mapas de incertidumbre relativa, $s_{R,pred}$.

3.5.3.1. Cálculo de las matrices Jacobianas

Los conjuntos de parámetros de los modelos LSS y NK (véanse las Tablas 3.1 y 3.2) se obtuvieron ajustando directamente los datos de retención experimentales en función de φ , para cada modelo y soluto. Debe tenerse en cuenta que, cuando el diseño no es el adecuado, los parámetros de los solutos obtenidos por regresión se hallarán sesgados, y el sesgo distorsionará el cálculo de las matrices jacobianas, enmascarando la influencia del diseño que se está evaluando. Por ello, en el tratamiento desarrollado para este trabajo se minimizó esta influencia mediante el uso de los parámetros del modelo obtenidos con el Diseño ISO1 para el cálculo de \mathbf{J}_{train} y \mathbf{J}_{pred} . Como ya se ha comentado en la Sección 3.2, los diseños isocráticos son los más fiables en términos de exactitud en las predicciones, pero se utilizan en menor medida debido a los largos tiempos de retención a los que dan lugar cuando se utilizan fases móviles de baja fuerza eluyente.

Para el modelo LSS, $\mathbf{J}_{\text{train}}$ y \mathbf{J}_{pred} pueden calcularse fácilmente con las Ecuaciones (3.13–3.15), válidas para aquellas condiciones (filas en las matrices jacobianas) en las que el soluto eluye a lo largo de la rampa del gradiente. Estas ecuaciones requieren modificaciones si el soluto eluye isocráticamente más allá del final de la rampa ($t_g > t_G$), o dentro del tiempo de permanencia. Para el modelo NK, no es posible el cálculo mediante el uso de derivadas algebraicas utilizando expresiones matemáticas razonablemente simples, incluso en el caso ideal en el que los solutos eluyen dentro de la rampa.

Debido a estas limitaciones, la evaluación de las matrices jacobianas se realizó mediante procedimientos numéricos, que pueden aplicarse a cualquier experimento isocrático o de gradiente, independientemente de la complejidad o el instante a lo largo del programa de elución en el que el soluto abandone la columna. En el procedimiento numérico, las derivadas se estimaron por el método de Ridders (véase la Sección 3.3.2.1), utilizando como variable el tiempo de retención de los gradientes predichos, \hat{t}_g . En el cálculo realizado, los valores de \hat{t}_g se despejaron de la Ecuación (3.10), utilizando antiderivadas analíticas (Ecuaciones (3.4) y (3.5)). Con este propósito, se aplicó el método de bisección de Newton, siguiendo el enfoque explicado en la Referencia [24]. En esa referencia, se da respuesta al cálculo eficiente de \hat{t}_g en situaciones donde las soluciones algebraicas de la Ecuación (3.10) no son factibles, debido a que se combinan modelos y programas de gradiente que carecen de antiderivadas.

En la estimación de las matrices jacobianas, las derivadas se obtuvieron mediante la disminución exponencial del parámetro h en la Ecuación (3.18), de acuerdo a la expresión $h(c_i) = c_i (1 + 0.5^\lambda)$, siendo λ una secuencia de números naturales. Se monitorizaron tres medidas a lo largo de la contracción de h (véanse las Ecuaciones (3.16) a (3.18)): (i) la derivada parcial ($\partial \hat{t}_g / \partial c_i$), (ii) el

tiempo de retención pronosticado \hat{t}_g , y (iii) el parámetro c_i . Siempre que las diferencias entre dos estimaciones consecutivas de $(\partial\hat{t}_g/\partial c_i)$, y entre los valores de \hat{t}_g y c_i en la misma iteración, fueran todas ellas superiores a 1000 veces la precisión de la máquina, se actualizó el nuevo valor de $(\partial\hat{t}_g/\partial c_i)$. El resultado final de la derivada parcial se tomó del último valor de $(\partial\hat{t}_g/\partial c_i)$ obtenido, para el que los tres parámetros indicados todavía cumplían la condición de validez.

3.5.3.2. Construcción de los mapas de incertidumbre

La evaluación de las incertidumbres implica dos tipos de diseño, a saber, de entrenamiento y de muestreo. Ambos diseños se combinan en la Ecuación (3.19), y cuando s_{pred}^2 se representa frente al factor escaneado en \mathbf{J}_{pred} (la composición de la fase móvil en los diseños isocráticos y la pendiente de la rampa en los diseños de gradiente), se obtiene un mapa de incertidumbre absoluta. Este tipo de mapa explora la capacidad predictiva de un diseño de entrenamiento, utilizando los experimentos del diseño de muestreo. En este trabajo, se han investigado cinco diseños de entrenamiento (véase la Sección 3.5.1). Puesto que, para un determinado mapa, el producto $(\mathbf{J}_{\text{train}}^T \cdot \mathbf{J}_{\text{train}})^{-1}$ en la Ecuación (3.19) es constante, sólo fue necesario calcularlo una única vez para cada diseño de entrenamiento, lo que permitió ahorrar operaciones. En cambio, \mathbf{J}_{pred} (asociado al diseño de muestreo) necesita ser calculado para cada una de las condiciones bajo evaluación. Téngase en cuenta que al referirse a más de una condición predicha, el jacobiano de la predicción es una matriz: \mathbf{J}_{pred} .

Para un determinado diseño de entrenamiento (cualquiera de los diseños de las Figuras 3.1 y 3.2), se estudiaron los efectos en las predicciones tanto

isocráticas como de gradiente, haciendo uso de diseños de muestreo específicos para cada tipo de elución, a partir del cual es calculado el término J_{pred} en la Ecuación (3.19). La Figura 3.3a muestra el diseño de muestreo isocrático, que está constituido por 16 fases móviles isocráticas. Este diseño permite inspeccionar la información proporcionada por el diseño de entrenamiento, cuando se predicen fases móviles isocráticas (en el intervalo de 10 a 25 % v/v de acetonitrilo, en incrementos de un 1 %).

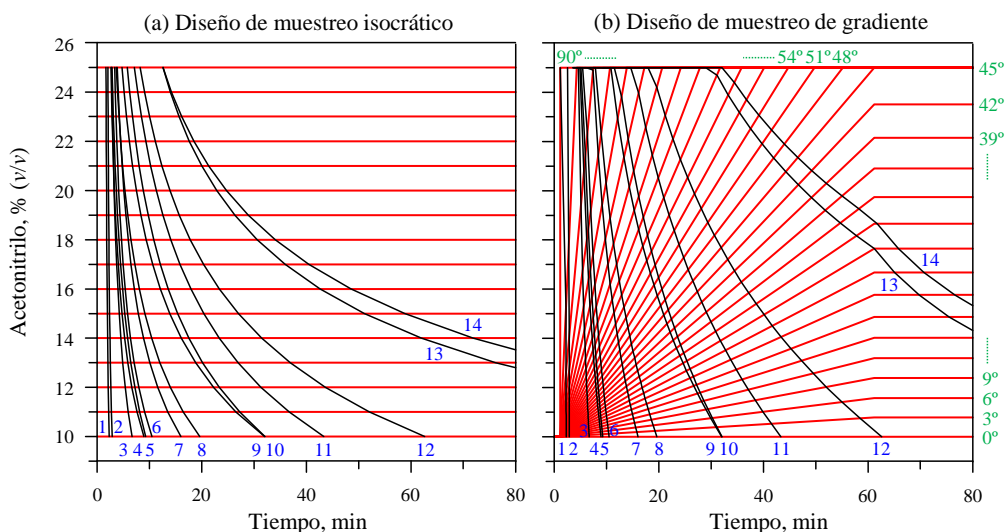


Figura 3.3. Diseños experimentales de muestreo isocrático (a) y de gradiente (b) (líneas rojas), utilizados para evaluar el rendimiento de los cinco diseños de entrenamiento mostrados en las Figuras 3.1 y 3.2. Las líneas negras superpuestas indican el tiempo de retención de cada soluto en cada experimento (se indica el código del soluto; para conocer su identidad consúltese la Sección 3.4.1). En (b), existen dos tipos de gradientes con pendientes que presentan un incremento angular constante de 3° : (i) gradientes con pendientes entre 0° y 45° , donde se varía la concentración objetivo (φ_F), siendo el tiempo de gradiente constante ($t_G = 60 + t_D$ min), y (ii) gradientes con pendientes entre 48° y 90° , donde φ_F es constante y t_G variable.

La Figura 3.3b muestra el diseño de muestreo de gradiente, que contiene 31 gradientes con pendientes progresivamente mayores distribuidas uniformemente, mostrando un incremento angular constante de 3° . De los 31 gradientes, 16 completan la rampa en $t_G = 60 + t_D$ min (cada uno de ellos a una concentración objetivo distinta, φ_F), mientras que los otros 15 gradientes tienen un t_G variable, aunque alcanzan el mismo valor de φ_F : 25 % de acetonitrilo. Las líneas superpuestas, que cruzan las fases isocráticas y programas de gradientes, representan los tiempos de retención de cada soluto para cada experimento de muestreo.

Las incertidumbres absolutas presentan una fuerte variación y dependen de la retención de los solutos, presentando un orden de magnitud similar a los valores de *SEP* (véase la Tabla 3.2). Por esta razón, a partir de este momento los resultados se representarán en términos de incertidumbre estándar relativa de predicción ($s_{R,pred}$), que proporciona resultados más significativos e interpretables, incluso cuando se comparan las incertidumbres para la elución isocrática y de gradiente. La Figura 3.4 muestra, a modo de ejemplo, los mapas de incertidumbre de gradiente para la sulfacoloropiridazina en el modo gradiente, para los cinco diseños de entrenamiento en estudio, utilizando el modelo NK y diversa precisión en la integración realizada para estimar \hat{t}_g .

Se muestran dos escalas de abscisas. La escala superior corresponde a los incrementos angulares de 3° , que dan lugar a una distribución uniforme de los gradientes (véase la Figura 3.3b). La escala inferior muestra los valores equivalentes cuando la pendiente se mide como un incremento en la concentración del modificador (v/v , expresada como fracción volumétrica) por minuto. Es interesante indicar que este eje de abscisas puede parecer exponencial, pero no lo es.

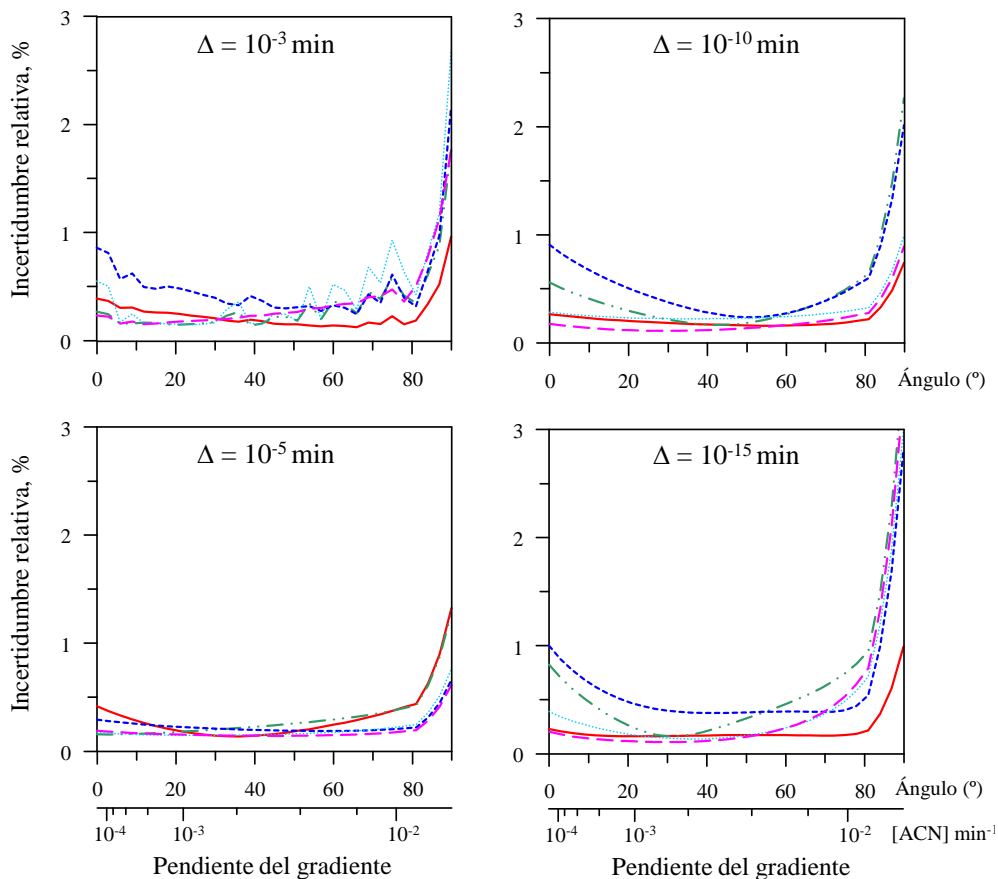


Figura 3.4. Efecto de la precisión (Δ) en el cálculo de t_g para la sulfacloropiridazina. Los mapas de incertidumbre relativa ($S_{R,\text{pred}}$) muestran las consecuencias de calcular las matrices jacobianas con tiempos de retención progresivamente más precisos ($\Delta = 10^{-3}$, 10^{-5} , 10^{-10} y 10^{-15} min), para cada uno de los cinco diseños de entrenamiento mostrados en las Figuras 3.1 y 3.2. Identificación de los diseños: ISO1 (línea roja continua), ISO2 (magenta de trazos largos), G1 (verde con una pauta de puntos y trazos), G2 (azul oscuro de trazos cortos) y G3 (azul claro punteado).

Los mapas de incertidumbre típicos para la elución en gradiente muestran patrones en U, con incrementos en ambos extremos (correspondientes al gradiente más plano y al más pronunciado), con errores bajos en el centro. Sin embargo, no siempre son visibles ambos extremos. En la siguiente sección, se realiza un análisis detallado de los mapas obtenidos (que incluye la polaridad del soluto, ambos modos de elución y los dos modelos de retención, LSS y NK). Un punto crítico para la elaboración de los mapas de incertidumbre es el nivel de precisión en el cálculo de \hat{t}_g , al que nos referimos como Δ . Cabe recordar que \hat{t}_g se obtiene por métodos numéricos, a partir de la Ecuación (3.10). La magnitud de Δ influye de forma muy importante en el tiempo de cálculo y afecta significativamente a la exactitud de las derivadas en las matrices jacobianas.

La Figura 3.4 ilustra las consecuencias de emplear valores de Δ progresivamente más pequeños. Cabe indicar que, para la simulación de cromatogramas y la estimación de la resolución, un valor de $\Delta = 10^{-3}$ min es perfectamente adecuado, y puede calcularse rápidamente sin necesidad del uso de métodos de búsqueda de raíces [24]. Sin embargo, como se puede observar, el patrón en U para el valor $\Delta = 10^{-3}$ min posee un ruido apreciable, y el ruido sigue siendo visible para $\Delta = 10^{-5}$. Más allá de $\Delta = 10^{-5}$, se necesitan métodos de búsqueda de raíces y las curvas se vuelven más suaves, aunque todavía algo sesgadas. Progresivamente (no se muestran los mapas para todos los niveles ensayados), las curvas aparecen mejor definidas hasta permanecer constantes más allá de $\Delta = 10^{-13}$. Se observaron resultados similares para el resto de solutos y diseños. En consecuencia, para los próximos estudios, se seleccionó el valor $\Delta = 10^{-15}$ min para la evaluación de \hat{t}_g .

3.5.4. Evaluación de los diseños

A partir de este punto, nos centraremos en el modelo NK (Ecuación (3.2)). Para la interpretación de los resultados, se debe tener en cuenta que un buen ajuste de un modelo de retención requiere:

- (i) valores suficientemente elevados y variados de los tiempos de retención (lo que está relacionado con la magnitud de las interacciones que se establecen con la columna, asociados al término k_w en el modelo NK), y
- (ii) valores variados de φ (asociados a los términos B y c que describen la dependencia con las composiciones) proporcionados por el diseño.

Por lo tanto, la participación de cada composición a lo largo de la migración de los solutos en los diseños de gradiente debe ser suficiente, y los intervalos de t_R y φ deben ser lo más amplios posible.

De los 14 solutos considerados inicialmente, dos de ellos (sulfaguanidina y sulfanilamida) no permitían un cálculo adecuado debido a su retención excesivamente corta. Para estos solutos, el número efectivo de grados de libertad en el ajuste de los datos de cada diseño experimental, al modelo NK, fue insuficiente (es decir, las diferencias en la retención entre los experimentos del mismo diseño eran cercanas a 0.1 min, e incluso menores). Por lo tanto, el cálculo de las incertidumbres fue menos fiable, con excepción del Diseño ISO1. Para el éxito de la metodología propuesta, solutos tan rápidos como la sulfaguanidina y la sulfanilamida requerirían diseños específicos, que incluyeran eluyentes más lentos, y por esta razón, se les excluyó de las discusiones que aparecen a continuación. En el caso de la sulfadiazina, el Diseño G1 también dio lugar a tiempos de retención insuficientemente distintos (marcados con una "a" en la Figura 3.2), lo que se tradujo en incertidumbres muy elevadas.

3.5.4.1. Comparación de las predicciones isocráticas frente a las predicciones en gradiente

Las Figuras 3.5 y 3.6 muestran los valores esperados de $s_{R,pred}$, cuando cada uno de los cinco diseños de entrenamiento (Figuras 3.1 y 3.2) se utilizan para predecir los tiempos de retención isocráticos (Figura 3.5) y de gradiente (Figura 3.6), mediante el modelo NK. Se han ordenado los gráficos de acuerdo a la polaridad decreciente de los solutos, desde sulfadiazina (con la mayor polaridad) hasta sulfaquinoxalina (con la más baja; el orden indicado en la Sección 3.4.1 sigue la secuencia de polaridad decreciente).

Una primera consideración que debe realizarse es que, cuando se inspecciona el rendimiento de un diseño experimental, para gráficos equivalentes (es decir, gráficos de incertidumbre para un mismo soluto y mismo tipo de diseño, en los modos isocrático y de gradiente), el primer punto de cada curva de incertidumbre es coincidente, debido a que la primera composición en el diseño de muestreo isocrático (10 %, Figura 3.3a) es también el primer experimento en el diseño de muestreo de gradiente (cuya pendiente es nula, véase la Figura 3.3b, gradiente 0°). Además, el valor de $s_{R,pred}$ obtenido para la concentración más elevada en el diseño de muestreo isocrático (25 % de acetonitrilo) tiende nuevamente a coincidir con la incertidumbre equivalente del gradiente que posee la mayor pendiente. La razón es que, para solutos suficientemente retenidos, el gradiente de 90° se halla cercano a una fase móvil isocrática que contiene 25 % de acetonitrilo (Figura 3.3b). En este gradiente, el efecto que produce una concentración de acetonitrilo del 10 %, a lo largo de la etapa isocrática asociada al tiempo de permanencia, es insignificante para solutos suficientemente retenidos.

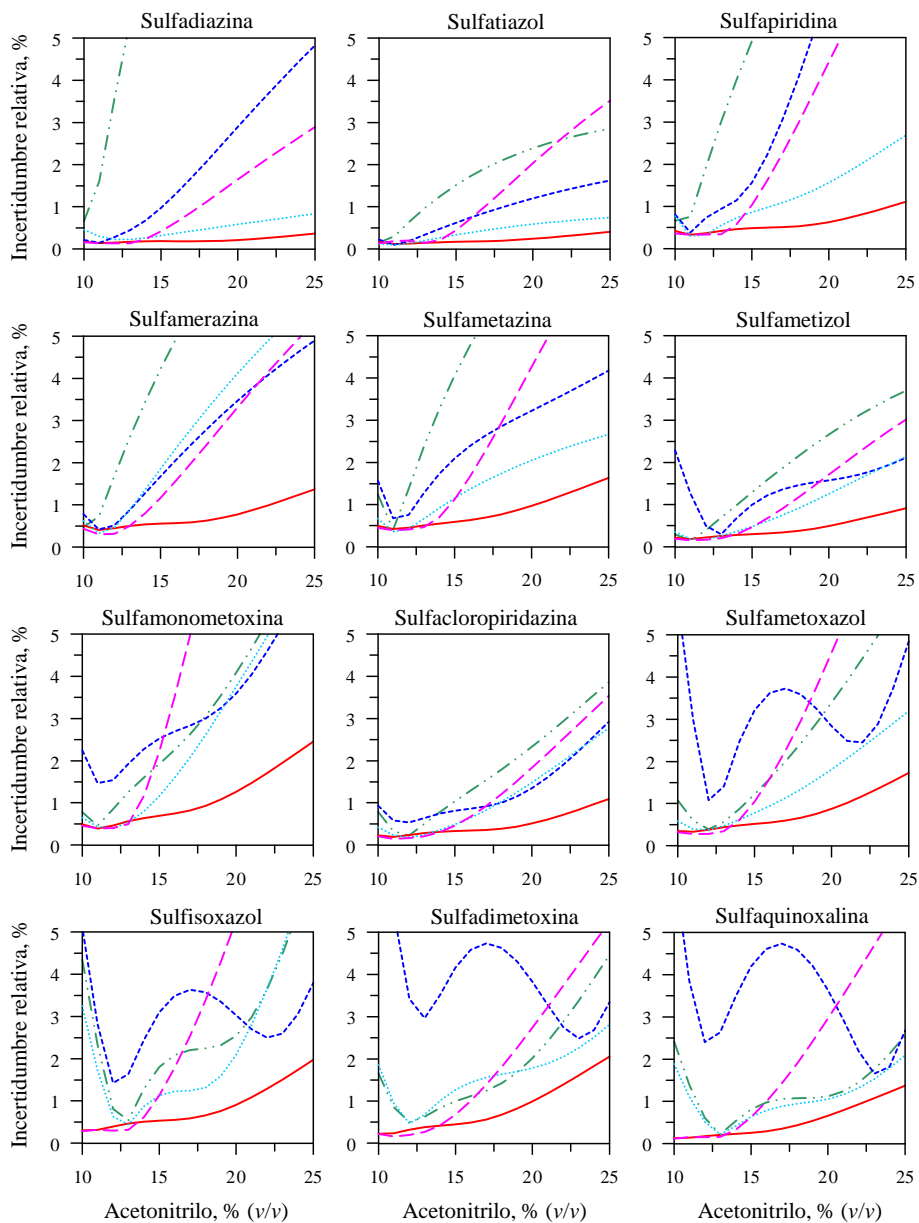


Figura 3.5. Mapas de incertidumbre relativa en la predicción de tiempos de retención, para condiciones experimentales isocráticas ($s_{R,pred}$ en función de la composición de la fase móvil), utilizando los cinco diseños experimentales de entrenamiento (Figuras 3.1 y 3.2) y el modelo NK. Consúltese la Figura 3.4 para la identificación de los diseños.

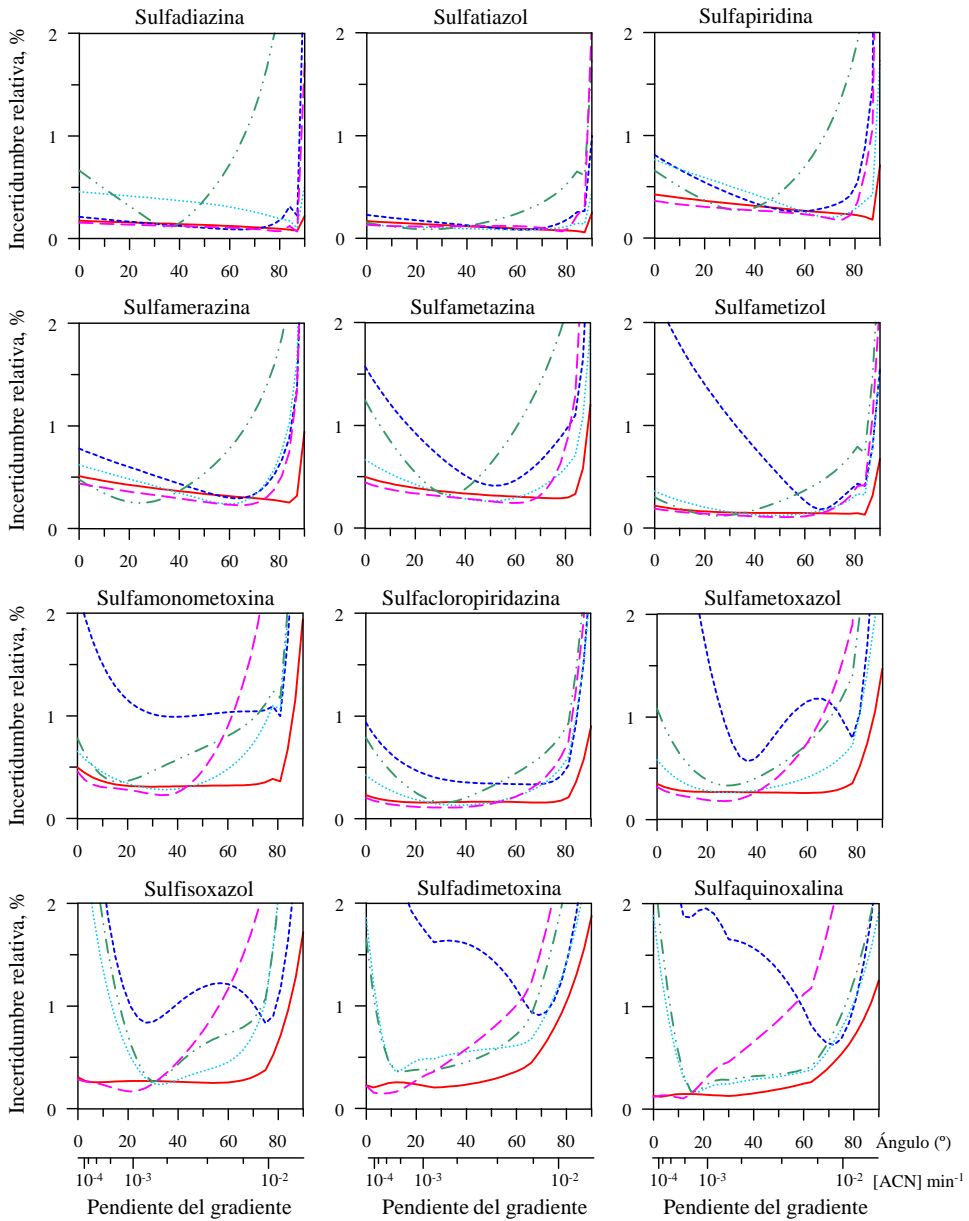


Figura 3.6. Mapas de incertidumbre relativa en la predicción de tiempos de retención para condiciones experimentales de gradiente ($J_{R,pred}$ en función de la pendiente de la rampa del gradiente), utilizando los cinco diseños experimentales de entrenamiento (Figuras 3.1 y 3.2) y el modelo de Neue-Kuss. Consúltese la Figura 3.4 para la identificación de los diseños.

En las regiones intermedias de los gráficos de incertidumbre, $s_{R,pred}$ varía de manera particular, dependiendo del soluto y el diseño de entrenamiento. A pesar de esta variabilidad, se puede observar que las curvas cambian sistemáticamente a medida que disminuye la polaridad del soluto, con una consistencia que da soporte tanto a los cálculos como a los resultados. Los valores de $s_{R,pred}$ son generalmente mayores cuando se pronostican condiciones isocráticas (debe tenerse en cuenta que la escala en los gráficos isocráticos y de gradiente no es la misma). La magnitud de la incertidumbre mínima en los gráficos isocráticos y de gradiente equivalentes tiende a ser similar, pero se deteriora más rápido para las predicciones isocráticas (incluso cuando el diseño de entrenamiento es adecuado). Por lo tanto, los gradientes se predicen con incertidumbres intrínsecamente más bajas para cualquier diseño experimental de entrenamiento.

3.5.4.2. Predicciones isocráticas

En términos generales y para el modelo NK, el Diseño ISO1 es el más adecuado para las predicciones isocráticas (curvas inferiores en la Figura 3.5), con un valor máximo de $s_{R,pred}$ de alrededor del 2 %. Además, considerando la dependencia con la composición del modificador a lo largo del escaneo, ISO1 es el menos afectado entre todos los diseños, con incertidumbres de predicción uniformes en todo el dominio.

Como era de esperar, el Diseño ISO2 presenta incertidumbres similares al Diseño ISO1, pero sólo con los eluyentes más lentos. Su rendimiento decae abruptamente cuando las predicciones se realizan excediendo el dominio cubierto (10–14 % de acetonitrilo). Más allá de este intervalo y para solutos lentos o intermedios (que implican extrapolaciones), el Diseño ISO2 muestra un comportamiento similar a los diseños de entrenamiento de gradiente (G1 a

G3), aunque con un deterioro más intenso, reflejando la ausencia de información para las composiciones más rápidas. Para los solutos más rápidos, los diseños de gradiente G1 y G2 funcionan incluso peor.

El Diseño G2 (constituido por un conjunto de gradientes con t_G variable) funciona mejor que el Diseño G1 (con φ_F variable) para los solutos más rápidos (sulfadiazina a sulfamonometoxina). Debe observarse que estos solutos eluyen en intervalos de concentración más amplios con el Diseño G2 (consúltese la etiqueta "b" en la Figura 3.2), lo que implica que los ajustes se benefician de una información más rica para las composiciones de modificador más altas, en comparación con el Diseño G1. Esta situación se invierte para los solutos más hidrofóbicos, ya que el Diseño G1 permite que los eluyentes más lentos participen en mayor medida en la retención de los solutos, antes de que los gradientes alcancen mayores concentraciones de eluyente. Por esta razón, el Diseño G1 es más informativo para los cinco solutos más retenidos (compárense las curvas de los Diseños G1 y G2 en la Figura 3.5). En comparación al Diseño ISO1, los Diseños G1 y G2 son insuficientemente informativos para los eluyentes más rápidos (aún peor en el caso del Diseño G1): cuanto más polar es el soluto, más grave es la situación.

Finalmente, se puede observar que el Diseño G3, propuesto para su estudio por reunir las buenas cualidades de los diseños isocráticos y de gradiente, proporciona un rendimiento razonablemente bueno para todo tipo de solutos. Por lo tanto, es la configuración más recomendada, entre las estudiadas en este capítulo, para la modelización de la retención, a excepción del Diseño ISO1.

3.5.4.3. Predicciones en gradiente

Para las predicciones en gradiente, la superioridad del Diseño ISO1 es igualmente notable. El Diseño ISO2 es, como era de esperar, incluso

ligeramente mejor que el ISO1 en la predicción de gradientes de baja pendiente. Esta ventaja se pierde a medida que aumenta la pendiente del gradiente, y la pérdida de capacidad predictiva se vuelve más importante para los solutos más hidrofóbicos, donde el peso de las composiciones más altas del eluyente es mayor. Esto se traduce en un desplazamiento progresivo del aumento de la incertidumbre hacia menores concentraciones de modificador.

Teniendo en cuenta el paralelismo entre el diseño de muestreo de gradiente (Figura 3.3b) y los diseños de entrenamiento G1 y G2 representados en la Figura 3.2, no sorprende que el Diseño G1 proporcione mejores predicciones para pendientes de la rampa del gradiente más pequeñas, para todos los solutos, mientras que el Diseño G2 es mejor para las pendientes más acusadas. El mejor diseño de entrenamiento de gradiente para la predicción de gradientes es nuevamente el Diseño G3, sólo superado por el Diseño ISO1. La magnitud de $s_{R,pred}$ para el Diseño G3 está por debajo del 1 % para los gradientes intermedios, siendo generalmente del 2 %, excepto en casos desfavorables.

3.5.5. Comparación del uso de los modelos de Neue-Kuss y de Snyder

Esta sección compara la capacidad de ambos modelos de retención, en términos de error de predicción. El modelo LSS presenta un inconveniente importante: la existencia de falta de ajuste (LOF, *lack of fit*). La Figura 3.7 representa el error relativo en la predicción de los tiempos de retención, para todos los solutos, calculado como:

$$ER = \frac{\hat{t}_g - t_{g,exp}}{t_{g,exp}} \times 100 \quad (3.23)$$

en función de la composición del eluyente ($ER = F(\varphi)$). En ausencia de falta de ajuste, las representaciones deberían exhibir patrones aleatorios. Sin embargo,

como se puede observar, las curvas obtenidas para el modelo LSS (Figura 3.7a), lejos de mostrar esa aleatoriedad (que es muy apreciable en las curvas para el modelo NK, Figura 3.7b), presentan correlaciones notables con φ , denotando un importante componente de falta de ajuste en la evaluación de la incertidumbre. La presencia de falta de ajuste confiere a los resultados una menor fiabilidad, y los errores reales pueden ser mayores de los esperados, con un patrón distinto. Cabe señalar que la metodología de propagación de errores no incluye las consecuencias de la falta de ajuste, por lo que no es prudente extraer conclusiones de tipo general (por ejemplo, relativo a la magnitud de la incertidumbre o el orden en la calidad de los diseños).

La simplicidad del modelo LSS origina mapas de incertidumbre relativa también más simples, con variaciones suaves, que aparentemente dan lugar a un máximo en las predicciones isocráticas (Figura 3.8), y un mínimo en las predicciones de gradiente (Figura 3.9). Sin embargo, estas variaciones pueden ser artefactos de cálculo, producidos como consecuencia de la falta de ajuste. Aun así, se puede concluir que, globalmente, las incertidumbres son mayores para el modelo LSS que para el modelo NK (compárense las Figuras 3.8 y 3.9 con las Figuras 3.5 y 3.6).

A lo largo de la secuencia de los solutos, ordenada de acuerdo a su polaridad decreciente, el rendimiento relativo del Diseño ISO1, respecto a los otros diseños, es notablemente distinto en ambos modelos de retención. Así, con el modelo LSS, el Diseño ISO1 comienza ofreciendo un rendimiento más deficiente que los conjuntos de entrenamiento de gradiente (G1 a G3), para los cuatro solutos más rápidos, mientras que presenta un rendimiento intermedio (similar para los tres diseños de gradiente) para los siguientes cinco solutos, y se convierte en el mejor diseño para los tres solutos restantes que poseen la polaridad más baja. Esta tendencia se puede observar de manera idéntica tanto

en las predicciones isocráticas como en las de gradiente (véanse las Figuras 3.8 y 3.9). Debe recordarse que el Diseño ISO1 es el mejor diseño para el modelo NK en todos los casos.

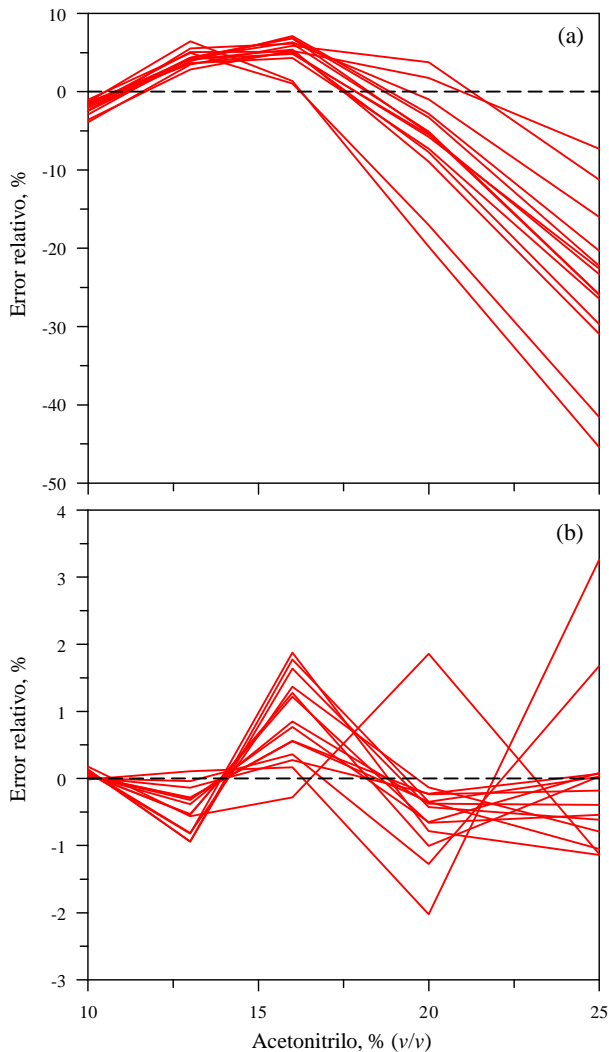


Figura 3.7. Gráfica de falta de ajuste, evidenciando posibles correlaciones entre el error relativo de predicción y la composición de la fase móvil (φ), cuando el ajuste se realiza con los modelos: (a) LSS y (b) NK.

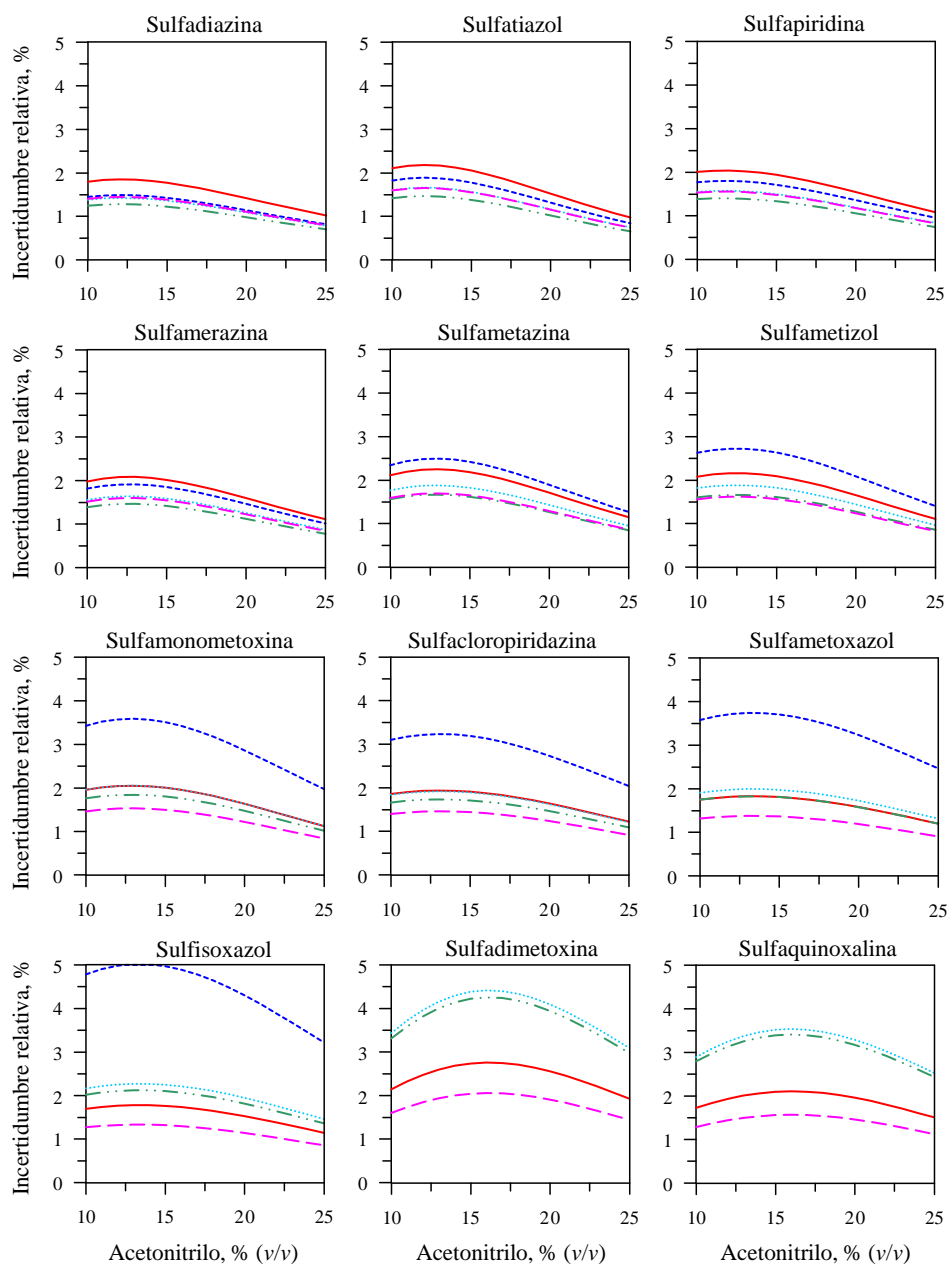


Figura 3.8. Mapas de incertidumbre relativa en la predicción de tiempos de retención para condiciones experimentales isocráticas, utilizando los cinco diseños experimentales de entrenamiento y el modelo LSS. Consúltese la Figura 3.4 para la identificación de los diseños.

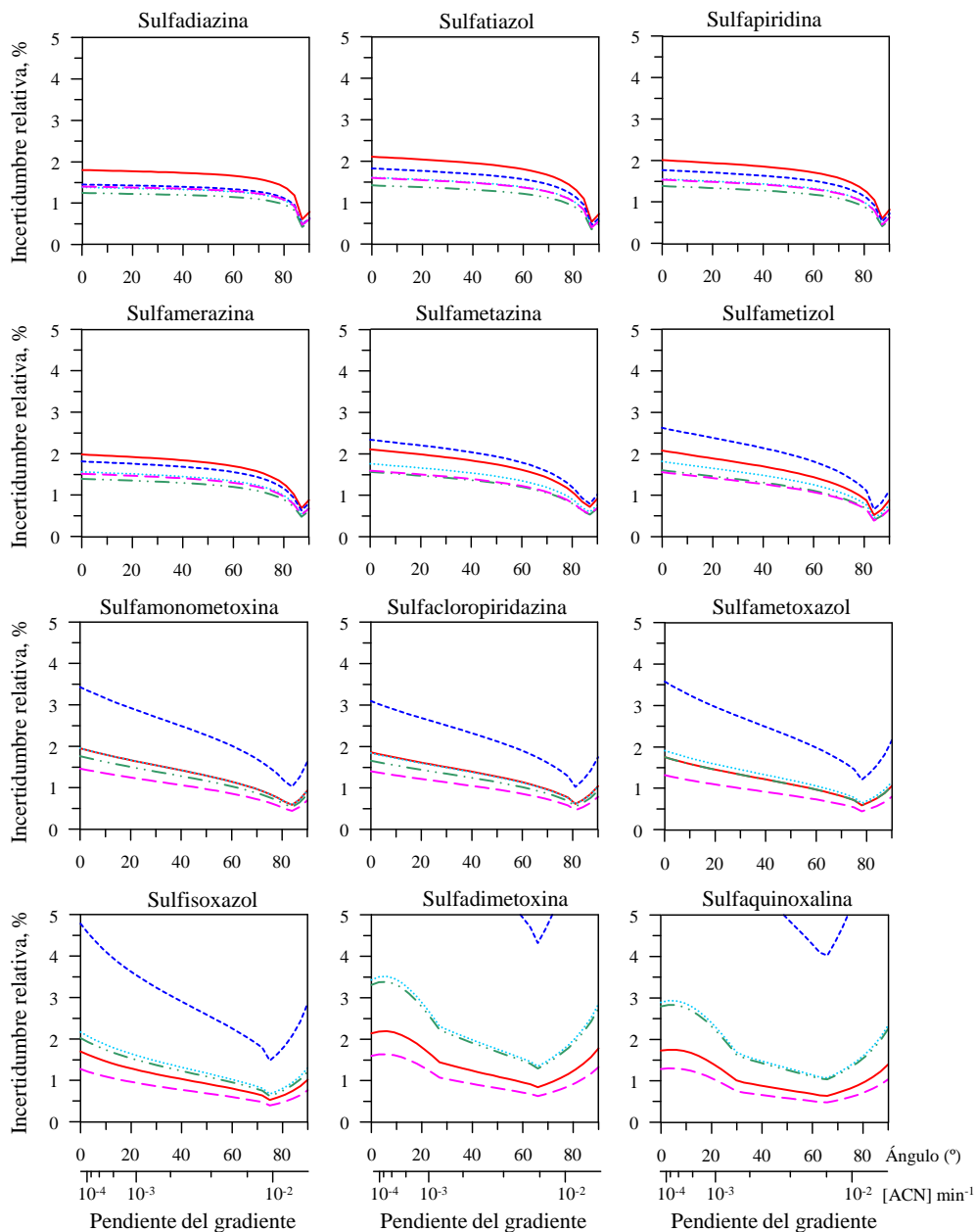


Figura 3.9. Mapas de incertidumbre relativa en la predicción de tiempos de retención para condiciones experimentales de gradiente, utilizando los cinco diseños experimentales de entrenamiento y el modelo LSS. Consúltese la Figura 3.4 para la identificación de los diseños.

El comportamiento anómalo observado para el modelo LSS es consecuencia de la falta de ajuste y puede explicarse analizando los resultados del Diseño ISO2 en las Figuras 3.8 y 3.9, donde las incertidumbres son comparables (o más favorables) a las encontradas con los tres diseños de gradiente (G1 a G3), en todos los casos. El Diseño ISO1, por el contrario, funciona peor que los tres diseños de gradiente para los solutos rápidos. De hecho, la obtención de predicciones exactas para los solutos más rápidos requiere un muestreo más exhaustivo de las composiciones más lentas y, cuando se realiza de este modo (Diseño ISO2), las incertidumbres son efectivamente más bajas. De hecho, el intervalo de concentraciones más estrecho del Diseño ISO2 hace que los efectos de la falta de ajuste sean menos visibles en el intervalo muestreado (10-14 % de acetonitrilo). Debe tenerse en cuenta que el Diseño ISO2 no proporciona información para composiciones de elevada fuerza eluyente (15-25 % de acetonitrilo). Por lo tanto, el peor rendimiento del Diseño ISO1 debe atribuirse a la falta de ajuste. Además, la incertidumbre pronosticada para contenidos de modificador orgánico más elevados no es segura.

Finalmente, entre los tres diseños de gradiente, el Diseño G2 es el que ofrece el peor rendimiento, mientras que el Diseño G3 se vuelve progresivamente equivalente al Diseño G1 (el mejor de los tres gradientes), a lo largo de la secuencia de polaridad. Estos resultados son consistentes con los encontrados en las Secciones 3.5.4.2 y 3.5.4.3.

A lo largo del estudio, se interpretaron las diferentes curvas de incertidumbre asociadas a cada diseño y soluto, en base a tres principios fundamentales:

- (i) cada composición a lo largo de un diseño de gradiente debe participar significativamente en la retención de los solutos,

- (ii) el intervalo de composiciones muestreadas a lo largo de la elución debe ser amplio, y
- (iii) los tiempos de retención encontrados a lo largo de un cierto diseño deben reflejar esa diversidad.

Para los dos solutos más polares (sulfaguanidina y sulfanilamida), que apenas se retuvieron para las composiciones incluidas en la mayoría de los diseños de entrenamiento, estas premisas no se cumplieron.

3.6. Conclusiones

Este capítulo presenta una metodología orientada a estudiar integralmente el rendimiento de diseños experimentales arbitrarios (diseños de gradientes isocráticos simples o multi-isocráticos, lineales o multi-lineales, o mixtos conteniendo gradientes y experiencias isocráticas, entre otros), basada en la determinación de las incertidumbres relativas obtenidas en la predicción de la retención ($S_{R,pred}$). La metodología se comprueba con cinco diseños experimentales de amplio uso, mostrados en las Figuras 3.1 y 3.2 (ISO1, ISO2, G1, G2 y G3), utilizados en la elución de 14 sulfonamidas de polaridad diversa, considerando dos modelos de retención: el modelo de Snyder o de la fuerza lineal del disolvente (LSS) y el modelo de Neue-Kuss (NK). Se comprobó que el modelo NK es más exacto y no muestra evidencias de falta de ajuste.

La evaluación del rendimiento predictivo de un diseño se basa en el cálculo sistemático de las incertidumbres asociadas a la estimación de los tiempos de retención, cuando se predice una distribución regular de condiciones experimentales isocráticas y de gradiente (diseños de exploración o de muestreo), a partir del diseño utilizado para modelizar la retención de los solutos eluidos (diseño de entrenamiento). El tratamiento requiere el cálculo de matrices jacobianas asociadas a los diseños de entrenamiento y de muestreo.

Las matrices jacobianas se pueden calcular algebraicamente para diseños que consideran el modelo LSS y experiencias isocráticas o gradientes lineales simples. Este cálculo no es práctico para el modelo NK en elución en gradiente, incluso para los gradientes más simples. Estas limitaciones obligaron a realizar la evaluación de las matrices jacobianas mediante procedimientos numéricos. Los resultados se presentan en términos de $s_{R,pred}$, que proporciona resultados más significativos e interpretables que las incertidumbres absolutas (s_{pred} en la Ecuación (3.19)), incluso cuando se comparan los experimentos isocráticos y de gradiente. Los cambios sistemáticos observados en los gráficos de incertidumbre, a medida que disminuye la polaridad del soluto, respaldan la fiabilidad de los cálculos, así como de los propios resultados. Debe tenerse en cuenta que los mapas de incertidumbre relativa son válidos en ausencia de falta de ajuste. En su presencia, los resultados deberían analizarse con cautela y sólo podrían derivarse conclusiones generales.

Se confirma que los diseños de entrenamiento isocráticos, que cubren todo el dominio de modificador (ISO1), son ampliamente aceptados como los más informativos, son los mejores tanto para realizar predicciones isocráticas como de gradiente, cuando el modelo no está sesgado (caso del modelo NK). En cambio, en presencia de falta de ajuste (caso del modelo LSS), el Diseño ISO1 funciona peor que los diseños de gradiente para los solutos más rápidos. Un diseño de entrenamiento enfocado a fases móviles de baja fuerza eluyente (ISO2) da lugar a incertidumbres similares al Diseño ISO1, en la región bien muestreada. Sin embargo, produce mayores incrementos en el error de predicción que los diseños de entrenamiento de gradiente (G1 a G3), cuando las predicciones implican extrapolaciones. Una característica interesante es que el mejor rendimiento del Diseño ISO1, cuando se utiliza el modelo NK, no se observa con el modelo LSS para los solutos más rápidos. Dentro del intervalo

más estrecho cubierto por el Diseño ISO2, no existe falta de ajuste apreciable y las predicciones en ese dominio serían correctas. Mientras tanto, el intervalo de modificador más amplio en ISO1 da lugar a falta de ajuste y una inversión anormal en el orden de la curva.

Entre los diseños de entrenamiento de gradiente, el Diseño G2 (con t_G variable) funcionó mejor para los solutos más rápidos que el Diseño G1 (con φ_F variable). Esta situación se invierte para los solutos con menor polaridad, ya que el Diseño G1 permite que las composiciones con menor fuerza eluyente participen en mayor medida en la migración de los solutos, traduciéndose en una mayor riqueza informativa sobre el efecto de estas composiciones. Ambos diseños de gradiente (G1 y G2) mostraron peor rendimiento que el diseño ISO1, para las composiciones con mayor contenido de modificador orgánico, debido a que estas composiciones participan en el diseño en menor medida. Se propuso un tercer diseño de gradiente (G3) para reunir las ventajas de los diseños ISO1, G1 y G2, y resolver algunos de sus inconvenientes. De hecho, el Diseño G3 se puede considerar como una extensión de los otros dos diseños de gradiente, al inicio del gradiente, donde falta información para los solutos rápidos. Por lo tanto, representa un equilibrio entre los diseños isocráticos y de gradiente para estos solutos. Como se esperaba, se encontró que el Diseño G3 ofrece un rendimiento comparable al Diseño G2 para los solutos más rápidos y al Diseño G1 para los más lentos. En el caso del modelo LSS, las diferencias fueron menos relevantes, aunque más notorias para los solutos más rápidos.

Como conclusión general, los gradientes se predicen con incertidumbres intrínsecamente menores, independientemente del diseño experimental de entrenamiento. Además, las predicciones de gradientes son más insensibles que las predicciones isocráticas respecto al tipo de diseño de entrenamiento utilizado. El error mínimo es comparable en magnitud, pero se ve menos

afectado por los cambios en la variable explorada cuando se predicen los gradientes. Por el contrario, en los diseños de muestreo isocráticos, la incertidumbre empeora rápidamente cuando las condiciones se apartan de las composiciones con error mínimo. Como consecuencia de los menores errores y la mayor extensión de la región de error mínimo, el patrón en U característico es más visible en los mapas de incertidumbre para las predicciones de gradiente que en las isocráticas. Todas estas consideraciones explican el rendimiento apreciablemente más satisfactorio de las predicciones de gradientes, incluso utilizando modelos sesgados. En cualquier caso, la magnitud de las incertidumbres es mayor para el modelo LSS que para el modelo NK.

En el Capítulo 4, la metodología desarrollada se aplicará a la evaluación de un número elevado de diseños, incluidas familias de diseños relacionados y mixtos. Se investigarán los efectos sobre el rendimiento predictivo, cuando los diseños varían sistemáticamente (por ejemplo, mediante un cambio en las pendientes, o en la distribución y número de los experimentos). El objetivo final es encontrar patrones de gradiente óptimamente informativos.

3.7. Referencias

- [1] S.N. Deming, S.L. Morgan, Experimental Design: A Chemometric Approach, en *Data Handling in Science and Technology*, Vol. 11, Elsevier Science, Amsterdam, 2ª ed., 1993.
- [2] D.B. Hibbert, Experimental design in chromatography: A tutorial review, *J. Chromatogr. B* 910 (2012) 2–13.
- [3] E. Rozet, P. Lebrun, P. Hubert, Design spaces for analytical methods, *Trends Anal. Chem.* 42 (2013) 157–167.

- [4] G. Vishnu Vardhana Rao, V. Srinivasa Rao, K. Ramakrishna, Experimental design based development of a simple and robust RP-HPLC technique for the related substances of dapoxetine hydrochloride in bulk drugs, *J. Chem. Pharm. Res.* 7 (2015) 639–646.
- [5] L.K. Shekhawat, A. Godara, V. Kumar, A.S. Rathore, Design of experiments applications in bioprocessing: Chromatography process development using split design of experiments, *Biotechnol. Prog.* 35 (2018) e2730 (doi:10.1002/btpr.2758, Wiley Online Library).
- [6] C. Muscat Galea, D. Didion, D. Clicq, D. Mangelings, Y. Vander Heyden, Method optimization for drug impurity profiling in supercritical fluid chromatography: Application to a pharmaceutical mixture, *J. Chromatogr. A* 1526 (2017) 128–136.
- [7] A.P. Schellinger, P.W. Carr, Isocratic and gradient elution chromatography: A comparison in terms of speed, retention reproducibility and quantitation, *J. Chromatogr. A* 1109 (2006) 253–266.
- [8] L.R. Snyder, J.J. Kirkland, J.L. Glajch, *Practical HPLC Method Development*, John Wiley & Sons, Nueva York, 2^a ed., 1997.
- [9] R. Carlson, J.E. Carlson, Design and optimization in organic synthesis, en *Data Handling in Science and Technology*, Vol. 24, Elsevier Science, Nueva York, 1992.
- [10] P.K. Sahu, N.R. Ramiseti, T. Cecchi, S. Swain, C. Patroa, J. Panda, An overview of experimental designs in HPLC method development and validation, *J. Pharm. Biomed. Anal.* 147 (2018) 590–611.
- [11] Y.B. Ji, Q.S. Xu, Y.Z. Hu, Y. Vander Heyden, Development, optimization and validation of a fingerprint of Ginkgo biloba extracts by high-performance liquid chromatography, *J. Chromatogr. A* 1066 (2005) 97–104.

-
- [12] W. Li, H.T. Rasmussen, Strategy for developing and optimizing liquid chromatography methods in pharmaceutical development using computer-assisted screening and Plackett-Burman experimental design, *J. Chromatogr. A* 1016 (2003) 165–180.
- [13] T. Sivakumar, R. Manavalan, C. Muralidharan, K. Valliappan, Multi-criteria decision making approach and experimental design as chemometric tools to optimize HPLC separation of domperidone and pantoprazole, *J. Pharm. Biomed. Anal.* 43 (2007) 1842–1848.
- [14] V.R. de Almeida Borges, A.F. Ribeiro, C. de Souza Anselmo, L.M. Cabral, V.P. de Sousa, Development of a high performance liquid chromatography method for quantification of isomers β -caryophyllene and α -humulene in copaiba oleoresin using the Box-Behnken design, *J. Chromatogr. B* 940 (2013) 35–41.
- [15] M.J. Cózar Bernal, A.M. Rabasco, M.L. González Rodríguez, Development and validation of a high performance chromatographic method for determining sumatriptan in niosomes, *J. Pharm. Biomed. Anal.* 72 (2013) 251–260.
- [16] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. DeJong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, en *Data Handling in Science and Technology*, Vol. 20B, Elsevier Science, Amsterdam, 1997.
- [17] P.F. de Aguiar, B. Bourguignon, D.L. Massart, Comparison of models and designs for optimisation of the pH and solvent strength in HPLC, *Anal. Chim. Acta* 356 (1997) 7–17.

- [18] J.R. Torres Lapasió, S. Pous Torres, J.J. Baeza Baeza, M.C. García Álvarez-Coque, Optimal experimental designs in RPLC at variable solvent content and pH based on prediction error surfaces, *Anal. Bioanal. Chem.* 400 (2011) 1217–1230.
- [19] G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Error analysis and performance of different retention models in the transference of data from/to isocratic/gradient elution, *J. Chromatogr. A* 1018 (2003) 169–181.
- [20] G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Estimation of significant solvent concentration ranges and its application to the enhancement of the accuracy of gradient predictions, *J. Chromatogr. A* 1057 (2004) 31–39.
- [21] J.A. Navarro Huerta, A. Gisbert Alonso, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Benefits of solvent concentration pulses in retention time modelling of liquid chromatography, *J. Chromatogr. A* 1597 (2019) 76–88.
- [22] U.D. Neue, H.J. Kuss, Improved reversed-phase gradient retention modeling, *J. Chromatogr. A* 1217 (2010) 3794–3803.
- [23] P. Nikitas, A. Pappa-Louisi, Expressions of the fundamental equation of gradient elution and a numerical solution of these equations under any gradient profile, *Anal. Chem.* 77 (2005) 5670–5677.
- [24] S. López Ureña, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Enhancement in the computation of gradient retention times in liquid chromatography using root-finding methods, *J. Chromatogr. A* 1600 (2019) 137–147.

- [25] W.H. Press, B.P. Flannery, S. Teukolsky, W.T. Vetterling, *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*, Cambridge University Press, Nueva York, 1992.
- [26] C.J.F. Ridder, Accurate computation of $F'(x)$ and $F'(x)F''(x)$, *Adv. Eng. Softw.* 4 (1978) 75–76.
- [27] J.R. Torres Lapasió, M.C. García Álvarez-Coque, J.J. Baeza Baeza, Global treatment of chromatographic data with MICHROM, *Anal. Chim. Acta* 348 (1997) 187–196.

CAPÍTULO 4

DISEÑOS EXPERIMENTALES EN CROMATOGRAFÍA LÍQUIDA: INFLUENCIA DE LA GEOMETRÍA DEL DISEÑO SOBRE LA CAPACIDAD PREDICTIVA DE LOS MODELOS DE RETENCIÓN

4.1. Resumen

En cromatografía líquida, la fiabilidad de las predicciones realizadas, utilizando modelos de retención, depende críticamente de la calidad del diseño experimental de entrenamiento. La búsqueda del mejor diseño es más compleja cuando se utilizan experimentos de gradiente, en lugar de isocráticos. En el Capítulo 3, se mostró el desarrollo y la validación de una metodología general basada en la teoría de propagación de errores, para evaluar la calidad de diseños de entrenamiento que utilizan gradientes. El tratamiento relaciona las propiedades matemáticas de un modelo de retención con la geometría de los diseños de entrenamiento y sus posteriores predicciones. En ese estudio, sólo se consideraron cinco diseños habituales. En este capítulo, se investiga en detalle el efecto sobre las predicciones de la retención cuando se varían sistemáticamente las características del diseño de entrenamiento (el número y la distribución de los experimentos, el contenido de modificador orgánico inicial y final, las pendientes del gradiente y la ubicación de sus nodos, así como la presencia de pulsos). Se evaluaron varios grupos de diseños relacionados, que contienen una o más etapas isocráticas, gradientes lineales o multi-lineales, y experimentos mixtos isocráticos y de gradiente, entre otros (en total 38 diseños). Para evidenciar las diferencias en la capacidad predictiva de los diferentes diseños de entrenamiento, se utilizaron diagramas de caja y bigotes, así como diagramas triples de incertidumbre relativa. El propósito de este estudio es proporcionar recomendaciones para construir diseños con una buena capacidad predictiva. Los mejores diseños son capaces de muestrear (considerando todos los experimentos incluidos en el diseño) concentraciones tan diversas como sea posible, a cualquier tiempo de gradiente.

4.2. Introducción

A menudo, la optimización de la resolución en cromatografía líquida se lleva a cabo con la información proporcionada por un conjunto reducido de experimentos, cuidadosamente planificados, a partir de los cuales se obtienen modelos que describen el comportamiento de retención para cada analito [1–4]. Estos experimentos deben ser lo más informativos posibles y cubrir un amplio dominio de los factores involucrados. Los modelos están destinados a predecir los tiempos de retención en nuevas condiciones arbitrarias. A partir de ellos, se pueden seleccionar las mejores condiciones de separación mediante metodologías asistidas por ordenador, en la denominada optimización interpretativa. La fiabilidad de las predicciones depende de la calidad del diseño experimental de entrenamiento, que debe obtenerse con el menor esfuerzo posible y proporcionar una incertidumbre mínima [1,5,6]. La búsqueda del mejor diseño es apreciablemente más compleja en elución en gradiente, en comparación con el modo isocrático, debido al gran número de posibles perfiles involucrados.

Encontrar el diseño más conveniente para cada soluto de interés en una muestra puede implicar el examen de un número prohibitivo de condiciones experimentales, lo que es inviable en la práctica. Una forma razonable de inspeccionar los diseños experimentales de manera integral es mediante la realización de estudios *in silico*, basados en modelos de retención, con el fin de utilizar un diseño común para todos los solutos objetivo, que supongan el uso de un número reducido de experimentos [7,8].

Hace una década, el grupo investigador llevó a cabo un estudio con el objetivo de obtener diseños experimentales óptimos en cromatografía líquida de fase inversa (RPLC, *Reversed-phase liquid chromatography*), que se aplicó a la separación isocrática de un conjunto de compuestos ionizables [7]. El estudio se

basó en el Principio de Optimalidad G [9,10], con el que se optimizó el contenido de modificador y el pH, a varios niveles de temperatura. Este enfoque relaciona las propiedades matemáticas del modelo de retención utilizado, y las características del diseño experimental de entrenamiento, con sus predicciones posteriores.

Recientemente, el grupo investigador se ha interesado en la evaluación de diseños experimentales formados por gradientes de cualquier perfil en RPLC, haciendo uso del contenido de modificador como factor. El problema radica en que encontrar un diseño con una distribución óptima de experimentos de gradiente no es sencillo. Además, la combinación de modelos de retención y programas de elución en gradiente no suele proporcionar expresiones algebraicas fáciles de manejar [11,12]. Ello obliga al uso de integración numérica para el cálculo de las derivadas parciales (ordenadas en matrices jacobianas), necesarias para evaluar las incertidumbres asociadas a los diseños experimentales. El cálculo de la capacidad de predicción es masivo y muy poco práctico, a menos que se sacrifique la exactitud. En nuestro laboratorio, se ha desarrollado un enfoque basado en métodos de búsqueda de raíces [13], que reduce significativamente el tiempo de cálculo, manteniendo una elevada exactitud. Con ello, se ha desarrollado una metodología universal competitiva (mostrada en el Capítulo 3) [14], para clasificar diseños experimentales de acuerdo a su capacidad predictiva, que permite incluir gradientes.

El cálculo implica dos tipos de diseños experimentales:

- (i) Diseños de entrenamiento, que contienen conjuntos de experimentos isocráticos y/o gradientes, y que ofrecen la información necesaria para modelizar la retención de los solutos. El objetivo de la metodología desarrollada es evaluar la capacidad de predicción de estos diseños.

- (ii) Diseños de muestreo, que consisten en condiciones isocráticas o de gradiente a explorar, dispuestas en distribuciones regulares, que se utilizan para monitorizar la capacidad de predicción de los diseños de entrenamiento, en una amplia gama de condiciones.

Las predicciones se llevaron a cabo utilizando el modelo propuesto por Neue-Kuss (modelo NK) [15], que se ha comprobado es muy superior al modelo LSS (*Linear Solvent Strength*) [16]. Hasta donde sabemos, actualmente no existe un método alternativo de características comparables para la evaluación integral de diseños de gradiente. La metodología propuesta se comprobó con cinco diseños experimentales de amplio uso en RPLC [17]. Se evaluó su capacidad predictiva en elución isocrática y en gradiente, mostrando en ambos casos resultados coherentes y fiables.

En este capítulo, se inspecciona una amplia variedad de diseños de entrenamiento (comunes e inusuales), con el fin de investigar en detalle el efecto sobre la capacidad predictiva, cuando sus características (número y distribución de los experimentos, contenido de modificador inicial y final, pendientes de los gradientes y ubicación de sus nodos, así como presencia de pulsos) se hacen variar sistemáticamente. A partir de la observación de los resultados, se ofrecen algunas recomendaciones para construir diseños con buena capacidad predictiva.

4.3. Parte experimental

4.3.1. Reactivos y experimentos isocráticos de referencia

Para evaluar la calidad de los diseños experimentales de entrenamiento, se analizaron las siguientes sulfonamidas (Sigma, Rödermark, Alemania): (1) sulfaguanidina, (2) sulfanilamida, (3) sulfadiazina, (4) sulfatiazol, (5) sulfapiridina, (6) sulfamerazina, (7) sulfametazina, (8) sulfametizol, (9) sulfamonometoxina, (10) sulfacloropiridazina, (11) sulfametoxazol, (12) sulfisoxazol, (13) sulfadimetoxina y (14) sulfaquinoxalina (ordenadas de acuerdo a su polaridad decreciente). Para la separación se utilizó una columna Zorbax Eclipse XDB-C18 (150 mm × 4.6 mm), con un tamaño de partícula de 5 µm (Agilent, Waldbronn, Alemania).

Para los 14 solutos, los tres parámetros del modelo NK (k_w , c y B):

$$k = k_w (1 + c \varphi)^2 e^{-\frac{B \varphi}{1 + c \varphi}} \quad (4.1)$$

se obtuvieron a partir de las experiencias de un diseño isocrático de cinco fases móviles de acetonitrilo-agua, concentradas gradualmente a bajos contenidos de modificador (Diseño ISO1 en la Figura 4.1). Los parámetros ajustados (14×3) se utilizaron para evaluar la capacidad de cualquier diseño de entrenamiento arbitrario, cuando se llevan a cabo predicciones isocráticas y de gradiente. Todas las inyecciones se realizaron en condiciones de temperatura controlada, a 25 °C, y un flujo de fase móvil de 1 mL·min⁻¹. El tiempo de permanencia fue de 1.16 min y el tiempo extra-columnar, 0.12 min. Se realizó el tratamiento de los datos con Matlab 2020a (The MathWorks Inc., Natick, MA, EE.UU.). Para otros detalles, debe consultarse el Capítulo 3 [14].

4.3.2. Diseños de entrenamiento

Para las 14 sulfonamidas, los diseños de entrenamiento muestrearon el intervalo de 10 a 25 % (v/v) de acetonitrilo en agua. Los experimentos con diseños isocráticos se caracterizan por el contenido de acetonitrilo en la fase móvil, mientras que los experimentos con diseños de gradiente se definen mediante al menos tres parámetros para gradientes lineales simples: contenido inicial de modificador, pendiente y tiempo de gradiente (t_G). Para gradientes multi-lineales, las coordenadas de uno o más nodos (el punto de intersección entre dos segmentos lineales consecutivos de distinta pendiente) son parámetros adicionales que modulan la elución.

El estudio se ha llevado a cabo analizando la capacidad predictiva de seis grupos de diseños de entrenamiento (Figura 4.1):

- (i) *ISO1–ISO5*: Diseños isocráticos donde se modificó el número y la distribución de las fases móviles. El Diseño ISO1 se ejecutó experimentalmente para obtener los datos necesarios que permitieran ajustar los parámetros del modelo, dando lugar a incertidumbres relativas en las predicciones por debajo del 0.5 %. Los parámetros obtenidos con ISO1 se utilizaron para evaluar todos los diseños a lo largo de este trabajo. Este diseño también se utilizó como referencia para escalar los resultados de los otros diseños estudiados en las gráficas de incertidumbre relativa predicha.
- (ii) *P1–P7*: Diseños isocráticos, en los que algunas fases incluyen incrementos breves y repentinos de contenido de modificador orgánico (pulsos), con una ubicación y duración variables.
- (iii) *SL1–SL6*: Diseños de gradientes lineales con cinco experimentos cada uno. Todos los diseños incluyeron rampas que se iniciaban al 10 % de acetonitrilo, con diferentes pendientes y tiempos de gradiente.

- (iv) *ML1–ML6*: Diseños lineales mixtos, cada uno de los cuales incluye de cuatro a seis gradientes lineales simples, junto a experimentos isocráticos. En todos ellos, los experimentos comienzan con una concentración inicial de gradiente en dos o tres niveles, con o sin un segmento isocrático inicial, mientras que la concentración final de los gradientes se mantuvo fija.
- (v) *ML7–ML13*: Diseños lineales mixtos, que incluyen cuatro o cinco gradientes lineales. La concentración inicial de los gradientes consta de tres a cinco niveles, con o sin un segmento isocrático inicial, y diversas concentraciones finales.
- (vi) *NI–N7*: Conjunto de diseños seleccionado como representativo de los diseños de entrenamiento multi-lineales. Cada diseño incluye un nodo, excepto N4 que contiene tres nodos, con o sin concentraciones de gradiente inicial y final comunes. Los gradientes N5 y N6 también incluyen segmentos isocráticos y un gradiente lineal.

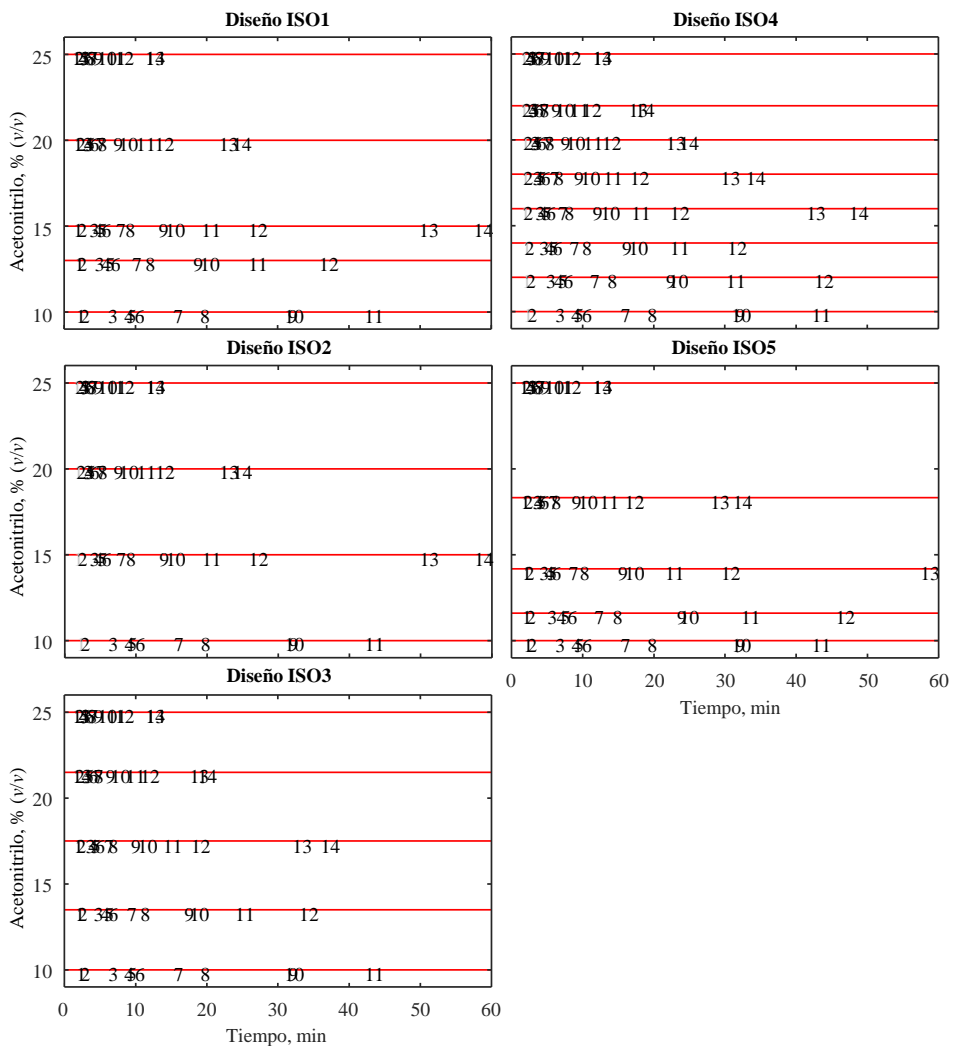


Figura 4.1. Diseños experimentales de entrenamiento isocráticos y de gradiente bajo evaluación. En cada diseño, se indican (superpuestos) los tiempos de retención de cada uno de los solutos, en las condiciones de cada experimento. En rojo, aparecen los solutos que no tienen medidas suficientemente diferenciadas para ser modelizados (grados de libertad < 0). Los solutos marcados en gris tienen tres medidas válidas (grados de libertad = 0). Consulte la Sección 4.3.1 para la identificación de los solutos. Para más detalles, véase la Sección 4.3.2.

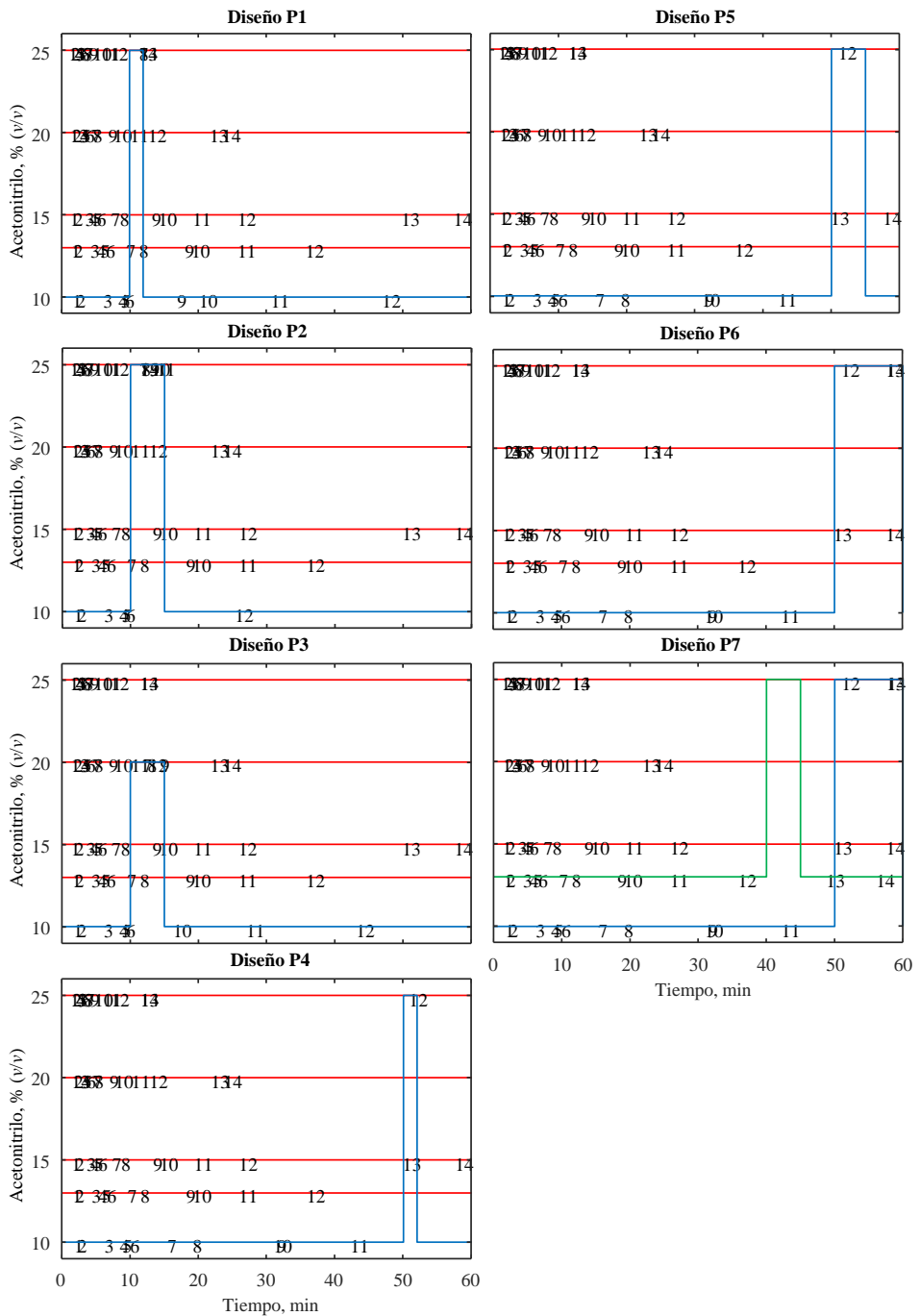


Figura 4.1 (continuación).

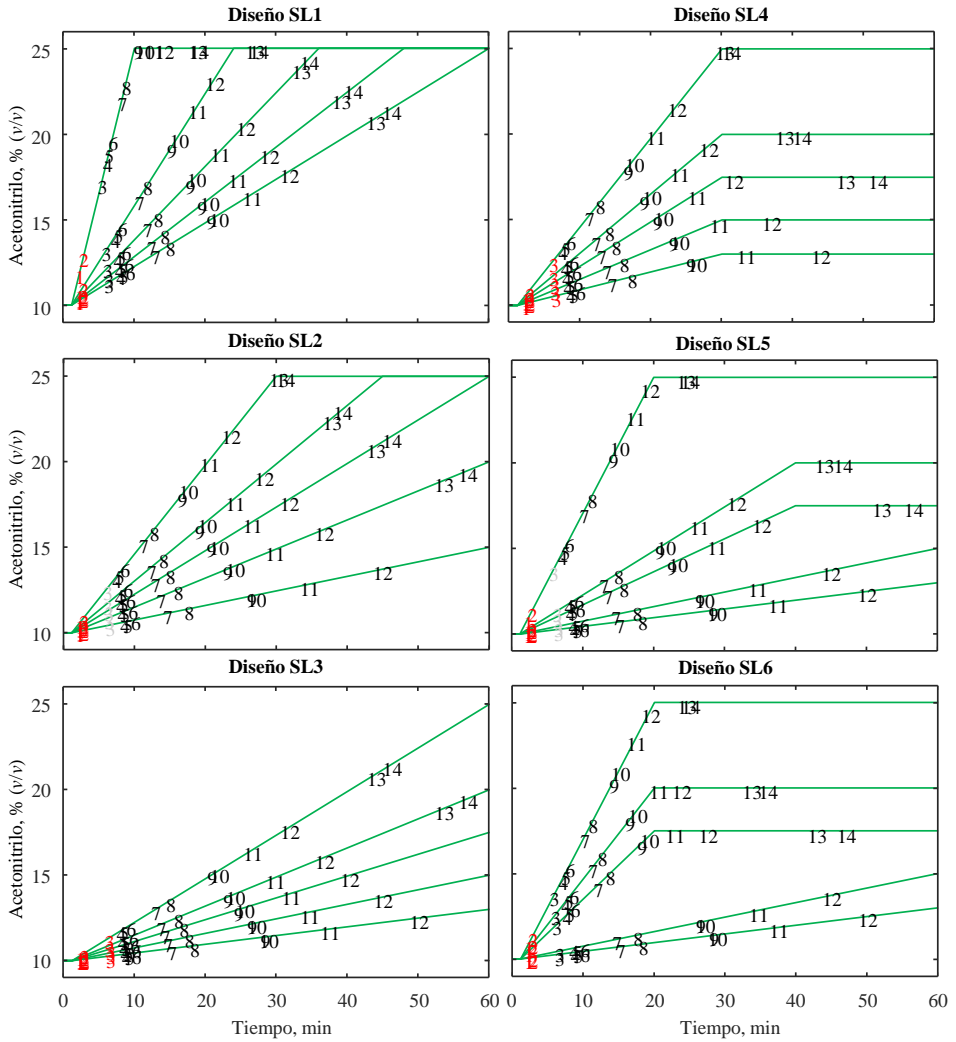


Figura 4.1 (continuación).

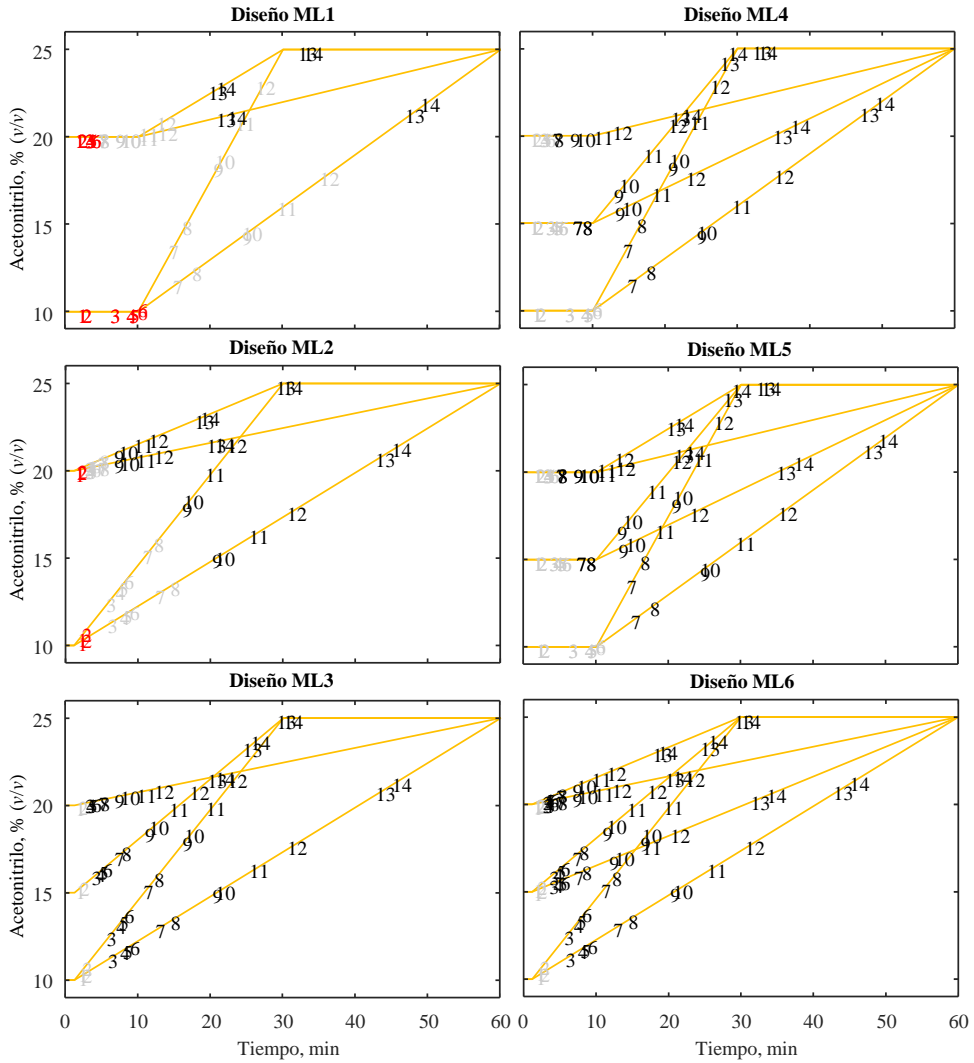


Figura 4.1 (continuación).

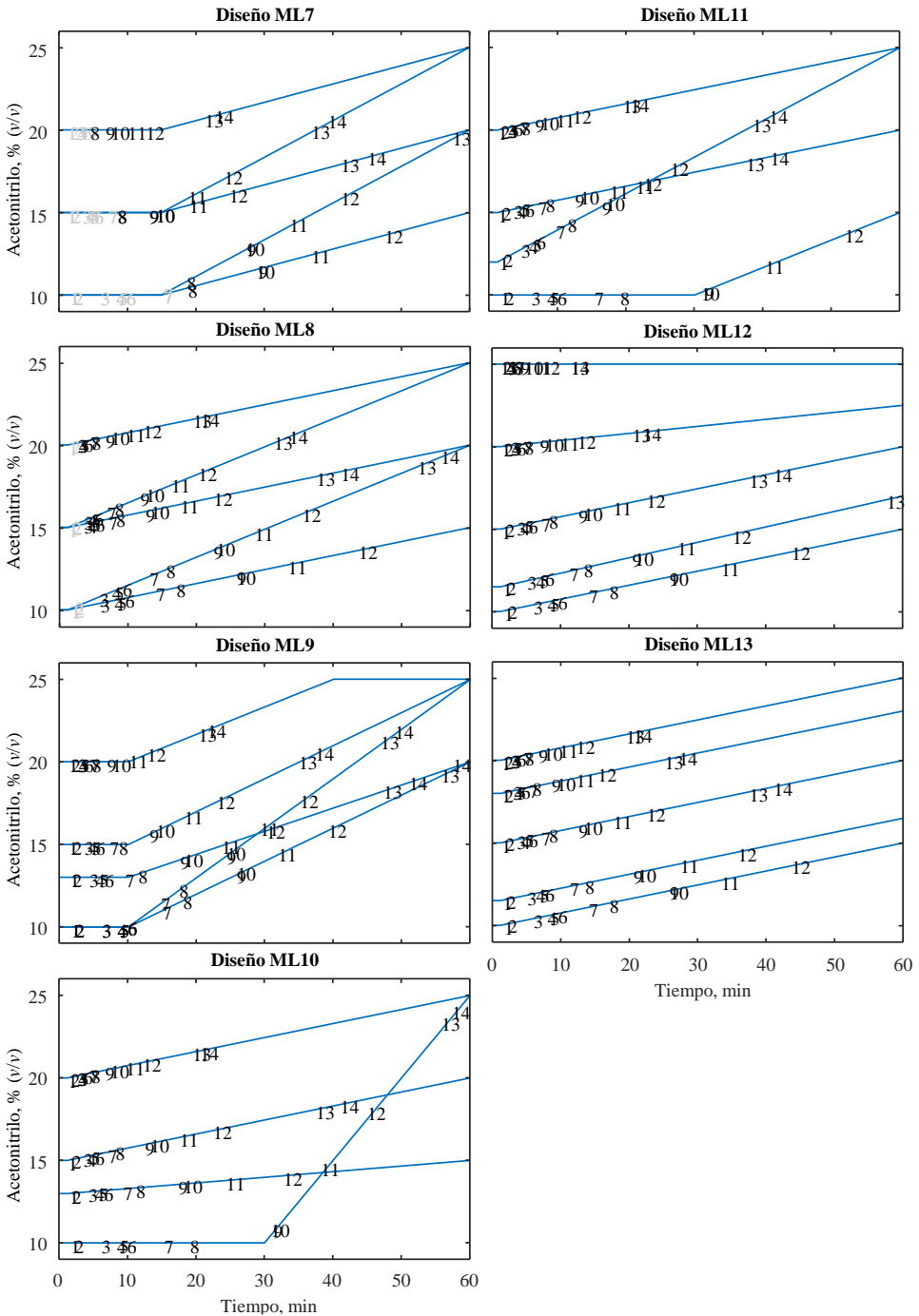


Figura 4.1 (continuación).

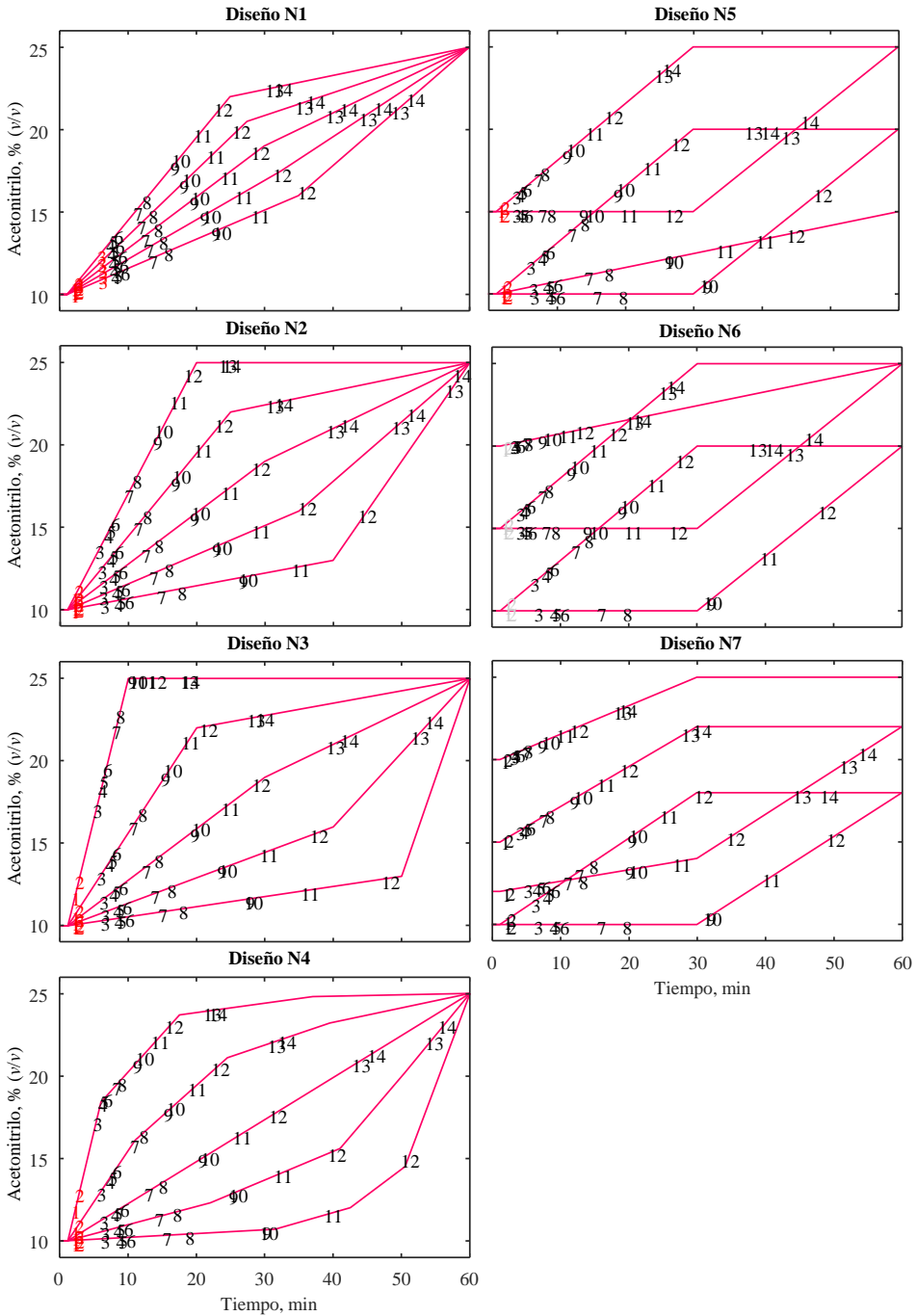


Figura 4.1 (continuación).

4.3.3. Diseños de muestreo

Con el fin de evaluar la capacidad predictiva de los 38 diseños de entrenamiento propuestos, tanto en elución isocrática como en gradiente, se consideraron dos diseños de muestreo:

- (i) *Diseño de muestreo isocrático* (Figura 3.3a en el Capítulo 3), para inspeccionar la calidad de las predicciones isocráticas, que constó de 16 experimentos en el intervalo del 10 al 25 % de acetonitrilo, distribuidos en incrementos del 1 %.
- (ii) *Diseño de muestreo de gradiente* (Figura 3.3b en el Capítulo 3), para inspeccionar la calidad de las predicciones en gradiente, que constó de 31 gradientes lineales simples distribuidos uniformemente. Todos los gradientes del diseño comienzan a un 10 % de acetonitrilo, con pendientes que aumentan progresivamente, mostrando un incremento angular constante de 3° en el intervalo de 0 a 90° , con el que se muestrea regularmente el espacio de los factores. Entre los gradientes, 16 alcanzan la concentración final a un $t_G = 60$ min, mientras que los otros 15 gradientes alcanzan la misma concentración final (25 % de acetonitrilo) a t_G variable. Más allá de la concentración máxima, la elución es isocrática.

4.4. Resultados y discusión

El estudio incluye 38 diseños de entrenamiento organizados en seis grupos, cada uno de los cuales contiene de cinco a siete diseños (Figura 4.1 en la Sección 4.3.2). Inicialmente, se consideraron más de 100 diseños de entrenamiento, de los cuales se seleccionaron 38 para ilustrar los diferentes comportamientos. Los diferentes diseños se clasificaron en grupos, ya que interesaba jerarquizar la importancia de los factores que hacen que un diseño

sea satisfactorio o deficiente en términos geométricos y, a partir de ahí, sugerir los cambios que pueden conducir a nuevas mejoras. Para comprender los resultados, es importante conocer el tiempo al que cada soluto abandona la columna en cada experimento incluido en los diseños, así como la composición instantánea correspondiente. Por ello, aparecen superpuestos, en cada uno de los experimentos de los diseños de la Figura 4.1, los códigos de los solutos estudiados (véanse sus identidades en la Sección 4.3.1). En los gráficos, la proyección en el eje de abscisas proporciona el tiempo de retención, y en el eje de ordenadas, la composición en el instante en el que el soluto abandona la columna, para cada diseño y experimento.

En el Capítulo 3, se recogen los detalles sobre el cálculo de las incertidumbres asociadas a la predicción de los tiempos de retención. Es importante señalar que el número de experimentos (n), en los diseños de entrenamiento, debe ser mayor que el número de parámetros, para que existan suficientes grados de libertad en el ajuste de la retención ($n > 3$ para el modelo NK). Además, para poder realizar un ajuste adecuado del modelo de retención, cada composición debe participar suficientemente en la migración de los solutos, a medida que progresa el gradiente, garantizando así la obtención de suficiente información.

4.4.1. Gráficos de incertidumbre relativa en las predicciones

En el Capítulo 3, se evalúa la capacidad predictiva de los diseños de entrenamiento con la ayuda de gráficos, donde se representa la incertidumbre relativa esperada en las predicciones, frente a las condiciones del eluyente (que en elución isocrática es una distribución regular de composiciones de fase móvil, y en elución en gradiente una distribución de pendientes para gradientes lineales, véase la Figura 3.3 en el Capítulo 3). Los gráficos describen cambios

graduales en las incertidumbres relativas asociados a los tiempos de retención previstos en diversas condiciones, cuando se utilizan los datos de diferentes diseños de entrenamiento para obtener modelos de retención. Dado que el comportamiento de retención depende de la polaridad y de otras propiedades de los solutos, se tuvo que construir una gráfica particular para cada soluto. En la primera parte de este estudio (Capítulo 3), sólo se analizaron cinco diseños de entrenamiento y los diagramas de incertidumbre utilizados permitieron una inspección detallada de la capacidad predictiva. El presente estudio realiza la evaluación de 38 diseños de entrenamiento, por lo que fue necesario construir otro tipo de gráficos, que condensan la información disponible y con ello facilitan que la información sea interpretable.

4.4.1.1. Diagramas de incertidumbre relativa de caja y bigotes para cada diseño

Este tipo de representación reúne y clasifica los resultados de todo el conjunto de solutos como una distribución de incertidumbre simplificada. Cada diagrama resume toda la información obtenida para cada diseño, considerando las predicciones para todos los solutos en conjunto, utilizando elución isocrática o en gradiente (Figuras 4.2 y 4.3, respectivamente). La representación conjunta de varios diagramas de caja y bigotes, utilizando una misma ventana, facilita la comparación de la capacidad predictiva de múltiples diseños. Aquellos diseños que comparten características geométricas similares se enmarcan en conjunto (véase la Figura 4.1 y la Sección 4.3.2). El diagrama de caja y bigotes para el Diseño ISO1 se ha incluido como referencia en cada fila para apreciar las diferencias de escala. En los gráficos, las incertidumbres relativas se dividieron en dos categorías: "dominio de muestreo bien descrito" (que comprende el "75 % de las experiencias con mejores incertidumbres"), representado en la

columna izquierda, y "dominio mal descrito" (con el "25 % de las experiencias con peores incertidumbres"), en la columna derecha.

Los diagramas de caja y bigotes ofrecen una visión general de la capacidad de cada diseño (véanse los resultados en la Sección 4.4.2). Sin embargo, la evaluación de los diseños no aparece diferenciada por solutos, ya que todos ellos se consideran en conjunto. Por lo tanto, no se revelan los solutos que hacen que un diseño falle. Otro inconveniente es que no se ofrece información sobre cuáles son las condiciones de exploración en el diseño de muestreo donde fallan las predicciones.

4.4.1.2. Diagramas triples de incertidumbre relativa que consideran cada diseño, soluto y composición

Para un análisis más completo de los resultados, se necesitaba un segundo tipo de representación que proporcionara la información que no ofrecen los diagramas de caja y bigotes (Sección 4.4.1.1). Para ello se optó por utilizar diagramas triples (Figuras 4.4 y 4.5), donde el eje x representa a los solutos (en columnas estrechas), el eje y indica las condiciones de elución (los experimentos isocráticos y de gradiente descritos en la Figura 3.3 del Capítulo 3), y el eje z proporciona las incertidumbres relativas en una escala de colores. Dichos gráficos ofrecen detalles sobre la capacidad predictiva, para cada soluto asociado a cada diseño, así como la calidad de las predicciones en función de la fuerza eluyente.

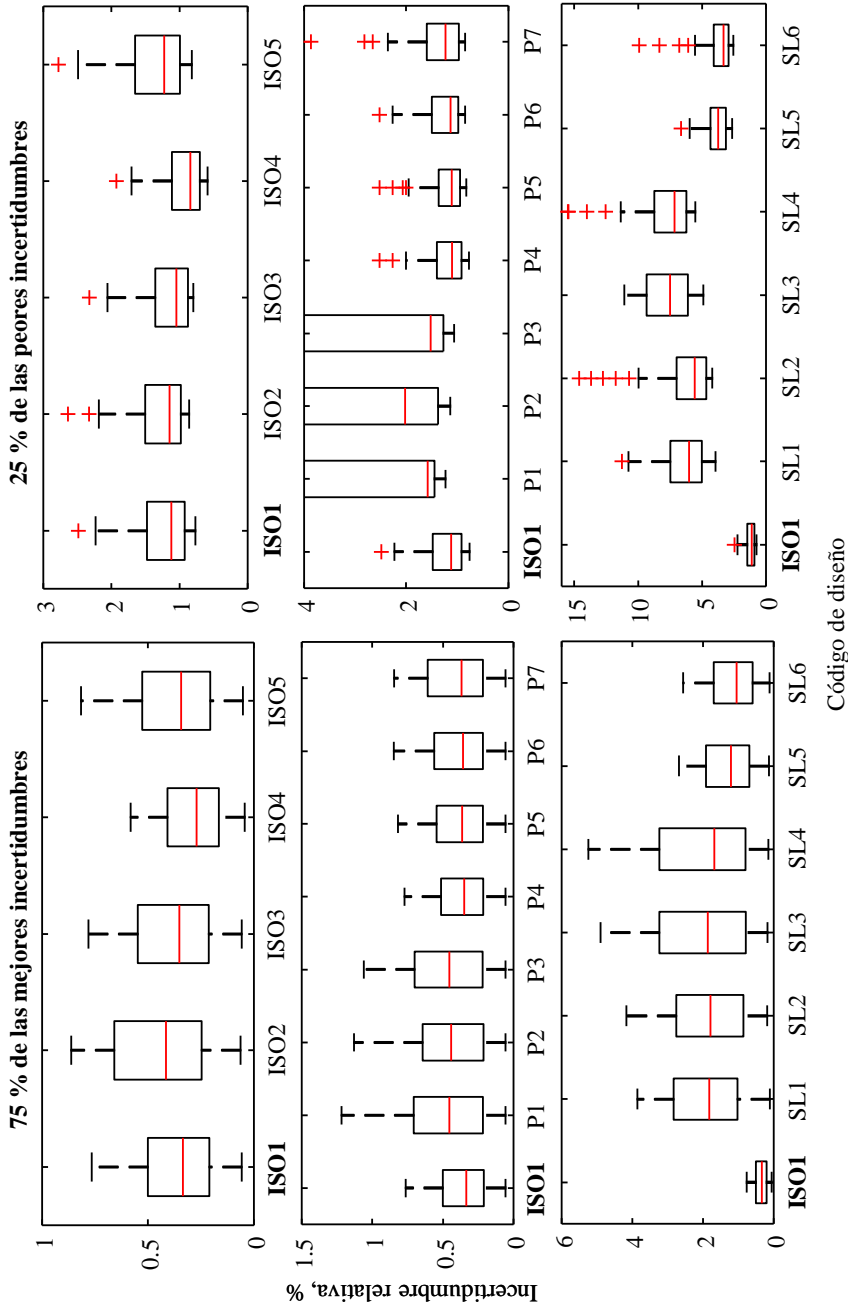
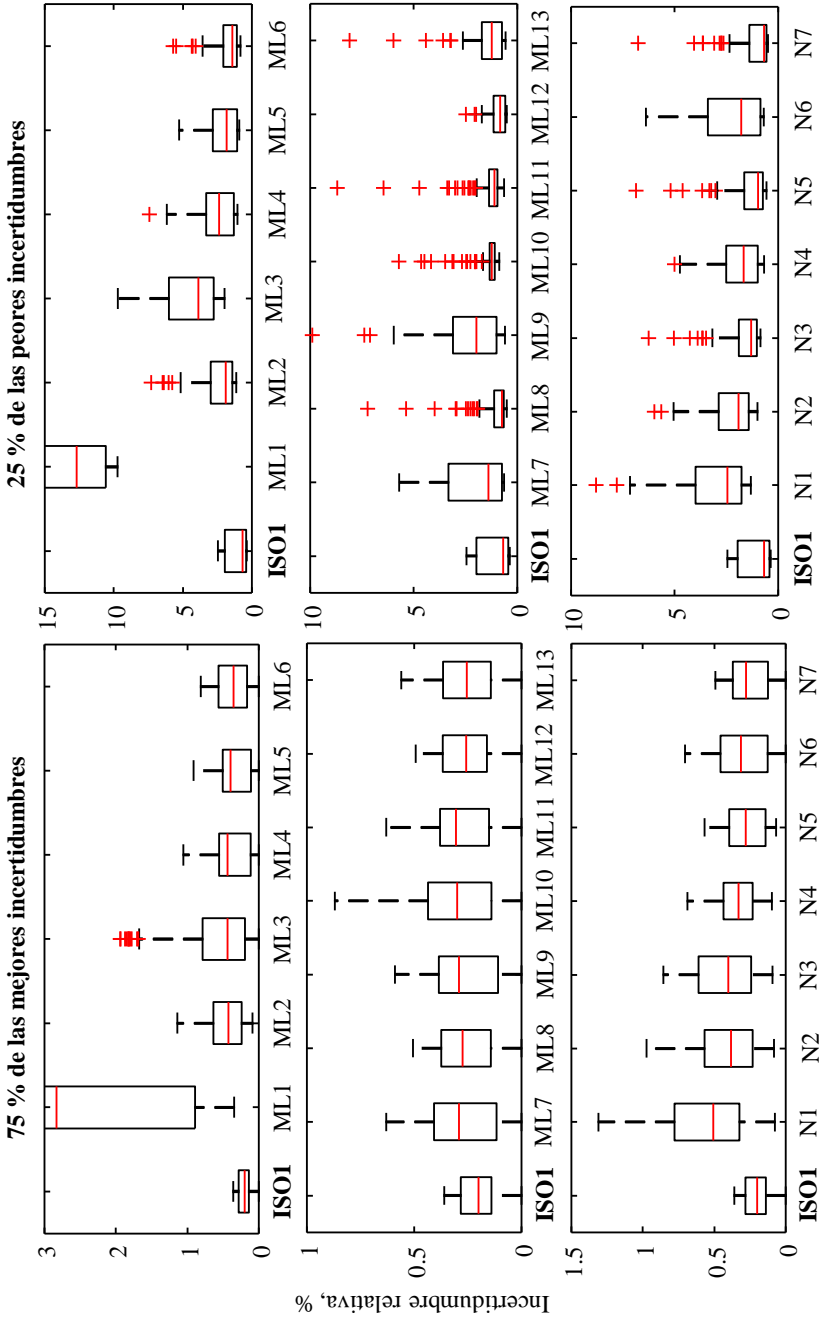


Figura 4.2. Diagramas de caja y bigotes representando las incertidumbres relativas esperadas en las predicciones isocráticas, considerando los 16 niveles de composición del diseño de muestreo (Figura 3.3a en el Capítulo 3). Para más detalles consúltase la Sección 4.4.1.1.



Código de diseño

Figura 4.2 (continuación).

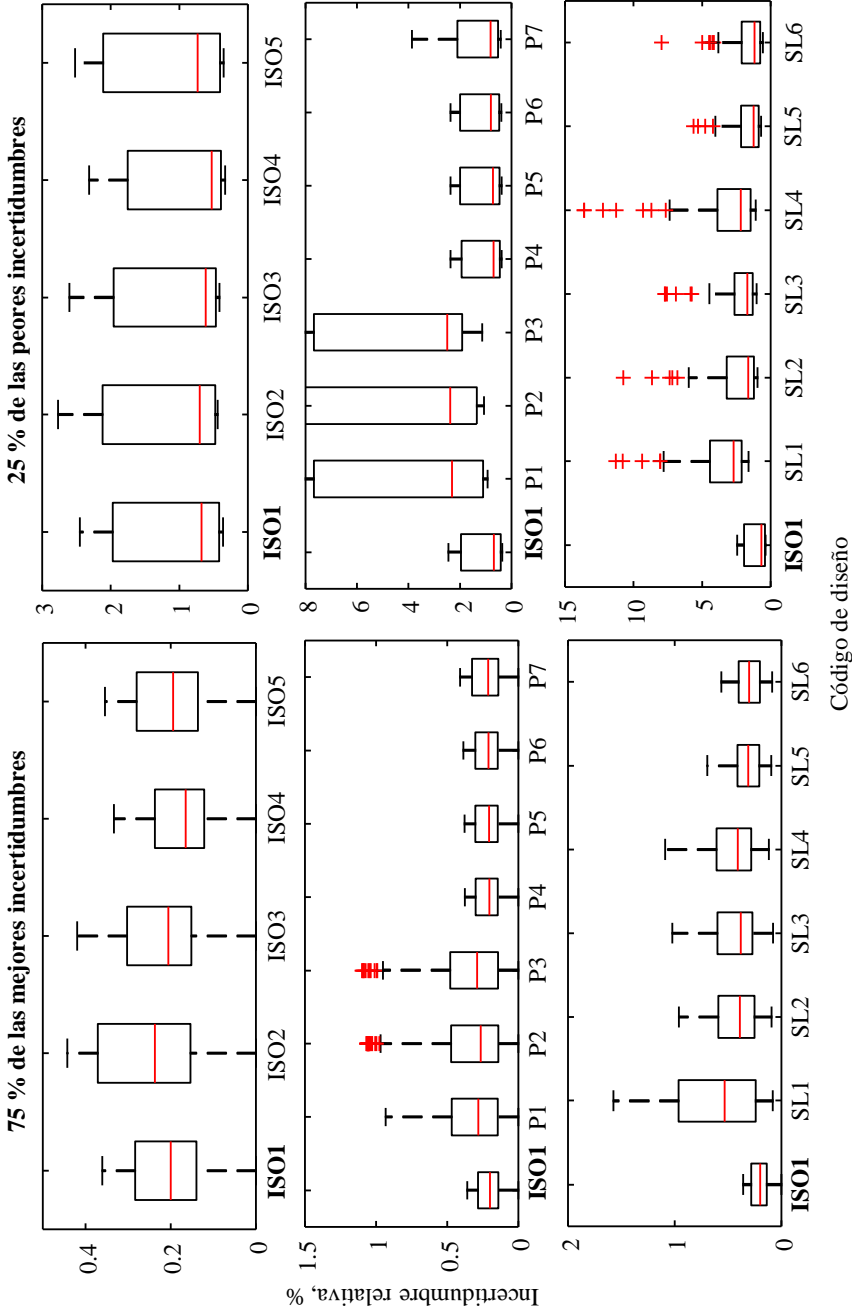


Figura 4.3. Diagramas de caja y bigotes representando las incertidumbres relativas esperadas en las predicciones de gradiente, considerando los 31 gradientes del diseño de muestreo (Figura 3.3b en el Capítulo 3). Para más detalles consúltese la Sección 4.4.1.1.

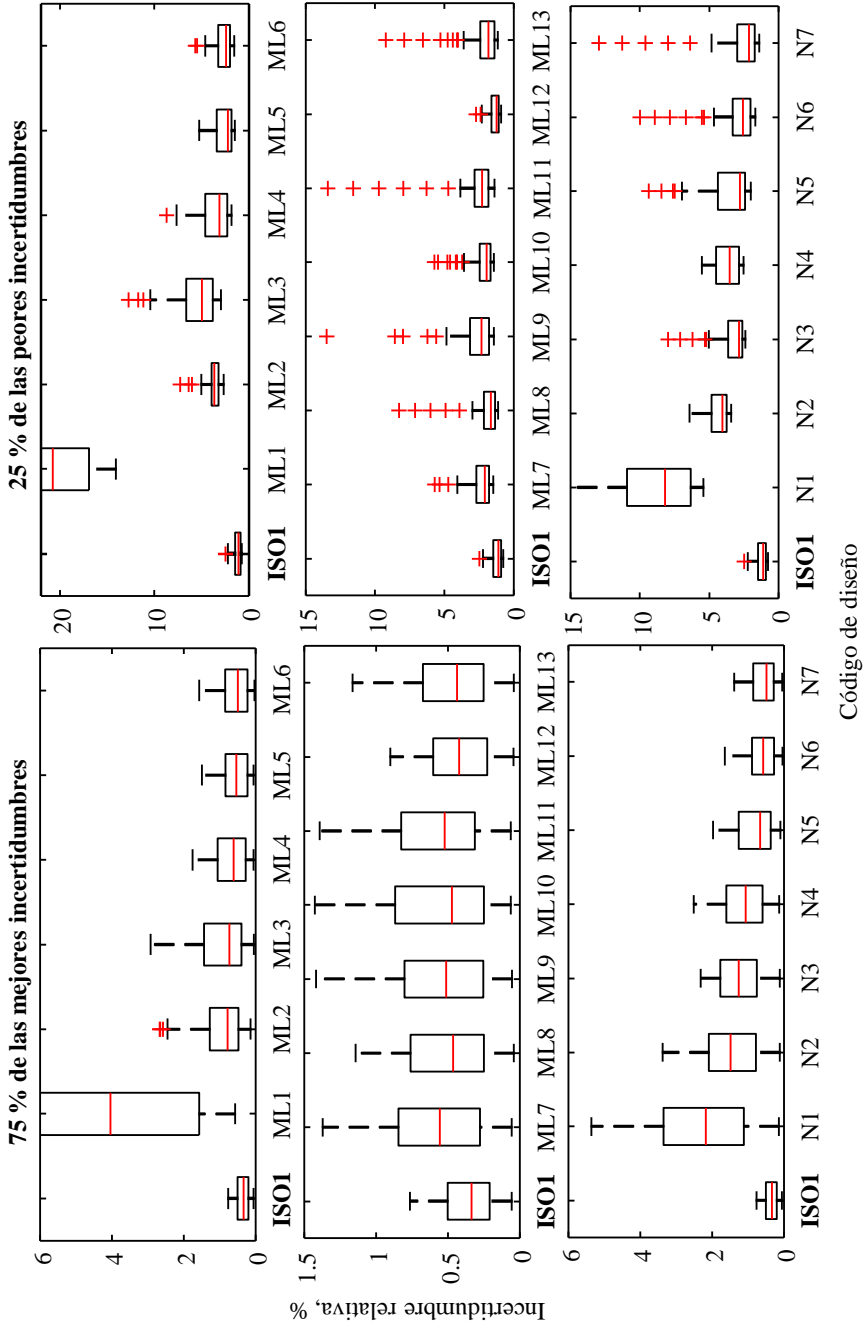


Figura 4.3 (continuación).

De manera similar a los diagramas de caja y bigotes, con fines comparativos, los diagramas triples para diseños relacionados se incluyen en la misma ventana. La escala de colores es distinta para la elución isocrática y de gradiente, lo que refleja las diferencias en la capacidad media de predicción, para ambos modos de elución. En la elución isocrática (Figura 4.4), la escala de colores osciló entre el negro para un valor del 0 % de incertidumbre relativa, al color blanco al que se asignó un 2.5 % y color rojo para el 10 %, que se intensificó a medida que aumentaba la incertidumbre relativa. Dado que las incertidumbres medias en la elución en gradiente son más bajas, en este caso, los límites representados fueron del 0 %, 1 % y 5 %, respectivamente. Más allá del 10 % y el 5 % de incertidumbre relativa en elución isocrática y de gradiente, respectivamente, se decidió que la intensidad del color rojo no variara. La razón de esta decisión se debe a que, a los valores más altos, las incertidumbres son lo suficientemente elevadas para que la magnitud sea irrelevante. Todos estos niveles son naturalmente arbitrarios y fueron seleccionados de acuerdo con las características del modelo de retención. El color azul de algunos solutos indica una variabilidad insuficiente de los datos para llevar a cabo un ajuste adecuado. Cabe indicar que, para el cálculo de los grados de libertad, dos tiempos de retención se consideraron idénticos cuando diferían en menos de 0.1 min.

Los diagramas triples proporcionan información bastante completa sobre las incertidumbres de predicción esperadas. A pesar de que se pierde algo de información, debido a que se representan colores en lugar de valores numéricos, la evaluación de los diseños es más sencilla e intuitiva, y el alto nivel de compresión de los datos permite visualizar una gran cantidad de información fácilmente. Los diagramas de caja y bigotes no brindan información tan detallada, pero son un complemento útil para ordenar la capacidad de los diseños, ya que indican su calidad general y se pueden utilizar

para tener una perspectiva global para un gran conjunto de diseños, de forma visual y rápida.

En las Figuras 4.1 a 4.4, cada fila corresponde a un grupo diferente de diseños. En las representaciones triples, si nos centramos en un diseño en concreto, los solutos rápidos se ubican a la izquierda y los solutos lentos a la derecha, en una secuencia de polaridad decreciente, mientras que las regiones de alta y baja fuerza eluyente se hallan en las partes superior e inferior, respectivamente. Así, por ejemplo, la región superior izquierda corresponde a las incertidumbres asociadas con las predicciones para solutos rápidos a fuerzas eluyentes elevadas.

4.4.2. Análisis de la capacidad predictiva de los diseños de entrenamiento

Las Figuras 4.2 y 4.4 representan el análisis de los resultados para las predicciones en elución isocrática, y las Figuras 4.3 y 4.5 corresponden a la elución en gradiente, utilizando diagramas de caja y bigotes (Sección 4.4.1.1) y representaciones triples de incertidumbre relativa (Sección 4.4.1.2) (para la interpretación de los resultados, véanse también la Figura 4.1 y la Figura 3.3 del Capítulo 3). Cabe señalar que las incertidumbres para las predicciones de gradiente son intrínsecamente más bajas, lo que refleja los intervalos más estrechos de composiciones que realmente participan en este modo de elución [18]. Esto también da lugar a que la calidad del modelo de retención sea menos exigente, lo que de nuevo puede explicarse por los intervalos de modificador más estrechos que participan en la migración. A continuación, se expone detalladamente un estudio de la capacidad predictiva de cada uno de los grupos de diseños de entrenamiento.

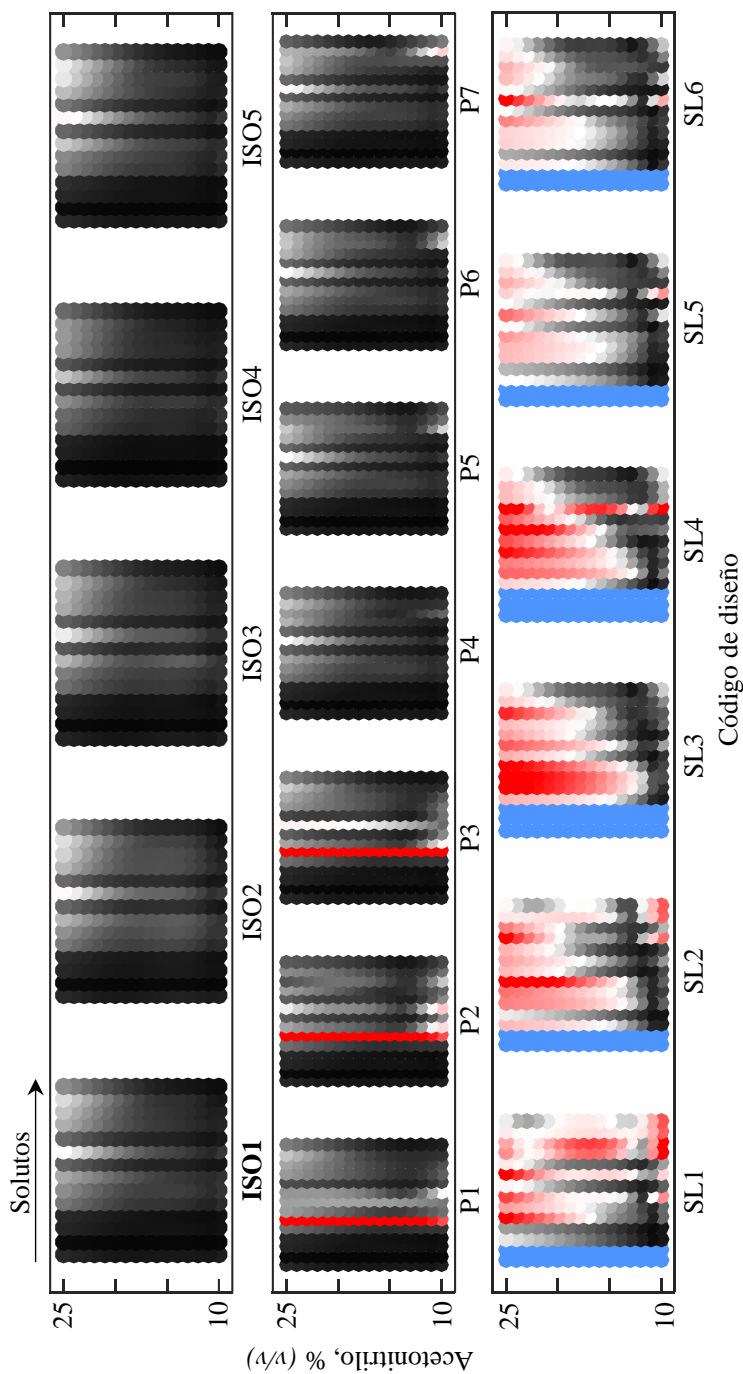


Figura 4.4. Diagramas triples mostrando las incertidumbres relativas esperadas en la predicción de condiciones isocráticas. Los resultados se muestran para cada composición predicha y soluto individual, según el diseño isocrático de muestreo. Los detalles relevantes para la interpretación de los diagramas triples se encuentran en la Sección 4.4.1.2.

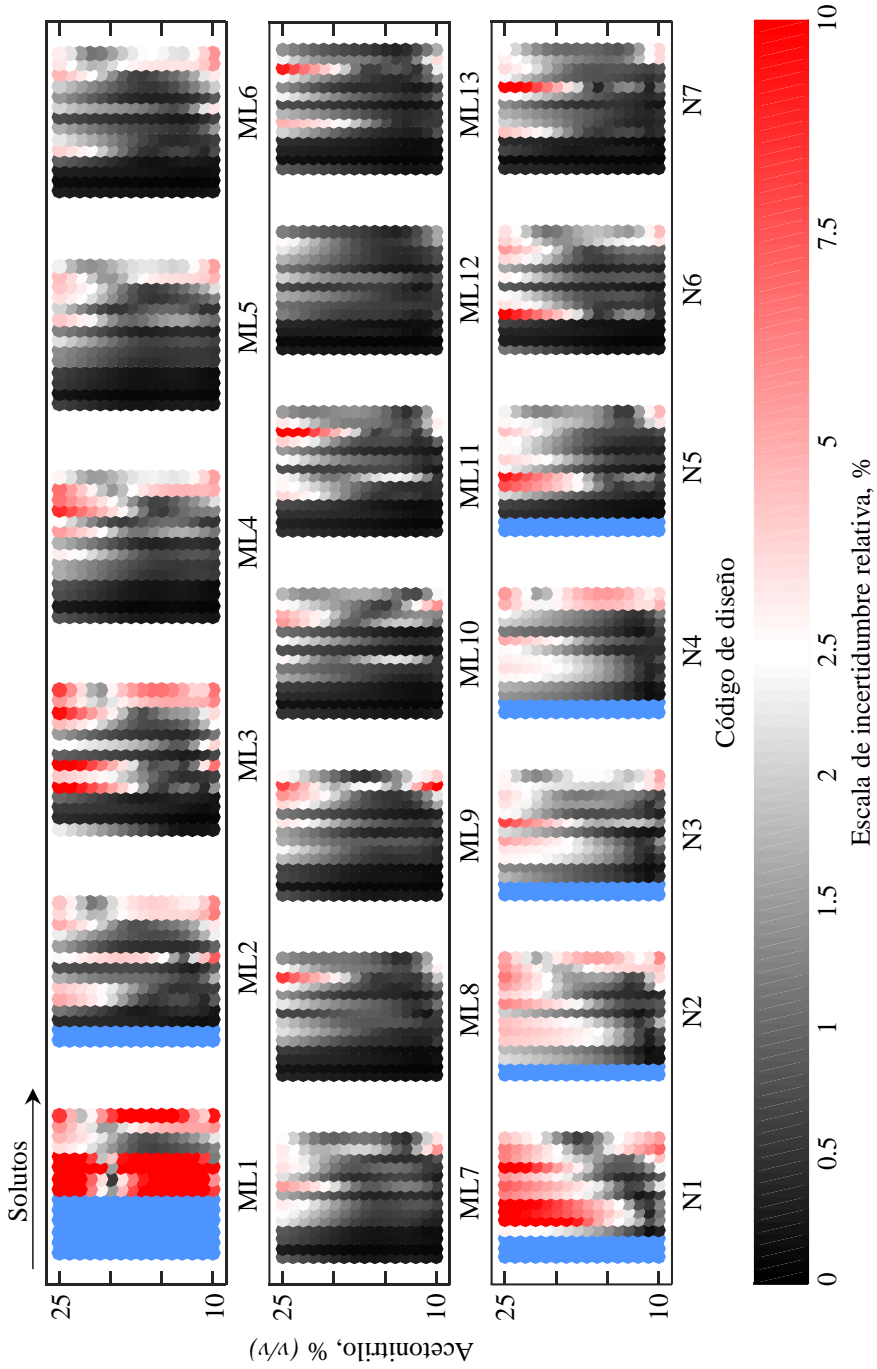


Figura 4.4 (continuación).

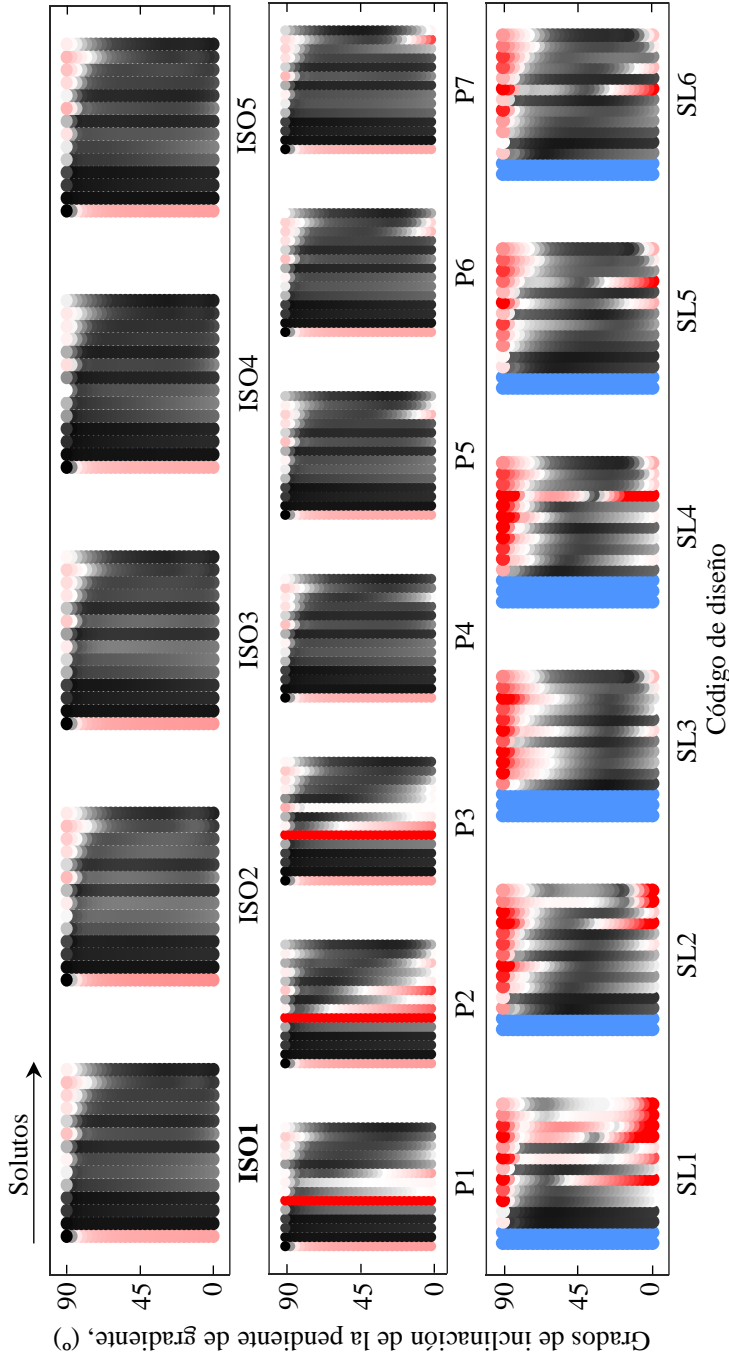


Figura 4.5. Diagramas triples mostrando las incertidumbres relativas esperadas en la predicción de condiciones de gradiente. Los resultados se muestran para cada pendiente de gradiente predicha y soluto individual, según el diseño de gradiente de muestreo. La escala de color de incertidumbre relativa se ha adaptado a la elución en gradiente, donde las incertidumbres son menores. Se pueden encontrar más detalles en la Sección 4.4.1.2.

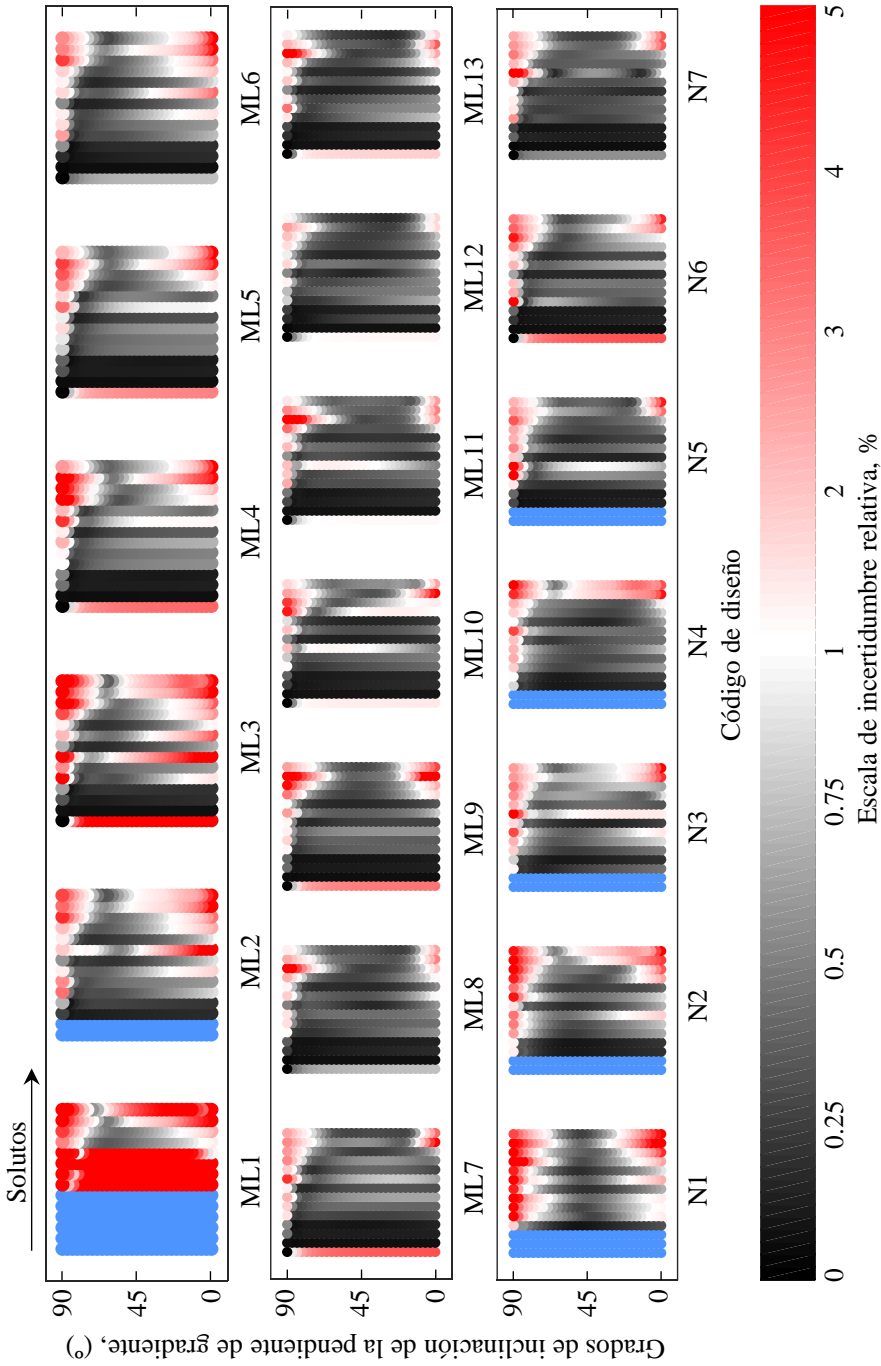


Figura 4.5 (continuación).

4.4.2.1. Diseños isocráticos: ISO1–ISO5

Este primer grupo incluye cinco diseños de entrenamiento, conteniendo cada uno de cuatro a ocho experimentos isocráticos distribuidos en el intervalo del 10 al 25 % de acetonitrilo. De acuerdo con lo esperado, el Diseño ISO4, con el mayor número de experimentos isocráticos, ofrece la mejor capacidad predictiva (véase la primera fila en las Figuras 4.2 y 4.3). Tampoco sorprende que el Diseño ISO2, con el menor número de fases isocráticas, produzca las mayores incertidumbres relativas (este diseño garantiza un grado de libertad para el modelo NK). Por otro lado, la capacidad predictiva de los Diseños ISO1, ISO3 e ISO5 (todos con cinco fases isocráticas) es similar para las predicciones isocráticas y de gradiente. Concentrar el esfuerzo experimental a bajas fuerzas eluyentes (ISO1 e ISO5), en lugar de distribuir los experimentos regularmente (ISO3), parece beneficioso para reducir las incertidumbres, ya que esta distribución de experimentos permite un mejor muestreo del dominio de modificador con mayor influencia en la retención (fuerzas eluyentes débiles).

A pesar de las aparentes diferencias, la capacidad predictiva de los cinco diseños isocráticos es bastante similar. Así, por ejemplo, para el dominio bien descrito (75 % de las condiciones que proporcionan las mejores incertidumbres), el intervalo intercuartílico en los diagramas de caja y bigotes para ISO2 (el peor diseño) e ISO4 (el mejor) es: 0.15–0.40 % y 0.25–0.65 %, respectivamente, para las predicciones isocráticas (Figura 4.2), y 0.12–0.25 % y 0.15–0.37 %, respectivamente, para las predicciones de gradiente (Figura 4.3). Análogamente, los diagramas triples (Figuras 4.4 y 4.5) indican sólo pequeñas diferencias entre todos los diseños isocráticos, en ambos modos de elución.

Un criterio secundario para calificar los diseños es el tiempo de adquisición, que puede estimarse como la suma de los tiempos de retención para los solutos

más retenidos utilizando cada experimento de un diseño (véase la Tabla 4.1). El tiempo de adquisición para el Diseño ISO1 es una hora menor en comparación con el ISO5. El Diseño ISO4, que ofrece las incertidumbres más pequeñas, requiere tres experimentos más y, en consecuencia, un mayor gasto de reactivos y tiempo de adquisición. Por lo tanto, y teniendo en cuenta lo anterior, ISO1 puede considerarse como el mejor entre los diseños isocráticos.

4.4.2.2. Diseños isocráticos con pulsos: P1–P7

Los diseños con pulsos (véase el Capítulo 2) se construyeron mediante la inserción de incrementos repentinos transitorios en el contenido de modificador, en la o las fases móviles de baja fuerza eluyente en el Diseño ISO1 (Figura 4.1). Como se ha visto en el Capítulo 2, la adición de pulsos aumenta temporalmente la fuerza eluyente, lo que reduce el tiempo de adquisición con respecto al diseño isocrático, pero manteniendo una capacidad de predicción similar. Este grupo de diseños se ha incluido para estudiar el efecto sobre la incertidumbre en la predicción de las variables que definen un pulso (ubicación, duración y contenido máximo de modificador que se alcanza). La ubicación de los Diseños P1 y P2 es la misma, pero su duración es distinta. El Diseño P3 mantiene la ubicación y duración del pulso respecto a P2, pero aumenta la concentración de modificador del pulso del 10 al 20 %, en lugar de utilizar un 25 % de acetonitrilo. Estos tres diseños tienen la característica común de la ubicación del pulso al comienzo del cromatograma, correspondiendo al Diseño P2 las peores incertidumbres tanto en elución isocrática como de gradiente (véase la columna derecha “25 % de las peores incertidumbres” en las Figuras 4.2 y 4.3).

Tabla 4.1. Tiempos de adquisición para los 38 diseños de entrenamiento ensayados, calculados como la suma del tiempo de retención del soluto más retenido, para todos los experimentos incluidos en los diseños.

Diseño	Tiempo (min)	Diseño	Tiempo (min)	Diseño	Tiempo (min)
Isocráticos		Lineales simples		Lineales mixtos II	
ISO1	365.2	SL1	167.6	ML7	258.8
ISO2	276.3	SL2	255.4	ML8	237.8
ISO3	329.6	SL3	356.6	ML9	223.3
ISO4	501.5	SL4	291.8	ML10	192.2
ISO5	415.8	SL5	315.7	ML11	196.6
Isocráticos + Pulsos		SL6	295.5	ML12	224.1
P1	337.5	Lineales mixtos I		ML13	241.1
P2	295.9	ML1	129.3	Con nodos	
P3	333.3	ML2	118.7	N1	211.2
P4	337.5	ML3	125.6	N2	210.9
P5	295.9	ML4	175.7	N3	208.5
P6	244.1	ML5	197.6	N4	221.4
P7	212.5	ML6	180.0	N5	263.5
				N6	203.1
				N7	227.2

Los diagramas triples de incertidumbre (Figuras 4.4 y 4.5) indican que los solutos que eluyen cerca del pulso son los responsables de las malas predicciones en P1–P3 (recuérdese que en la Figura 4.1 puede consultar el tiempo de elución de cada soluto para cada experimento y diseño de entrenamiento). La incertidumbre es máxima (por encima del 10 y 5 % para la elución isocrática y de gradiente, respectivamente) para el soluto 6, que eluye al inicio del pulso (véase la Figura. 4.1). Este aumento súbito de la incertidumbre es consecuencia del cálculo de las derivadas parciales para la construcción de las matrices jacobianas (véase la Sección 3.3.2.1 en el Capítulo 3). Cuando un soluto eluye demasiado cerca de un pulso y se modifican ligeramente los parámetros de su modelo para calcular las derivadas, la retención sufre un fuerte cambio debido al aumento repentino del contenido de modificador. Esto se traduce en una variación extrema en los valores de las derivadas y, en consecuencia, en un aumento de la incertidumbre. Para los solutos que eluyen a lo largo del pulso (solutos 7–9, véase su ubicación en la Figura 4.1), los diagramas triples también muestran un aumento en las incertidumbres, especialmente en la región de menor fuerza eluyente. Esto es coherente, ya que estos solutos eluyen principalmente a un 25 % de acetonitrilo, siendo bastante pequeña la contribución de la fase móvil del 10 %.

A continuación, se consideran las consecuencias de ubicar el pulso al final de los cromatogramas, previamente a la elución de los solutos más retenidos (P4–P6), que pueden verse afectados por el pulso. Para evaluar los efectos, debe tenerse en cuenta que los solutos retenidos difícilmente eluirán cerca de un pulso de corta duración, debido al aumento exponencial en la retención para estos solutos, que tienden a crear grandes regiones vacías al final de los cromatogramas isocráticos. Por lo tanto, las incertidumbres para P4–P6 serán similares a las de ISO1. El pequeño aumento en las incertidumbres de

predicción, en la región de menor fuerza eluyente, se puede atribuir a los solutos que abandonan la columna a lo largo del pulso (solute 12 en P4 y P5, y solutos 13 y 14 en P6, véase la Figura 4.1). Las predicciones en el modo gradiente exhiben patrones de incertidumbre similares, que son más acusadas para los gradientes con menores pendientes. El Diseño P4 es el mejor, como puede esperarse para el pulso de menor duración. Sin embargo, el analista podría preferir el P6, debido a la reducción considerable en el tiempo de adquisición, en comparación con ISO1 (30, 70 y 120 min para los Diseños P4, P5 y P6, respectivamente, consúltese la Tabla 4.1).

Finalmente, el Diseño P7 es una variante del P6, en el que se introduce un segundo pulso en la fase móvil isocrática del 13 % de acetonitrilo. Con este diseño, las incertidumbres de predicción se deterioran ligeramente respecto a diseños anteriores en ambos modos de elución, pero el tiempo de adquisición se reduce en 2.5 horas respecto al Diseño ISO1.

4.4.2.3. Diseños de gradiente con una única rampa y la misma composición inicial: SL1–SL6

A partir de este punto, los diseños que se analizan estarán formados exclusivamente por gradientes. La característica común del grupo de diseños *SL1–SL6* es que todos los gradientes comparten la misma composición inicial (10 % de acetonitrilo, véase la Figura. 4.1). En el Diseño SL1, la composición final también es la misma para todos los gradientes (25 %), pero el tiempo al cual se alcanza dicha composición es distinto. En SL2, las pendientes de los gradientes disminuyen progresivamente, alcanzando sólo tres de los gradientes el 25 % de acetonitrilo; más allá del t_G la elución continúa isocráticamente hasta los 60 min. En SL3, las pendientes son todavía menores, siendo todas las composiciones finales diferentes, pero alcanzándose éstas a un mismo tiempo

de 60 min. El Diseño SL4 es una variante de SL3, donde el tiempo de gradiente (t_G) se ha reducido a 30 min, continuando la elución isocráticamente para tiempos más allá del valor de t_G .

Los Diseños SL1–SL4 dan lugar a las incertidumbres de predicción más deficientes del grupo, para ambos modos de elución (Figuras 4.2 y 4.3), siendo apreciablemente peores respecto a los diseños de entrenamiento isocráticos (donde debe tenerse en cuenta que la escala de incertidumbre es diferente, véase la Sección 4.4.2.1). Para las predicciones isocráticas, los diagramas triples de incertidumbre muestran regiones más estrechas con una buena descripción. El Diseño SL1 sólo muestrea satisfactoriamente solutos rápidos eluidos a baja fuerza eluyente. Las incertidumbres de predicción aumentan significativamente a fuerzas eluyentes elevadas, debido a que los solutos rápidos abandonan la columna antes de que los gradientes de entrenamiento muestreen dicha región. Por lo tanto, no existe suficiente información, teniéndose que extrapolar los tiempos de retención para estos solutos. Por su parte, para los solutos lentos no se dispone de información a baja fuerza eluyente, debido a la geometría del diseño, que hace que las predicciones se obtengan por extrapolación. Se puede observar que SL2, siendo aún un diseño pobre, cubre mejor que SL1 la elución de los solutos lentos con las fases móviles lentas. El diseño SL3 implica pendientes de gradiente más suaves, lo que mejora las predicciones a baja fuerza eluyente, pero las empeora para fuerzas eluyentes elevadas. Finalmente, SL4 mejora ligeramente las predicciones para los solutos más lentos, debido a los segmentos isocráticos finales.

Los Diseños SL5 y SL6 se construyeron modificando las pendientes de los gradientes de SL4, pero manteniendo las mismas composiciones finales de gradiente. Como se observa en los diagramas de caja y bigotes y en los

diagramas triples (Figuras 4.2–4.5), estos diseños exhiben la mejor capacidad de predicción del grupo. Para predicciones isocráticas, SL6 ofrece incertidumbres algo más bajas en comparación con SL5, en el dominio de composiciones bien descrito, mientras que, en el dominio mal descrito, las incertidumbres son mayores (columnas izquierda y derecha, respectivamente, en la Figura 4.2). El diseño SL6 describe ligeramente mejor la retención de los solutos rápidos, eluidos con fases móviles rápidas, debido a que la pendiente para los gradientes de entrenamiento intermedios es mayor respecto a SL5 (véase la Figura 4.1). Para las predicciones en gradiente, tanto SL5 como SL6 ofrecen una incertidumbre de predicción superior al 3 % en la región de alta fuerza eluyente. Finalmente, el tiempo de adquisición para SL6 es 20 min más corto, en comparación con SL5 (Tabla 4.1).

4.4.2.4. Diseños de gradiente con una única rampa y diferente composición inicial: ML1–ML6

Como se ha visto para el grupo de diseños anterior, la inclusión de experimentos que compartan la misma concentración inicial (en este caso, 10 % de acetonitrilo) no conduce a una capacidad predictiva suficientemente satisfactoria, para diseños de gradiente que contienen una única rampa. Esto es debido principalmente a que:

- (i) Para los solutos más rápidos, el número de medidas de calidad utilizadas para ajustar el modelo de retención es insuficiente (el número de grados de libertad se encuentra por debajo de las necesidades de ajuste del modelo NK).
- (ii) Realizar predicciones con baja incertidumbre no es posible para los solutos más rápidos, eluidos a fuerzas eluyentes elevadas.

El problema con los solutos más rápidos puede resolverse utilizando gradientes con un contenido inicial de modificador a distintos niveles. De acuerdo a ello, se construyeron nuevos diseños (ML1–ML6), en los que se adoptaron dos o tres niveles iniciales diferentes, mientras que la concentración final del modificador era la misma para todos los gradientes. Tanto los diagramas de caja y bigotes, como los diagramas triples, muestran que el Diseño ML1 da lugar, con diferencia, a las incertidumbres más elevadas del grupo. Este diseño consta de cuatro gradientes que se inician a dos niveles. Esto proporciona sólo dos medidas válidas para los solutos más rápidos (solutos 1–6, marcados en rojo en la Figura 4.1), lo que es insuficiente para ajustar el modelo NK, que contiene tres parámetros. Además, ML1 sólo proporciona tres medidas válidas para los solutos de retención intermedia, haciendo posible el ajuste, aunque con cero grados de libertad.

El Diseño ML2 es análogo al ML1, pero carece de segmento isocrático inicial, dando comienzo las rampas de gradiente justo después del tiempo de permanencia. El Diseño ML3 se basa en el Diseño ML2, pero el inicio de uno de los gradientes superiores se traslada desde el 20 al 15 % de acetonitrilo, dando lugar a tres niveles de gradiente iniciales. Sin embargo, ML2 ofrece una capacidad predictiva más satisfactoria, en comparación con ML3, ya que su geometría lo hace más informativo en la región de alta fuerza eluyente. Por su parte, ML3 funciona mejor para los solutos rápidos, ya que la existencia de niveles iniciales a tres valores distintos aumenta la información para estos solutos, que abandonan rápidamente la columna.

El Diseño ML4 incluye cinco gradientes (a diferencia de ML1–ML3, que contienen sólo cuatro) y tres niveles iniciales con segmentos isocráticos. En general, ML4, con un experimento más, mejora la calidad de las predicciones, haciéndolas más informativas. Este diseño también mejora las predicciones

isocráticas para los solutos rápidos, cuando se utilizan fases móviles rápidas, debido a la inclusión de los segmentos isocráticos. Finalmente, los Diseños ML5 y ML6 son los mejores de este grupo, ya que constan de seis gradientes que se inician a tres niveles distintos, presentando ML5 segmentos isocráticos iniciales de 10 min antes del inicio de las rampas. Los diagramas de caja y bigotes son parecidos para ambos diseños y ambos modos de elución. Se extraen conclusiones similares de los diagramas triples, siendo la capacidad de predicción de ML5 mejor para los solutos más rápidos, eluidos a fuerzas eluyentes altas. La ausencia de sectores isocráticos iniciales en ML6 hace que los gradientes exhiban pendientes más suaves. Por razones geométricas, este efecto tiene mayores consecuencias en los gradientes que parten del 20 % de acetonitrilo (para los que la diferencia de pendiente es mayor). Por lo tanto, las rampas varían la composición más lentamente, haciendo que los solutos con retención intermedia experimenten cada composición a lo largo del gradiente durante más tiempo. Los tiempos de adquisición también muestran una pequeña diferencia para ML5 y ML6 (alrededor de 20 min más cortos para ML6).

4.4.2.5. Diseños de gradiente con una única rampa en la que se modifican ambas composiciones inicial y final: ML7–ML13

El análisis de los diseños constituidos por gradientes que comparten el contenido de modificador inicial (SL1–SL6), o final (ML1–ML6), indica que los primeros no predicen bien la elución de los solutos con fases móviles y gradientes de elevada fuerza eluyente, mientras que los segundos originan incertidumbres algo más elevadas (pero aún aceptables) para los solutos lentos, especialmente a baja fuerza eluyente. Por lo tanto, en ambos casos, la geometría del diseño no proporciona suficiente información y algunas predicciones requieren extrapolaciones. Debido a esto, los Diseños ML7–ML13, en los que

se varían las composiciones inicial y final de los gradientes, se incorporaron al estudio para investigar si era posible mejorar la calidad de las predicciones.

Ambos Diseños ML7 y ML8 poseen tres niveles iniciales y finales, diferenciándose en que ML7 incluye segmentos isocráticos al comienzo de los gradientes (con el objetivo de introducir, en cierta medida, las ventajas de los experimentos isocráticos en los diseños contruidos con gradientes). Cuando se predicen gradientes y fases móviles isocráticas rápidas con ML7, las incertidumbres para los solutos 5–12 son ligeramente mayores (especialmente para los solutos intermedios, véanse las Figuras 4.4 y 4.5). Esto puede explicarse nuevamente por el hecho de que estos solutos abandonan la columna a lo largo de la región isocrática inicial de 15 min, lo que implica que las concentraciones superiores al 20 % no participan en la migración de los solutos. El Diseño ML8 carece del segmento isocrático inicial y las pendientes del gradiente son más pequeñas. Esto hace que la incertidumbre máxima se desplace hacia los solutos más retenidos (solutos 11–13), que experimentan concentraciones de modificador mayores antes de abandonar la columna. El Diseño ML9, que contiene cuatro niveles iniciales y dos finales, no predice bien la retención de los solutos lentos eluidos con fases móviles isocráticas lentas o gradientes suaves, debido a la falta de información en la región de fuerza eluyente muy baja. Esto sugiere la conveniencia de incluir gradientes con pendientes más suaves.

Los Diseños ML10 y ML11 incluyen cuatro gradientes con cuatro niveles iniciales y tres finales. Se observa que la capacidad predictiva de estos dos diseños ofrece un rendimiento similar al de ML7–ML9. Los resultados también deben compararse con los obtenidos con ML1, que contiene cuatro experimentos. Las regiones peor descritas para ML10 y ML11 corresponden a

la elución de los solutos lentos eluidos con fases móviles isocráticas o gradientes rápidos.

Los diseños que exhiben la mejor capacidad de predicción del grupo son el ML12 (especialmente) y el ML13, ambos con cinco gradientes lineales, y distintas concentraciones iniciales y finales. Nótese que el Diseño ML12 incluye una fase móvil isocrática al 25 % de acetonitrilo y la separación entre los niveles iniciales es mayor respecto al Diseño ML13. Por lo tanto, con ML12 se cubre un dominio más extenso. Una mayor separación entre niveles implica pendientes más suaves, por lo que cada soluto experimenta un intervalo variado de composiciones para el ajuste, considerando todos los gradientes en conjunto. No sorprende que ML12 y ML13 sean los mejores diseños de entrenamiento formados por gradientes, ya que son los más similares a los diseños isocráticos.

4.4.2.6. Diseños que contienen gradientes multi-lineales: N1–N7

Los gradientes en este grupo de diseños incluyen, al menos, un nodo que conecta dos segmentos con distinta pendiente. Los primeros cuatro diseños (N1–N4) se planificaron para conseguir la separación de solutos en un intervalo de polaridad extremadamente amplio. Los Diseños N1–N3 constan de cinco gradientes, cada uno con un solo nodo y composiciones iniciales y finales comunes. Estos diseños tienen un perfil romboédrico, progresivamente ensanchado debido a la mayor diferencia de pendientes entre los segmentos lineales consecutivos. El Diseño N1 cubre un intervalo de modificador relativamente pequeño, que puede ser insuficiente para los solutos rápidos en la región de fuerza eluyente alta y para los solutos lentos para fuerzas eluyentes bajas. Por lo tanto, se esperan incertidumbres elevadas en ambos casos. Estos dos inconvenientes se corrigen para N2 y N3, ya que el “rombo” cubre un dominio de modificador progresivamente mayor. En general, los Diseños N1–

N3 pueden ser útiles en los casos en los que las predicciones se limiten a los gradientes incluidos en el área romboédrica.

El diseño N4 es equivalente a N3, pero en lugar de un único nodo, los gradientes de menor y mayor fuerza eluyente incluyen tres nodos que le confieren un perfil casi curvilíneo. Este diseño también incluye un gradiente lineal central. Los diagramas triples muestran que N4 mejora las predicciones respecto a N3. Las predicciones realizadas con N1–N4 para los solutos lentos, analizados con eluyentes rápidos, son más pobres, ya que estos solutos abandonan la columna antes de alcanzar estas composiciones. Con gradientes lentos, las predicciones también se ven comprometidas, debido a que las concentraciones más bajas no se hallan adecuadamente muestreadas.

Los diagramas triples para el Diseño N1 (última fila en las Figuras 4.4 y 4.5), indican que el número de condiciones bien descritas en la predicción de gradientes es claramente mayor, en comparación con las predicciones isocráticas. Este efecto también se observa para N2–N4 y puede explicarse por la extensión del intervalo de composiciones que cubre el diseño. Debe aquí comentarse que en el Capítulo 5 se hace uso, con excelentes resultados, de un diseño similar a N3 para obtener un modelo global que permite describir la retención de un gran número de componentes en extractos de hierbas medicinales [20].

Los últimos diseños del grupo (N5–N7) intentan utilizar niveles lo más diversos posibles, para cada tiempo de medida, cubriendo en conjunto un intervalo de composiciones muy amplio. El diseño N5 no ofrece tiempos de retención significativamente diferentes para los solutos 1 y 2 (véase la Figura 4.1). Sin embargo, en comparación con diseños anteriores (N1–N4), las predicciones para la mayoría de los solutos eluidos con fases móviles lentas o gradientes suaves mejoran, lo que es debido a la existencia de dos niveles

iniciales. Estas mejoras son más importantes en N6 y N7 con tres y cuatro niveles iniciales, respectivamente. Por otro lado, la capacidad de predicción mejora para los solutos lentos, debido al mayor intervalo de composiciones cubierto en cualquier instante, a lo largo del gradiente, siendo los resultados más pobres los que ofrece el Diseño N6. Finalmente, el mejor diseño del grupo es el N7, para ambos modos de elución, con incertidumbres que superan ocasionalmente el 5 %, especialmente en el modo isocrático.

4.4.3. Comparación de los mejores diseños encontrados en cada grupo

Tras el estudio detallado de los diferentes grupos de diseños de entrenamiento realizado en la Sección 4.4.2, se procedió a comparar los mejores diseños de cada grupo (ISO4, P4, SL6, ML6, ML12 y N7), tanto en condiciones isocráticas como de gradiente. Con el fin de disminuir el volumen de datos y obtener gráficos más prácticos, se seleccionaron cuatro sulfonamidas representativas de diferente polaridad, que en orden creciente de retención son (véase la Sección 4.3.1): sulfadiazina (solute 3, escasamente retenido), sulfamerazina (solute 6, con una retención relativamente baja), sulfametoxazol (solute 11, con retención intermedia) y sulfaquinoxalina (solute 14, altamente retenido). Los diagramas de caja y bigotes, junto a los diagramas triples de incertidumbre para estos compuestos, tanto para las predicciones isocráticas como de gradiente, se muestran en las Figuras 4.6 y 4.7, respectivamente.

Los resultados para ambos modos de elución revelan conclusiones generales comunes. Además, y en términos comparativos, la Figura 4.8 muestra la calidad predictiva de los diseños, a través de la representación de las curvas o mapas de incertidumbre para cada soluto (utilizadas en el Capítulo 3), tanto para realizar predicciones isocráticas como de gradiente.

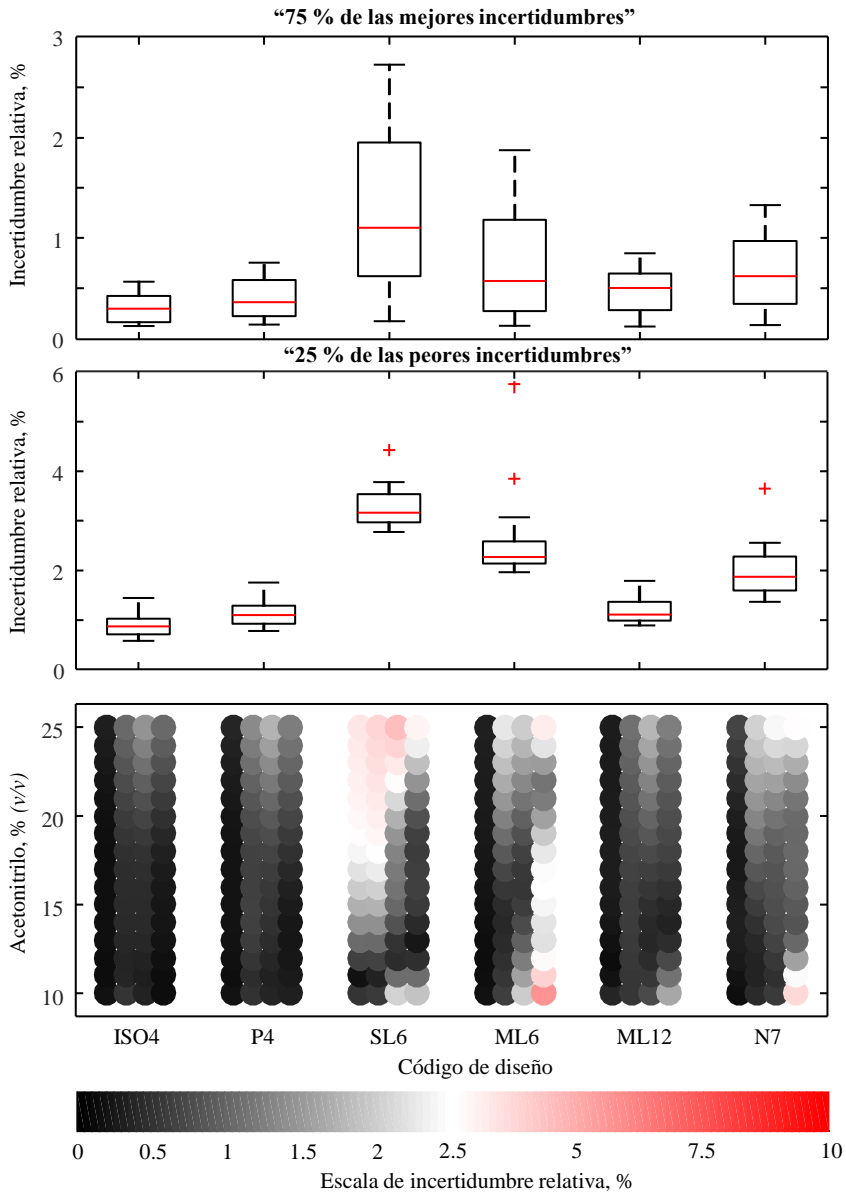


Figura 4.6. Diagramas de caja y bigotes y triples para los mejores diseños encontrados en cada uno de los grupos de diseños de la Figura 4.1. Se representa la incertidumbre relativa esperada para las predicciones en condiciones isocráticas, para los cuatro solutos seleccionados. Para más detalles, véase la Sección 4.4.3.

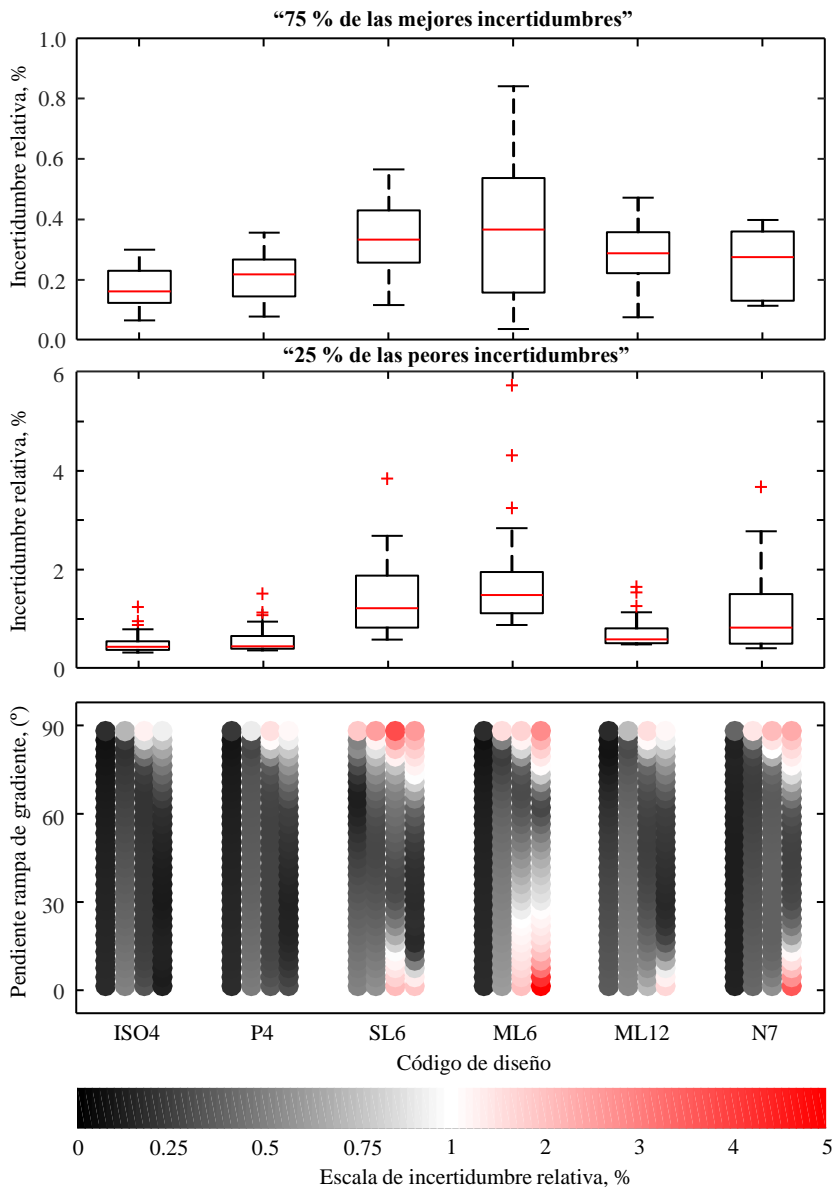


Figura 4.7. Diagramas de caja y bigotes y triples para los mejores diseños encontrados en cada uno de los grupos de diseños de la Figura 4.1. Se representa la incertidumbre relativa esperada para las predicciones en condiciones de gradiente para los cuatro solutos seleccionados. Para más detalles, véase la Sección 4.4.3.

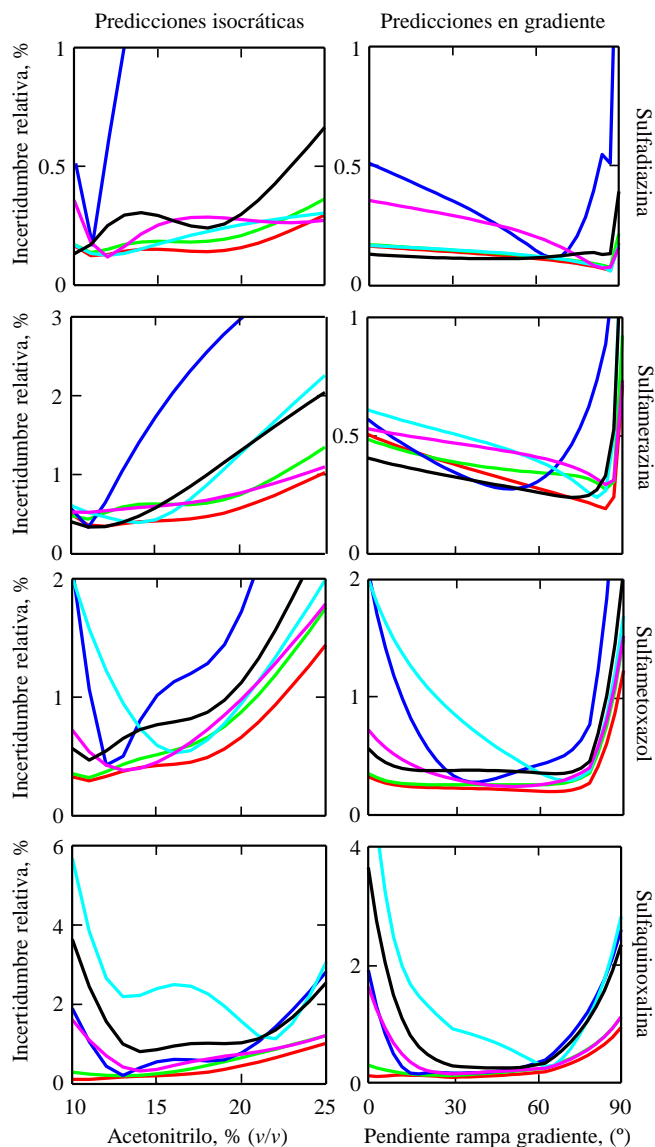


Figura 4.8. Mapas que ilustran, para cuatro sulfonamidas representativas, la incertidumbre relativa en la predicción de las condiciones isocráticas (columna izquierda) y de gradiente (columna derecha), en función de la composición de la fase móvil, y el incremento angular en la pendiente del gradiente, respectivamente (véase la Figura 3.3 en el Capítulo 3). Diseños: ISO4 (línea roja), P4 (verde), SL6 (azul oscuro), ML6 (azul claro), ML2 (magenta) y N7 (negro).

Los resultados recogidos en las Figuras 4.6 y 4.8 indican que los Diseños SL6 y ML6 ofrecen la capacidad de predicción más deficiente (al comparar los mejores diseños de cada grupo), especialmente el SL6 para predicciones isocráticas. Para este diseño (compuesto por gradientes con pendientes variables y diferentes niveles de concentración final), la incertidumbre en las predicciones se deteriora para los solutos rápidos en condiciones de elevada fuerza eluyente, ya que estos solutos ya han abandonado la columna antes de que se alcancen las composiciones finales del gradiente. Como se observa, el número de condiciones bien descritas a fuerza eluyente elevada aumenta al incrementarse la hidrofobicidad de los solutos. Para el diseño ML6 (que incluye gradientes con pendientes entre moderadas y altas, partiendo de diferentes niveles de concentración inicial), las incertidumbres máximas se obtienen en la predicción de los solutos lentos para condiciones de baja fuerza eluyente. Cuando se ejecuta este diseño, los solutos lentos experimentan cambios en la retención, principalmente a las composiciones más elevadas de los gradientes de entrenamiento, mientras que las composiciones más bajas apenas participan (véanse los tiempos de retención en la Figura 4.1). En las predicciones realizadas para la elución en gradiente, este efecto se intensifica para los gradientes de menor pendiente, haciendo que los resultados obtenidos con el Diseño ML6 sean, generalmente, más pobres en comparación con SL6, ya que la región de baja fuerza eluyente no es muestreada por ML6.

Los Diseños ISO4 y P4 son los más informativos y producen las mejores predicciones, para todos los solutos, en todo el dominio considerado. El Diseño ML12, constituido por rampas de gradiente de pequeña pendiente, también ofrece predicciones con baja incertidumbre, aunque a medida que disminuye la polaridad del soluto, la incertidumbre se incrementa ligeramente a baja fuerza eluyente. Este comportamiento puede explicarse por el hecho de que, utilizando

el gradiente de menor fuerza eluyente en ML12, los solutos lentos abandonan la columna sólo al final del gradiente o una vez éste ha finalizado (Figura 4.1). Debido a su geometría, la capacidad predictiva de N7 es algo peor respecto a la de ML12, ya que aumentan las incertidumbres en las predicciones para los solutos lentos a baja fuerza eluyente y para todos los solutos a alta fuerza eluyente (ver Figura 4.1).

4.5. Conclusiones

Este capítulo complementa el estudio realizado en el Capítulo 3, donde se presenta el desarrollo de una metodología basada en la teoría de propagación de errores, utilizando el Principio de Optimalidad G. Su implementación práctica necesitó la resolución de algunos problemas, como lograr el cálculo de las matrices jacobianas con buena precisión, en tiempos suficientemente bajos. El propósito del Capítulo 4 ha sido comparar, utilizando dicha metodología, la capacidad de predicción de una amplia variedad de diseños relacionados geoméricamente y revelar el efecto de diferentes factores. Idealmente, el diseño de entrenamiento debería ser común para un grupo de solutos de diferente naturaleza y contener un número razonablemente pequeño de fases móviles o gradientes.

En este capítulo, se han estudiado 38 diseños de diferentes tipos, construidos para conseguir una retención adecuada y descubrir las características que debe cumplir un diseño para obtener predicciones con incertidumbres bajas. Se da respuesta a algunas preguntas que generalmente se resuelven de forma intuitiva, como cuál es la geometría más correcta para un diseño, así como el número de experiencias necesarias. El análisis de los resultados sugiere principios generales y recomendaciones para construir diseños de entrenamiento con una buena capacidad predictiva, tanto en elución isocrática como en gradiente.

Para visualizar los resultados, se han utilizado diagramas de caja y bigotes, así como diagramas triples que incorporan códigos de colores, permitiendo así una evaluación simultánea de la capacidad predictiva de una gran cantidad de diseños, atendiendo a las incertidumbres relativas esperadas. Los diagramas de caja y bigotes brindan una perspectiva rápida de la capacidad de predicción de los diseños, mientras que los gráficos triples resumen los resultados para cada soluto y fase móvil, o programa de gradiente. La combinación de ambos tipos de gráficos permite realizar múltiples comparaciones para un gran número de diseños y un análisis sencillo de los resultados, aunque sacrificando algunos detalles con respecto a los gráficos de incertidumbre utilizados en el Capítulo 3.

Los diseños isocráticos ofrecen la mejor capacidad predictiva. Un pequeño, pero suficiente número de experiencias (que proporcionen al menos un grado de libertad) es satisfactorio, siempre que el dominio de composiciones del eluyente esté bien cubierto por el diseño. El número y la distribución de los experimentos isocráticos no resultó crítico. Sin embargo, los diseños puramente isocráticos poseen el inconveniente de originar largos tiempos de análisis, que pueden reducirse si se implementa un pulso de modificador orgánico (acetonitrilo) en la fase móvil de menor fuerza eluyente del diseño isocrático. En el Capítulo 2, se ha visto que este tipo de diseño produce buenas predicciones para los solutos lentos, cuando se eluyen con fases móviles con bajos contenidos de modificador. Los largos tiempos de adquisición se acortan sin que se deteriore la capacidad predictiva, siempre y cuando la ubicación del pulso sea adecuada.

Los diseños de gradiente ofrecen la ventaja de muestrear la región de menor fuerza eluyente, sin producir largos tiempos de análisis. Sin embargo, cuando la retención se modeliza utilizando diseños de entrenamiento compuestos exclusivamente por gradientes, cada gradiente sólo examina intervalos

estrechos del dominio de modificador. Por lo tanto, obtener buenas predicciones para condiciones isocráticas, utilizando diseños de gradiente, sólo es factible si la elución del soluto se encuentra bien muestreada en la región del modificador inspeccionada por el conjunto de gradientes, lo que no suele ocurrir. Por el contrario, en los diseños de entrenamiento que consisten únicamente de fases isocráticas, todas las fases móviles muestreadas participan completamente (al 100 %) en la migración de los solutos. En cuanto a las predicciones realizadas, se ha encontrado que los gradientes se predicen con incertidumbres intrínsecamente más bajas, independientemente del diseño utilizado para modelizar la retención, siendo más insensibles al tipo de diseño de entrenamiento que las predicciones isocráticas.

Se ha encontrado que diseños de gradiente ampliamente utilizados, como SL1 y SL3 (Figura. 4.1), no proporcionan predicciones con baja incertidumbre en ciertas regiones del dominio. Esto se debe a que estos gradientes, que presentan los mismos niveles de concentración de modificador inicial o final, tienen una capacidad de muestreo intrínsecamente peor. Por el contrario, cuando se implementan tres o más niveles de modificador al principio y al final de los gradientes, la capacidad predictiva mejora. Éste es el caso del Diseño ML12, que ofrece incertidumbres comparables a las que brindan los diseños isocráticos, con un tiempo de adquisición de los datos sustancialmente más corto (Tabla 4.1). Por lo tanto, y en virtud de los resultados, puede decirse que es altamente recomendable separar los niveles de concentración inicial y final de los gradientes.

La posibilidad de muestrear, considerando todos los gradientes de un diseño, composiciones tan diversas como sea posible y a cualquier tiempo de gradiente, también es compartida por el Diseño multi-lineal N7, que es el siguiente en términos de capacidad predictiva (tras el Diseño ML12), ofreciendo un tiempo

de adquisición similar. Este diseño requeriría un nivel de concentración adicional en la región de baja fuerza eluyente para mejorar su capacidad predictiva. En el estudio realizado, se ha comprobado que el uso de gradientes multi-lineales mejora la calidad de las predicciones, respecto a los diseños de gradientes lineales que poseen una única rampa. Sin embargo, los diseños de gradientes multi-lineales son más propensos a depender de los requisitos de información de cada soluto.

Además de las incertidumbres relativas, hay otras consideraciones a tener en cuenta, como el intervalo de hidrofobicidad de los solutos a modelizar. En el caso de que dicho intervalo fuese demasiado amplio, otros diseños con limitaciones predictivas como los diseños en forma de rombo (N1 a N3), pueden constituir una buena opción.

A continuación, se proporcionan algunas recomendaciones generales para construir diseños experimentales de gradientes, con una buena capacidad predictiva:

- (i) El intervalo de composición muestreado debe ser amplio y las concentraciones deberían ser tan diversas como sea posible, a cualquier tiempo de gradiente (considerando todos los gradientes del diseño).
- (ii) Cada composición a lo largo del gradiente debe participar significativamente en la retención de los solutos eluidos. Esto implica el uso de intervalos moderados de modificador en cada gradiente del diseño, es decir, el uso de gradientes con pendientes suaves.
- (iii) El diseño no debe dejar concentraciones de modificador sin muestrear: la exactitud se verá comprometida si un intervalo de modificador carece de información, ya que las predicciones requerirán extrapolaciones.
- (iv) Para solutos rápidos, deben evitarse especialmente las situaciones en las que el soluto ha abandonado la columna cuando los gradientes de

entrenamiento alcanzan las composiciones de modificador más altas, o en los que algunas concentraciones no participan en la retención del soluto (por ejemplo, 25 % en el Diseño ML1 para los solutos más rápidos, que se encuentran bien muestreados en ML12).

- (v) Los segmentos isocráticos situados con anterioridad al inicio del gradiente, o los gradientes con múltiples niveles iniciales, constituyen alternativas a los diseños isocráticos. En ambos casos, es conveniente separar los experimentos para garantizar un buen muestreo (véase, por ejemplo, el Diseño ML12).
- (vi) El analista debe tener en cuenta el tipo de predicciones a realizar, antes de seleccionar el diseño. Algunos diseños que dan lugar a predicciones isocráticas muy pobres, pueden generar predicciones de gradiente aceptables. Sin embargo, el mejor diseño es el que no compromete el tipo de predicción.

La metodología y pautas de construcción de diseños de gradiente propuestas se pueden utilizar en cualquier aplicación que requiera la modelización de la retención, como es la optimización de la selectividad, el desarrollo de métodos o la caracterización de las propiedades de los solutos.

4.6. Referencias

- [1] M.C. García Álvarez-Coque, J.R. Torres Lapasió, J.J. Baeza Baeza, Models and objective functions for the optimisation of selectivity in reversed-phase liquid chromatography, *Anal. Chim. Acta* 579 (2006) 125–145.
- [2] J.R. Torres Lapasió, M.C. García Álvarez-Coque, Levels in the interpretive optimisation of selectivity in high-performance liquid

- chromatography: A magical mystery tour, *J. Chromatogr. A* 1120 (2006) 308–321.
- [3] R. Cela, E.Y. Ordoñez, J.B. Quintana, R. Rodil, Chemometric-assisted method development in RPLC, *J. Chromatogr. A* 1287 (2013) 2–22.
- [4] J.R. Torres Lapasió, M.C. García Álvarez-Coque, Liquid chromatography: Strategies for optimization, en *Encyclopedia of Analytical Science*, 3^a ed. (editado por P.J. Worsfold, A. Townshend, C. Poole, M. Miró), Elsevier, Waltham, MA, Vol. 6, 2019, págs. 252–260.
- [5] D.B. Hibbert, Experimental design in chromatography: A tutorial review, *J. Chromatogr. B* 910 (2012) 2–13.
- [6] E. Rozet, P. Lebrun, P. Hubert, Design spaces for analytical methods, *Trends Anal. Chem.* 42 (2013) 157–167.
- [7] J.R. Torres Lapasió, S. Pous Torres, J.J. Baeza Baeza, M.C. García Álvarez-Coque, Optimal experimental designs in RPLC at variable solvent content and pH based on prediction error surfaces, *Anal. Bioanal. Chem.* 400 (2011) 1217–1230.
- [8] A. Kensert, G. Collaerts, K. Efthymiadis, G. Desmet, D. Cabooter, Deep Q-learning for the selection of optimal isocratic scouting runs in liquid chromatography, *J. Chromatogr. A* 1638 (2021) 461900.
- [9] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. DeJong, P.J. Lewi, J. Smeyers-Verbeke, en *Handbook of Chemometrics and Qualimetrics, Data Handling in Science and Technology* (Vol. 20B), Elsevier Science, Amsterdam, 1997.
- [10] A.C. Atkinson, R.D. Tobias, Optimal experimental design in chromatography, *J. Chromatogr. A* 1177 (2008) 1–11.

- [11] P. Nikitas, A. Pappa-Louisi, Expressions of the fundamental equation of gradient elution and a numerical solution of these equations under any gradient profile, *Anal. Chem.* 77 (2005) 5670–5677.
- [12] J.J. Baeza Baeza, M.C. García Álvarez-Coque, Extension of the linear solvent strength retention model including a parameter that describes the elution strength changes in liquid chromatography, *J. Chromatogr A* 1615 (2020) 460757.
- [13] S. López Ureña, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Enhancement in the computation of gradient retention times in liquid chromatography using root-finding methods, *J. Chromatogr. A* 1600 (2019) 137–147.
- [14] J.A. Navarro Huerta, A. Gisbert Alonso, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Testing experimental designs in liquid chromatography (I): Development and validation of a method for the comprehensive inspection of experimental designs, *J. Chromatogr. A* 1624 (2020) 461180.
- [15] U.D. Neue, H.J. Kuss, Improved reversed-phase gradient retention modeling, *J. Chromatogr. A* 1217 (2010) 3794–3803.
- [16] L.R. Snyder, J.J. Kirkland, J.L. Glajch, *Practical HPLC Method Development*, 2^a ed., John Wiley & Sons, Nueva York, 1992.
- [17] L.R. Snyder, *High-Performance Gradient Elution*, Wiley-Interscience, Nueva York, 2007.
- [18] G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Estimation of significant solvent concentration ranges and its application to the enhancement of the accuracy of gradient predictions, *J. Chromatogr. A* 1057 (2004) 31–39.

- [19] J.A. Navarro Huerta, A. Gisbert Alonso, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Benefits of solvent concentration pulses in retention time modelling of liquid chromatography, *J. Chromatogr. A* 1597 (2019) 76–88.
- [20] A. Gisbert Alonso, J.A. Navarro Huerta, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Global retention models and their application to the prediction of chromatographic fingerprints, *J. Chromatogr. A* 1637 (2021) 46184.

Parte 2

DESARROLLO DE MODELOS GLOBALES DE RETENCIÓN PARA SU APLICACIÓN A MUESTRAS COMPLEJAS

CAPÍTULO 5

**DESARROLLO DE METODOLOGÍAS PARA OBTENER
MODELOS GLOBALES DE RETENCIÓN
QUE DESCRIBAN GRUPOS DE COMPUESTOS**

5.1. Resumen

La resolución de muestras que contienen compuestos desconocidos de distinta naturaleza, o sin estándares disponibles, como es el caso de las plantas medicinales, sigue siendo un reto. Posiblemente, el aspecto más problemático que impide el desarrollo sistemático de métodos, en cromatografía líquida, es la construcción de modelos que describan sin sesgo el comportamiento de retención de los constituyentes de las muestras. En este capítulo, se propone el uso de modelos globales (capaces de describir el comportamiento de todos los compuestos de la muestra), como alternativa al uso de modelos individuales para cada soluto. Los modelos globales contienen parámetros que son específicos de cada soluto, mientras que otros parámetros –relacionados con la columna y el modificador– son comunes para todos los solutos. Se presenta un procedimiento de regresión especial para la construcción de modelos globales, que se aplican a la predicción de cromatogramas altamente complejos, como es el caso de las huellas dactilares cromatográficas, haciendo uso de diversas condiciones experimentales, tanto en elución isocrática como de gradiente. Otra aplicación interesante es la predicción de propiedades moleculares, tales como el coeficiente de reparto octanol/agua, $\log P_{o/w}$, a partir de los parámetros específicos de los solutos en los modelos globales. Los modelos examinados para describir la retención se basan en las ecuaciones propuestas por Snyder, Schoenmakers, Neue-Kuss (NK), Jandera y Bosch-Rosés. En todos los casos, la capacidad predictiva es muy satisfactoria. Los modelos globales se aplican a dos casos de estudio: cromatogramas de extractos de manzanilla eluidos con gradientes de acetonitrilo, y un conjunto de 145 compuestos conocidos, con una amplia gama de estructuras y funcionalidades, eluidos isocráticamente con fases móviles de acetonitrilo/agua.

5.2. Introducción

En cromatografía líquida de fase inversa (RPLC, *reversed-phase liquid chromatography*), el analista dispone de varios factores experimentales que puede modificar a voluntad con el fin de lograr una separación selectiva [1]. Entre los factores disponibles, los relacionados con la composición de la fase móvil son los más útiles para lograr la separación. Esto justifica la importancia de contar con modelos fiables para describir la retención considerando como factor el contenido de modificador en la fase móvil. Generalmente, los modelos se ajustan a partir de los datos proporcionados por un conjunto de experimentos cuidadosamente diseñados (el denominado diseño experimental, véanse los Capítulos 3 y 4). Una vez que se construyen los modelos, éstos pueden aplicarse para predecir la calidad de la separación bajo nuevas condiciones experimentales [2–11]. Se han desarrollado metodologías interpretativas muy eficientes para encontrar las condiciones óptimas de separación en cromatografía líquida (LC, *liquid chromatography*), en base al comportamiento descrito por modelos de retención. Este capítulo aborda el desarrollo y ajuste de modelos globales de retención, que describen el comportamiento de un conjunto de solutos, en lugar del de solutos específicos.

La fiabilidad de las predicciones de retención depende de la calidad de la información proporcionada por el diseño experimental (número y distribución de los experimentos dentro del dominio de los factores), para cada soluto. Otros aspectos decisivos son la existencia de ecuaciones que describan el comportamiento de los compuestos con una elevada precisión y sin sesgo, y el procedimiento de ajuste utilizado. El intervalo de valores del factor experimental considerado y el modo de elución (isocrático o de gradiente) también pueden influir en la exactitud de los ajustes del modelo y las predicciones realizadas a partir de él.

Recientemente, el grupo de investigación ha desarrollado un método para la evaluación de la capacidad de pico en elución en gradiente, basado en la simulación de picos cromatográficos. El método es válido en una amplia gama de situaciones no ideales: elución haciendo uso de gradientes complejos, presencia de picos asimétricos y volúmenes de conexión apreciables, entre otras situaciones [12]. El método propuesto se basa en la construcción de un modelo global, a partir de las correlaciones que pueden establecerse entre los parámetros del modelo cuadrático logarítmico [13,14], para algunos solutos. Anteriormente se utilizó una correlación similar entre la pendiente y la ordenada en el origen para la ecuación lineal logarítmica [15], con el fin de caracterizar columnas y clasificar compuestos estructuralmente relacionados [16–19].

Cabe mencionar que los modelos individuales son capaces de predecir el comportamiento de retención de un único compuesto, con gran exactitud, ya que se particulariza la descripción al soluto, mientras que un modelo global obtiene una tendencia general, y es capaz de predecir la retención de varios compuestos, aunque pagando el precio de sacrificar la exactitud.

En el método desarrollado para el cálculo de la capacidad de pico [12], el modelo global se construyó relacionando los parámetros del modelo individual para compuestos estructuralmente relacionados, con hidrofobicidad variable. Sin embargo, debido a la necesidad de la existencia de similitud química entre los solutos, dicho modelo global no es aplicable a muestras que contengan compuestos que presenten diversas propiedades químicas o compuestos desconocidos. Éste es el caso de los componentes de los extractos de plantas medicinales.

En este capítulo, se comprueba que la construcción de un modelo global basado en correlaciones no es aplicable a la gran diversidad de compuestos

contenidos en los extractos de plantas medicinales. Como solución, se proponen otras alternativas para construir modelos globales de retención que describan el comportamiento de una muestra y, finalmente, permitan la predicción de los cromatogramas de las muestras. En los modelos planteados, algunos parámetros son específicos de cada soluto, mientras que otros explican los efectos generales de los cambios en la concentración del modificador sobre la columna, y son comunes para todos los solutos. El método se desarrolla y aplica a extractos de manzanilla, analizados mediante elución en gradiente. Para validar el método, por un lado, se comparan los cromatogramas experimentales y predichos de los extractos de manzanilla, que contienen una gran cantidad de componentes, y por otro, se comprueba la correspondencia entre la retención isocrática experimental y predicha de un gran conjunto de compuestos de los que se conoce su identidad, y que poseen una gran diversidad de estructuras.

La predicción de picos cromatográficos basada en el uso de modelos globales puede ser de interés en la optimización de cromatogramas que incluyan un número muy elevado de solutos, incluso de naturaleza desconocida o sin patrones disponibles. Éste es el caso de las huellas dactilares cromatográficas de plantas medicinales [20–26], cuya optimización es aún un desafío. Los modelos globales poseen también la capacidad de separar las contribuciones comunes asociadas a la columna y a los cambios de concentración del modificador, de los efectos específicos de los solutos. A su vez, estos efectos pueden correlacionarse con propiedades moleculares.

5.3. Teoría

5.3.1. Ecuaciones para describir la retención

Como se ha comentado, el contenido de modificador orgánico en la fase móvil es el factor experimental controlado, con mayor frecuencia, en LC para modular la fuerza eluyente, la selectividad y el tiempo de análisis, consiguiendo así la separación más adecuada. En la bibliografía, se ha propuesto una amplia variedad de ecuaciones para describir la retención cromatográfica, basadas en dicho factor [7,9–11]. Estas ecuaciones permiten predecir el factor de retención (k), en función de la fracción volumétrica de modificador (φ) en la fase móvil. En este capítulo, se consideran algunas de las ecuaciones más comúnmente utilizadas en RPLC para modelizar la retención, las cuales se toman como base para desarrollar los modelos globales. Las ecuaciones examinadas fueron propuestas por Snyder *et al.* [15], Schoenmakers *et al.* [13,14] (ambas ampliamente utilizadas e históricamente relevantes), Neue y Kuss [27] (que ofrece un rendimiento excepcionalmente bueno en los ajustes para solutos individuales), Jandera *et al.* [28] (con la ventaja de conducir a una solución analítica cuando se aplica a elución en gradiente, siendo válida para otros modos de elución, con buena capacidad predictiva, y cuyo rendimiento aún es poco conocido en RPLC). Finalmente, se estudia la ecuación de Bosch y Rosés [29] (un modelo de carácter global, que se utiliza aquí con fines comparativos).

5.3.1.1. Modelo logarítmico lineal de Snyder

Una de las ecuaciones más simples, utilizadas para modelizar la retención en RPLC, es el modelo logarítmico lineal (también llamado modelo lineal de fuerza del disolvente), que fue propuesto por Snyder *et al.* [15]:

$$\ln k = \ln \frac{t_R - t_0}{t_0 - t_{\text{ext}}} = \ln k_w - S \varphi \quad (5.1)$$

donde t_R , t_0 y t_{ext} son el tiempo de retención, tiempo muerto y tiempo extra-columnar (correspondiente a la migración de los solutos a través de las conexiones del equipo), respectivamente; $\ln k_w$ es el logaritmo del factor de retención cuando se utiliza agua pura como eluyente; y S cuantifica la fuerza eluyente del modificador. La ecuación de Snyder sólo es fiable para describir la retención cuando el diseño cubre un intervalo estrecho de modificador orgánico. Sin embargo, posee la ventaja de originar expresiones algebraicas que permiten la predicción directa de la retención cuando se utilizan gradientes lineales.

5.3.1.2. Modelo logarítmico cuadrático de Schoenmakers

Para intervalos más amplios de modificador orgánico, la Ecuación (5.1) puede adaptarse añadiendo un término cuadrático [13,14]:

$$\ln k = \ln k_w - S \varphi + T \varphi^2 \quad (5.2)$$

donde T corrige la desviación de la linealidad de $\ln k$ frente a φ . La ecuación cuadrática logarítmica posee la desventaja de que no existe una expresión algebraica que prediga directamente el tiempo de retención en elución en gradiente.

5.3.1.3. Modelo de Bosch-Rosés

Esta ecuación describe la retención utilizando una función lineal, donde se aíslan las contribuciones a la polaridad de los tres agentes involucrados en el proceso de separación (solute, fase estacionaria y fase móvil) [29]:

$$\ln k = (\ln k)_0 + p_S (P_M^N - P_S^N) \quad (5.3)$$

donde p_S y P_M^N son descriptores de polaridad para el soluto y la fase móvil, respectivamente, y $(\ln k)_0$ y P_S^N están relacionados con la hidrofobicidad de la fase estacionaria. El término $(\ln k)_0$ representa la retención en una fase móvil hipotética, con la misma polaridad que la fase estacionaria ($P_M^N = P_S^N$). Para mezclas de acetonitrilo-agua, P_M^N se calcula mediante la siguiente expresión:

$$P_M^N = 1.00 - \frac{2.068\varphi}{1 + 1.341\varphi} \quad (5.4)$$

5.3.1.4. Modelo de Neue-Kuss

En 2010, Neue y Kuss propusieron una ecuación [27], cuyo rendimiento descriptivo ha demostrado ser excelente en elución isocrática, incluso en intervalos de modificador orgánico muy amplios:

$$k = k_0(1 + c\varphi)^2 e^{\frac{b\varphi}{1+c\varphi}} \quad (5.5)$$

donde k_0 es el factor de retención extrapolado para una fase móvil constituida únicamente por agua, c es un parámetro de curvatura y b una medida de la fuerza eluyente, relacionada con la distribución del soluto entre ambas fases (estacionaria y móvil), debido a las interacciones hidrofóbicas. Además, la Ecuación (5.5) permite soluciones algebraicas en elución en gradiente.

5.3.1.5. Modelo de Jandera

Como se ha visto, la mayoría de las ecuaciones publicadas para describir la retención implican funciones logarítmicas. Sin embargo, existen otros enfoques. Éste es el caso de una ecuación propuesta por Jandera *et al.* para cromatografía líquida de fase normal (NPLC, *normal phase liquid chromatography*) [28], que también ofrece una excelente exactitud en RPLC:

$$\frac{1}{k} = [a + b\varphi]^m \quad (5.6)$$

donde a , b y m son los parámetros del modelo. Una ventaja interesante de la ecuación de Jandera es que la resolución de la ecuación fundamental de elución en gradiente conduce a soluciones algebraicas.

5.3.2. Modelos globales basados en la correlación de los parámetros de los modelos para los solutos individuales

Como se ha comentado, el grupo de investigación propuso recientemente un modelo global, que relaciona los parámetros de los modelos de retención individuales de un conjunto de estándares químicamente relacionados [12]. El enfoque propuesto se basó en el modelo de Schoenmakers (Ecuación (5.2)). Los parámetros de retención ($\log k_{w,i}$, S_i y T_i), para solutos i estructuralmente relacionados, se obtuvieron a partir de las correlaciones $S = f_1(\ln k_w)$ y $T = f_2(\ln k_w)$. Por otro lado, el perfil de los picos cromatográficos se predijo a partir de la correlación entre las semianchuras izquierda, A , y derecha, B , de los picos (medidas al 10 % de altura de pico) y el tiempo de retención isocrático (t_{RI}): $A = f_3(t_{RI}(\varphi))$ y $B = f_4(t_{RI}(\varphi))$.

La correspondencia entre las semianchuras y el tiempo de retención isocrático es algo más compleja, cuando los datos experimentales a procesar se

obtienen a partir de gradientes de elución. En este caso, el soluto i abandona la columna en un determinado instante de composición φ_i , a lo largo del programa de gradiente. Utilizando la aproximación de Jandera [30], se pueden predecir las semianchuras de los picos en ese instante, a partir del tiempo de retención isocrático para esa composición (t_{RI}), utilizando las siguientes funciones: $A_i = f_3(t_{RI,i}(g_\varphi))$ y $B_i = f_4(t_{RI,i}(g_\varphi))$, donde g_φ es el programa de gradiente: $g_\varphi = f_5(t)$. En el Capítulo 2 (Sección 2.3.2), se expone detalladamente la metodología de predicción de la retención en elución en gradiente, utilizando métodos de búsqueda de cero e integración analítica [31].

Las ecuaciones de la Sección 5.3.1 describen la retención de un determinado soluto. En este capítulo, dichas ecuaciones se modifican para que respondan globalmente a conjuntos de solutos, estableciendo parámetros comunes para todos ellos. En la Sección 5.5.2 se explica en detalle el desarrollo de un enfoque alternativo para construir modelos globales de retención. Finalmente, se muestran dos importantes aplicaciones que demuestran su utilidad.

5.4. Parte experimental

5.4.1. Análisis de muestras de manzanilla

La metodología propuesta se aplicó a muestras de manzanilla, adquiridas en una tienda local. Los extractos de esta planta se prepararon siguiendo las recomendaciones descritas por Álvarez Segura *et al.* [25]. Para ello, después de triturar y homogeneizar la muestra original, se pesaron porciones de 1.0 g del polvo resultante para proceder a la extracción de sus componentes, con ayuda de un baño de ultrasonidos, durante 60 min a una temperatura de 80 °C, haciendo uso de 15 mL de disolución extractante al 70 % (v/v) de metanol (Scharlau, Barcelona). Los extractos se centrifugaron a 3000 rpm durante 5 min

y los sobrenadantes se filtraron a través de membranas de Nylon con un tamaño de poro de 0.45 μm de Micron Separations (Westboro, MA, EE.UU.), para su posterior inyección en el cromatógrafo.

Las huellas dactilares cromatográficas de los extractos de manzanilla se obtuvieron con un cromatógrafo HPLC (Serie Infinity 1200 de Agilent, Waldbronn, Alemania), compuesto por los siguientes módulos: bomba cuaternaria, autoinyector provisto de viales de 2 mL, compartimento de columna termostaticado y detector de fila de diodos UV-visible. Las señales cromatográficas se registraron en el intervalo de longitud de onda de 210 a 280 nm. El volumen de inyección fue de 10 μL y el flujo de fase móvil se mantuvo constante a 1.0 $\text{mL}\cdot\text{min}^{-1}$.

La columna analítica fue una Zorbax Eclipse XDB-C18 (4.6×150 mm, Agilent), con un tamaño de partícula de 5 μm . Todas las inyecciones se realizaron en condiciones de temperatura controlada, a 25.0 $^{\circ}\text{C}$. La separación se llevó a cabo en elución en gradiente, con mezclas hidro-orgánicas compuestas por agua nanopura obtenida utilizando un sistema de purificación de Adrona B30 Trace (Burladingen, Alemania) y acetonitrilo grado HPLC (Scharlau). Estas mezclas se tamponaron a un valor de pH cercano a 3.0, mediante la adición de ácido fórmico al 0.1 % (v/v) (Acros Organics, Fair Lawn, NJ). Todas las soluciones se desgasificaron utilizando un baño ultrasónico de Elmasonic (Singen, Alemania), y posteriormente se filtraron a través de las membranas de Nylon de 0.45 μm mencionadas anteriormente.

El tiempo muerto del sistema se determinó mediante la inyección de KBr (Acros Organics, Geel, Bélgica). El tiempo de permanencia se midió utilizando un gradiente de acetona (Scharlau), siendo $t_D = 1.17$ min. El tiempo extra-columnar poseía un valor de $t_{\text{ext}} = 0.052$ min.

Para la modelización de los extractos de manzanilla, se hizo uso de un diseño experimental de seis gradientes (Figura 5.1). Con el fin de seleccionar las condiciones operativas, se ejecutó en primer lugar un gradiente de exploración que osciló entre el 5 y el 100 % (v/v) de acetonitrilo en 60 min (G1 en la Figura 5.1). A continuación, se diseñaron cinco gradientes más añadiendo un nodo intermedio al gradiente de exploración en diferentes posiciones (G2 a G6 en la Figura 5.1).

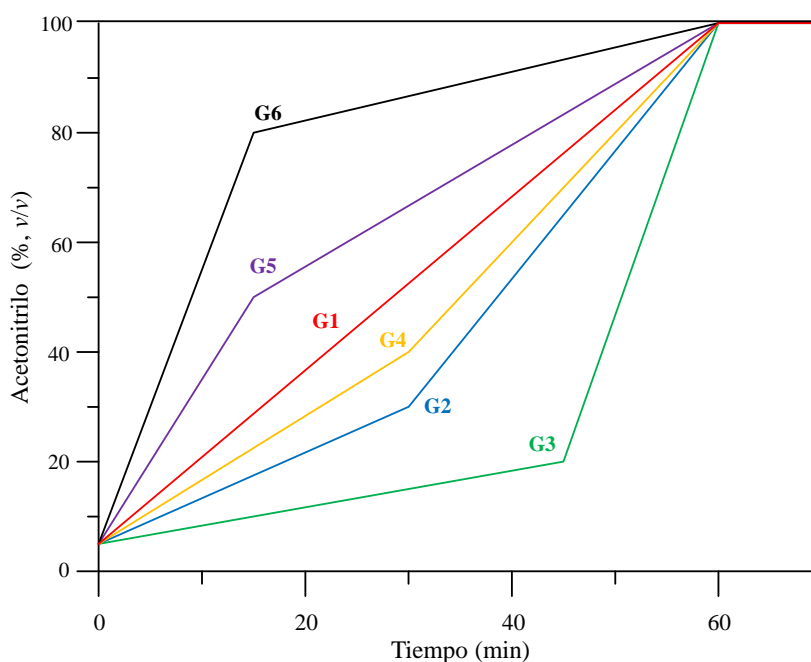


Figura 5.1. Diseño experimental en gradiente, utilizado para eluir los extractos de manzanilla y modelizar 15 picos seleccionados, trazables en las huellas dactilares cromatográficas para cada gradiente, utilizando una columna C18 de 15 cm. Los códigos G1 a G6 indican el orden de adquisición de los cromatogramas. Para más detalles, véase la Sección 5.4.1.

5.4.2. Conjunto de 15 sulfonamidas

Con fines comparativos, se seleccionó un conjunto de solutos que se habían utilizado en estudios previos (Referencia [12]), para construir modelos basados en la correlación de los parámetros S y T vs. $\log k_w$ en la ecuación de Schoenmakers (Ecuación (5.2)). El conjunto de compuestos consistió en 15 sulfonamidas: sulfacetamida, sulfacoloropiridazina, sulfadiazina, sulfadimetoxina, sulfaguanidina, sulfamerazina, sulfametazina, sulfametizol, sulfametoxazol, sulfamonometoxina, sulfanilamida, sulfapiridina, sulfaquinoxalina, sulfatiazol y sulfisoxazol. El comportamiento cromatográfico se obtuvo con una columna de 9 cm construida acoplando columnas C18 más cortas (5 + 2 + 2 cm, ACE, Aberdeen, Escocia, Reino Unido). Las sulfonamidas se eluyeron isocráticamente utilizando mezclas de acetonitrilo-agua tamponadas a pH 3.5 con dihidrógeno fosfato anhidro 0.01 M y HCl. Los porcentajes de acetonitrilo en el diseño experimental fueron 10, 13, 15, 17 y 20 % (v/v). Otros detalles se pueden consultar en la Referencia [12].

5.4.3. Conjunto de 145 compuestos de diversa naturaleza

La metodología desarrollada se validó utilizando un conjunto de datos extraídos de la bibliografía, para un gran número de compuestos. Inicialmente, el conjunto de datos incluía 173 compuestos de carácter diverso (alquilbencenos, fenoles, fenonas, halobencenos, nitrobenenos, amidas aromáticas, aldehídos, ésteres, éteres y nitrilos, entre otros), cuyos datos de retención se publicaron en varios trabajos de Smith y Burr [32–37]. Las condiciones experimentales con las que se obtuvieron los datos son siete mezclas isocráticas de acetonitrilo y agua, con las siguientes composiciones:

30, 40, 50, 60, 70, 80 y 90 % (v/v), utilizando una columna Spherisorb ODS-2 (100 mm × 5 mm).

Anteriormente, el grupo utilizó el conjunto de compuestos indicados para evaluar la calidad de la transferencia de los datos de retención entre columnas y examinar la dependencia de los parámetros de polaridad con diferentes descriptores de solvatación [38–40]. De los 173 compuestos, se excluyeron aquéllos para los que no había datos de retención para al menos cuatro fases móviles, y/o para los que no había valores $\log P_{o/w}$ disponibles. Bajo tales restricciones, se mantuvieron un total de 145 compuestos para el estudio (Tabla 5.1). Para 126 de los 145 compuestos, se dispuso de medidas de retención para cinco o más fases móviles. En algunos casos, la ausencia de datos se debió a la baja polaridad de los compuestos, lo que se tradujo en tiempos de retención prohibitivos cuando se eluyeron con fases móviles de baja fuerza eluyente.

El tratamiento de datos se llevó a cabo con funciones desarrolladas por el grupo de investigación, escritas en Matlab 2020a (The Mathworks, Natick, MA).

5.5. Resultados y discusión

El objetivo del trabajo realizado es proponer una metodología de modelización, aplicable a muestras complejas, para las que no se dispone de información detallada de sus componentes. El modelo propuesto responde a varios solutos simultáneamente, aun cuando no exista una relación química entre ellos. La metodología fue desarrollada para ser aplicada a plantas medicinales, con el objetivo de diseñar, mejorar y validar modelos globales de retención que, en último término, permitan la predicción de huellas dactilares cromatográficas completas de este tipo de muestras.

Tabla 5.1. Conjunto de 145 compuestos eluidos con fases móviles isocráticas de acetonitrilo-agua, obtenidos de informes sucesivos de Smith y Burr (Referencias [32]–[37]).

3-Fenil-1-propionamida	4-Metilbenzaldehido	3-Bromotolueno
Fenilacetamida	2-Hidroxibenzaldehido	4-Bromotolueno
Benzamida	3-Hidroxibenzaldehido	2-Clorotolueno
2-Metilbenzamida	4-Hidroxibenzaldehido	3-Clorotolueno
3-Metilbenzamida	3-Fenil-1-propionitrilo	4-Clorotolueno
4-Metilbenzamida	4-Fenil-1-butironitrilo	2-Nitrotolueno
N,N-Dimetilbenzamida	Benzonitrilo	3-Nitrotolueno
N-Metilbenzamida	2-Metilbenzonitrilo	4-Nitrotolueno
2-Hidroxibenzamida	3-Metilbenzonitrilo	Benceno
4-Hidroxibenzamida	4-Metilbenzonitrilo	Etilbenceno
Phenacilbromuro	2-Hidroxibenzonitrilo	1,2-Dimetilbenceno
2-Feniletilbromuro	3-Hydroxibenzonitrilo	1,3-Dimetilbenceno
3-Fenil-1-propilbromuro	4-Hidroxibenzonitrilo	1,4-Dimetilbenceno
Benzilbromuro	2-Metilanisol	n-Propilbenceno
Benzilcianuro	3-Metilanisol	Isopropilbenceno

Tabla 5.1. (continuación)

	3-Hidroxiacetofenona	Fenol
1-Bromo-2-nitrobenceno		
Metoxibenceno	4-Hidroxiacetofenona	2-Metilfenol
1,2-Dihidroxibenceno	Dimetilftalato	3-Metilfenol
1,3-Dihidroxibenceno	Metil-3-fenilpropionato	4-Metilfenol
1,4-Dihidroxibenceno	3-Etilfenilpropionato	2,4-Dimetilfenol
Anilina	Etilbenzoato	2,5-Dimetilfenol
2-Metilaniлина	Metilbenzoato	2-Fenilfenol
3-Metilaniлина	Metil-2-metilbenzoato	3-Fenilfenol
4-Metilaniлина	Metil-3-metilbenzoato	4-Fenilfenol
N-Etilaniлина	Metil-4-metilbenzoato	4-t-Butilfenol
2-Bromoaniлина	Metil-2-hidroxi benzoato	2-Bromofenol
3-Bromoaniлина	Metil-3-hidroxi benzoato	3-Bromofenol
2-Nitroaniлина	Metil-4-hidroxi benzoato	4-Bromofenol
3-Nitroaniлина	Metil-4-fenilbutirato	2-Aminofenol
1-Fenil-2-butanona	Etilfenilacetato	3-Aminofenol
4-Fenil-2-butanona	Metilfenilacetato	4-Aminofenol

5.5.1. Construcción de modelos de retención individuales para muestras de manzanilla

5.5.1.1. Selección de picos de referencia

El primer paso para modelizar el comportamiento cromatográfico de un conjunto de compuestos consiste en recopilar información sobre los tiempos de retención y el perfil de las señales, en condiciones de elución controladas, según un diseño experimental específico, como el de la Figura 5.1 (véase la Sección 5.4.1). En los casos más favorables, donde se conoce la identidad de los componentes y se dispone de estándares para todos ellos, éstos se pueden inyectar individualmente o en grupos, utilizando las condiciones experimentales de un diseño. Desafortunadamente, las plantas medicinales contienen una cantidad muy elevada de compuestos, y es poco probable que haya estándares disponibles para todos ellos. Además, a menudo se desconoce la identidad de la mayoría de los componentes. En consecuencia, el estudio de la retención y el perfil de las señales debe realizarse directamente a partir de los picos cromatográficos de los componentes de la muestra.

En este capítulo, se propone una metodología que consiste en seleccionar una serie de picos de las huellas dactilares, preferiblemente los que poseen mayor intensidad, o al menos, picos que sean trazables en todas las condiciones de elución ensayadas. En la Sección 5.5.2, a partir de estos picos, se derivan parámetros comunes del modelo que recogen la influencia de la columna y el eluyente sobre la retención. Como se verá más adelante, estos parámetros globales permiten el cálculo de parámetros relacionados con el comportamiento de los solutos, tanto para los picos seleccionados, como para todos los demás componentes del cromatograma.

Con este objetivo, y con el fin de ser utilizados como referencia, se han seleccionado arbitrariamente 15 picos de las huellas dactilares de los extractos de manzanilla, distribuidos a lo largo del cromatograma (véase la Figura 5.2). El uso del cromatograma total, es decir, de un cromatograma que muestra la absorbancia máxima para cada instante de tiempo, en un intervalo de longitudes de onda determinado, junto al desarrollo del diseño experimental de entrenamiento de forma escalonada, facilitó la trazabilidad de los picos. Con ello, no se produjo ambigüedad alguna en la identificación de los picos de referencia en las distintas condiciones de gradiente ensayadas.

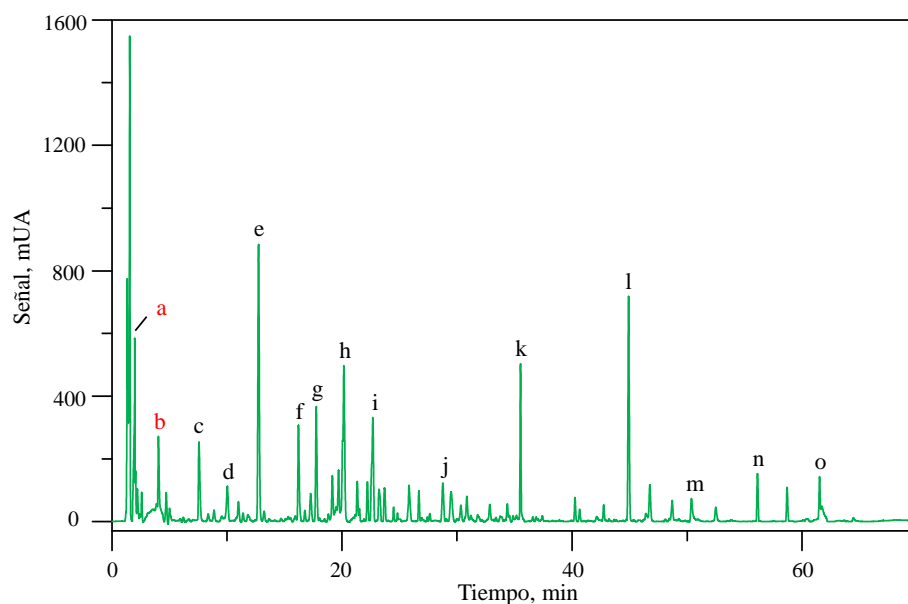


Figura 5.2. Huella cromatográfica correspondiente a la elución de un extracto de manzanilla, utilizando el gradiente más lento del diseño experimental de la Figura 5.1 (gradiente G3). Se indican los 15 picos que pudieron trazarse a lo largo de los seis gradientes, lo que permitió su uso para el ajuste satisfactorio de las diferentes ecuaciones de retención.

Los picos seleccionados pueden corresponder a un solo compuesto, o a más de uno, pero a lo largo de la secuencia de gradientes se observó que se movían siempre de acuerdo a patrones consistentes.

5.5.1.2. Ajuste de modelos de retención individuales para los picos de referencia

Con el fin de obtener modelos de retención individuales, de acuerdo a diversas ecuaciones, los datos de retención de cada uno de los 15 picos de referencia en los cromatogramas de los extractos de manzanilla (Figura 5.2) se ajustaron soluto a soluto. Para realizar los ajustes, se utilizaron las medidas recogidas de acuerdo al diseño experimental de seis gradientes mostrado en la Figura 5.1.

La Tabla 5.2 contiene los estadísticos de regresión para los 15 picos de referencia, correspondientes al modelo de Schoenmakers (Ecuación (5.2)) y a su simplificación lineal (Ecuación (5.1)), así como a los modelos NK (Ecuación 5.5), y de Jandera (Ecuación (5.6)). Todas estas ecuaciones se ajustaron a partir de sus respectivas primitivas, utilizando el método de Newton [31] aplicado a la ecuación fundamental de elución en gradiente para gradientes multi-lineales, y siguiendo la metodología presentada en la Sección 2.3.2 del Capítulo 2. Todos los modelos ofrecieron un excelente rendimiento, incluso el de Snyder (Ecuación (5.1)), pero el modelo NK destaca por su fiabilidad y excelentes resultados.

Tabla 5.2. Estadísticos de ajuste para diferentes modelos de retención, considerando los 15 picos de referencia seleccionados en los extractos de manzanilla, utilizando como fuente de datos los tiempos de retención de gradiente obtenidos con el diseño experimental mostrado en la Figura 5.1.

Solutos	Estadísticos de ajuste ^a	Modelo de retención			
		Lineal	Cuadrático	Neue-Kuss	Jandera
a	R^2_{aj}	0	0	0	0
	ER	0.12	2.05	0.50	0.18
	F	0	3	2	1
	S_{pred}	0	0.07	0.02	0.01
b	R^2_{aj}	0.85152	0	0.84091	0.81986
	ER	1.13	2.73	1.01	1.09
	F	31	9	29	26
	S_{pred}	0.06	0.18	0.07	0.07
c	R^2_{aj}	0.99992	0.93628	0.99713	0.99322
	ER	0.10	2.87	0.74	1.12
	F	64932	99	1718	703
	S_{pred}	0.01	0.34	0.07	0.11
d	R^2_{aj}	0.99819	0.99708	0.99997	0.99998
	ER	1.11	1.21	0.13	0.09
	F	2844	1796	153192	279944
	S_{pred}	0.13	0.16	0.02	0.01
e	R^2_{aj}	0.99919	0.99909	0.99999	0.99998
	ER	0.90	0.83	0.10	0.12
	F	6308	5113	374593	242905
	S_{pred}	0.14	0.15	0.02	0.02
f	R^2_{aj}	0.99963	0.99995	0.99989	0.99964
	ER	0.76	0.27	0.38	0.65
	F	13748	100363	47808	13694
	S_{pred}	0.15	0.06	0.08	0.15
g	R^2_{aj}	0.99974	0.99968	0.99996	0.99999
	RE	0.63	0.68	0.20	0.10
	F	19586	15887	118486	531358
	S_{pred}	0.14	0.15	0.06	0.03
h	R^2_{aj}	0.99979	0.99942	0.99999	0.99985
	ER	0.64	0.95	0.38	0.42
	F	24198	8487	52842	32930
	S_{pred}	0.16	0.27	0.11	0.14

Tabla 5.2 (continuación).

Solutos	Estadísticos de ajuste ^a	Modelo de retención			
		Lineal	Cuadrático	Neue-Kuss	Jandera
i	R^2_{aj}	0.99983	0.99989	0.99993	0.99998
	ER	0.6	0.46	0.33	0.16
	F	30372	46006	72911	266849
	S_{pred}	0.17	0.13	0.11	0.06
j	R^2_{aj}	0.99984	0.99980	0.99901	0.99967
	ER	0.51	0.60	0.83	0.66
	F	31471	24814	5222	15028
	S_{pred}	0.18	0.21	0.46	0.27
k	R^2_{aj}	0.99996	0.99993	0.99999	0.99987
	ER	0.26	0.32	0.12	0.42
	F	117233	69175	346715	38239
	S_{pred}	0.10	0.13	0.06	0.17
l	R^2_{aj}	0.99981	0.99997	0.99896	0.99837
	ER	0.42	0.14	0.86	1.00
	F	26133	157207	4848	3107
	S_{pred}	0.21	0.09	0.49	0.61
m	R^2_{aj}	0.98555	0.99816	0.99146	0.99969
	ER	2.79	0.77	1.61	0.38
	F	350	2891	617	16123
	S_{pred}	1.72	0.61	1.32	0.25
n	R^2_{aj}	0.99991	0.99983	0.99989	0.99873
	ER	0.22	0.22	0.17	0.70
	F	57621	29030	45635	4004
	S_{pred}	0.13	0.19	0.15	0.50
o	R^2_{aj}	0.99986	0.99909	0.99735	0.99727
	ER	0.18	0.30	0.55	0.65
	F	35845	5606	2012	1827
	S_{pred}	0.13	0.34	0.58	0.59

^a n (número de medidas) = 6; R^2_{aj} = coeficiente de determinación ajustado; ER = error relativo medio; F = coeficiente F de Snedecor; S_{pred} = error estándar en las predicciones (min).

Dos de los solutos (“a” y “b” en la Figura 5.2) eluyen tan rápido que su retención apenas varía con la composición de los gradientes, lo que conduce a que no se puedan establecer correlaciones de suficiente calidad. El soluto etiquetado como “a” eluye en aproximadamente 1.9 min (prácticamente dentro del tiempo de permanencia, que fue de 1.2 min). De ahí la imposibilidad de ajustar los datos para este soluto.

5.5.1.3. Construcción de modelos globales basados en correlaciones entre los parámetros para todo el conjunto de picos de referencia

Como se ha mencionado, en la Referencia [12] se estableció con éxito, para un conjunto de sulfonamidas (véase la Sección 5.4.2), un modelo global basado en la existencia de correlaciones entre los parámetros de los modelos individuales. En contraste, en los extractos de plantas medicinales, las variaciones extremas en la naturaleza de los compuestos, que los gradientes hacen compatibles, se traducen en ajustes problemáticos donde incluso encontrar soluciones iniciales comunes es difícil. Como consecuencia, se requiere la aplicación de estrategias de prueba y error, con varios intentos a partir de diferentes estimaciones iniciales, para obtener el modelo.

En el caso de la manzanilla, los gradientes abarcaron entre el 5 y 100 % (v/v) de acetonitrilo. Por lo tanto, se necesitaban ecuaciones de retención capaces de manejar amplias variaciones de composición del eluyente. Ésta y otras consideraciones hicieron recomendable el uso de la ecuación cuadrática logarítmica de Schoenmakers (Ecuación (5.2)). Por lo tanto, se intentó en primer lugar la implementación de un modelo global, basado en esta ecuación, estableciendo correlaciones entre los parámetros S y T vs. $\ln k_w$.

La obtención de la correlación de los parámetros de los modelos individuales para los picos de referencia de la manzanilla condujo a los siguientes resultados:

- (i) Existe una buena correlación entre los tiempos de retención de los picos de referencia y los valores de $\ln k_w$ (Tabla 5.3), que es un parámetro que mide la hidrofobicidad de los solutos en la Ecuación (5.2): a medida que aumenta la retención, también lo hace el valor de $\ln k_w$.
- (ii) Por el contrario, las correlaciones de S y T vs. $\ln k_w$ (Figuras 5.3a y 5.3c) son muy pobres, pudiéndose distinguir dos tendencias independientes, asociadas a solutos de menor y mayor hidrofobicidad. Los valores de S siguen un comportamiento descendente para los primeros 11 y últimos cuatro solutos, mientras que los valores de T se mantienen estables para los primeros 10 solutos con un valor próximo a 5, para seguir una tendencia descendente para los últimos cinco solutos. Debe tenerse en cuenta que uno de los solutos más hidrofóbicos se halla anormalmente ubicado en la correlación de S vs. $\ln k_w$.

Tabla 5.3. Parámetros de los modelos de retención de Snyder y Schoenmakers (individuales o globales), obtenidos para cada pico de referencia seleccionado en los extractos de manzanilla. Los diferentes tratamientos se describen en la Sección 5.5.2.

Solutos	Parámetros de retención ^a	Tratamiento ^c					
		A	B	C	D	E	F
a	p_1	-0.66	0.62	-0.68	0.62	2.93	0.54
	p_2	0.03	-23.61	0.00	-26.31	-25.31	-21.96
	p_3	—	4.74	5.00	5.02	4.86	—
b	p_1	1.32	2.29	1.33	2.29	3.93	1.99
	p_2	-9.93	-25.95	-10.29	-26.31	-25.31	-21.96
	p_3	—	5.00	2.88	5.02	4.86	—
c	p_1	2.68	3.46	2.69	3.46	4.18	3.10
	p_2	-15.83	-26.00	-16.12	-26.31	-25.31	-21.96
	p_3	—	5.02	2.25	5.02	4.86	—
d	p_1	3.90	4.10	3.86	4.10	4.93	3.76
	p_2	-23.71	-26.01	-23.73	-26.31	-25.31	-21.96
	p_3	—	5.00	4.85	5.02	4.86	—
e	p_1	4.57	4.71	4.52	4.71	5.07	4.36
	p_2	-24.20	-25.89	-24.20	-26.31	-25.31	-21.96
	p_3	—	4.97	4.98	5.02	4.86	—
f	p_1	6.04	5.50	5.45	5.50	6.09	5.15
	p_2	-29.58	-26.55	-25.58	-26.31	-25.31	-21.96
	p_3	—	5.00	4.80	5.02	4.86	—
g	p_1	5.73	5.78	5.65	5.78	1.86	5.38
	p_2	-24.80	-25.99	-24.81	-26.31	-25.31	-21.96
	p_3	—	5.58	4.80	5.02	4.86	—
h	p_1	6.82	6.47	6.72	6.47	16.86	5.91
	p_2	-28.35	-26.90	-28.59	-26.31	-25.31	-21.96
	p_3	—	5.02	5.81	5.02	4.86	—

Tabla 5.3 (continuación).

Solutos	Parámetros de retención ^a	Tratamiento ^c					
		A	B	C	D	E	F
i	p_1	6.40	6.37	6.42	6.37	7.71	7.38
	p_2	-17.74	-18.21	-18.46	-26.31	-25.31	-21.96
	p_3	-	2.44	2.64	5.02	4.86	-
j	p_1	8.98	10.14	8.64	10.14	10.18	8.82
	p_2	-22.35	-27.65	-22.37	-26.31	-25.31	-21.96
	p_3	-	5.01	3.39	5.02	4.86	-
k	p_1	7.05	10.24	7.29	10.24	10.25	13.28
	p_2	-9.70	-22.81	-11.08	-26.31	-25.31	-21.96
	p_3	-	12.98	1.73	5.02	4.86	-
l	p_1	18.68	11.28	6.69	11.28	10.93	15.82
	p_2	-26.07	-21.56	-7.23	-26.31	-25.31	-21.96
	p_3	-	11.08	0.41	5.02	4.86	-
m	p_1	8.63	13.68	9.30	13.68	13.96	18.77
	p_2	-8.50	-22.42	-10.36	-26.31	-25.31	-21.96
	p_3	-	9.42	1.28	5.02	4.86	-
n	p_1	10.96	17.35	7.24	17.35	17.91	21.13
	p_2	-9.94	-24.21	-9.46	-26.31	-25.31	-21.96
	p_3	-	7.88	5.01	5.02	4.86	-
Estadísticos de ajuste global ^b	R^2	0.99958	0.99991	0.99495	0.89845	0.99907	0.99851
	ER	0.72	0.55	1.70	14.52	1.89	2.20
s_{pred}	F	213421	1012113	17451	617	93959	58308
		0.38	0.18	1.33	5.94	0.57	0.72

^a Modelo lineal (Ecuación (5.1)): $p_1 = \log k_w$, $p_2 = S$; modelo cuadrático (Ecuación (5.2)): $p_1 = \log k_w$, $p_2 = S$, $p_3 = T$.

^b n (número de medidas de retención) = 90; R^2 = Coeficiente de determinación; ER = Error relativo medio; F = F Snedecor; s_{pred} = Error estándar en las predicciones (min).

^c Ver texto para los diferentes tratamientos.

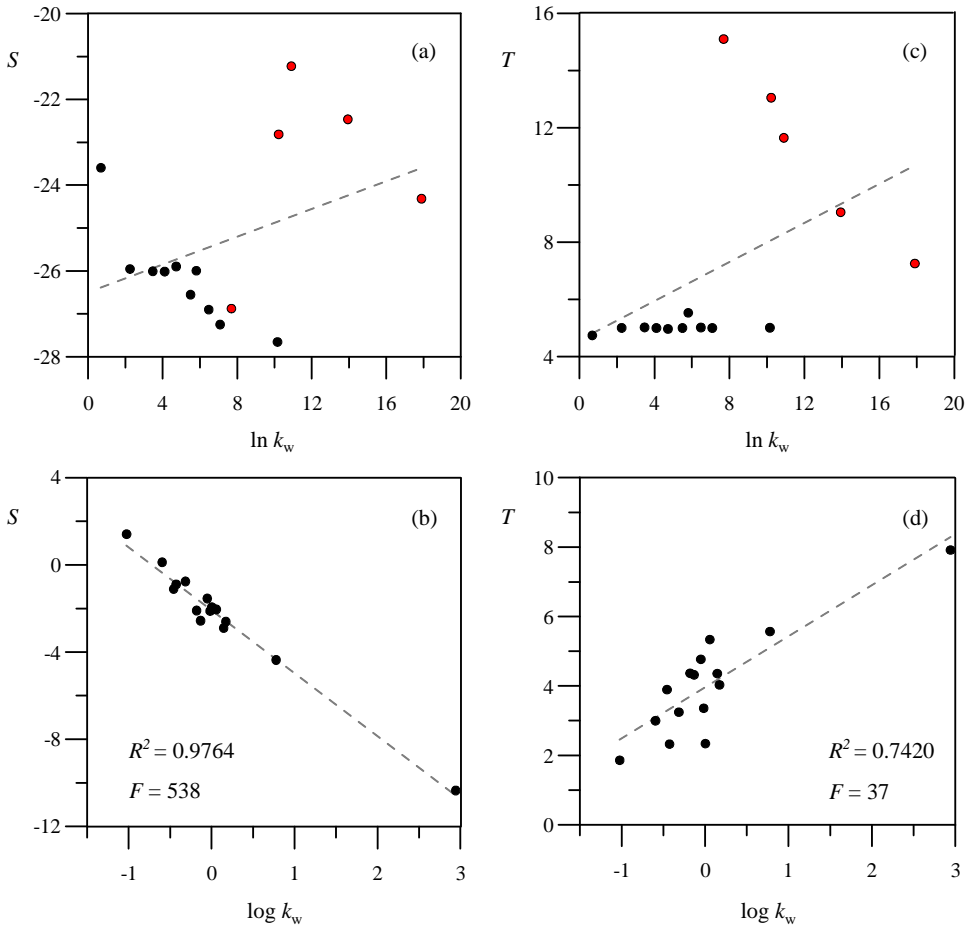


Figura 5.3. Correlación entre los parámetros S y T (modelo de Schoenmakers, Ecuación (5.2)), respecto al logaritmo de k_w , para: (a,c) los 15 picos en los extractos de manzanilla obtenidos con los seis gradientes de acetonitrilo/agua pertenecientes al diseño experimental (véanse las Figuras 5.1 y 5.2), y (b, d) los picos para un conjunto de 15 sulfonamidas eluidas con cinco fases móviles isocráticas de acetonitrilo/agua (ver Referencia [12]). No se dan estadísticas para (a) y (c), debido a la falta de correlación; Se pueden observar dos tendencias independientes para los solutos más lentos (marcados como puntos rojos) y más rápidos (puntos negros). En (b) y (d) se han utilizado logaritmos decimales.

Por lo tanto, debido a la dispersión de los valores y la presencia de tendencias diferenciadas, no es posible estimar S y T a partir de correlaciones con $\ln k_w$. Obsérvese la gran diferencia en la calidad de las correlaciones encontradas en las Figuras 5.3a y 5.3c, en comparación a las excelentes correlaciones que se muestran en las Figuras 5.3b y 5.3d, obtenidas para 15 solutos pertenecientes a la misma familia (sulfonamidas). Las excelentes correlaciones para el conjunto de sulfonamidas pueden explicarse porque comparten una estructura química básica común. El comportamiento cromatográfico para estos compuestos se modelizó, además, utilizando tiempos de retención medidos en elución isocrática, mediante la inyección de estándares, lo que permite obtener modelos más exactos y evita posibles problemas en la identificación de picos en diferentes experimentos.

El peor rendimiento de las correlaciones observadas para los picos de los extractos de manzanilla, que se repite para los extractos de otras plantas medicinales, se puede explicar considerando que las señales modelizadas corresponden a compuestos pertenecientes a diferentes familias, probablemente con estructuras químicas y tamaños extremadamente variables. Además, los gradientes ensayados abarcan un intervalo de modificador extremadamente amplio (del 5 al 100 % v/v de acetonitrilo). Los productos naturales, en general, contienen un elevado número de compuestos de muy diversa naturaleza, que incluye a polímeros y biomoléculas de gran tamaño, lo que da lugar a diferentes interacciones de los componentes de la muestra con la columna.

Cuando no se consiguen buenas correlaciones entre los parámetros del modelo de retención, para un conjunto de compuestos, una posible alternativa podría ser establecer correlaciones locales para conjuntos de solutos con tendencias similares, o que pertenezcan a la misma familia (si se han establecido sus identidades mediante cualquier técnica de detección selectiva).

Para los extractos de manzanilla, se pueden establecer dos tendencias independientes basadas en las diferencias de polaridad (que aparecen resaltadas en las Figuras 5.3a y 5.3c con puntos negros y rojos). Sin embargo, esta estrategia no es útil en la práctica, ya que las dos tendencias locales arrojarían predicciones ambiguas dependiendo de la serie utilizada para ajustar el modelo.

5.5.2. Desarrollo de un nuevo enfoque para construir modelos globales de retención para los picos de referencia de la manzanilla

Lamentablemente, no es posible desarrollar un modelo global basado en la correlación de los parámetros S y T vs. $\ln k_w$ en la ecuación de Schoenmakers, que sea representativo de todos los compuestos en los extractos de manzanilla. Por lo tanto, era necesario desarrollar un enfoque distinto. El objetivo de los siguientes estudios ha sido construir un modelo válido para todos los compuestos presentes en los extractos de manzanilla, que se pudiera utilizar en lugar de las correlaciones entre los parámetros individuales de los solutos. Una forma de construir tales modelos globales es desarrollar los ajustes, utilizando estimaciones iniciales de los parámetros lo más similares posibles. Con estos valores iniciales y deteniendo prematuramente el proceso, se obtienen predicciones menos adaptadas a cada soluto para lograr un ajuste perfecto. Ello da como resultado parámetros ajustados más similares entre los solutos. La idea de detener prematuramente el proceso de convergencia permite captar en mayor medida el comportamiento de retención, antes de que se alcance una adaptación completa a cada soluto.

5.5.2.1. Fijación de parámetros para desarrollar un modelo global

Como se ha mencionado, los gradientes utilizados para la elución de los extractos de manzanilla oscilaron entre el 5 y 100 % v/v de acetonitrilo, intervalo necesario para lograr la elución adecuada de todos los componentes. El modelo global desarrollado para este tipo de muestras debe ser una ecuación capaz de manejar variaciones extremas en la composición del eluyente. Una de dichas ecuaciones es la logarítmica cuadrática propuesta por Schoenmakers *et al.* (Ecuación (5.2)), que bajo ciertas condiciones se puede simplificar a la relación logarítmico lineal de Snyder (Ecuación (5.1)). Esta última ecuación puede ser útil para encontrar estimaciones iniciales de S y $\ln k_w$ para los modelos basados en la ecuación de Schoenmakers, construidos para cada soluto.

Así, el proceso de ajuste se inició ajustando la ecuación de Snyder (Ecuación (5.1)), usando los valores de $\ln k_w = 0.1$ y $S = -20$ como estimaciones iniciales, comunes para todos los solutos (adecuados para moléculas de tamaño intermedio). La ventaja de la ecuación de Snyder es que, al depender sólo de dos parámetros, la elección de las estimaciones iniciales no presenta consecuencias graves sobre la convergencia y los efectos de sobreajuste son mínimos. Además, es una ecuación muy simple que describe los efectos principales del modificador (es decir, la disminución de $\ln k$ con φ). La Tabla 5.3 incluye los parámetros ajustados y los estadísticos de regresión, de acuerdo con este tratamiento (Tratamiento A), y la Figura 5.4a muestra el gráfico de correlación entre los tiempos de retención experimental y predicho, para los 15 picos de referencia. Estos resultados deben compararse con los encontrados para los ajustes individuales a la ecuación de Schoenmakers (Tratamiento B en la Tabla 5.3 y Figura 5.4d).

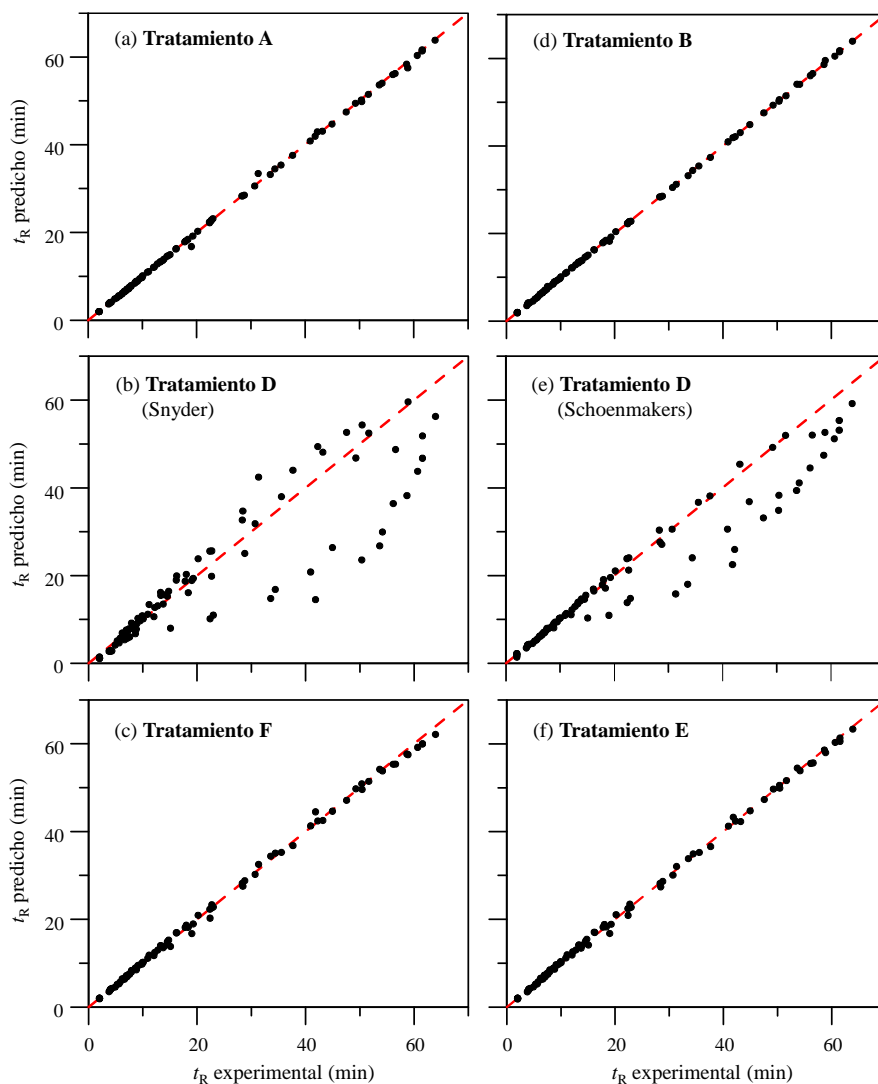


Figura 5.4. Exactitud en la predicción de tiempos de retención para los 15 picos trazables de manzanilla, obtenidos con programas de gradiente utilizando las ecuaciones de Snyder (a-c), y Schoenmakers (d-f). Las predicciones se obtuvieron: (a,d) realizando ajustes individuales para cada soluto, (b,e) reemplazando los valores individuales de S y T en la ecuación de Schoenmakers por sus medianas, y (c,f) realizando un ajuste global de ambas ecuaciones de retención empleando la regresión alterna (Secciones 5.5.2.2 y 5.5.2.3). La Tabla 5.3 proporciona datos correspondientes a los distintos tratamientos.

Es de resaltar que, a pesar de la reducción en la adaptabilidad y la gran magnitud del intervalo de modificador orgánico cubierto, la descripción proporcionada por la ecuación de Snyder es satisfactoria. Se puede observar que los valores de $\ln k_w$ para los picos de referencia cubren un intervalo muy amplio, reflejando la diversidad estructural de los componentes extraídos, mientras que los valores de S poseen cierta similitud, aunque menor que cuando los compuestos pertenecen a la misma familia. Dado que no se observa correlación alguna entre los valores de S y $\ln k_w$, no es posible proponer un modelo global más allá de reemplazar los valores individuales de S en la ecuación de Snyder por su media (o mediana).

Se decidió utilizar los valores ajustados de S y $\ln k_w$ en la ecuación de Snyder, junto con valores específicos de T (por ejemplo, -10 , 0 o 5), como estimaciones iniciales para ajustar la ecuación de Schoenmakers, utilizando el algoritmo de Powell [41]. El objetivo era averiguar qué combinación de valores iniciales de $\ln k_w$, S y T conducía a soluciones más cercanas de S y T , para todos los solutos tras alcanzar la convergencia, dando lugar finalmente a un modelo global. Los valores positivos de T generalmente condujeron a una convergencia adecuada y buenas correlaciones. Para $T = 5$, se logró una convergencia correcta para todos los solutos, en el primer intento, lo que condujo a conjuntos coherentes de parámetros óptimos (Tratamiento C), como los encontrados con el Tratamiento B para los ajustes individuales. Cabe señalar que los parámetros ajustados obtenidos con la ecuación de Schoenmakers (Tratamiento B) fueron más consistentes entre solutos que los encontrados con la ecuación de Snyder (Tratamiento A). Por lo tanto, la ecuación de Schoenmakers resultó más prometedora para derivar un modelo global.

Los resultados obtenidos con el Tratamiento B muestran cierta similitud para los valores de S y T , para los compuestos estudiados, a lo largo de la

secuencia de hidrofobicidad creciente. Esto sugiere que se podría desarrollar un procedimiento de regresión simultánea para todo el conjunto de solutos, en el que los parámetros S y T se vean forzados a ser idénticos para todos los solutos durante el proceso de ajuste. A priori, se puede perder cierta capacidad predictiva cuando dos de los tres parámetros del modelo se vean obligados a ser comunes para todos los solutos, pero esto permitiría estimar la retención de solutos de hidrofobicidad intermedia.

Una primera propuesta de modelo global, aprovechando la similitud de los parámetros S y T , es utilizar sus medianas (extendidas a todos los solutos): $S_m = -26.31$ y $T_m = 5.02$ (Tabla 5.3), junto a los valores de $\ln k_w$ encontrados en el Tratamiento B para cada soluto. Los resultados obtenidos con las medianas para los modelos de Schoenmakers (Tratamiento D) se muestran en la Figura 5.4e y la Tabla 5.3, y para los de Snyder, en la Figura 5.4b. Se puede observar que los solutos que poseen valores de S y T cercanos a las medianas se predicen con errores bajos, mientras que los solutos con valores alejados de las medianas muestran una dispersión considerable. La sustitución de los valores específicos (S_i y T_i) para cada soluto i , por las medianas, naturalmente afecta a la calidad de las predicciones.

5.5.2.2. Regresión alterna basada en la ecuación de Schoenmakers

Para mejorar los malos resultados obtenidos con el Tratamiento D (Ecuación de Schoenmakers utilizando las medianas, con $R^2 = 0.8985$, véanse la Tabla 5.3 y Figura 5.4e), se desarrolló un nuevo procedimiento de regresión que permite optimizar ambos parámetros S y T (que dependen de la columna y modificador, y que son comunes para todos los solutos), y al mismo tiempo, encuentra valores óptimos de $\ln k_{w,i}$ (específicos para cada soluto y relacionados con su hidrofobicidad). Para ello, se propone restringir en etapas separadas los

parámetros que se ajustan, atendiendo a todos los solutos y condiciones simultáneamente. En ciertas etapas, los valores de S y T son comunes y se mantienen constantes para todos los solutos, mientras que el proceso de regresión ajusta los valores de $\ln k_{w,i}$ para cada soluto. En otras etapas, los valores de S y T se modifican para que sean comunes para todos los solutos, mientras que los valores de $\ln k_{w,i}$ son constantes.

Para los 15 picos seleccionados para el extracto de manzanilla, el procedimiento consiste en optimizar un vector que contiene 17 elementos: $[S T \ln k_{w,1} \ln k_{w,2} \dots \ln k_{w,15}]$, y se comienza utilizando las medianas (S_m y T_m , véase la Sección 5.5.2.1) como estimaciones iniciales de los parámetros de la columna en el vector, mientras que los valores de $\ln k_{w,i}$ se obtienen de la regresión individual de cada soluto i . Esta regresión se puede implementar de diferentes maneras. A continuación, se describe el procedimiento que proporcionó los mejores resultados.

En general, durante el proceso de ajuste se observan grandes cambios en las primeras iteraciones, que rápidamente tienden a atenuarse para luego evolucionar más lentamente. Para acelerar el proceso de convergencia en la optimización de grandes series de datos, se llevó a cabo la siguiente secuencia:

- (i) Ajustar $\ln k_{w,i}$ independientemente (soluto a soluto), atendiendo únicamente a los datos de retención de cada soluto i , y manteniendo inalteradas las medianas de S y T .
- (ii) Ajustar los valores de S y T , obligando a que sean idénticos para todos los solutos, y manteniendo inalterados los valores de $\ln k_{w,i}$ para cada soluto, optimizados en la etapa anterior.
- (iii) Ajustar simultáneamente todos los parámetros de los solutos y columna: $[S T \ln k_{w,1} \ln k_{w,2} \dots \ln k_{w,15}]$.
- (iv) Volver a la etapa (ii), si es necesario.

Los resultados de este procedimiento (Tratamiento E) se muestran en la Tabla 5.3 y Figura 5.4f. Los valores de los estadísticos R^2 y F de Snedecor indican que la calidad del ajuste global es muy buena, al menos comparable a la basada en las correlaciones entre parámetros utilizadas en el estudio recogido en la Referencia [12], para solutos químicamente relacionados. El modelo global así encontrado ofrece una calidad predictiva muy satisfactoria, tan buena como la encontrada con los modelos individuales, y muy superior a la encontrada utilizando las medianas (Tratamiento D). Es así posible, utilizando el procedimiento de regresión restringida, separar las contribuciones de los solutos y columna ($\ln k_w$), y del modificador y columna (S y T), utilizando el modelo de Schoenmakers, con una excelente calidad predictiva.

5.5.2.3. Regresión alterna basada en la ecuación de Snyder

En este punto, se consideró interesante inspeccionar la calidad que presentaría un modelo global similar al descrito en el apartado anterior, pero basado en la ecuación de Snyder, que ha sido ampliamente utilizada en los laboratorios de rutina. El procedimiento de ajuste sería análogo al seguido para construir el modelo global con la ecuación de Schoenmakers, utilizando etapas alternas (Sección 5.5.2.2). Los resultados se muestran en la Tabla 5.3 y Figura 5.4c (Tratamiento F). Se puede observar que, aunque el modelo global de Schoenmakers ofrece un mejor rendimiento, la calidad del modelo global basado en la ecuación de Snyder es totalmente comparable, siendo más perceptible la diferencia de rendimiento con los solutos más retenidos.

Esta conclusión parece contradecir la premisa de que los gradientes que poseen un amplio intervalo requieren un modelo capaz de describir las concentraciones de modificador en grandes dominios. Debe tenerse en cuenta que modelizar un soluto en elución en gradiente sólo aprovecha intervalos

relativamente estrechos de concentración de modificador. De ahí seguramente el inesperado buen resultado del modelo global de Snyder.

5.5.3. Aplicación de los modelos globales a grandes conjuntos de solutos conocidos de diversa naturaleza eluidos isocráticamente

Una validación adecuada de los modelos global de retención requiere verificar si también es posible construir modelos equivalentes para conjuntos de solutos conocidos, con una gran diversidad química. Esto confirmaría que los modelos globales se pueden aplicar de manera fiable a otros tipos de muestras, lo que implicaría que potencialmente poseen un intervalo de aplicación mayor. Además, se debe comprobar si la excelente calidad de los modelos globales sólo es posible con datos de gradiente (cuyo intervalo de concentración efectiva para un soluto es más estrecho), o se podría extender a la elución isocrática (donde todas las composiciones del diseño experimental participan en la migración de los solutos).

Dejando de lado las consideraciones matemáticas, se puede esperar que el uso de datos de gradiente facilite la obtención de modelos globales válidos, y que la ventaja de la ecuación de Schoenmakers sea más evidente cuando se deba responder a intervalos de concentración efectivos más amplios para cada soluto. Todo ello se comprobó utilizando un conjunto de 145 compuestos de muy diversa naturaleza, cuyos datos de retención en elución isocrática fueron publicados por Smith y Burr (Tabla 5.1 y Sección 5.4.3). De esta forma, el modelo global tenía que afrontar el reto de dar respuesta a un conjunto muy amplio de solutos, muy distintos entre sí, eluidos a intervalos extremos de concentración efectiva de modificador orgánico.

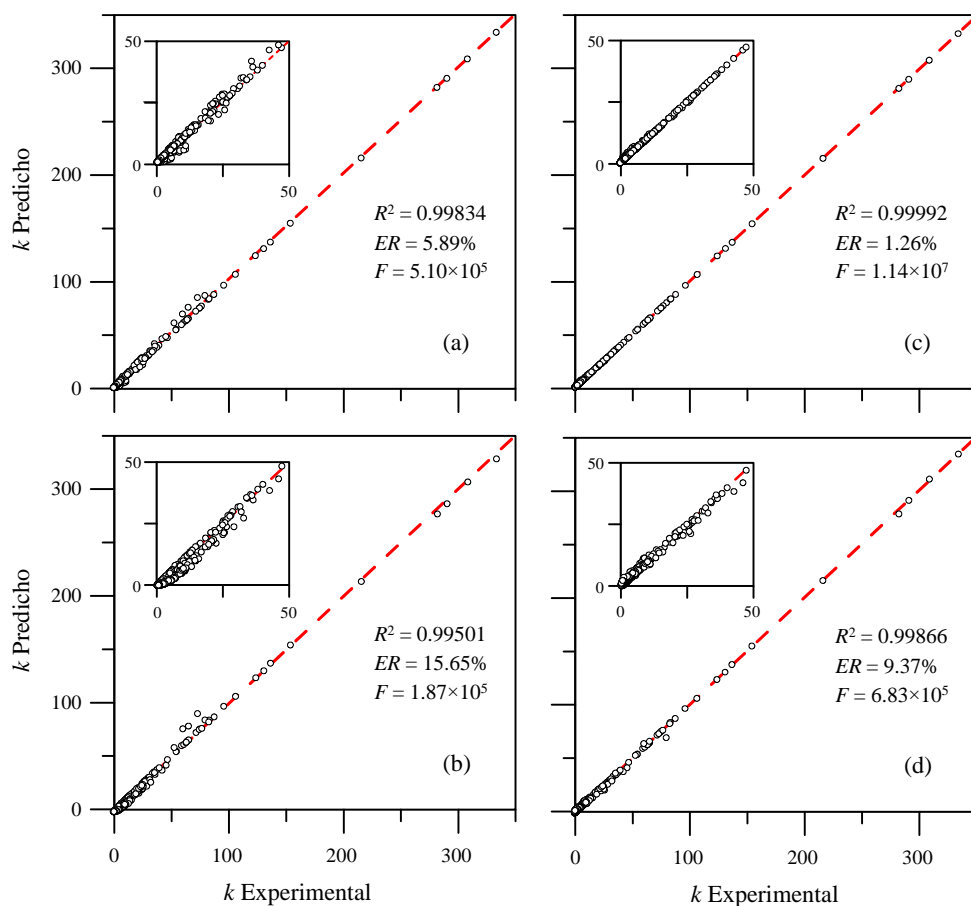


Figura 5.5. Gráficos de correlación que muestran la calidad en la predicción de los factores de retención (k), considerando 145 compuestos de diversa naturaleza eluidos isocráticamente, utilizando los modelos de: (a,b) Snyder y (c,d) Schoenmakers. Los datos se ajustaron considerando 145 modelos individuales, uno por soluto (a,c), y un modelo global único para todos los solutos (b,d).

Los datos de retención experimentales de Smith y Burr se ajustaron, soluto a soluto, a los modelos individuales de Snyder y Schoenmakers (Tratamientos A y B, respectivamente), y también forzando parámetros comunes para todos los solutos (utilizando los modelos globales de Snyder y Schoenmakers, Tratamientos F y E, respectivamente). El rendimiento predictivo se comprobó mediante la correlación entre los valores experimentales y predichos del factor de retención (véase la Figura 5.5).

Se obtuvieron las siguientes conclusiones:

- (i) El modelo de Snyder aplicado individualmente a todos los solutos (Figura 5.5a) muestra correlaciones con $R^2 = 0.9983$, un valor sorprendentemente alto considerando el amplio intervalo de concentraciones cubiertas por el diseño (30–90% *v/v* de acetonitrilo), que podría pensarse excesivo para su capacidad de modelización.
- (ii) El modelo global basado en la ecuación de Snyder (Figura 5.5b) muestra un rendimiento prácticamente idéntico ($R^2 = 0.9950$) a los modelos individuales. Este resultado también sorprende, teniendo en cuenta que el número total de parámetros en el modelo global se ha reducido prácticamente a la mitad.
- (iii) El modelo de Schoenmakers aplicado individualmente predice, con extraordinaria exactitud (Figura 5.5c), los datos experimentales para todos los solutos y fases móviles ensayados ($R^2 = 0.99992$).
- (iv) El modelo global basado en la ecuación de Schoenmakers (Figura 5.5d) muestra una capacidad predictiva algo más deficiente ($R^2 = 0.9987$), respecto a los ajustes individuales, aunque sigue siendo satisfactoria, teniendo en cuenta que dos de los tres parámetros del modelo (*S* y *T*) se han visto obligados a ser idénticos para todos los solutos, y que el número total de parámetros es casi un tercio del número de parámetros

requeridos en los modelos individuales. En comparación con el modelo global de Snyder, su rendimiento es algo superior.

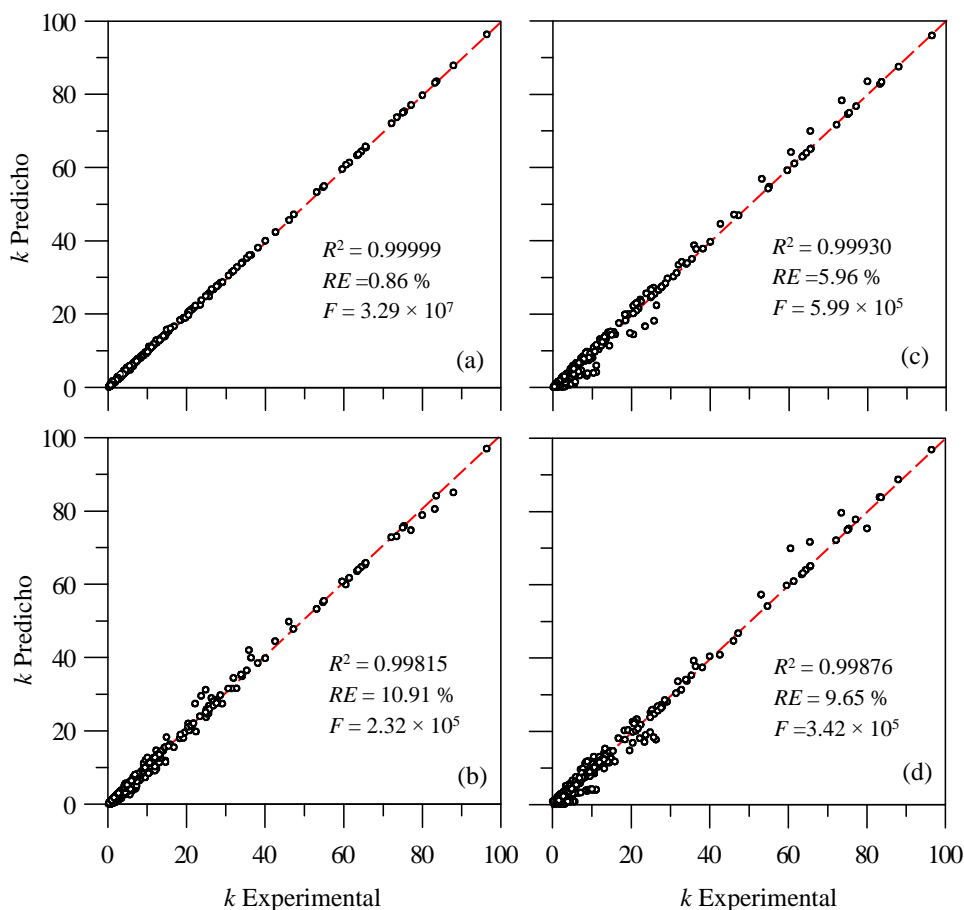


Figura 5.6. Gráficos de correlación que muestran la calidad en la predicción de los factores de retención, considerando 145 compuestos de diversa naturaleza química (Tabla 5.1), eluidos isocráticamente, para las ecuaciones de: (a,b) Neue-Kuss y (c,d) Jandera, ajustando los datos considerando: (a,c) 145 modelos individuales (uno por soluto), y (b,d) un modelo global para todos los solutos.

Seguidamente se proporcionan los gráficos de correlación para las predicciones proporcionadas por los modelos individuales y globales, utilizando los modelos NK (Figuras 5.6a y 5.6b) y de Jandera (Figuras 5.6c y 5.6d). Los parámetros específicos de los solutos para estos modelos son k_0 en la Ecuación (5.5) y m en la Ecuación (5.6). El modelo NK, cuyo ajuste individual ofreció los mejores resultados, mostró un rendimiento modesto cuando se ajustó globalmente.

5.5.4. Separación de las contribuciones polares con el modelo global de Bosch-Rosés y predicciones de $\log P_{o/w}$

En la bibliografía, se pueden encontrar algunos modelos que separan las contribuciones a la retención de los diferentes agentes responsables del proceso cromatográfico (solute, modificador y columna). Éste es el caso del modelo propuesto por Bosch y Rosés [38] (Ecuación (5.3)). En su formulación original [42], los autores desarrollaron un modelo global de la siguiente manera:

- (i) Ajustar, soluto a soluto, los valores de retención de acuerdo a:

$$\log k = q_s + p_s \left(1.00 - \frac{2.068\varphi}{1 + 1.341\varphi} \right) = q_s + p_s P_M^N \quad (5.7)$$

(véanse también las Ecuaciones (5.3) y (5.4)). La calidad de las predicciones del modelo individual de Bosch-Rosés (Figura 5.7a, $R^2 = 0.9997$) es comparable a la del modelo de Schoenmakers (Figura 5.5c, $R^2 = 0.99992$).

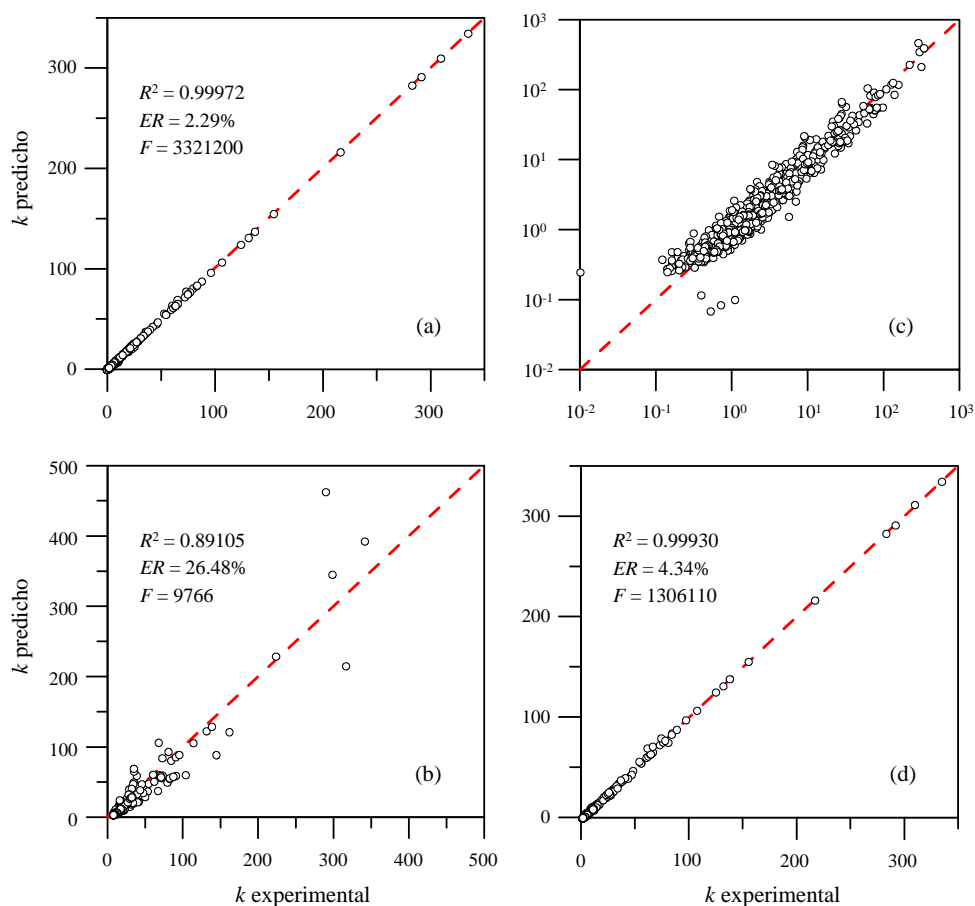


Figura 5.7. Gráficos de correlación que muestran la calidad en la predicción de los factores de retención, considerando 145 compuestos de diversa naturaleza y elución isocrática, para la: (a) Ecuación (5.7) ajustada de forma independiente (soluto a soluto), (b,c) Ecuación (5.9) ajustada globalmente con el procedimiento original propuesto por Bosch y Rosés (estableciendo una correlación entre q_s y p_s para todos los solutos, como se explica en la Sección 5.5.4, Ecuación (5.8)), y (d) Ecuación (5.9) ajustada globalmente con el tratamiento propuesto en este capítulo. Para los ajustes de Bosch-Rosés, los resultados se muestran en escala normal (b), donde los residuos son heterocedásticos debido a la transformación logarítmica, y en escala logarítmica (c), donde los residuos son homocedásticos. Como los valores de k experimentales y predichos son idénticos para (b) y (c), las estadísticas son idénticas para ambas figuras.

- (ii) Correlacionar los parámetros q_s y p_s , obtenidos del ajuste de la Ecuación (5.7) para el conjunto de solutos, según la siguiente relación lineal:

$$q_s = a + b p_s \quad (5.8)$$

- (iii) Sustituir la Ecuación (5.8) en la Ecuación (5.7), para obtener el modelo global:

$$\log k = a + b p_s + p_s P_M^N = a + p_s (P_M^N + b) = (\log k)_0 + p_s (P_M^N - P_S^N) \quad (5.9)$$

Debe indicarse que el algoritmo original de Bosch-Rosés se ve afectado de dos problemas:

- (i) el sesgo observado para los valores de p_s , obtenidos del ajuste de la Ecuación (5.7), que son sensibles a la naturaleza particular de los solutos, y
(ii) las pobres correlaciones obtenidas para la Ecuación (5.8).

Debido a ello, las predicciones de retención son excesivamente pobres (véanse las Figuras 5.7b y 5.7c, donde se muestra la misma correlación en escala normal y logarítmica, respectivamente). A pesar de las modestas predicciones obtenidas, el modelo de Bosch-Rosés resulta de gran interés para la transferencia de datos de retención a otras columnas o modificadores, o para la interpretación de mecanismos de retención.

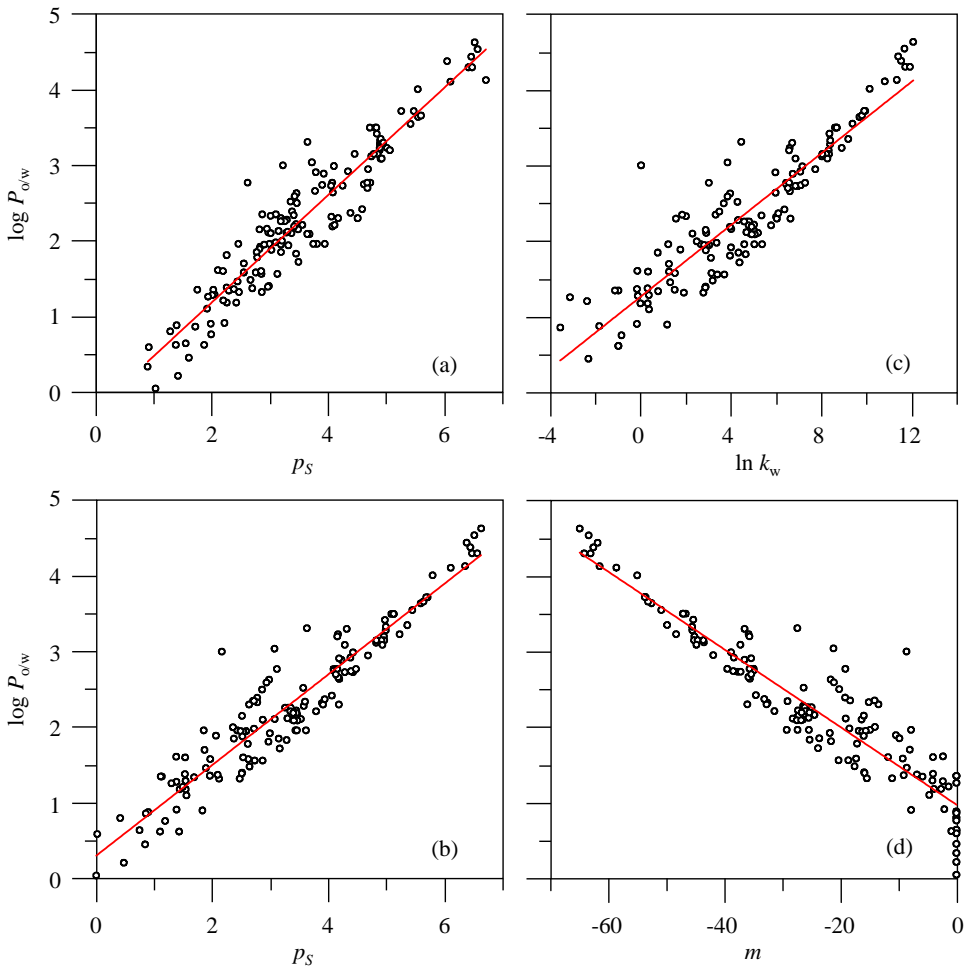


Figura 5.8. Gráficos de correlación del logaritmo del coeficiente de reparto octanol-agua ($\log P_{o/w}$) vs. el parámetro individual que describe la retención asociada a cada soluto, para el conjunto de 145 compuestos de diversa naturaleza, eluidos isocráticamente, según: (a) el procedimiento original de Bosch-Rosés, y (b,c,d) tras aplicar el procedimiento de regresión alterna propuesto. Se utilizaron las ecuaciones de (se indica el parámetro del soluto individual): (a,b) Bosch-Rosés (p_s), (c) Schoenmakers ($\ln k_w$), y (d) Jandera (m).

Se consideró que los resultados obtenidos de la ecuación de Bosch-Rosés podrían mejorarse, aplicando el procedimiento de regresión alterna a la construcción de modelos globales basados en dicha ecuación. La Figura 5.7d muestra la calidad de los resultados que, como puede observarse, son ligeramente mejores ($R^2 = 0.9993$) que los que ofrece el modelo global de Schoenmakers ($R^2 = 0.9987$). En elución isocrática estos resultados son excelentes, pero hay que tener en cuenta que la ecuación de Bosch-Rosés carece de función primitiva para ser aplicada en elución en gradiente, haciendo uso de métodos algebraicos.

Una aplicación interesante de la ecuación de Bosch-Rosés es la predicción del logaritmo del coeficiente de reparto octanol-agua, a partir de correlaciones de $\log P_{o/w}$ vs. p_s . Las Figuras 5.8a y 5.8b muestran dichas correlaciones, obtenidas según el procedimiento original de Bosch-Rosés, y aplicando el procedimiento de regresión alterna, respectivamente. Se proporcionan también otras dos predicciones equivalentes de $\log P_{o/w}$, usando como predictores los valores de $\ln k_{w,i}$ del modelo global de Schoenmakers (Figura 5.8c), y los valores de m_i del modelo global de Jandera (Figuras 5.8d).

5.5.5. Simulación de huellas dactilares cromatográficas utilizando modelos globales

Los modelos globales de retención, basados en parámetros comunes, permiten la predicción de cromatogramas tanto en elución isocrática como en gradiente, siendo aplicables a muestras con cientos de componentes. A continuación, se detalla el protocolo seguido en la simulación de cromatogramas:

- (i) Ejecutar un conjunto de 4 a 6 gradientes, el primero de los cuales (gradiente de exploración) se utiliza para establecer los márgenes operativos de concentración del modificador.
- (ii) Identificar los candidatos a picos de referencia trazables en los cromatogramas obtenidos para todo el diseño experimental y medir sus propiedades (tiempo de retención, semianchuras y área de pico).
- (iii) Construir el modelo global para los picos de referencia, ajustando los datos de retención para los gradientes del diseño experimental (en el ejemplo de la manzanilla, 15×6 picos considerando los seis gradientes ensayados), siguiendo el procedimiento descrito en la Sección 5.5.2.2. Para este propósito, se recomienda el modelo global de Schoenmakers, pero también se pueden utilizar otros modelos, siempre que los parámetros se puedan clasificar como dependientes del soluto/columna o modificador/columna. De esta forma, se obtienen los parámetros globales (S y T) y los parámetros individuales ($\ln k_{w,i}$) de los 15 picos de referencia.
- (iv) Desarrollar los modelos globales para las semianchuras (tendencias parabólicas de A y B , en función del tiempo de retención isocrático, consúltese la Sección 5.3.2), utilizando las medidas de gradiente experimentales y la aproximación de Jandera [30]. En este capítulo, se pretende llevar a cabo predicciones de gradientes arbitrarios que involucren los 15 picos de referencia, pero el objetivo final es la predicción de un cromatograma completo con cientos de componentes (que es el objetivo del Capítulo 7).
- (v) Hacer uso de un método integral para la sustracción de la línea base, tal como el algoritmo BEADS [43], para preprocesar el cromatograma obtenido con el gradiente más lento del diseño experimental (Gradiente

G3 en la Figura 5.1). Para medir las propiedades de todos los picos visibles, se ha utilizado un programa de detección automática de picos (Chromscan, desarrollado en nuestro laboratorio [24]). Los parámetros utilizados son la posición, área y tiempos de inicio y final, para cada pico en el cromatograma.

- (vi) Utilizar la posición (t_R) de cada pico detectado en el gradiente más lento (etapa (v)), y los parámetros globales S y T (encontrados en la etapa (iii)), para obtener los valores de $\ln k_{w,i}$ para todos los picos presentes en el cromatograma, sean o no picos de referencia, considerando todos los gradientes de interés (véase también el procedimiento descrito en la Sección 5.5.2.2, etapa (ii)). El cromatograma para el gradiente más lento (G3) incluía 279 picos (Figura 5.2). En consecuencia, el modelo global de la huella dactilar cromatográfica será el vector $[S \ T \ \ln k_{w,1} \ \ln k_{w,2} \ \dots \ \ln k_{w,279}]$. Es conveniente realizar estos ajustes de forma ordenada, siguiendo la secuencia de tiempos de retención crecientes, de modo que el valor de $\ln k_{w,i}$ encontrado para el soluto anterior se utilice como estimación inicial para el siguiente soluto.
- (vii) El vector resultante (conteniendo los parámetros del modelo global) puede utilizarse para construir los modelos individuales para los n picos detectados en el cromatograma, reorganizando el modelo global en una colección de vectores $[S \ T \ \ln k_{w,i}]$, que describan cada uno de los 279 componentes del cromatograma. Esto permite predecir la huella dactilar cromatográfica para cualquier otro gradiente. Si en la etapa (vi) se utiliza cualquier otro gradiente (en lugar del más lento), se obtendrían predicciones menos exactas para aquellos gradientes más lentos que el seleccionado, ya que implicarían llevar a cabo una extrapolación de la retención.

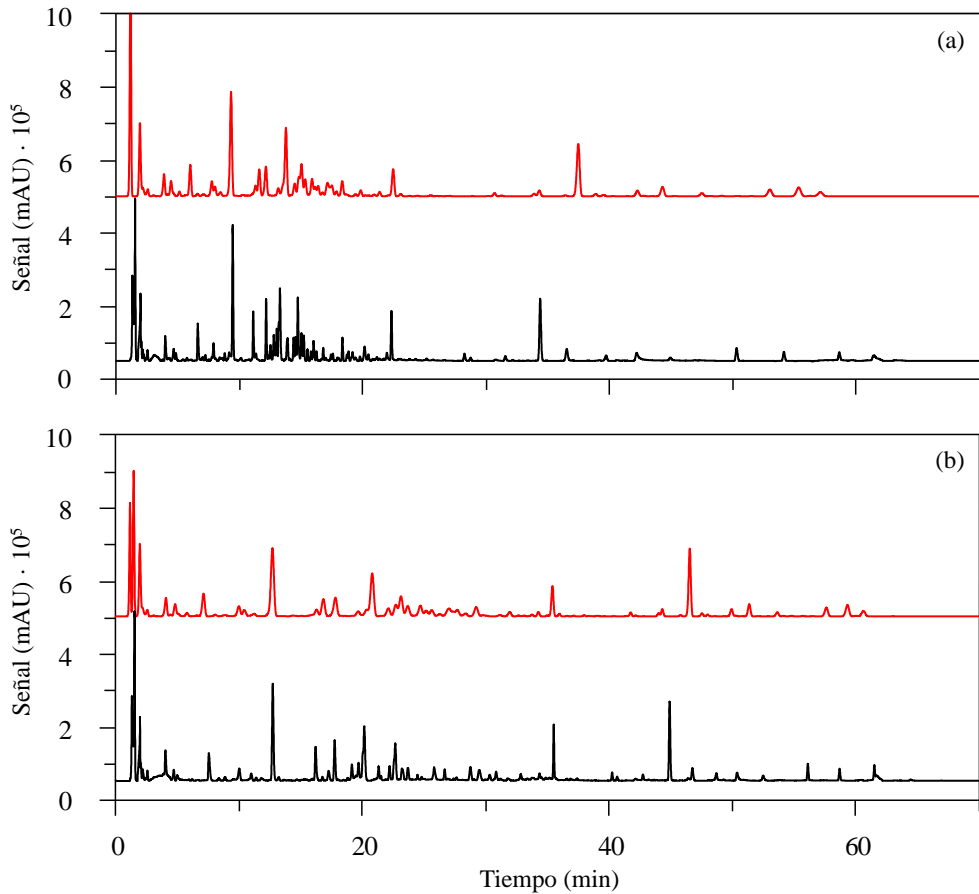


Figura 5.9. Cromatogramas predichos (rojo) y experimentales (negro), obtenidos utilizando el gradiente más lento del diseño (G3), como condición experimental de origen. Una vez medidos los tiempos de retención de los picos minoritarios en G3, se calculan sus respectivos valores de $\ln k_{w,i}$, manteniendo los valores S y T encontrados previamente con los seis cromatogramas y 15 picos de referencia (etapa (iii) en la Sección 5.5.5). Se muestran los cromatogramas para los otros gradientes del diseño experimental: (a) G1, (b) G2, (c) G4, (d) G5 y (e) G6 (véase la Figura 5.1).

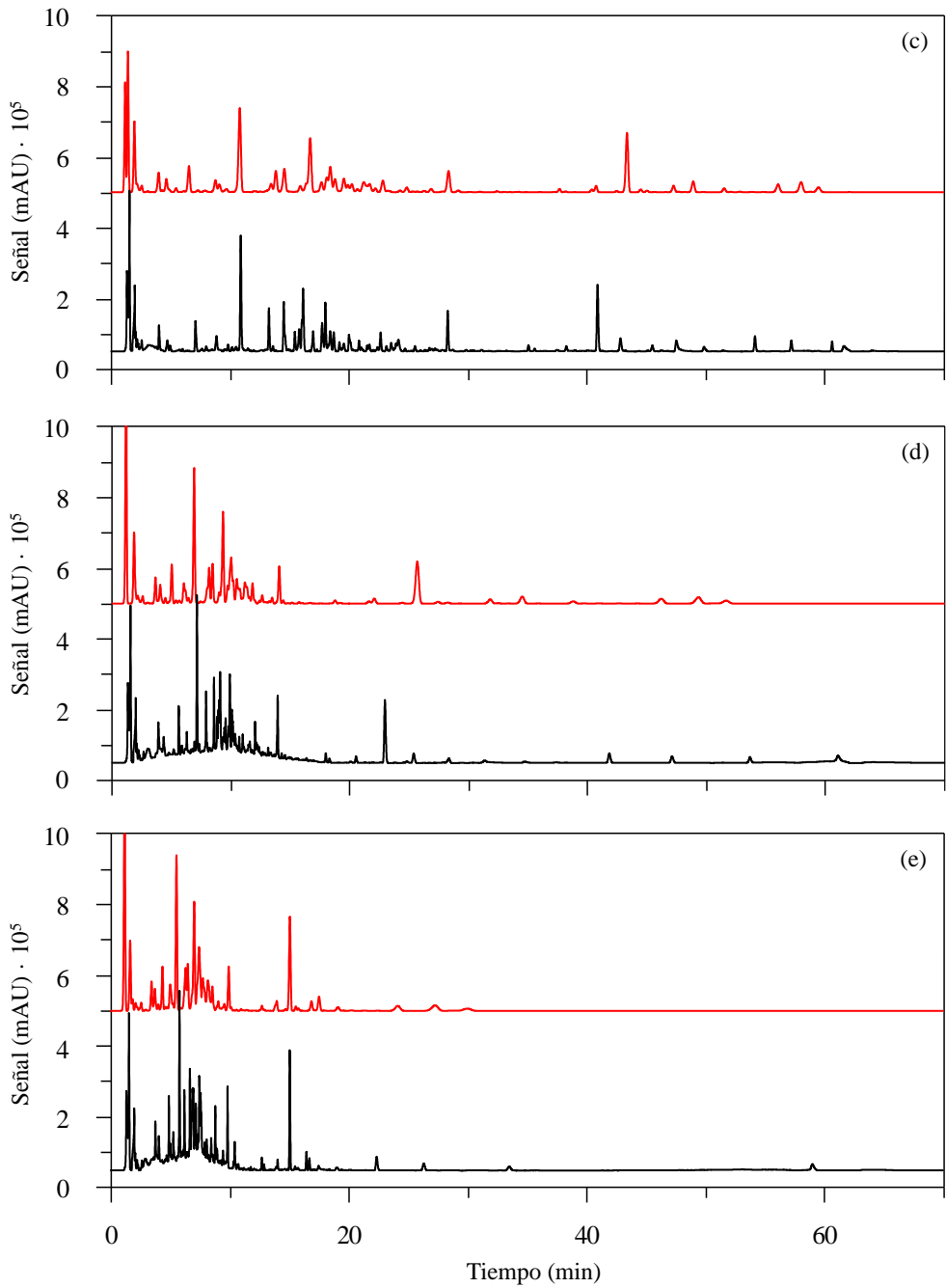


Figura 5.9 (continuación).

Como se ha comentado, la Figura 5.2 muestra el cromatograma experimental obtenido con el Gradiente G3, que es el más lento del diseño experimental (Figura 5.1). Las Figuras 5.9a – 5.9e muestran los cromatogramas experimentales (negro) y predichos (rojo) para los otros gradientes de entrenamiento del diseño (G1, G2, G4, G5 y G6), haciendo uso del modelo global basado en el método de regresión alterna para realizar las predicciones. Dado que los gradientes son multi-lineales y la sección final del Gradiente G3 presenta la pendiente más elevada, la incertidumbre de las predicciones será mayor al final del gradiente.

5.6. Conclusiones

Los modelos globales de retención representan una alternativa al uso de los modelos individuales para describir la retención en LC. En los modelos globales propuestos, algunos parámetros son específicos de cada soluto, mientras que otros parámetros describen los efectos generales de los cambios de modificador sobre la retención en la columna. Aunque los modelos individuales ofrecen descripciones más exactas (atendiendo a las singularidades de cada soluto), sólo son aplicables a los solutos con los que se obtuvieron los modelos.

Uno de los objetivos de este trabajo ha sido desarrollar un procedimiento que incluye una regresión particular, que permite el ajuste simultáneo de los datos de varios solutos a una ecuación de retención común, con el fin de describir el comportamiento de conjuntos de compuestos. Se presentan modelos globales de retención basados en varias ecuaciones, propuestas por Snyder, Schoenmakers, Neue y Kuss, Jandera y Bosch y Rosés.

Tanto el modelo global como el procedimiento de ajuste necesario para derivarlo se probaron en dos casos de estudio:

- (i) huellas dactilares cromatográficas (que implican a cientos de compuestos desconocidos, sin estándares disponibles), obtenidas con gradientes multi-lineales, y
- (ii) un conjunto de 145 compuestos con estructuras químicas y funcionalidades muy diferentes, eluidos isocráticamente.

Con las huellas dactilares cromatográficas, el procedimiento se inicia con la selección de picos característicos (picos de referencia), de tal manera que se puedan rastrear sin ambigüedades en las diferentes condiciones contempladas en el diseño experimental. En este estudio, se han seleccionado 15 picos de referencia, pero puede ser cualquier otro número. La función de los picos de referencia es la determinación de parámetros columna-modificador (tal como S y T en la Ecuación (5.2), o a y b en la Ecuación (5.6)), comunes para todos los solutos. El procedimiento de regresión utilizado permite ajustar estos parámetros y, al mismo tiempo, encontrar los parámetros específicos (por ejemplo, $\ln k_{w,i}$ en la Ecuación (5.2), o m en la Ecuación (5.6)), para cada soluto. La regresión restringida realiza el ajuste en etapas consecutivas (Sección 5.5.2.2), en las que se ajustan alternadamente los parámetros comunes y específicos, ofreciendo un modelo global con una calidad predictiva satisfactoria, comparable a la encontrada con los modelos obtenidos para los solutos individuales, utilizando las cinco ecuaciones de retención ensayadas. Tras establecer los parámetros columna-modificador, se obtienen los parámetros específicos de los solutos para aquellos compuestos contenidos en la huella dactilar que no participan en la etapa anterior. Como consecuencia, el modelo global predice no sólo los picos de referencia, sino la huella en su totalidad.

Las experiencias isocráticas obligan a una participación completa de cada fase móvil en la construcción de modelos de retención, mientras que los

gradientes se benefician de ventanas de modificador más estrechas. En consecuencia, las predicciones son más exactas en elución en gradiente y las ecuaciones necesarias son más simples. Estas ecuaciones suelen funcionar peor en las predicciones isocráticas, como es el caso de la ecuación lineal logarítmica, pero muestran un rendimiento excelente en las predicciones de gradientes. Esto también sucede con los modelos globales, con la ventaja adicional de que el mismo modelo cumple con los requisitos de compuestos de naturaleza hidrofílica e hidrofóbica.

El estudio muestra, al menos, dos interesantes aplicaciones:

- (i) La predicción de cromatogramas de gran complejidad, para diversas condiciones experimentales en elución isocrática y en gradiente. Ésta es una característica clave que abre la posibilidad de simular y optimizar cromatogramas muy complejos, tales como las huellas dactilares que se muestran en este trabajo.
- (ii) La predicción de propiedades moleculares, como $\log P_{o/w}$, a partir de los parámetros específicos de los solutos en los modelos globales.
- (iii) Otra aplicación de interés, no mostrada en este trabajo, es la evaluación de la capacidad de pico.

El rendimiento de los ajustes individuales fue excelente, especialmente para las ecuaciones de Neue-Kuss y Schoenmakers, con $R^2 > 0.9999$, siendo peor para la ecuación de Snyder ($R^2 = 0.9983$). Los cinco modelos globales de retención ensayados en este trabajo muestran menores diferencias de rendimiento que los modelos individuales, y una capacidad de predicción similar o mejor que la ofrecida por la ecuación de Snyder, cuando se ajusta soluto a soluto. No hay motivo para pensar que otras ecuaciones que modelicen el efecto del modificador orgánico sobre la retención no puedan adaptarse y

ajustarse mediante el procedimiento de regresión propuesto en este trabajo, dando lugar a un modelo global válido.

Una particularidad de la ecuación de Bosch-Rosés es que permite correlaciones aceptables entre $\log P_{o/w}$ y p_s . Se pueden establecer correlaciones de calidad similar para otros modelos globales de retención, usando como predictores los parámetros específicos de los solutos en el modelo global (por ejemplo, $\ln k_{w,i}$ o m en las Ecuaciones (5.2) y (5.6), respectivamente).

El objetivo final de este trabajo ha sido la simulación de huellas dactilares cromatográficas. En la metodología propuesta se construye, en primer lugar, un modelo global (por ejemplo, el basado en la Ecuación (5.2)), utilizando la información obtenida de picos de referencia a partir de un diseño experimental. Esto permite encontrar los parámetros comunes para todos los solutos en la muestra (por ejemplo, S y T en la Ecuación (5.2)). A continuación, se determinan los parámetros específicos de los solutos ($\ln k_{w,i}$), para un número elevado de componentes en las huellas dactilares, utilizando los tiempos de retención experimentales encontrados en el cromatograma más lento del diseño experimental y los parámetros comunes asociados a la columna y modificador encontrados previamente. Esta información completa el modelo global para todos los picos del cromatograma, incluidos aquéllos que no participan en el cálculo de los parámetros columna-modificador (por ejemplo, $[S T \ln k_{w,1} \dots \ln k_{w,n}]$ para la Ecuación (5.2); donde para la manzanilla $n = 279$). El modelo global final permite predecir con éxito el cromatograma para cualquier gradiente arbitrario. Todos estos resultados indican que los modelos globales pueden constituir una vía prometedora para afrontar la optimización de problemas de separación no dirigidos (*untargeted*). Además, son interesantes para la clasificación de compuestos desconocidos relacionados estructu-

ralmente, pero también no relacionados o desconocidos [17]. También pueden ser útiles para caracterizar columnas [19].

Cabe señalar, sin embargo, que los modelos globales tienen limitaciones:

- (i) Para compuestos ionizables, los cambios de pH debidos a la variación de la concentración del modificador orgánico, a lo largo del gradiente, pueden afectar a la retención [44]. Por lo tanto, solutos con valores de pK_a de acidez cercanos al pH del tampón pueden sufrir desviaciones en sus predicciones, como consecuencia del incremento de modificador orgánico a lo largo del gradiente. Sin embargo, la comparación de los cromatogramas predichos y experimentales en el ejemplo de la huella dactilar sugiere que el número de solutos afectados incidentalmente por tales cambios de pK_a debe ser minoritario. Además, la inclusión del pH como factor implicaría un incremento importante del esfuerzo experimental, lo que es poco práctico.
- (ii) La selectividad específica de los solutos individuales y la inversión de la posición de los picos no puede atenderse con precisión.

Sin embargo, los modelos globales pueden predecir la distribución de los picos en un cromatograma. Esto podría ser interesante como base para el desarrollo de una función objetivo que oriente la optimización de la resolución cromatográfica, cuando el número de picos hace imposible atender a la selectividad específica de cada soluto.

Las medidas basadas en modelos globales son potencialmente más informativas que la capacidad de pico, debido a que permiten la predicción de la distribución de los picos y del desorden interno en ciertas regiones del cromatograma, a diferencia de la medida de la capacidad de pico. Cuando no hay estándares disponibles, la posibilidad de fortificar un componente dado

para establecer la identidad de un pico no es factible, y las posibilidades de alcanzar el éxito en una optimización, son muy reducidas.

Las huellas dactilares cromatográficas representan un caso extremo de cromatogramas complejos con compuestos de identidad desconocida. Cabe señalar que los modelos individuales son inviables para muestras con varios cientos de componentes, y un modelo basado en la información obtenida de estándares para cada compuesto no es práctico.

5.7. Referencias

- [1] M.C. García Álvarez-Coque, J.J. Baeza Baeza, G. Ramis Ramos, Reversed phase liquid chromatography, en *Analytical Separation Science Series* (J.L. Anderson, A. Stalcup, A. Berthod, V. Pino), Vol. 1, Wiley-VCH, Nueva York, 2015, págs. 159–197.
- [2] R. Cela, C.G. Barroso, J.A. Pérez-Bustamante, Objective functions in experimental and simulated chromatographic optimization: Comparative study and alternative proposal, *J. Chromatogr. A* 485 (1989) 477–500.
- [3] L.R. Snyder, The future of chromatography: Optimizing the separation, *Analyst* 116 (1991) 1237–1244.
- [4] P.H. Lukulay, V.L. McGuffin, Evolution from univariate to multivariate optimization methods in liquid chromatography, *J. Microcolumn Sep.* 8 (1996) 211–224.
- [5] A.M. Siouffi, R. Phan-Tan-Luu, Optimization of methods in chromatography and capillary electrophoresis, *J. Chromatogr. A* 892 (2000) 75–106.
- [6] V. Harang, A. Karlsson, M. Josefson, Liquid chromatography method development and optimization by statistical experimental design and chromatogram simulations, *Chromatographia* 54 (2001) 703–709.

-
- [7] M.C. García Álvarez-Coque, J.R. Torres Lapasió, J.J. Baeza Baeza, Models and objective functions for the optimisation of selectivity in reversed-phase liquid chromatography, *Anal. Chim. Acta* 579 (2006) 125–145.
- [8] J.R. Torres Lapasió, M.C. García Álvarez-Coque, Levels in the interpretive optimisation of selectivity in high-performance liquid chromatography: A magical mystery tour, *J. Chromatogr. A* 1120 (2006) 308–321.
- [9] P. Nikitas, A. Pappa-Louisi, Retention models for isocratic and gradient elution in reversed-phase liquid chromatography, *J. Chromatogr. A* 1216 (2009) 1737–1755.
- [10] E. Tyteca, A. Périat, S. Rudaz, G. Desmet, D. Guillarme, Retention modeling and method development in hydrophilic interaction chromatography, *J. Chromatogr. A* 1337 (2014) 116–127.
- [11] M.R. Euerby, J. Hulse, P. Petersson, A. Vazhentsev, K. Kassam, Retention modelling in hydrophilic interaction chromatography, *Anal. Bioanal. Chem.* 407 (2015) 9135–9152.
- [12] J.A. Navarro Huerta, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Estimation of peak capacity based on peak simulation, *J. Chromatogr. A* 1574 (2018) 101–113.
- [13] R.M. López Marques, P.J. Schoenmakers, Modelling retention in reversed-phase liquid chromatography as a function of pH and solvent composition, *J. Chromatogr. A* 592 (1992) 157–182.
- [14] P.J. Schoenmakers, H.A.H. Billiet, R. Tussen, L. de Galan, Gradient selection in reversed-phase liquid chromatography, *J. Chromatogr. A* 149 (1978) 519–537.
-

- [15] L.R. Snyder, J.J. Kirkland, J.L. Glajch, *Practical HPLC Method Development*, 2^a ed., John Wiley & Sons, Nueva York, 1997.
- [16] K. Valkó, L.R. Snyder, J.L. Glajch, Retention in reversed-phase liquid chromatography as a function of mobile-phase composition, *J. Chromatogr. A* 656 (1993) 501–520.
- [17] H.B. Xiao, X.M. Liang, R.C. Lu, Classification of structurally related compounds from Astragalus extract by correlation of the log k_w and S, *Chromatographia* 51 (2000) 212–220.
- [18] C.F. Poole, *The Essence of Chromatography*, Elsevier, Amsterdam, 2003, págs. 303–304.
- [19] J.A. Navarro Huerta, E.J. Carrasco Correa, J.R. Torres Lapasió, J.M. Herrero Martínez, M.C. García Álvarez-Coque, Modelling retention and peak shape of small polar solutes analysed by nano-HPLC using methacrylate monolithic columns, *Anal. Chim. Acta* 1086 (2019) 142–155.
- [20] P. Xie, S. Chen, Y. Liang, X. Zeng, X. Wang, R. Tian, R. Upton, Chromatographic fingerprint analysis: A rational approach for quality assessment of traditional chinese plantal medicine, *J. Chromatogr. A* 1112 (2006) 171–180.
- [21] G. Jin, X. Xue, F. Zhang, Y. Jin, X. Liang, Computer-aided target optimization for traditional chinese medicine by ultra-performance liquid chromatography, *Talanta* 78 (2009) 278–283.
- [22] C. Tistaert, B. Dejaegher, Y. Vander Heyden, Chromatographic separation techniques and data handling methods for herbal fingerprints, *Anal. Chim. Acta* 690 (2011) 148–161.
- [23] G. Alaerts, S. Pieters, H. Logie, J. Van Erps, M. Merino-Arévalo, B. Dejaegher, J. Smeyers-Verbeke, Y. Vander Heyden, Exploration and

- classification of chromatographic fingerprints as additional tool for identification and quality control of several artemisia species, *J. Pharm. Biomed. Anal.* 95 (2014) 34–46.
- [24] T. Álvarez Segura, A. Gómez Díaz, C. Ortiz Bolsico, J.R. Torres Lapasió, M.C. García Álvarez-Coque, A chromatographic objective function to characterise chromatograms with unknown compounds or without standards available, *J. Chromatogr. A* 1409 (2015) 79–88.
- [25] T. Álvarez Segura, E. Cabo Calvet, J.R. Torres Lapasió, M.C. García Álvarez-Coque, An approach to evaluate the information in chromatographic fingerprints: Application to the optimisation of the extraction and conservation conditions of medicinal herbs, *J. Chromatogr. A* 1422 (2015) 178–185.
- [26] J.A. Navarro Huerta, T. Álvarez Segura, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Study of the performance of a resolution criterion to characterise complex chromatograms with unknowns or without standards, *Anal. Methods* 9 (2017) 4293–4303.
- [27] U.D. Neue, H.J. Kuss, Improved reversed-phase gradient retention modeling, *J. Chromatogr. A* 1217 (2010) 3794–3803.
- [28] P. Jandera, T. Hájek, M. Ružičková, Retention models on core-shell columns, *J. AOAC International* 100 (2017) 1636–1646.
- [29] E. Bosch, P. Bou, M. Rosés, Linear description of solute retention in reversed-phase liquid chromatography by a new mobile phase polarity parameter, *Anal. Chim. Acta* 299 (1994) 219–229.
- [30] P. Jandera, Predictive calculation methods for optimization of gradient elution using binary and ternary solvent gradients, *J. Chromatogr.* 485 (1989) 113–141.

- [31] S. López Ureña, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Enhancement in the computation of gradient retention times in liquid chromatography using root-finding methods, *J. Chromatogr. A* 1600 (2019) 137–147.
- [32] R.M. Smith, C.M. Burr, Retention prediction of analytes in reversed-phase high-performance liquid chromatography based on molecular structure: I. Monosubstituted aromatic compounds, *J. Chromatogr.* 475 (1989) 57–74.
- [33] R.M. Smith, C.M. Burr, Retention prediction of analytes in reversed-phase high-performance liquid chromatography based on molecular structure: II. Long term reproducibility of capacity factors and retention indices, *J. Chromatogr.* 475 (1989) 75–83.
- [34] R.M. Smith, C.M. Burr, Retention prediction of analytes in reversed-phase high-performance liquid chromatography based on molecular structure: III. Monosubstituted aliphatic compounds, *J. Chromatogr.* 481 (1989) 71–84.
- [35] R.M. Smith, C.M. Burr, Retention prediction of analytes in reversed-phase high-performance liquid chromatography based on molecular structure: IV. Branched and unsaturated alkylbenzenes, *J. Chromatogr.* 481 (1989) 85–95.
- [36] R.M. Smith, C.M. Burr, Retention prediction of analytes in reversed-phase high-performance liquid chromatography based on molecular structure: V. Cripes (chromatographic retention index prediction expert system), *J. Chromatogr.* 485 (1989) 325–340.
- [37] R.M. Smith, C.M. Burr, Retention prediction of analytes in reversed-phase high-performance liquid chromatography based on molecular

- structure: VI. Disubstituted aromatic compounds, *J. Chromatogr.* 550 (1991) 335–356.
- [38] J.R. Torres Lapasió, M.C. García Álvarez-Coque, M. Rosés, E. Bosch, Prediction of the retention in reversed-phase liquid chromatography using solute-mobile phase-stationary phase polarity parameters, *J. Chromatogr. A* 955 (2002) 19–34.
- [39] R. Bosque, J. Sales, E. Bosch, M. Rosés, M.C. García Álvarez-Coque, J.R. Torres Lapasió, A QSPR study of the p solute polarity parameter to estimate retention in HPLC, *J. Chem. Inform. Comput. Sci.* 43 (2003) 1240–1247.
- [40] J.R. Torres Lapasió, M.C. García Álvarez-Coque, M. Rosés, E. Bosch, A.M. Zissimos, M.H. Abraham, Analysis of a solute polarity parameter in reversed-phase liquid chromatography on a linear solvation relationship basis, *Anal. Chim. Acta* 515 (2004) 209–227.
- [41] M.J.D. Powell, An efficient method for finding the minimum of a function of several variables without calculating derivatives, *Computer J.* 7 (1964) 155–162.
- [42] E. Bosch, P. Bou, M. Rosés, Linear description of solute retention in reversed-phase liquid chromatography by a new mobile phase polarity parameter, *Anal. Chim. Acta* 299 (1994) 219–229.
- [43] J.A. Navarro Huerta, J.R. Torres Lapasió, S. López Ureña, M.C. García Álvarez-Coque, Assisted baseline subtraction in complex chromatograms using the BEADS algorithm, *J. Chromatogr. A* 1507 (2017) 1–10.
- [44] A. Andrés, A. Téllez, M. Rosés, E. Bosch, Chromatographic models to predict the elution of ionizable analytes by organic modifier gradient in reversed phase liquid chromatography, *J. Chromatogr. A* 1247 (2012) 71–80.

CAPÍTULO 6

**ANÁLISIS DE HUELLAS DACTILARES
CROMATOGRÁFICAS DE EXTRACTOS DE
TÉ VERDE, MELISA Y TILA:
DESARROLLO DE MODELOS GLOBALES DE RETENCIÓN
SIN EL USO DE ESTÁNDARES**

6.1. Resumen

Se presenta una metodología que permite obtener modelos globales, capaces de describir, simultáneamente, el comportamiento de retención de varios componentes de una muestra. Los modelos globales ensayados incluyen parámetros que modelizan los efectos generales de la columna y el modificador orgánico sobre la retención, que son comunes para todos los componentes, mientras que otros parámetros son específicos para cada componente de la muestra. Los modelos se ajustan mediante regresión alterna y ofrecen una calidad en las predicciones comparable a la ofrecida por los modelos de retención individuales. El enfoque es adecuado para muestras de productos naturales que incluyen una gran cantidad de componentes, en concentraciones extremadamente diversas y en ausencia de estándares. Se proporcionan pautas para el desarrollo de diseños experimentales constituidos por gradientes multilineales, orientados al análisis de muestras complejas como las obtenidas de productos naturales. Los diseños propuestos también facilitan la trazabilidad de picos para diversas condiciones experimentales. Se ha encontrado que el modelo de retención propuesto por Neue y Kuss proporciona las mejores predicciones. La metodología propuesta se ha aplicado a los extractos de muestras de té verde, melisa y tila, y ofrece excelentes predicciones de la retención para los componentes seleccionados.

6.2. Introducción

Las estrategias de optimización interpretativa son capaces de encontrar las mejores condiciones de separación para cualquier tipo de muestra, de forma fiable y eficaz [1], ya sea utilizando modelos [2] o estableciendo una optimización jerárquica [3]. En estas estrategias, el primer paso consiste en recopilar información experimental (generalmente tiempos de retención y perfiles de pico) sobre el comportamiento de los constituyentes de la muestra. En las situaciones más favorables, se conoce la identidad de los componentes y se dispone de estándares para todos ellos. Esto permite desarrollar modelos de calidad para cada componente de la muestra. Sin embargo, algunas muestras son demasiado complejas para permitir un número suficientemente pequeño de inyecciones, o no existen estándares para todos los componentes. Además, se desconoce la identidad de muchos de ellos [4,5]. Por lo tanto, no se pueden construir modelos de retención individuales para cada componente, y el comportamiento debe inferirse de alguna manera a partir de los cromatogramas de las muestras.

Para optimizar la separación, Jin *et al.* propusieron la predicción de la retención de un grupo de compuestos desconocidos en huellas dactilares cromatográficas de plantas medicinales, basándose en una ecuación lineal con sólo dos parámetros [6]. Para construir los modelos, los autores seleccionaron 24 picos de los cromatogramas resultantes de la elución de extractos de *Rhizoma Corydalis*, usada en la medicina tradicional china, obtenidos mediante la elución con cinco gradientes lineales.

El estudio que se presenta aquí propone también el uso de la información de picos seleccionados, pero utilizando un modelo global de retención (en lugar de modelos individuales) para predecir la retención (ver Capítulo 5). Ello permitirá extender el modelo global obtenido para los compuestos seleccionados, a todos

los componentes de la muestra. Los modelos de retención individuales ofrecen predicciones de retención más perfectas que los modelos globales. En los modelos globales, algunos parámetros describen los efectos generales de la columna y modificador orgánico sobre la retención, que son comunes para todos los componentes, mientras que otros parámetros son específicos para cada componente [7]. Por lo tanto, los modelos globales de retención ofrecen tendencias válidas para varios componentes, a costa de una cierta pérdida en la predicción de la retención. Sin embargo, si se ajustan correctamente, los resultados pueden ser útiles para aplicaciones prácticas. Se pueden desarrollar modelos globales para componentes muy diversos, en la medida en que sean susceptibles al modo de separación y se pueda sacrificar cierta exactitud. Dichos modelos pueden ser de interés para la optimización de cromatogramas con un alto número de componentes desconocidos.

En la bibliografía, existen otras propuestas para construir modelos globales de retención, como la realizada por Bosch *et al.* [8], basada en descriptores de polaridad (véase la Sección 5.3.1.3 en el Capítulo 5). Este modelo es particularmente interesante para transferir datos de retención entre columnas o modificadores [9]. Otro ejemplo es el establecimiento de correlaciones entre las pendientes y las intersecciones de la ecuación lineal logarítmica propuesta por Snyder para diferentes compuestos [10], que ha sido utilizada para clasificar compuestos estructuralmente relacionados [11,12], así como para la caracterización de columnas [13]. El grupo de investigación ha desarrollado también recientemente una metodología para estimar la capacidad de pico (máximo número de picos que idealmente pueden resolverse completamente en una ventana de tiempo pre-establecida), en cromatogramas obtenidos con gradientes multi-lineales, en base a la simulación de cromatogramas [14]. La propuesta parte del establecimiento de correlaciones entre los parámetros de la

ecuación de retención logarítmico cuadrática vs. $\log k_w$ [15], para un grupo de compuestos relacionados estructuralmente, de hidrofobicidad variable. Sin embargo, los resultados fueron muy deficientes cuando se aplicaron a muestras complejas que contenían compuestos con propiedades químicas demasiado diversas (pertenecientes a diferentes familias, o con estructuras químicas y tamaños muy distintos). Además, el cálculo de la capacidad de pico mediante el método de simulación requería conocer la identidad de los componentes de la muestra.

En el Capítulo 5, se ha propuesto un procedimiento de regresión alterna para optimizar un modelo global que contiene parámetros comunes para la columna y el modificador, así como un parámetro específico para cada compuesto eluido. Los parámetros se ajustan en etapas consecutivas, atendiendo simultáneamente a todas las condiciones experimentales y compuestos de la muestra. El enfoque propuesto se validó usando varios conjuntos de datos tomados de la bibliografía, incluyendo una gran cantidad de compuestos de carácter diverso, eluidos bajo varias condiciones isocráticas y de gradiente. También se analizó un extracto de manzanilla mediante elución en gradiente. Los modelos globales de retención se aplicaron con éxito, con un rendimiento comparable al ofrecido por los modelos individuales, cumpliendo con los requisitos de componentes de diversa polaridad que contenía la muestra.

En este capítulo, se ofrecen pautas para el desarrollo de diseños experimentales multi-lineales adaptados a la elución de los componentes de las huellas dactilares cromatográficas de diversas plantas medicinales. Estos diseños también facilitan la trazabilidad de picos en los cromatogramas en distintas condiciones de elución, lo que permite la construcción de modelos globales para componentes seleccionados, aplicando un procedimiento de regresión alterna. El rendimiento predictivo de los modelos globales ofrece

buenos resultados, en comparación a los modelos individuales. Se muestra que el modelo de retención propuesto por Neue y Kuss (NK) arrojó las mejores predicciones con respecto a los resultados ofrecidos por las ecuaciones de Snyder, Schoenmakers y Jandera. Se consideraron tres plantas medicinales no relacionadas, que ofrecen características de separación muy diversas: *Camellia sinensis* (té verde), *Melissa officinalis* (melisa) y *Tilia platyphyllos* (tila).

El objetivo final de esta investigación es proponer una metodología fiable para obtener modelos globales de retención, adecuados para predecir cromatogramas completos que contengan cientos de componentes (lo que se desarrollará en el Capítulo 7). Se pretende que, en el futuro, el uso de los modelos globales permita la predicción de huellas dactilares cromatográficas para el tipo de muestra tratado en este trabajo, así como de otras muestras complejas, sin requerir el uso de estándares, y todo ello con fines de optimización de la resolución.

6.3. Teoría

6.3.1. Modelos individuales de retención

En este capítulo, se investiga el rendimiento de los modelos de retención utilizados para describir el comportamiento en cromatografía líquida en fase inversa (RPLC, *reversed-phase liquid chromatography*), de compuestos seleccionados pertenecientes a cromatogramas de tres plantas medicinales. Las ecuaciones estudiadas, ampliamente utilizadas para describir compuestos en un amplio intervalo de polaridades son las siguientes:

$$\log k_i = \log k_{w,i} - S_i \varphi \quad (6.1)$$

$$\log k_i = \log k_{w,i} - S_i \varphi + T_i \varphi^2 \quad (6.2)$$

$$k_i = k_{0,i} (1 + c_i \varphi)^2 e^{\frac{-b_i \varphi}{1+c_i \varphi}} \quad (6.3)$$

propuestas por Snyder [10], Schoenmakers et al. [15] y NK [17], respectivamente. El término “modelo individual” especifica que los datos de retención se ajustan de forma independiente, soluto a soluto. Por lo tanto, todos los parámetros en cada modelo obtenido son específicos para cada soluto, a pesar de que los parámetros asociados a la columna y modificador (S_i , T_i , b_i y c_i en las Ecuaciones (6.1) a (6.3)), idealmente deberían ser comunes para el sistema separador. Los ajustes se pueden llevar a cabo con los datos obtenidos a partir de un conjunto de experimentos, mediante elución isocrática o en gradiente. Dado que no se disponía de estándares, e incluso se desconocía la identidad de muchos componentes en las muestras de plantas medicinales, los modelos de retención tuvieron que construirse sólo con los datos de algunos compuestos, que podían trazarse en todos los gradientes ensayados del diseño experimental, a los que denominaremos “compuestos de referencia” para resaltar la diferencia con los estándares convencionales (véase la Sección 6.5.3). Como se ha comentado, el enfoque se extenderá en el Capítulo 7 para describir todos los picos visibles en los cromatogramas, de acuerdo a los modelos desarrollados para los compuestos de referencia. La condición que decide la idoneidad de un componente de una muestra, para ser utilizado como compuesto de referencia, reside principalmente en su trazabilidad entre experimentos. Por lo tanto, la selección se realiza atendiendo a las características que hacen que los picos sean fácilmente identificables, como el tamaño de pico o la ubicación cercana a grupos de picos con patrones reconocibles.

Los diseños de entrenamiento se construyeron considerando gradientes multi-lineales. Los tiempos de retención se calcularon resolviendo la ecuación

fundamental para elución en gradiente, mediante integración analítica [18]. Los modelos de retención se ajustaron atendiendo a la suma de residuos al cuadrado (*SSR*, *sum of squared residuals*, Ecuación (6.4)), como función objetivo [19], extendida a todos los solutos i y gradientes g en el diseño experimental, que minimiza la diferencia entre los tiempos de retención predichos ($\hat{t}_{Ri,g}$) y experimentales ($t_{Ri,g}$):

$$SSR = \sum_{i=1}^{ns} \sum_{g=1}^{ng} (t_{Ri,g} - \hat{t}_{Ri,g})^2 \quad (6.4)$$

donde ns y ng son el número de compuestos de referencia y gradientes en el diseño experimental, respectivamente.

6.3.2. Modelos globales de retención

Los modelos globales se formularon de manera similar a los modelos individuales, pero con parámetros comunes asociados a la columna y modificador (S , T , b y c), para todos los solutos:

$$\log k_i = \log k_{w,i} - S \varphi \quad (6.5)$$

$$\log k_i = \log k_{w,i} - S \varphi + T \varphi^2 \quad (6.6)$$

$$k_i = k_{0,i} (1 + c\varphi)^2 e^{\frac{-b\varphi}{1+c\varphi}} \quad (6.7)$$

El término “modelo global” pretende indicar que el mismo modelo es capaz de predecir la retención de un grupo de solutos. La obtención de parámetros comunes requiere el ajuste simultáneo de los datos para todos los solutos de interés y todos los gradientes del diseño experimental. Con este propósito, se desarrolló un procedimiento de regresión alterna (véase la Sección 6.3.3), que separa el ajuste de los parámetros específicos del soluto ($\log k_{w,i}$ y $k_{0,i}$), de los

parámetros de la columna y modificador en etapas consecutivas (Capítulo 5). En cada etapa de la regresión, se aborda el conjunto completo de datos disponibles, alterando algunos parámetros mientras que otros permanecen constantes. A lo largo de las iteraciones, se comprueban las consecuencias de los cambios en los valores de los parámetros, en términos de diferencia entre los tiempos de retención predicho y experimental (Ecuación (6.4)).

El modelo global se puede representar mediante un vector de parámetros que describen la ecuación de cada soluto. Así, por ejemplo, el tiempo de retención del compuesto de referencia i obtenido con la Ecuación (6.6) se puede calcular seleccionando los elementos requeridos del vector $[S, T, \log k_{w,1}, \log k_{w,2}, \dots, \log k_{w,ns}]$ (con $ns + 2$ elementos) y reorganizándolos en un subvector $[S, T, \log k_{w,i}]$, que se puede aplicar como un modelo de retención individual.

6.3.3. Regresión alterna para ajustar modelos globales

El procedimiento de regresión alterna utilizado para obtener los parámetros del modelo global comienza con el ajuste, soluto a soluto, de los datos de retención de los ns compuestos seleccionados (los compuestos de referencia) a los modelos de retención individuales. A continuación, se calculan las medianas de los parámetros comunes en estos modelos para todos los compuestos (los parámetros que describen el sistema columna/modificador). Los parámetros se organizan a continuación en un vector, que constituye la estimación inicial de la retención: para la ecuación de Schoenmakers (Ecuación (6.2), el vector es $[S_m, T_m, \log k_{w,1}, \log k_{w,2}, \dots, \log k_{w,ns}]$, donde S_m y T_m son las medianas de los valores de S_i y T_i , y $\log k_{w,i}$ es el parámetro específico para cada soluto i , obtenido de los ajustes individuales. El procedimiento es similar para los otros modelos de retención: $[S_m, \log k_{w,1}, \log k_{w,2}, \dots, \log k_{w,ns}]$ para el modelo de

Snyder (Ecuación (6.5)), y $[b_m, c_m, k_{0,1}, k_{0,2}, \dots, k_{0,ns}]$ para el modelo NK (Ecuación. (6.7)).

Hay varias formas de llevar a cabo la regresión y obtener los parámetros de los modelos globales, que ofrecen diferentes rendimientos. Para acelerar el proceso de convergencia, y conseguir un cálculo más seguro, se ha aplicado la siguiente secuencia a los datos obtenidos con los gradientes de entrenamiento. Se ilustrará nuevamente con la ecuación de Schoenmakers:

- (i) En primer lugar, se realiza un ajuste independiente para obtener el conjunto de valores $\log k_{w,i}$, soluto a soluto, manteniendo las medianas S_m y T_m sin cambios.
- (ii) A continuación, se ajustan los parámetros S y T comunes en el modelo global, utilizando el conjunto de datos para todos los solutos y gradientes. Los parámetros S y T se fuerzan a ser idénticos para todos los solutos y experimentos, mientras que los valores de $\log k_{w,i}$, optimizados en la etapa (i), permanecen sin cambios.
- (iii) Los parámetros comunes S y T optimizados en la etapa (ii), y los valores de $\log k_{w,i}$ de la etapa (i) se toman en conjunto como valores iniciales para el ajuste simultáneo de todos los parámetros en el modelo global, obteniéndose el vector $[S, T, \log k_{w,1}, \log k_{w,2}, \dots, \log k_{w,ns}]$.
- (iv) Tras llevar a cabo las etapas (i)-(iii), a menudo no es necesario continuar con el proceso de optimización. Sin embargo, cuando la mejora es insuficiente o el proceso avanza con demasiada lentitud, puede ser útil reiniciar la búsqueda en la etapa (ii), tomando como punto de partida el modelo encontrado en la etapa (iii).

En general, a lo largo del proceso de ajuste, se observan grandes cambios en los valores de los parámetros del modelo y el valor de SSR (Ecuación (6.4)),

durante las primeras iteraciones, que rápidamente tienden a progresar más lentamente.

6.4. Parte experimental

Para la obtención de los extractos que se inyectaron en el cromatógrafo, se procesaron tres plantas medicinales según el procedimiento descrito en la Referencia [16]. Las huellas dactilares cromatográficas se obtuvieron utilizando un instrumento de cromatografía líquida de alta resolución (HPLC) (Serie 1200 Infinity, Agilent, Waldbronn, Alemania), que constaba de los siguientes módulos: bomba cuaternaria, autoinyector con viales de 2 mL, compartimento termostatzado para la columna y detector de fila de diodos UV-visible. La monitorización de los picos se llevó a cabo a 230 nm.

La separación se realizó con una columna Zorbax Eclipse XDB-C18 (4.6 × 150 mm de Agilent), que contenía partículas de 5 µm. La temperatura de la columna se fijó a 25 °C. Para los análisis, se utilizó elución en gradiente con mezclas hidro-orgánicas preparadas con agua nanopura obtenida con un sistema de purificación B30 Trace (Adrona, Burladingen, Alemania) y acetonitrilo grado HPLC (Scharlau, Barcelona). Para determinar el intervalo de modificador orgánico adecuado para construir el diseño experimental para cada muestra, se utilizó un gradiente de exploración, en el que se incrementó la concentración de acetonitrilo del 5 al 100 % (v/v) en 60 min. El pH del eluyente se fijó a 3.0 con ácido fórmico al 0.1 % (v/v) (Acros Organics, Fair Lawn, NJ, EE.UU.). El volumen de inyección fue de 10 µL y el flujo de fase móvil de 1 mL·min⁻¹.

Todas las disoluciones se desgasificaron en un baño ultrasónico (Elmasonic, Singen, Alemania) y se filtraron a través de membranas de Nylon de 0.45 µm (Micron Separations, Westboro, MA, EE.UU.), antes de inyectarlas en el sistema cromatográfico. El tiempo muerto se determinó con una disolución de

KBr (Acros Organics, Geel, Bélgica), siendo de 1.35 min (aunque cambió de valor ligeramente con la viscosidad de la fase móvil). El tiempo de permanencia (1.167 min) se midió utilizando un gradiente de acetona (Scharlau). El tiempo extra-columnar fue de 0.052 min.

El tratamiento de datos se ha realizado utilizando varias funciones desarrolladas por el grupo de investigación en Matlab 2020a (The Mathworks, Natick, MA).

6.5. Resultados y discusión

6.5.1. Plantas medicinales y gradientes de exploración

Se eluyeron los extractos de varias plantas medicinales utilizando un gradiente generalista, al que se denominará “gradiente de exploración”, con el fin de observar la distribución de los picos en los cromatogramas y las características particulares de las separaciones. Inicialmente se consideraron cinco plantas: té verde, melisa, tila, menta y manzanilla. Se aplicó el mismo gradiente de exploración a todas ellas, utilizando acetonitrilo como modificador orgánico, iniciándolo al 5 % de modificador y aumentando su concentración hasta alcanzar el 100 % en 60 min. En la Figura 6.1, se muestran los cromatogramas obtenidos para cada planta, ordenados según el tamaño de la región donde eluyen la mayoría de los picos.

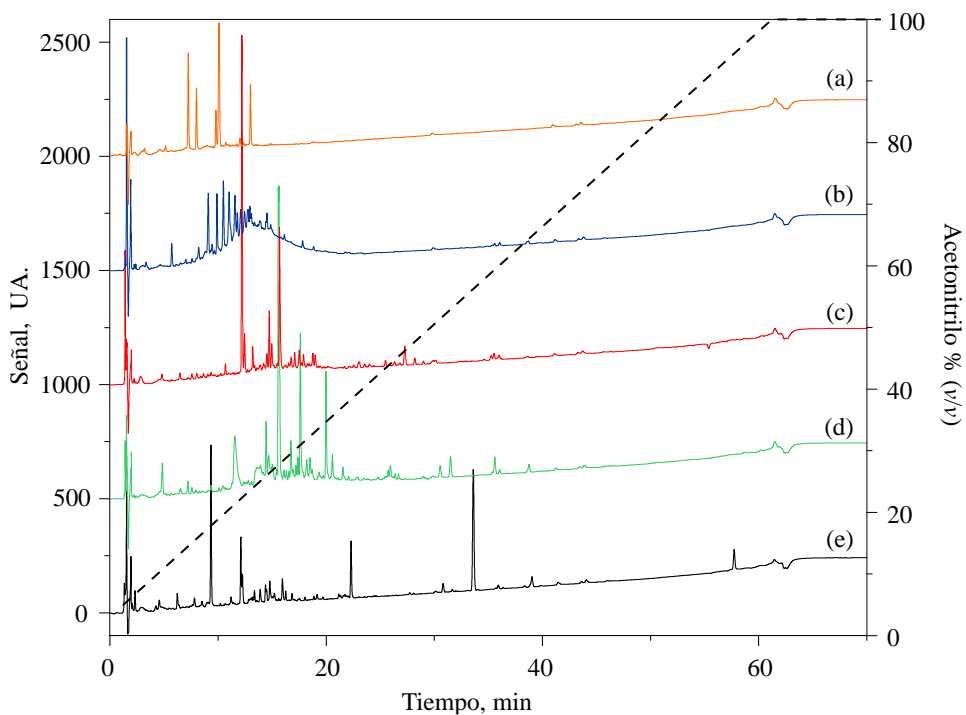


Figura 6.1. Cromatogramas de extractos en mezclas de metanol-agua de plantas medicinales, utilizando el gradiente de exploración superpuesto en la figura (en el texto se proporcionan más detalles): (a) té verde, (b) tila, (c) menta, (d) melisa, y (e) manzanilla. Se ha alterado el tamaño de la señal para que coincida con la de los distintos picos.

Una vez completado el gradiente de exploración, la concentración de modificador se mantuvo constante al 100 % durante 10 min más, con el fin de eliminar los compuestos hidrofóbicos que hubieran quedado en la columna, evitando así problemas de contaminación con la muestra anterior. Tras ello, la concentración de modificador se restauró al 5 %. Sin embargo, antes de llevar a cabo la siguiente inyección, se consideró necesaria una etapa adicional, en la que se mantuvo la concentración del modificador al 5 % durante 30 min adicionales, con el fin de reequilibrar la columna. Con tiempos más cortos, la

región inicial del cromatograma aún mostraba evidencias de no hallarse completamente reequilibrada, mostrando perturbaciones residuales de cromatogramas anteriores, que se superponían a los picos de interés. Además, al realizar inyecciones repetidas, los picos tendían a desplazarse a tiempos más cortos. Por ello se consideró necesario añadir una etapa de limpieza con acetonitrilo al 100 %, seguida de un largo periodo de reequilibrado para obtener cromatogramas reproducibles a lo largo de las inyecciones, independientemente del gradiente aplicado.

Las plantas medicinales seleccionadas para este estudio han sido:

- (i) Té verde (Figura 6.1a), que posee un cromatograma relativamente simple, representativo de muestras con un número limitado de picos. En un estudio anterior, esta planta mostró resultados insatisfactorios, por lo que se utilizó nuevamente con el objetivo de mejorarlos (véase la Sección 6.5.2).
- (ii) La melisa y la tila (Figuras 6.1d y 6.1b, respectivamente), con cromatogramas más complejos, que presentan algunas características interesantes que vale la pena estudiar, tales como agrupaciones de picos y distribuciones de picos particulares, incluidas prominencias en la línea base. Utilizando el gradiente de exploración de acetonitrilo en el intervalo 5–100 %, las prominencias se observan entre los 10 y 25 min para la melisa, y entre 5 y 25 min para la tila (más intensa en este caso).
- (iii) La menta (Figura. 6.1c), con un cromatograma similar al del té verde, se descartó para este estudio, ya que no proporcionaba ninguna característica especial desde una perspectiva de la separación.
- (iv) Manzanilla (Figura. 6.1e), de la que se informó en un estudio preliminar con resultados satisfactorios (incluida en el Capítulo 5).

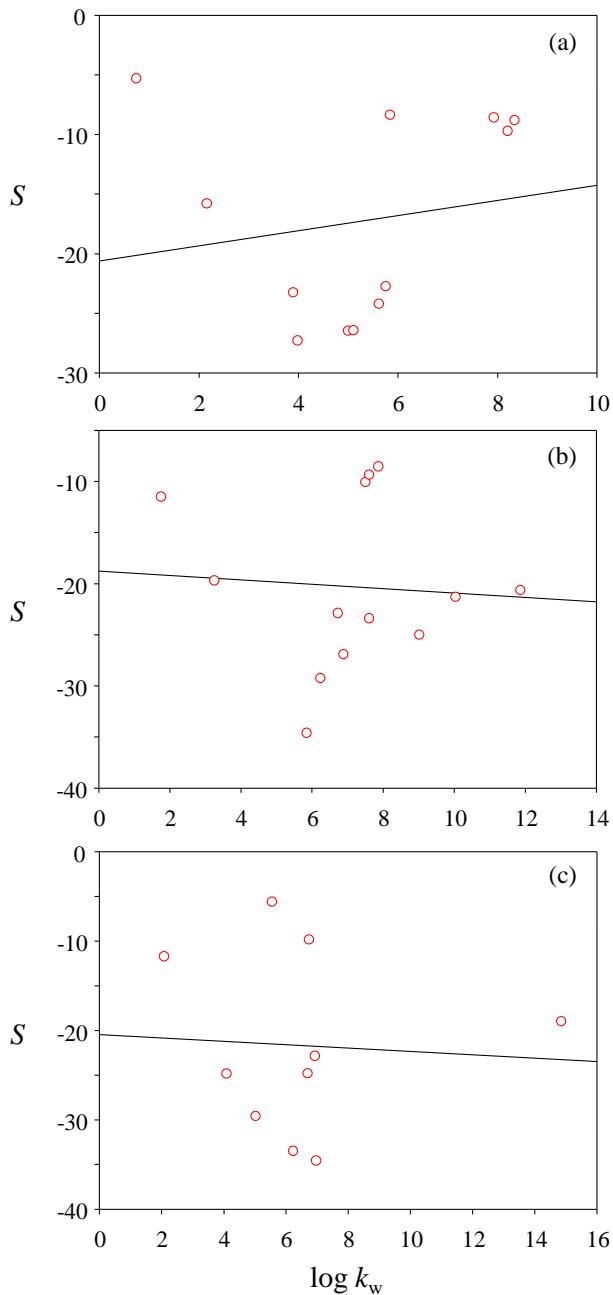


Figura 6.2. Correlación entre los parámetros de la Ecuación de Snyder para los compuestos de referencia de las tres plantas medicinales seleccionadas: (a) té verde, (b) melisa y (c) tila.

Los coeficientes de correlación para los modelos individuales obtenidos con la ecuación de Snyder, para los compuestos de referencia, considerando todos los gradientes y solutos simultáneamente, fueron $R = 0.99993$, 0.99983 y 0.99521 para té verde, melisa y tila, respectivamente. Para la tila, los coeficientes de correlación para los tres compuestos más retenidos fueron más bajos (0.9981 , 0.9774 y 0.8652), ya que uno de los gradientes eluyó dichos compuestos dentro de una rampa con una pendiente muy acusada, no permitiendo alcanzar el equilibrio (véase la Sección 6.5.3.3). Con esta excepción, la calidad de los modelos de retención individuales fue satisfactoria.

De la observación de los cromatogramas obtenidos con el gradiente de exploración, se puede concluir que las tres muestras seleccionadas incluyen componentes que exhiben una diversidad química extrema. La variabilidad de los comportamientos puede evidenciarse simplemente por la falta de correlación, para los compuestos de referencia, entre los parámetros S_i y $\log k_{w,i}$ obtenidos de los modelos individuales para la ecuación de Snyder (Ecuación (6.1)) [11]. Los resultados que se muestran en la Figura 6.2 corresponden a todo el conjunto de gradientes en el diseño experimental de entrenamiento. Obsérvese que, a pesar del buen rendimiento de los modelos individuales, la correlación entre los parámetros del modelo (S vs. $\log k_w$) puede considerarse inexistente.

6.5.2. Estudio preliminar haciendo uso de un diseño experimental de gradientes no relacionados

Con el diseño experimental inicialmente utilizado (Figura 6.3), el modelo de retención obtenido para el té verde fue insatisfactorio, debido a errores en la identificación de los picos para algunos compuestos de referencia eluidos con algunos gradientes. Sin embargo, los errores pudieron detectarse y corregirse parcialmente observando los valores atípicos en las correlaciones de los tiempos de retención previstos frente a los experimentales.

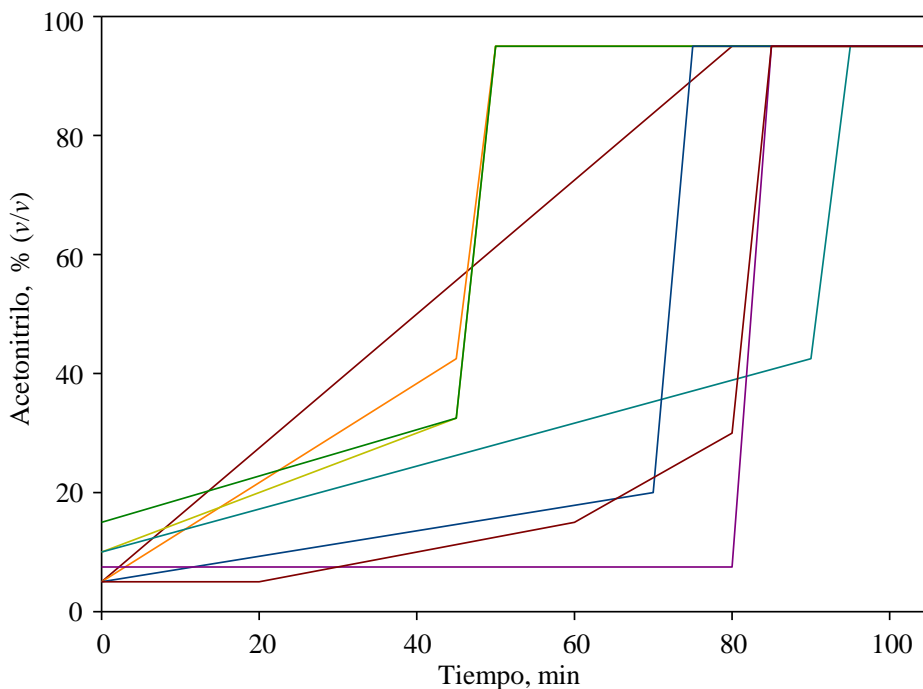


Figura 6.3. Diseños de entrenamiento utilizados para eluir extractos de té verde en un estudio preliminar

Sin embargo, a pesar de haber realizado todas las correcciones aparentemente necesarias, los modelos aún mostraban una dispersión excesiva (Figura 6.4), revelando con ello que la identificación de todos los picos aún no era correcta. Por lo tanto, el uso de un diseño experimental basado en gradientes simples no relacionados no resultaba apropiada, no sólo por las ambigüedades en la identificación de los picos, sino también porque no se disponía de la información necesaria para modelizar componentes de alta y baja hidrofobicidad.

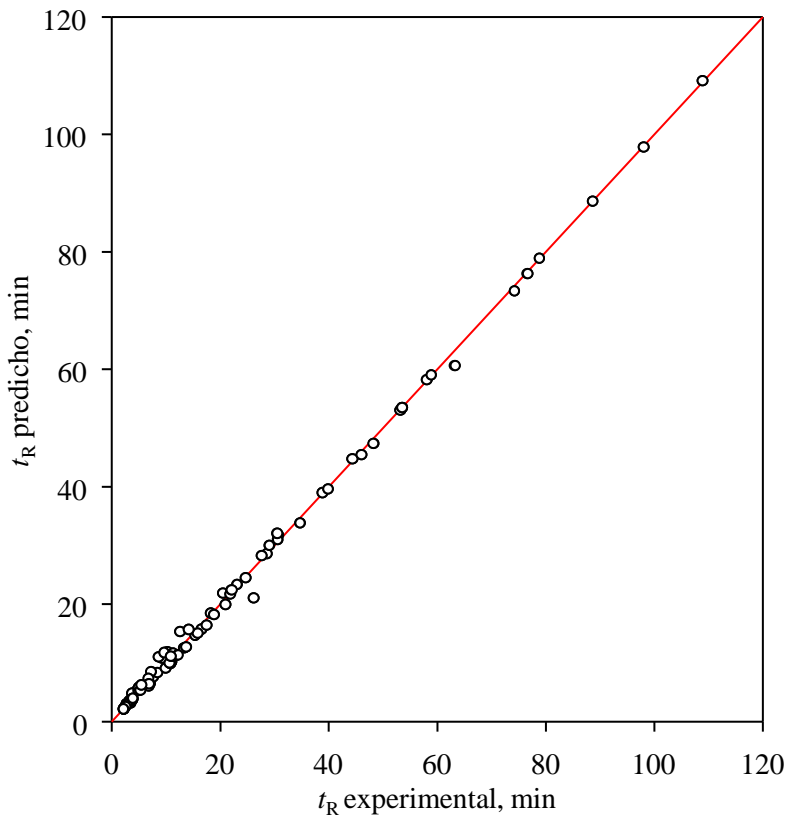


Figura 6.4. Rendimiento de predicción de la retención para el té verde, haciendo uso de la ecuación de Schoenmakers, ajustada individualmente (soluto a soluto), usando los gradientes ilustrados en la Figura 6.3.

El estudio anterior sugirió que el diseño de gradiente utilizado para ajustar los modelos de retención debía incluir variaciones ordenadas en la composición del eluyente, como las propuestas en la Sección 6.5.3. Este protocolo facilitaría la identificación de los picos de un mismo componente, eluido bajo diferentes gradientes. Además, los gradientes aplicados también reducirían las diferencias de retención entre los componentes de la muestra.

Por otro lado, la similitud entre algunos parámetros del modelo (por ejemplo, los valores de S_i en la Ecuación (6.1) para diferentes grupos de solutos), observada en el Capítulo 5, sugirió la viabilidad de un modelo con algunos parámetros comunes para todos los solutos (es decir, la posibilidad de construir un modelo global), válido para predecir la retención de todos los componentes en una planta medicinal.

6.5.3. Desarrollo de diseños de entrenamiento en forma de diamante que hacen uso de gradientes

Con el fin de facilitar la identificación de los picos, para utilizarlos como referencia, y que colateralmente la retención de todos los componentes en una muestra se halle dentro de un intervalo de modificador razonable, se propuso un tipo especial de diseño experimental para modelizar la retención, formado por gradientes ordenados [7]. Los gradientes en estos diseños son multi-lineales, con dos segmentos lineales consecutivos que comparten un nodo intermedio de posición variable. Los nodos se ubican siguiendo una diagonal, lo que da lugar a un perfil característico (Figura 6.5). Este tipo de diseños se aplicó por primera vez a modelizar la retención de compuestos de referencia seleccionados para extractos de manzanilla (véase la Figura 5.1 en el Capítulo 5), siendo definidos con el término de “diseños en forma de diamante” o “diseños de diamante”.

Los diseños de diamante permiten explorar composiciones extremas, sin que los tiempos de retención resulten excesivamente elevados para los compuestos más hidrofóbicos, o demasiado cortos para los más hidrofílicos. La calidad de estos diseños se estudió en el Capítulo 4, donde se analizó en detalle el contenido informativo de un gran número de diseños experimentales, midiendo su rendimiento según el Principio de Optimalidad G [20,21]. Un inconveniente de estos diseños es la coincidencia de composiciones para diferentes gradientes, al principio o final de la elución. Esto limita las predicciones al dominio cubierto por los gradientes extremos, aunque no implica consecuencias graves.

Nótese que, con este tipo de diseños, el intervalo de composiciones muestreado es muy diverso, sin que ello dé lugar a tiempos de retención inaceptablemente largos para los componentes más lentos, o demasiado cortos para los más rápidos.

El uso de diseños de gradiente en forma de diamante, para modelizar la retención, requiere su adaptación a cada tipo de muestra. El gradiente de exploración de la Figura 6.1 se ha utilizado para determinar el dominio de modificador adecuado para té verde, melisa y tila, con el fin de obtener una elución adecuada en la región más significativa de los cromatogramas de los extractos. Los intervalos de acetonitrilo fueron diferentes para cada planta medicinal, reflejando la diversidad en la distribución de los componentes.

A continuación, se describe la construcción de los diseños de gradiente apropiados para cada planta medicinal.

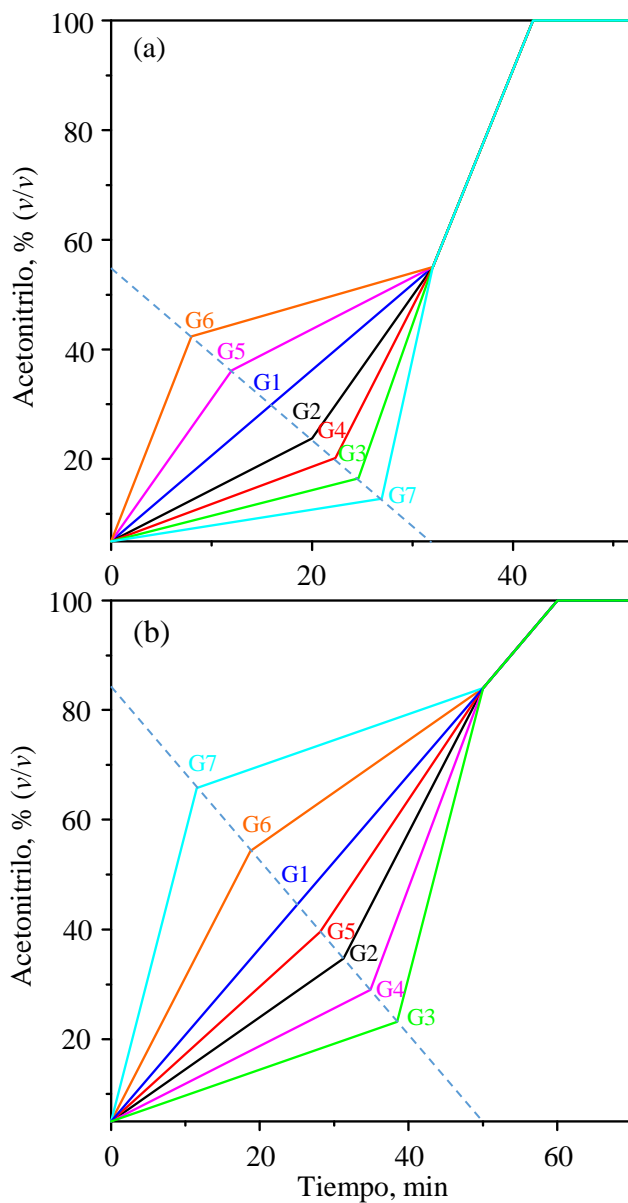


Figura 6.5. Diseños experimentales en forma de diamante, utilizados para obtener los modelos individuales y globales de: (a) té verde, (b) melisa y (c) tila. La secuencia de adquisición de los gradientes es G1 a G6 (o G7). La diagonal a lo largo de la cual se varió la posición del nodo se superpone como una línea de puntos.

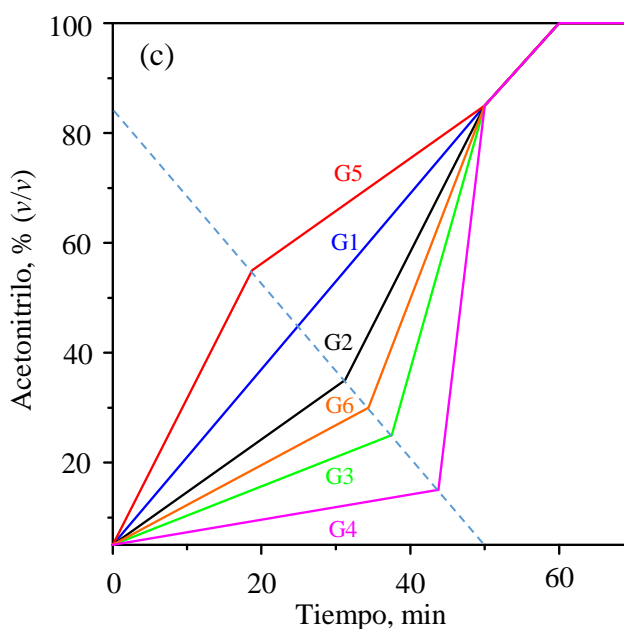


Figura 6.5 (continuación).

6.5.3.1. Té verde

La mayoría de los picos en el cromatograma del extracto de té verde aparecen durante los primeros 30 min del gradiente de exploración. Después de ese tiempo, sólo son visibles unos pocos picos de baja intensidad muy separados entre sí. Los gradientes multi-lineales de entrenamiento comienzan a un 5 % de acetonitrilo, e incluyen un nodo intermedio para maximizar el intervalo de variación de las composiciones en la región que contiene un mayor número de picos. Todos los gradientes alcanzan la misma composición que el gradiente de exploración a los 30 min (55 % de acetonitrilo), con el propósito de favorecer que la mayoría de los picos eluyan antes del final de cada gradiente multi-lineal aplicado. Más allá de este punto, la concentración de acetonitrilo se incrementa hasta alcanzar el 100 % en 10 min, permaneciendo

constante durante otros 10 min para efectuar la limpieza de la columna. A continuación, se reduce el contenido de modificador al 5 % y se mantiene esta concentración para volver a equilibrar la columna antes de la siguiente inyección. Se realizaron experimentos con tiempos de reequilibrado progresivamente más largos, partiendo de 10 min, lo que evidenció la necesidad de mantener esta etapa al menos 30 min (como se menciona en la Sección 6.5.1 para el gradiente de exploración).

El primer gradiente del diseño de entrenamiento (G1 en la Figura 6.5a) coincide con el gradiente de exploración durante los primeros 30 min, por lo que los cromatogramas son similares, pero con un tiempo de análisis más corto que para el gradiente de exploración: 50 min vs. 70 min (a los que se les ha sumado un tiempo adicional para la limpieza de la columna). Los demás gradientes del diseño (G2 a G7) incluyen un nodo intermedio, cuya posición se varió a lo largo de la diagonal que une las coordenadas (0 min, 55 % de acetonitrilo) y (30 min, 5 % de acetonitrilo). Para construir estos gradientes, el nodo se trasladó siguiendo la secuencia G1 a G7 a lo largo de la diagonal. Esta secuencia se construyó paso a paso, de acuerdo a los resultados de cada experimento anterior. Dado que más allá de los 30 min los picos se encontraban lo suficientemente distantes y su tamaño era bastante pequeño, se decidió aumentar la pendiente del segundo segmento del gradiente tras el nodo, para reducir el tiempo de análisis y aumentar la visibilidad de los picos.

6.5.3.2. *Melisa*

En el cromatograma de la melisa obtenido con el gradiente de exploración (Figura 6.1d), la mayoría de los picos eluyeron en los primeros 50 min, en los que el gradiente alcanza el 84 % de acetonitrilo. Más allá de este tiempo, el cromatograma contiene sólo unos pocos picos de muy baja intensidad.

Teniendo en cuenta la distribución de los picos y el hecho de que el acetonitrilo alcanza una concentración elevada a los 50 min, no fue necesario incrementar la pendiente para el segmento de gradiente final, como sucedió con el té verde para reducir la duración de la etapa de limpieza.

Para la melisa, el primer gradiente del diseño de entrenamiento, G1 en la Figura 6.5b, coincidió con el gradiente de exploración hasta el 100 % de acetonitrilo. Para los gradientes G2 y G3, el nodo intermedio se movió hacia una concentración de acetonitrilo progresivamente más baja a lo largo de la diagonal. Sin embargo, a diferencia del comportamiento observado con el té verde, los picos obtenidos a lo largo del segmento inicial del gradiente G3 tendieron a ensancharse en exceso, lo que disminuyó significativamente su intensidad. Por lo tanto, se decidió dejar de trasladar el nodo a concentraciones más bajas, más allá del gradiente G3, y en cambio, explorar gradientes intermedios más rápidos (G4 y G5). La adición de estos gradientes intermedios mejoró los modelos y permitió rastrear las identidades de los picos más fácilmente, en comparación con los gradientes más cercanos de mayor y menor fuerza eluyente. Después de ejecutar los gradientes G1 a G5, se aplicaron dos gradientes adicionales más rápidos (G6 y G7), con el nodo intermedio situado por encima del gradiente de exploración.

6.5.3.3. *Tila*

El cromatograma para el extracto de tila, correspondiente al primer gradiente en el diseño, G1 (Figura 6.1c), coincide con el gradiente de exploración (Figura 6.5c). De forma similar a la melisa, la mayoría de los picos eluyen en los primeros 50 min, donde se alcanza el 84 % de acetonitrilo. Como se ha explicado para la melisa, el desarrollo del diseño se inició añadiendo gradientes con un nodo intermedio a lo largo de la diagonal, debajo del gradiente de

exploración (G2 a G4, añadiendo más tarde el gradiente G6). En el gradiente G4, el incremento final en la concentración de acetonitrilo fue tan pronunciado que la separación se llevó a cabo en condiciones alejadas del equilibrio, lo que dio lugar a desviaciones para los solutos más lentos. Por esa razón, no se implementaron gradientes con nodos por debajo del G4, y en su lugar, se consideraron nodos por encima del gradiente de exploración. Sin embargo, sólo se pudo ejecutar uno de estos gradientes (el G5), debido a que la retención de los picos se estrechó demasiado en los primeros minutos del cromatograma. Un gradiente más rápido magnificaría este problema, favoreciendo la co-elución de los picos y dificultando su identificación.

6.5.4. Selección de los picos de los compuestos de referencia

Como se ha comentado, en el enfoque propuesto, la caracterización de los cromatogramas se basa en modelizar la retención para los picos de unos pocos componentes de las muestras, de los que es probable se desconozca su identidad, a los que se ha denominado compuestos de referencia. Los picos se seleccionan de acuerdo con los siguientes criterios:

- (i) Deben estar distribuidos a lo largo del cromatograma.
- (ii) Debe ser posible rastrear su identidad observando en los cromatogramas la similitud de los patrones de distribución de los picos, cuando se cambian las condiciones del gradiente, o prestando atención a su intensidad.
- (iii) Los picos deben ser lo suficientemente intensos para permitir su detección, incluso en las condiciones de menor fuerza eluyente.

Como se ha discutido en el Capítulo 5, los diseños en forma de diamante resultan muy adecuados para rastrear las identidades de los picos de las plantas

medicinales. La selección de los picos de referencia en cada muestra se realizó una vez se obtuvo la información de al menos tres gradientes. Con estos gradientes, fue posible realizar una verificación preliminar del correcto seguimiento de los picos, continuando el proceso mediante la incorporación de más experimentos. La confirmación final del seguimiento de los picos provino del ajuste del conjunto completo de datos de que se disponía, a los modelos de retención individuales.

La Figura 6.6 muestra los picos de referencia que cumplieron con las condiciones anteriores, para té verde, melisa y tila, utilizados para construir los modelos de retención para cada planta medicinal. Los cromatogramas de la figura se obtuvieron con el gradiente más lento de entre los diseños en forma de diamante, ensayados para cada muestra (véase la Figura 6.5), tras la sustracción de la línea base.

El número de picos de referencia seleccionados fue 12, 13 y 10 para los extractos de té verde (Figura 6.6a), melisa (Figura 6.6b) y tila (Figura 6.6c), respectivamente. Debe prestarse también atención a la presencia de picos de origen refractométrico, al principio y final de los cromatogramas (R1 y R2 en la Figura 6.6). Los picos R1 son el resultado de cambios en el índice de refracción producidos por la inyección de extractos que contienen un 70 % de metanol, cuando se mezclan con el eluyente de acetonitrilo-agua. Por su parte, los picos R2 se producen debido a la parada brusca del gradiente, una vez alcanzado el 100 % de acetonitrilo.

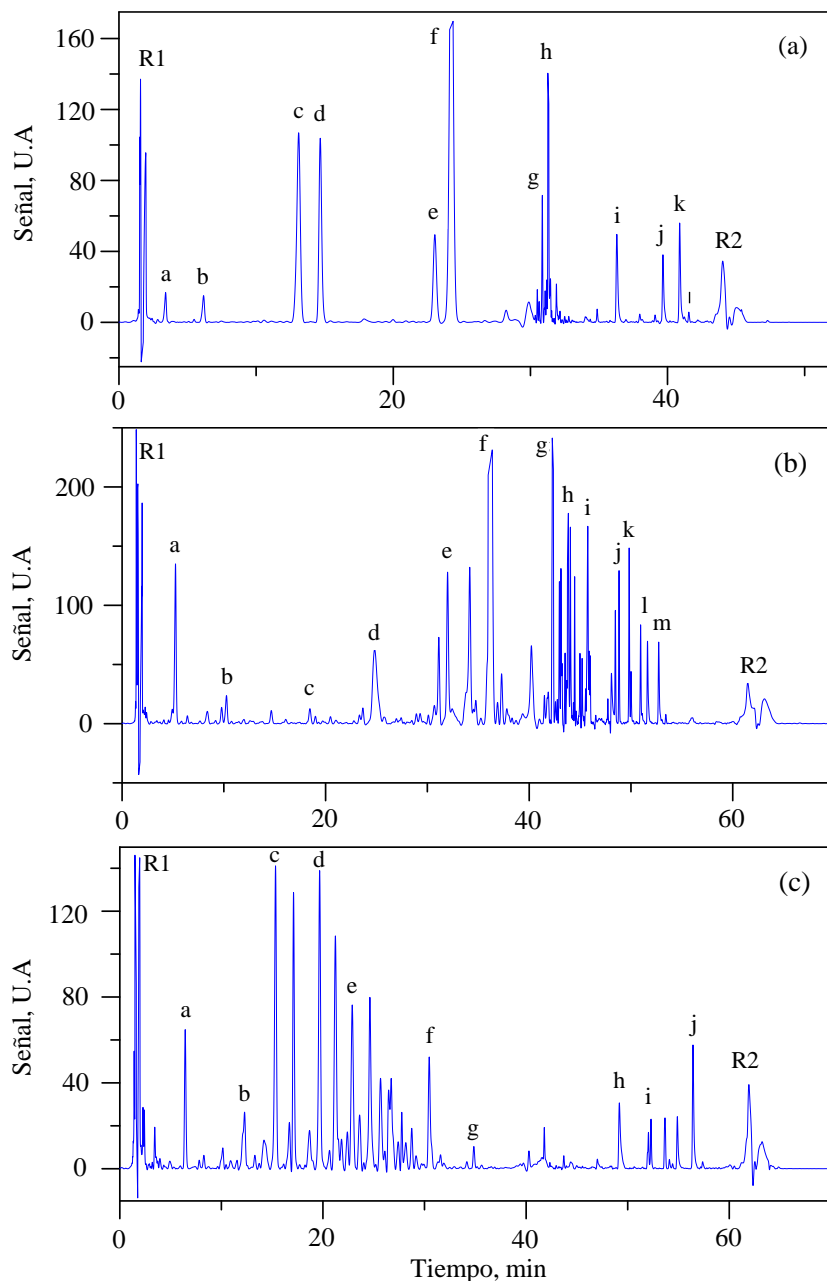


Figura 6.6. Cromatogramas para: (a) té verde, (b) melisa y (c) tila, obtenidos con los gradientes más lentos de cada diseño: G7, G3 y G4, respectivamente (véase la Figura 6.5). En cada caso, se indican los picos de los compuestos de referencia seleccionados. Los picos R1 y R2 tienen un origen refractométrico.

6.5.5. Descripción del comportamiento de retención para los compuestos de referencia

6.5.5.1. Modelos individuales de retención

Las Tablas 6.1, 6.2 y 6.3 muestran los estadísticos de ajuste para los modelos de retención individuales, de acuerdo a las ecuaciones de Snyder, Schoenmakers y NK (Ecuaciones (6.1) a (6.3)), respectivamente, para los compuestos de referencia seleccionados para las tres plantas medicinales (té verde, melisa y tila, respectivamente). Como se observa, la calidad de los ajustes a los modelos individuales es excelente (véanse también las Figuras 6.7a-6.7c). Este comportamiento, junto con el seguimiento satisfactorio de los picos cuando se modifican los gradientes, así como la consistencia de los perfiles, indican que los picos seleccionados corresponden, probablemente, a compuestos individuales. Aunque el ajuste de los datos a las tres ecuaciones es satisfactorio, el modelo NK destaca como el mejor, seguido del modelo de Schoenmakers.

La ecuación de Snyder ofrece resultados sorprendentemente buenos para el té verde y la melisa, lo que puede atribuirse en parte al uso de datos de gradiente, ya que esta ecuación funciona bien para intervalos de modificador orgánico significativamente estrechos [22]. Los extractos de tila se eluyeron utilizando gradientes de entrenamiento con pendientes más extremas, en comparación con el té verde y la melisa, lo que dio lugar a un sesgo significativo en las predicciones realizadas con la ecuación de Snyder, para los tres compuestos de referencia más hidrofóbicos, como se ha mencionado anteriormente. Para el té verde y la melisa, la ecuación de Snyder funcionó incluso mejor que la ecuación de Schoenmakers, probablemente debido a la convergencia prematura de esta última.

Tabla 6.1. Estadísticos de regresión para el ajuste de modelos de retención individuales para los compuestos de referencia seleccionados para el té verde, basados en los modelos de Snyder (Ecuación (6.1)), Schoenmakers (Ecuación (6.2)) y NK (Ecuación (6.3)).

Pico	Estadístico ^a	Ecuación (6.1)	Ecuación (6.2)	Ecuación (6.3)
a	<i>R</i>	– ^b	– ^b	– ^b
	<i>SEP</i>	0.09	0.13	0.05
	<i>F</i>	7	1	2
	<i>ER (%)</i>	1.61	1.69	0.81
b	<i>R</i>	0.986	0.9915	0.9931
	<i>SEP</i>	0.13	0.11	0.10
	<i>F</i>	149	176	221
	<i>ER (%)</i>	1.54	1.24	1.19
c	<i>R</i>	0.9986	0.9983	0.99967
	<i>SEP</i>	0.18	0.23	0.10
	<i>F</i>	1534	854	4529
	<i>ER (%)</i>	1.31	1.32	0.58
d	<i>R</i>	0.9980	0.99991	0.99972
	<i>SEP</i>	0.24	0.06	0.10
	<i>F</i>	1068	17598	5285
	<i>ER (%)</i>	1.89	0.39	0.59
e	<i>R</i>	0.99940	0.99951	0.99999
	<i>SEP</i>	0.24	0.26	0.04
	<i>F</i>	3469	3100	118829
	<i>ER (%)</i>	1.35	1.20	0.19

^a *R* (coeficiente de correlación), *SEP* (error estándar en la predicción), *F* (*F* de Snedecor), y *ER* (error relativo referido a la retención media). El número de gradientes de entrenamiento es $n = 7$.

^b El soluto apenas se retiene por lo que su ajuste es deficiente.

Tabla 6.1 (continuación).

Pico	Estadístico ^a	Ecuación (6.1)	Ecuación (6.2)	Ecuación (6.3)
f	<i>R</i>	0.99936	0.99955	0.99989
	<i>SEP</i>	0.27	0.26	0.13
	<i>F</i>	3213	3438	13680
	<i>ER (%)</i>	1.43	1.23	0.58
g	<i>R</i>	0.99956	0.99953	0.99989
	<i>SEP</i>	0.30	0.36	0.18
	<i>F</i>	4566	3134	13125
	<i>ER (%)</i>	1.20	1.22	0.59
h	<i>R</i>	0.99967	0.99965	0.99997
	<i>SEP</i>	0.27	0.33	0.10
	<i>F</i>	6066	4134	48077
	<i>ER (%)</i>	1.07	1.04	0.34
i	<i>R</i>	0.99954	0.99991	0.99967
	<i>SEP</i>	0.25	0.13	0.24
	<i>F</i>	4110	16802	4607
	<i>ER (%)</i>	0.61	0.23	0.43
j	<i>R</i>	0.9988	0.9985	0.9982
	<i>SEP</i>	0.04	0.05	0.06
	<i>F</i>	1666	1004	843
	<i>ER (%)</i>	0.06	0.07	0.08
k	<i>R</i>	0.9979	0.9968	0.9961
	<i>SEP</i>	0.04	0.06	0.07
	<i>F</i>	942	463	378
	<i>ER (%)</i>	0.07	0.09	0.10
l	<i>R</i>	0.961	0.956	0.960
	<i>SEP</i>	0.13	0.15	0.14
	<i>F</i>	48	31	36
	<i>ER (%)</i>	0.19	0.21	0.19

Tabla 6.2. Estadísticos de regresión para el ajuste de modelos de retención individuales para los compuestos de referencia seleccionados para la melisa (para más detalles, véase la Tabla 6.1).

Pico	Estadístico ^a	Ecuación (6.1)	Ecuación (6.2)	Ecuación (6.3)
a	<i>R</i>	0.9927	0.99927	0.9923
	<i>SEP</i>	0.05	0.04	0.06
	<i>F</i>	330	444	229
	<i>ER (%)</i>	0.76	0.62	0.76
b	<i>R</i>	0.99927	0.99972	0.9982
	<i>SEP</i>	0.08	0.05	0.23
	<i>F</i>	2778	7148	1159
	<i>ER (%)</i>	0.67	0.36	1.53
c	<i>R</i>	0.9925	0.9924	0.9933
	<i>SEP</i>	0.60	0.67	0.63
	<i>F</i>	347	265	296
	<i>ER (%)</i>	3.63	3.68	3.24
d	<i>R</i>	0.99981	0.99992	0.99994
	<i>SEP</i>	0.14	0.10	0.09
	<i>F</i>	13165	26726	31085
	<i>ER (%)</i>	0.69	0.45	0.39
e	<i>R</i>	0.99985	0.99991	0.99996
	<i>SEP</i>	0.16	0.14	0.09
	<i>F</i>	17100	22882	54185
	<i>ER (%)</i>	0.67	0.52	0.33
f	<i>R</i>	0.99982	0.99990	0.99994
	<i>SEP</i>	0.20	0.17	0.13
	<i>F</i>	13830	20695	33911
	<i>ER (%)</i>	0.76	0.47	0.43

^a El número de gradientes de entrenamiento es $n = 7$.

Tabla 6.2 (continuación).

Pico	Estadístico ^a	Ecuación (6.1)	Ecuación (6.2)	Ecuación (6.3)
g	<i>R</i>	0.99989	0.99988	0.99997
	<i>SEP</i>	0.19	0.23	0.12
	<i>F</i>	22043	15773	60719
	<i>ER (%)</i>	0.60	0.57	0.28
h	<i>R</i>	0.99994	0.99996	0.99997
	<i>SEP</i>	0.16	0.14	0.12
	<i>F</i>	38487	47076	71332
	<i>ER (%)</i>	0.44	0.37	0.29
i	<i>R</i>	0.99999	0.99999	0.99999
	<i>SEP</i>	0.08	0.06	0.05
	<i>F</i>	172891	277991	373753
	<i>ER (%)</i>	0.17	0.12	0.09
j	<i>R</i>	0.9978	0.99999	0.99999
	<i>SEP</i>	0.92	0.04	0.04
	<i>F</i>	1142	569123	636689
	<i>ER (%)</i>	2.05	0.08	0.08
k	<i>R</i>	0.99993	0.99999	0.99999
	<i>SEP</i>	0.17	0.06	0.05
	<i>F</i>	34217	203160	365904
	<i>ER (%)</i>	0.33	0.11	0.09
l	<i>R</i>	0.99990	0.99999	0.99999
	<i>SEP</i>	0.18	0.05	0.05
	<i>F</i>	2610	301980	398374
	<i>ER (%)</i>	0.35	0.09	0.08
m	<i>R</i>	0.99988	0.99999	0.99999
	<i>SEP</i>	0.19	0.06	0.5
	<i>F</i>	20186	224073	318918
	<i>ER (%)</i>	0.36	0.9	0.08

Tabla 6.3. Estadísticos de regresión para el ajuste de modelos de retención individuales para los compuestos de referencia seleccionados para la tila (para más detalles, véase la Tabla 6.1).

Pico	Estadístico ^a	Ecuación (6.1)	Ecuación (6.2)	Ecuación (6.3)
a	<i>R</i>	0.9925	0.9922	0.9932
	<i>SEP</i>	0.07	0.09	0.08
	<i>F</i>	266	202	218
	<i>ER (%)</i>	0.88	0.90	0.90
b	<i>R</i>	0.9982	0.9986	0.99976
	<i>SEP</i>	0.23	0.23	0.10
	<i>F</i>	1159	1140	6185
	<i>ER (%)</i>	1.53	1.26	0.54
c	<i>R</i>	0.99945	0.99963	0.99997
	<i>SEP</i>	0.20	0.19	0.06
	<i>F</i>	3731	4110	44793
	<i>ER (%)</i>	1.10	0.95	0.27
d	<i>R</i>	0.99967	0.99950	0.99999
	<i>SEP</i>	0.26	0.36	0.05
	<i>F</i>	6148	2943	168749
	<i>ER (%)</i>	1.05	1.35	0.17
e	<i>R</i>	0.99975	0.99980	0.99992
	<i>SEP</i>	0.28	0.29	0.18
	<i>F</i>	8282	7553	19307
	<i>ER (%)</i>	1.02	0.98	0.58
f	<i>R</i>	0.99980	0.99982	0.99989
	<i>SEP</i>	0.30	0.33	0.25
	<i>F</i>	9777	7979	13410
	<i>ER (%)</i>	0.58	0.59	0.55

^a El número de gradientes de entrenamiento es $n = 6$.

Tabla 6.3 (continuación).

Pico	Estadístico ^a	Ecuación (6.1)	Ecuación (6.2)	Ecuación (6.3)
g	<i>R</i>	0.99963	0.99968	0.99988
	<i>SEP</i>	0.41	0.44	0.27
	<i>F</i>	5195	4537	11882
	<i>ER (%)</i>	0.73	0.61	0.51
h	<i>R</i>	0.989	0.9987	0.9972
	<i>SEP</i>	2.13	0.82	1.22
	<i>F</i>	135	1115	459
	<i>ER (%)</i>	3.53	1.13	1.77
i	<i>R</i>	0.977	0.9978	0.9962
	<i>SEP</i>	2.74	0.99	1.30
	<i>F</i>	58	614	356
	<i>ER (%)</i>	4.35	1.37	1.52
j	<i>R</i>	0.970	0.9989	0.9984
	<i>SEP</i>	2.59	0.57	0.70
	<i>F</i>	39	1293	852
	<i>ER (%)</i>	3.81	0.74	0.82

Los modelos individuales presentan un inconveniente intrínseco: los ajustes se adaptan excesivamente a las particularidades de cada soluto, lo que, unido a la existencia de correlaciones entre los parámetros del modelo, conduce a la existencia de soluciones con rendimientos de predicción casi equivalentes, implicando parámetros con valores muy diferentes. Este hecho se ilustra en la Tabla 6.4, donde se proporcionan los parámetros de los modelos NK para los 12 picos de referencia seleccionados para el té verde.

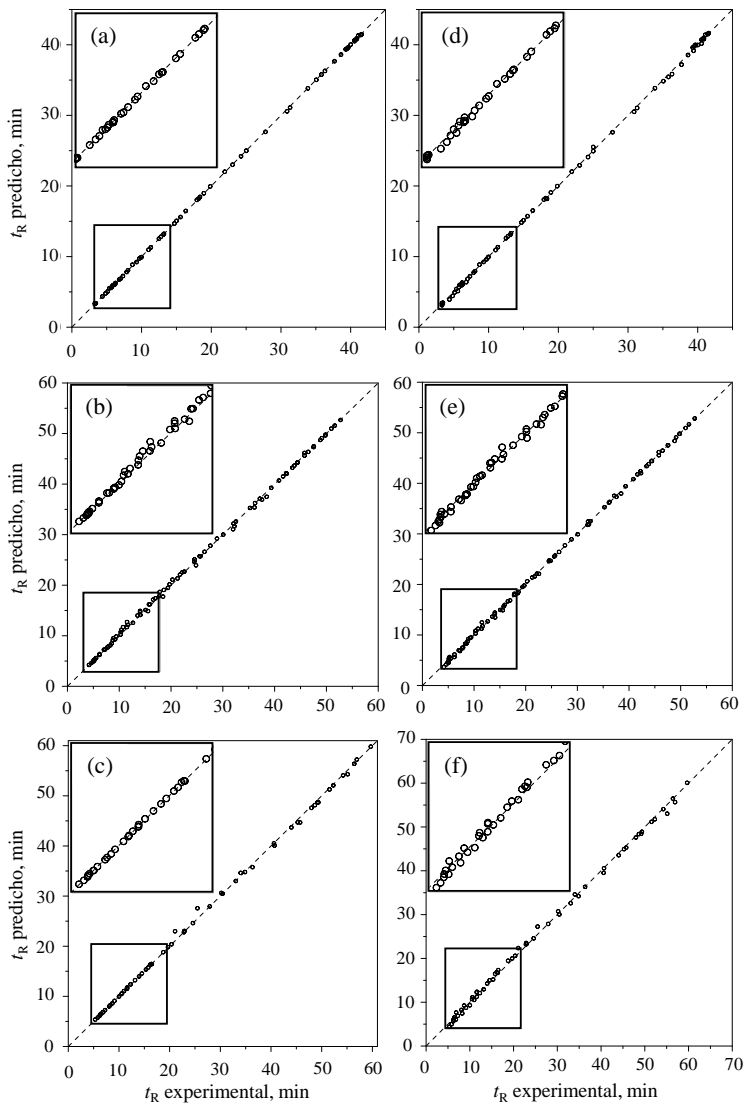


Figura 6.7. Comparación entre los valores experimentales y las correspondientes predicciones, tras ajustar la ecuación NK a los modelos individuales (a-c), y a los modelos globales obtenidos con regresión alterna (d-f) para: (a, d) té verde, (b, e) melisa, y (c, f) tila. Se ha superpuesto una ampliación de la región de baja retención.

Tabla 6.4. Parámetros del modelo individual NK para los 12 picos de referencia, para la muestra de té verde.

$\log k_{0,i}$	b_i	c_i
0.709	14.26	75.99
1.929	15.96	150.06
2.326	5.021	49.991
2.357	-0.839	0.704
2.383	6.609	83.131
2.423	6.402	88.100
3.081	5.958	91.723
3.082	5.387	85.509
4.313	5.866	112.23
7.061	4.333	108.32
8.839	3.036	93.196
8.992	2.988	95.661

Obsérvese que, para los solutos más rápidos, los valores de b y c son demasiado variables, mientras que para los demás solutos (a excepción del último soluto cuya retención varía escasamente en el diseño), los valores de los parámetros son bastante similares. A pesar de esta notable variabilidad aparente, el modelo global para los 12 solutos (donde se ha forzado la uniformidad de los parámetros, consúltese la Sección 6.5.5.2), proporciona predicciones de calidad comparable y valores coherentes de los parámetros.

6.5.5.2. Modelos globales de retención

Los modelos globales (Ecuaciones (6.5) a (6.7)) reducen significativamente el número de parámetros requeridos para describir un conjunto de compuestos (en más del 50 %), pero pagando el precio de una cierta pérdida de rendimiento

predictivo respecto a los modelos individuales. Previamente al desarrollo del modelo global, los datos para cada soluto deben ajustarse individualmente. De esta forma, tomando como ejemplo el modelo NK (Ecuación (6.7)), se obtendría un conjunto de vectores $[b_1, c_1, k_{0,1}]$, $[b_2, c_2, k_{0,2}]$, ..., $[b_n, c_n, k_{0,n}]$. Inicialmente se intentó desarrollar modelos globales reemplazando los parámetros asociados a la columna y modificador utilizados (b_i y c_i en los modelos individuales), por las respectivas medias o medianas (b_m y c_m), de nuevo para el modelo NK.

En las Figuras 6.8a – 6.8c, se pueden apreciar las correlaciones entre los tiempos de retención predichos y experimentales, para las tres plantas medicinales, cuando se utilizaron las medianas como parámetros comunes de los modelos globales (Ecuación (6.7)). En las representaciones, se incluyen los datos obtenidos para todos los compuestos de referencia y el conjunto completo de gradientes del diseño de entrenamiento para la muestra en cuestión. Como puede observarse, el uso de medianas en los modelos globales da lugar a correlaciones muy pobres. Es también interesante llamar la atención sobre las pobres correlaciones de S_i vs. $\log k_{w,i}$ obtenidas para la ecuación de Snyder en la Figura 6.2, donde la mediana del parámetro S está lejos de ser representativa. Sin embargo, a pesar de las grandes desviaciones de los valores individuales, respecto al valor de las medianas, la realización de un ajuste inicial utilizando medianas como una primera estimación de los parámetros comunes de los modelos globales, posee el interés de proporcionar valores iniciales adecuados para iniciar el proceso de regresión alterna (véase la Sección 6.3.3).

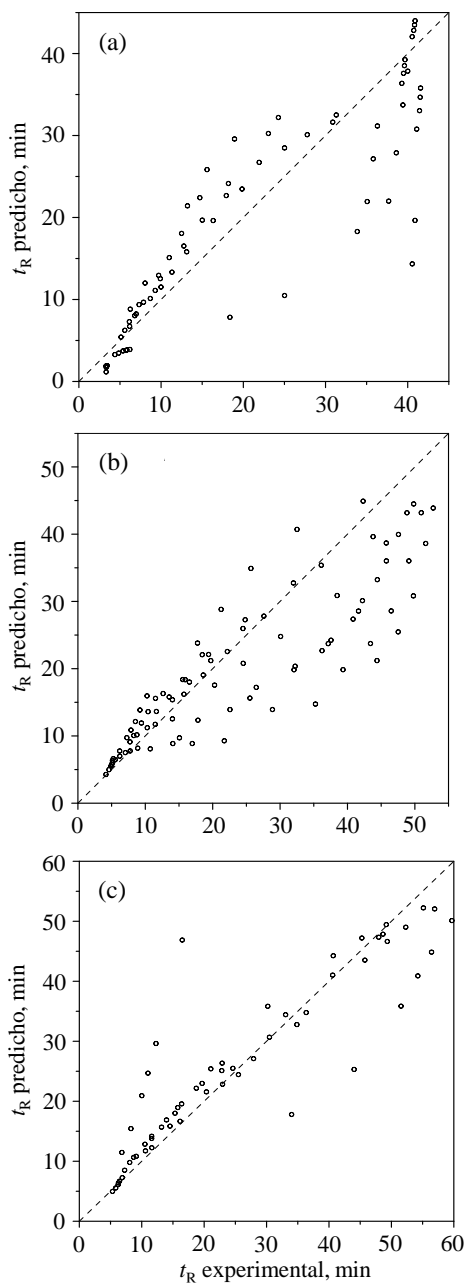


Figura 6.8. Consecuencias del uso de las medianas de los parámetros columna/disolvente, calculadas a partir de los modelos individuales (b_m y c_m en el modelo NK), para construir los modelos globales para: (a) té verde, (b) melisa y (c) tila.

En la regresión alterna, se combinan dos etapas: (i) el ajuste de los parámetros globales columna/disolvente para los compuestos de referencia, haciendo uso de todas las condiciones de entrenamiento y manteniendo fijos los parámetros específicos de los solutos; y (ii) el ajuste de los parámetros específicos de los solutos, manteniendo fijos los parámetros globales columna/disolvente. La Tabla 6.5 muestra los estadísticos de ajuste para los modelos globales de retención, obtenidos al aplicar la regresión alterna, de acuerdo a las tres ecuaciones: Snyder (Ecuación (6.5)), Schoenmakers (Ecuación (6.6)) y NK (Ecuación (6.7)). Los valores de R y F de Snedecor, para los modelos globales, generalmente denotan un rendimiento de predicción muy bueno, similar al encontrado con los modelos individuales (compárense las Tablas 6.1 - 6.3 con la 6.5).

La Figura 6.7 permite comparar las predicciones obtenidas con los modelos individuales, para las tres plantas medicinales, mediante la ecuación de NK (izquierda), con las que ofrece el modelo global obtenido mediante regresión alterna (derecha), considerando los resultados obtenidos para todos los solutos y gradientes. Como se puede observar, tras alcanzar la convergencia, los modelos globales correctamente ajustados ofrecieron un rendimiento de predicción similar al de los modelos individuales, a pesar de haber forzado que algunos parámetros fueran idénticos para todos los solutos.

Tabla 6.5. Estadísticos de regresión para el ajuste de modelos globales para los compuestos de referencia en tres plantas medicinales, basados en las ecuaciones de Snyder, Schoenmakers y NK.

Modelo	Estadísticos ^a	Té verde	Melisa	Tila	Tila ^b
Ecuación (6.5) ^b	<i>R</i>	0.99943	0.99919	0.9933	0.9994
	<i>SEP</i>	0.48	0.61	1.99	0.39
	<i>F</i>	60278	53995	3731	32320
	<i>ER (%)</i>	1.64	1.80	4.65	1.90
	<i>n</i>	84	91	60	42
Ecuación (6.6)	<i>R</i>	0.99984	0.99936	0.9938	0.9989
	<i>SEP</i>	0.25	0.55	1.92	0.53
	<i>F</i>	221923	68028	4064	17390
	<i>ER (%)</i>	0.85	1.56	4.36	2.12
	<i>n</i>	84	91	60	42
Ecuación (6.7)	<i>R</i>	0.99989	0.99917	0.9992	0.9993
	<i>SEP</i>	0.22	0.46	0.71	0.58
	<i>F</i>	306461	51589	33353	35160
	<i>ER (%)</i>	0.82	1.95	2.17	2.09
	<i>n</i>	84	91	60	50

^a *R* (coeficiente de correlación), *SEP* (error estándar en la predicción), *F* (F de Snedecor), *ER* (error relativo referido a la retención media) y *n* (número de medidas de retención).

^b El intervalo de hidrofobicidad de los compuestos de referencia era demasiado amplio para ajustar los modelos de Snyder y Schoenmakers (Ecuaciones (6.5) y (6.6)), por lo que fue necesario eliminar los tres compuestos más hidrofóbicos. Con el modelo NK, se comprobó el efecto de eliminar el gradiente más problemático (G4), manteniendo todos los compuestos.

Cabe señalar que la polaridad de los componentes de las plantas medicinales se extiende a valores extremos. Así, para los compuestos de referencia, k_0 en la Ecuación (6.7) oscila entre 10 y alrededor de 10^7 , mientras que los parámetros b y c adoptan valores mucho más pequeños (alrededor de 5 y 80, respectivamente). Estas grandes diferencias de escala entre k_0 y los otros dos parámetros (b y c) afectan al proceso de regresión y, por lo tanto, al rendimiento del ajuste y la velocidad de la convergencia. Dado que los cálculos utilizando la Ecuación (6.7) y regresión alterna son lentos, y su convergencia es complicada, es conveniente reformular la ecuación para que la magnitud de los parámetros no sea tan diferente entre sí. Una forma de lograr esto es utilizar valores logarítmicos de k_0 , en lugar de operar directamente con k_0 :

$$k_i = 10^{\log k_{0,i}} (1 + c\varphi)^2 e^{\frac{-b\varphi}{1+c\varphi}} \quad (6.8)$$

La Tabla 6.6 muestra los parámetros de ajuste para los modelos globales de retención, utilizando el modelo NK. Es interesante observar que cuando los datos de retención de los compuestos de referencia se ajustan individualmente, se producen inversiones ocasionales en la magnitud de sus parámetros específicos ($\log k_{0,i}$), considerando el orden de elución. Sin embargo, cuando los datos de retención se ajustan a los modelos globales, donde los parámetros columna/modificador (b y c) son comunes para todos los compuestos, la secuencia de valores $\log k_{0,i}$ para el modelo NK muestra una concordancia perfecta con el orden de elución.

Tabla 6.6. Coeficientes de regresión obtenidos tras aplicar regresión alterna a los datos de retención en gradiente, para las tres plantas medicinales, utilizando un modelo global basado en la ecuación de NK.

Planta medicinal	Té verde	Melisa	Tila
Parámetros comunes columna/modificador			
b	1.45	1.92	4.67
c	39.5	51.7	81.7
Parámetros individuales de los solutos			
$\log k_{0,1}$	0.97	1.58	2.05
$\log k_{0,2}$	1.40	2.14	2.67
$\log k_{0,3}$	1.91	2.75	2.95
$\log k_{0,4}$	2.01	3.18	3.31
$\log k_{0,5}$	2.39	3.62	3.54
$\log k_{0,6}$	2.44	3.86	3.89
$\log k_{0,7}$	2.79	4.21	4.10
$\log k_{0,8}$	2.94	4.58	5.11
$\log k_{0,9}$	5.05	5.27	5.37
$\log k_{0,10}$	5.94	5.85	5.62
$\log k_{0,11}$	6.12	6.18	–
$\log k_{0,12}$	6.24	6.41	–
$\log k_{0,13}$	–	6.76	–

Los diseños de entrenamiento para el té verde y la melisa proporcionaron en general buenos resultados. La reducida dispersión de los resultados, y la buena concordancia con la diagonal en los gráficos de correlación, indican un rendimiento satisfactorio y una ausencia de sesgo para las tres ecuaciones (Snyder, Schoenmakers y NK), ajustadas para el té verde y la melisa, pero no para la tila (Figuras 6.7d-6.7f). Con esta muestra, y haciendo uso del modelo NK, la dispersión fue más evidente, observándose además una disminución progresiva de los tiempos de retención predichos, en comparación con los valores experimentales. El sesgo es particularmente visible para el gradiente G4 (Figura 6.5c) y los componentes más hidrofóbicos. Este comportamiento puede explicarse por la extrema pendiente del gradiente tras el nodo en G4, lo que da lugar a un cierto retraso en el reequilibrado de la columna.

Se repitieron algunos ajustes, con el objetivo de comprobar el origen de las desviaciones para la tila, eliminando los compuestos más retenidos (solutos 8 a 10), o el gradiente más problemático (G4). Cabe señalar que, en la región donde eluyen los solutos 8 a 10, sólo existen unos pocos componentes. La Tabla 6.5 contiene una columna adicional con los resultados obtenidos después de eliminar los datos indicados. Se observa que las correlaciones para los modelos globales obtenidos con las ecuaciones de Snyder y Schoenmakers mejoran, proporcionando valores similares a los que ofrece el modelo NK. Sin embargo, es importante resaltar que este último modelo fue el único que proporcionó ajustes exactos, sin la necesidad de eliminar ningún dato (la mejora obtenida después de eliminar el gradiente G4 fue insignificante).

También se ha intentado construir modelos globales basados en una ecuación propuesta por Jandera [23]:

$$k_i = (a + b\varphi)^{-n_i} \quad (6.9)$$

Se ha encontrado que esta ecuación, desarrollada para cromatografía líquida de fase normal, ofrece un buen rendimiento de predicción para RPLC [24]. Sin embargo, con las plantas medicinales analizadas en este trabajo, la Ecuación (6.9) no mostró resultados satisfactorios. Cuando se ajustó individualmente para cada soluto, las predicciones fueron generalmente buenas (con $R > 0.999$), aunque fallaron para los solutos más hidrofóbicos. Por otro lado, al tratar de ajustar modelos globales basados en la ecuación de Jandera, el parámetro n_i (asociado a los solutos) adoptó valores no operables (raíces pares negativas), a lo largo de las iteraciones. Incluso evitando una convergencia incorrecta (caso de la tila), el rendimiento del ajuste global con el modelo de Jandera fue aún deficiente, en comparación a los otros tres modelos de retención ensayados (Snyder, Schoenmakers y NK), mostrando desviaciones significativas. Para el té verde y la melisa, fue imposible evitar la obtención de números complejos a lo largo de los cálculos.

6.6. Conclusiones

Esta investigación tiene como objetivo mejorar la calidad de la información obtenida de los cromatogramas de productos naturales, que contienen cientos de componentes, en ausencia de estándares o conteniendo picos de identidad desconocida. Como respuesta al problema de la modelización de muestras de extrema complejidad, se propone el uso de modelos globales y un tipo de diseño de gradiente de entrenamiento que facilita el análisis de muestras con componentes comprendidos en un intervalo extremo de hidrofobicidad y la coincidencia entre los picos para distintos gradientes. En la metodología descrita en este capítulo, se selecciona un número reducido de componentes en las muestras (a los que se ha denominado “compuestos de referencia”), para construir el modelo de retención. Un modelo exitoso requiere que, incluso si se

desconoce la identidad de los componentes de la muestra, su presencia pueda poderse rastrear correctamente en los cromatogramas obtenidos con todos los gradientes del diseño experimental.

Para poder realizar una asignación de picos correcta, se propone un diseño experimental que consiste en gradientes multi-lineales, que deben adaptarse a los tiempos de retención de los componentes de los extractos. Los gradientes incluyen dos segmentos lineales con un nodo de posición variable, así como las etapas necesarias para la limpieza y el reequilibrado de la columna, antes de la siguiente inyección. Aplicando el método de regresión alterna propuesto, se ha construido un modelo global de retención válido para los compuestos de referencia, que separa los efectos generales de la columna y el modificador sobre la retención, de los efectos asociados a la naturaleza química de cada compuesto eluido.

Se propone y valida una metodología para obtener modelos globales que proporciona resultados muy satisfactorios, utilizando tres muestras con diferentes características de separación (té verde, melisa y tila). El rendimiento es comparable al de los modelos individuales, a pesar de la gran dispersión en la hidrofobicidad de los solutos en los extractos de las plantas y el amplio intervalo de concentraciones del modificador, cubierto por los gradientes ensayados.

Entre las ecuaciones utilizadas para desarrollar modelos globales, tres de ellas resultaron adecuadas (Snyder, Schoenmakers y NK, particularmente la última), y una (propuesta por Jandera) tuvo que ser descartada. El modelo global basado en la ecuación de NK no presentó problemas de predicción con las plantas medicinales estudiadas, ofreciendo excelentes correlaciones de los tiempos de retención predichos frente a los experimentales, utilizando para ello datos provenientes de todos los gradientes y compuestos ensayados. Es

interesante recalcar que las predicciones con los modelos globales se realizan con un número de parámetros mucho menor que los involucrados en los modelos individuales (menos de la mitad). Además, no se requiere el uso de estándares.

El análisis de muestras altamente complejas debe llevarse a cabo idealmente a un valor de pH constante y controlado. En las separaciones isocráticas, la elución a pH constante no presenta ningún problema. Sin embargo, en la elución en gradiente, los valores de pK_a tanto del tampón como de los analitos con comportamiento ácido-base experimentan desplazamientos, debido a la variación en la composición del modificador de la fase móvil. Esto afecta al valor del pH a lo largo de la elución, así como a los tiempos de retención para aquellos solutos con un valor de pK_a cercano al pH instantáneo de la fase móvil, ya que su protonación puede verse afectada significativamente [25]. En este caso, las predicciones por sí mismas revelarán los solutos que probablemente están afectados por este problema. Sin embargo, en las tres muestras estudiadas, no fue destacable. Finalmente, cabe señalar que algunos picos que se apartan de las posiciones esperadas entre cientos bien pronosticados, en una muestra en la que se intenta mejorar la separación general, no son motivo de preocupación.

En el Capítulo 7, se amplía el desarrollo de los modelos globales de retención basados en la ecuación NK, para los compuestos de referencia, con el fin de incluir todos los componentes de las plantas medicinales (los compuestos de referencia y cualquier otro que dé lugar a picos visibles). Estas características, junto con la posibilidad de desarrollar modelos globales que describan la anchura de banda, permiten la predicción de cromatogramas completos con cientos de componentes, en condiciones de gradiente distintas a las utilizadas para entrenar los modelos globales. De esta forma se obtiene una notable coincidencia con los cromatogramas reales.

6.7. Referencias

- [1] J.R. Torres Lapasió, M.C. García Álvarez-Coque, Liquid chromatography: Strategies for optimization, en *Encyclopedia of Analytical Science* (editado por P.J. Worsfold, A. Townshend, C. Poole, M. Miró), 3^a ed., Elsevier, Amsterdam, 2019, págs. 252–260.
- [2] M.C. García Álvarez-Coque, J.R. Torres Lapasió, J.J. Baeza Baeza, Models and objective functions for the optimisation of selectivity in reversed-phase liquid chromatography, *Anal. Chim. Acta* 579 (2006) 125–145.
- [3] J.R. Torres Lapasió, M.C. García Álvarez-Coque, Levels in the interpretive optimisation of selectivity in high-performance liquid chromatography: A magical mystery tour, *J. Chromatogr. A* 1120 (2006) 308–321.
- [4] C. Tistaert, B. Dejaegher, Y. Vander Heyden, Chromatographic separation techniques and data handling methods for herbal fingerprints, *Anal. Chim. Acta* 690 (2011) 148–161.
- [5] R. Govindarajan, V. Tejas, P. Pushpangadan, High-performance liquid chromatography (HPLC) as a tool for standardization of complex herbal drugs, *J. AOAC Int.* 102 (2019) 993–1002.
- [6] G. Jin, X. Xue, F. Zhang, X. Zhang, Q. Xu, Y. Jin, X. Liang, Prediction of retention times and peak shape parameters of unknown compounds in traditional Chinese medicine under gradient conditions by ultra-performance liquid chromatography, *Anal. Chim. Acta* 628 (2008) 95–103.

-
- [7] A. Gisbert Alonso, J.A. Navarro Huerta, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Global retention models and their application to the prediction of chromatographic fingerprints, *J. Chromatogr. A* 1637 (2021) 461845.
- [8] E. Bosch, P. Bou, M. Rosés, Linear description of solute retention in reversed-phase liquid chromatography by a new mobile phase polarity parameter, *Anal. Chim. Acta* 299 (1994) 219–229.
- [9] J.R. Torres Lapasió, M.C. García Álvarez-Coque, M. Rosés, E. Bosch, Prediction of the retention in reversed-phase liquid chromatography using solute-mobile phase-stationary phase polarity parameters, *J. Chromatogr. A* 955 (2002) 19–34.
- [10] L.R. Snyder, J.J. Kirkland, J.L. Glajch, Practical HPLC Method Development, 2^a ed., John Wiley & Sons, Nueva York, 1997.
- [11] H.B. Xiao, X.M. Liang, R.C. Lu, Classification of structurally related compounds from Astragalus extract by correlation of the log k_w and S, *Chromatographia* 51 (2000) 212–220.
- [12] C.F. Poole, The Essence of Chromatography, Elsevier, Amsterdam, 2003, págs. 303–304.
- [13] J.A. Navarro Huerta, E.J. Carrasco Correa, J.R. Torres Lapasió, J.M. Herrero Martínez, M.C. García Álvarez-Coque, Modelling retention and peak shape of small polar solutes analysed by nano-HPLC using methacrylate-based monolithic columns, *Anal. Chim. Acta* 1086 (2019) 142–155.
- [14] J.A. Navarro Huerta, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Estimation of peak capacity based on peak simulation, *J. Chromatogr. A* 1574 (2018) 101–113.
-

- [15] P.J. Schoenmakers, H.A.H. Billiet, R. Tussen, L. de Galan, Gradient selection in reversed-phase liquid chromatography, *J. Chromatogr. A* 149 (1978) 519–537.
- [16] T. Álvarez Segura, E. Cabo Calvet, J.R. Torres Lapasió, M.C. García Álvarez-Coque, An approach to evaluate the information in chromatographic fingerprints: Application to the optimisation of the extraction and conservation conditions of medicinal herbs, *J. Chromatogr. A* 1422 (2015) 178–185.
- [17] U.D. Neue, H.J. Kuss, Improved reversed-phase gradient retention modeling, *J. Chromatogr. A* 1217 (2010) 3794–3803.
- [18] S. López Ureña, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Enhancement in the computation of gradient retention times in liquid chromatography using root-finding methods, *J. Chromatogr. A* 1600 (2019) 137–147.
- [19] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics*, Elsevier, Amsterdam, 1998.
- [20] J.A. Navarro Huerta, A. Gisbert Alonso, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Testing experimental designs in liquid chromatography (I): Development and validation of a method for the comprehensive inspection of experimental designs, *J. Chromatogr. A* 1624 (2020) 461180.
- [21] A. Gisbert Alonso J.A. Navarro Huerta, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Testing experimental designs in liquid chromatography (II): Influence of the design geometry on the prediction performance of retention models, *J. Chromatogr. A* 1654 (2021) 462458.

- [22] G. Vivó Truyols, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Error analysis and performance of different retention models in the transference of data from/to isocratic/gradient elution, *J. Chromatogr. A* 1018 (2003) 169–181.
- [23] P. Jandera, T. Hájek, M. Růžicková, Retention models on core-shell columns, *J. AOAC International* 100 (2017) 1636–1646.
- [24] J.J. Baeza Baeza, M.C. García Álvarez-Coque, Extension of the linear solvent strength retention model including a parameter that describes the elution strength changes in liquid chromatography, *J. Chromatogr. A* 1615 (2020) 460757.
- [25] A. Andrés, M. Rosés, E. Bosch, Gradient retention prediction of acid-base analytes in reversed phase liquid chromatography: A simplified approach for acetonitrile-water mobile phases, *J. Chromatogr. A* 1370 (2014) 129–134.

CAPÍTULO 7

**ANÁLISIS DE HUELLAS DACTILARES
CROMATOGRÁFICAS DE EXTRACTOS DE
TÉ VERDE, MELISA Y TILA:
SIMULACIÓN DE CROMATOGRAMAS HACIENDO USO
DE MODELOS GLOBALES DE RETENCIÓN**

7.1. Resumen

Las plantas medicinales contienen una gran variedad de compuestos químicos en concentraciones muy variables, por lo que su control de calidad es especialmente complejo. Con este propósito, las instituciones reguladoras han aceptado las huellas dactilares cromatográficas como una herramienta válida para realizar los análisis. Para mejorar el nivel informativo de los cromatogramas, se necesitan condiciones de separación que maximicen el número de picos detectados. Este capítulo propone la extensión de una estrategia de simulación a todos los picos detectados en el cromatograma, basada en modelos globales de retención desarrollados previamente para compuestos seleccionados (Capítulo 6),. Los modelos globales contienen parámetros específicos para cada componente de la muestra, mientras que otros parámetros son comunes a todos los componentes y describen los efectos combinados de la columna y el modificador. El tratamiento propuesto se inicia mediante la detección y medida automática de la posición de todos los picos en un cromatograma, obtenido preferentemente con el gradiente más lento de entre los utilizados en el diseño experimental. Seguidamente, se ajusta el tiempo de retención para cada componente detectado, con el fin de encontrar su parámetro específico en el modelo global que conduzca a la mejor concordancia con el valor experimental medido. El proceso se completa con el desarrollo de modelos de anchura de banda para los compuestos seleccionados utilizados para construir el modelo global de retención, basado en datos de gradiente, que se aplican a todos los picos del cromatograma. La utilidad de la metodología de simulación se demuestra mediante la predicción de las huellas dactilares cromatográficas de tres plantas medicinales que poseen problemas de separación específicos (té verde, melisa y tila), obtenidas utilizando varios gradientes multi-lineales que dan lugar a predicciones problemáticas.

7.2. Introducción

En la medicina tradicional, los preparados derivados de plantas medicinales se han utilizado desde hace miles de años en la prevención y tratamiento de enfermedades. Su actividad terapéutica se debe a la presencia de compuestos químicos biológicamente activos, que pueden actuar sinérgicamente [1,2]. Debido a la eficacia de los tratamientos basados en estos productos naturales y a su baja toxicidad, su uso se ha extendido en gran medida en los últimos años [3]. La calidad de las plantas medicinales puede verse afectada por varios factores, como son el tipo de suelo, su ubicación geográfica, las condiciones ambientales durante su crecimiento, la temporada y los métodos utilizados en su recogida, las condiciones de almacenamiento y los procedimientos empleados en su preparación. Por lo tanto, los productos deben seguir un control de calidad que certifique al consumidor su seguridad y eficacia farmacológica. Sin embargo, la gran diversidad química de los compuestos naturales, que se hallan en concentraciones muy diferentes, hace que el control de calidad sea extremadamente complicado [2]. Para solucionar los problemas encontrados en el control sanitario de las plantas medicinales, debido a su compleja composición, la Organización Mundial de la Salud (OMS), la Administración de Drogas y Alimentos de los Estados Unidos (FDA, *Food and Drug Administration*), y la Administración Estatal de Alimentos y Drogas de China, han reconocido a las huellas dactilares cromatográficas como una herramienta válida para garantizar su calidad [4–7].

Probablemente, el aspecto más problemático que impide el desarrollo de métodos para optimizar la resolución de los compuestos existentes, en este tipo de cromatogramas, es encontrar modelos de retención que describan a todos los componentes de las muestras, en situaciones en las que no existan estándares [8–11]. En el Capítulo 6, se ha desarrollado una metodología para describir el

comportamiento de retención de compuestos desconocidos en un cromatograma, mediante el uso de modelos globales [12,13]. El propósito es obtener un conjunto de parámetros que permitan predecir el comportamiento de un grupo de compuestos (conocidos o desconocidos), como alternativa al uso de parámetros enfocados a cada compuesto. En los modelos globales, algunos parámetros son específicos de cada soluto, mientras que otros describen los efectos combinados de la columna y el modificador, comunes para todos los solutos.

La propuesta que se presenta en este capítulo consiste en que, una vez obtenidos los cromatogramas de la muestra de acuerdo a un determinado diseño experimental (véase el Capítulo 6), se seleccionan los picos de varios compuestos (a los que se ha denominado “picos de referencia”), con el fin de obtener la información cromatográfica necesaria para construir el modelo global. Para ello no es necesario conocer la identidad de los compuestos que generan los picos.

Existen algunas reglas para la selección de los picos de referencia, siendo indispensable la condición de que los picos equivalentes sean fácilmente reconocibles en todos los gradientes. Por ejemplo, los picos de referencia pueden ser picos que destaquen de los demás por su intensidad o posición, o que den lugar a patrones fácilmente identificables respecto a sus picos vecinos, para que puedan ser trazables en todos los gradientes de entrenamiento. La presencia de valores atípicos o una dispersión anormal en los gráficos de correlación, obtenidos para los modelos individuales, permiten revelar posibles errores en la identificación de los picos.

El Capítulo 6 [13] desarrolla la construcción de modelos globales de retención para los compuestos seleccionados como referencia, en las huellas dactilares cromatográficas de extractos de algunas plantas medicinales,

utilizando la información obtenida al aplicar diseños experimentales apropiados. Los diseños aplicados se desarrollaron a partir de un gradiente de exploración lineal común y consisten en varios gradientes multi-lineales relacionados, lo que facilita el seguimiento de los picos [14].

En el presente estudio, los modelos globales de retención para los compuestos de referencia se amplían para incluir el resto de componentes en el cromatograma, que den lugar a picos detectables. La información requerida para actualizar el modelo global de retención se obtiene preferentemente del cromatograma correspondiente a la condición experimental más lenta del diseño de entrenamiento, tras la corrección de la línea base [15]. El modelo extendido que incluye a todos los picos detectables en el cromatograma permite predecir cromatogramas completos para cualquier nueva condición experimental arbitraria. Por su parte, la construcción de modelos de anchura de banda, para los compuestos de referencia, permite realizar predicciones completas de cromatogramas obtenidos mediante elución en gradiente. La metodología desarrollada se aplicó a la simulación de cromatogramas de extractos de *Camellia sinensis* (té verde), *Melissa officinalis* (melisa) y *Tilia platyphyllos* (tila), con resultados satisfactorios.

7.3. Teoría

7.3.1. Modelos globales de retención para los compuestos de referencia

La metodología propuesta en este capítulo, para simular huellas dactilares cromatográficas, requiere el ajuste previo de un modelo global para un conjunto de compuestos seleccionados con picos distribuidos a lo largo del cromatograma (los llamados “compuestos de referencia”). No es necesario conocer la naturaleza química de dichos compuestos, pero su identidad debe

establecerse inequívocamente en los cromatogramas obtenidos con todos los gradientes de entrenamiento. Además, los picos deben ser lo suficientemente intensos para efectuar una detección adecuada en condiciones de baja fuerza eluyente.

Las pautas para seleccionar los picos de los compuestos de referencia se proporcionan en la Sección 6.5.4 del Capítulo 6. Allí se compara también el rendimiento de los modelos globales de retención basados en las ecuaciones propuestas por Snyder [17], Schoenmakers [18] y Neue-Kuss [19]. Entre ellas, la ecuación de Neue-Kuss (NK):

$$k_i = k_{0,i} (1 + c\varphi)^2 e^{\frac{-b\varphi}{1+c\varphi}} \quad (7.1)$$

ofreció los mejores resultados. Por lo tanto, sólo esta ecuación será considerada para la simulación de las huellas dactilares cromatográficas, siendo reformulada como:

$$k_i = 10^{\log k_{0,i}} (1 + c\varphi)^2 e^{\frac{-b\varphi}{1+c\varphi}} \quad (7.2)$$

para obtener parámetros del modelo menos dispares en escala de magnitud, lo que facilita la convergencia [13].

El modelo global se puede representar mediante el vector $[b, c, \log k_{0,1}, \log k_{0,2}, \dots, \log k_{0,ns}]$, donde b y c son los parámetros comunes que describen a la columna y al modificador, y $\log k_{0,i}$, los parámetros específicos de los solutos. Las etapas necesarias para ajustar el modelo global se describen brevemente a continuación (consúltese el Capítulo 6 para obtener más detalles):

- (i) En primer lugar, los datos de retención para cada compuesto de referencia i se ajustan individualmente a la Ecuación (7.2). De ese modo, se obtienen los valores de los parámetros b_i , c_i y $\log k_{0,i}$, utilizándose el

- conjunto completo de tiempos de retención experimentales, medidos con todos los gradientes de entrenamiento.
- (ii) Con el conjunto de valores de los parámetros b_i y c_i ajustados para cada soluto en la etapa (i), se obtienen las medianas (b_m y c_m) de los parámetros que describen el comportamiento asociado a la columna y al modificador. Estos valores se toman como estimaciones iniciales de los parámetros globales, mientras que los valores de $\log k_{0,i}$ para cada compuesto i se obtienen mediante un ajuste.
 - (iii) A continuación, se vuelven a ajustar los parámetros b y c , esta vez manteniendo fijos los valores de $\log k_{0,i}$ encontrados en la etapa anterior, y atendiendo simultáneamente a la predicción de los valores para todos los solutos y gradientes de entrenamiento.
 - (iv) Finalmente, todos los parámetros que definen el vector $[b, c, \log k_{0,1}, \log k_{0,2}, \dots, \log k_{0,ns}]$ del modelo global de retención se optimizan utilizando todos los datos disponibles.
 - (v) Si es necesario, se repite el proceso a partir de la etapa (ii) hasta conseguir la convergencia.

7.3.2. Extensión de los modelos globales de retención a la totalidad de los picos detectados en las plantas medicinales

Los modelos globales de retención obtenidos con los compuestos de referencia permiten predicciones que involucran exclusivamente a los compuestos de referencia, para cualquier gradiente arbitrario. Sin embargo, el objetivo de esta investigación es la predicción de cromatogramas completos para las plantas medicinales, que pueden incluir varios cientos de compuestos. Por lo tanto, se desarrolló una metodología que permite extender los modelos globales, ajustados con los datos de los compuestos de referencia, a la

predicción de la retención para todos los picos detectados en los cromatogramas.

El modelo global de retención, establecido inicialmente con los compuestos de referencia, se modificó para incluir otros componentes en el cromatograma, de la siguiente manera:

- (i) En primer lugar, se selecciona un cromatograma obtenido con un gradiente perteneciente al diseño de entrenamiento, preferentemente aquél con el mayor número de picos detectables, que suele ser el correspondiente al gradiente que muestra la menor pendiente inicial en el diseño. Previamente a su procesamiento, la línea base se sustrae del cromatograma experimental utilizando un algoritmo adecuado. A este cromatograma se le ha denominado “cromatograma base”.
- (ii) A continuación, se mide la posición de todos los picos detectados en el cromatograma base, utilizando una función de análisis automático. Los picos medidos son aquéllos que superan ciertos umbrales de aceptabilidad, como una altura crítica o una determinada anchura de banda. Para ello, se aplicó el software de autodetección desarrollado por el grupo de investigación [16].
- (iii) Seguidamente, se obtienen los tiempos de retención de todos los picos detectados ($t_{R,i}$) (los picos de referencia o cualquier otro que supere los umbrales de detección), junto con otras medidas que definen el perfil de pico (anchuras de banda y áreas).
- (iv) El proceso seguido para extender el modelo global, a todos los picos detectados, consiste en un ajuste por mínimos cuadrados, donde los parámetros asociados a la columna y al modificador (b y c) se mantienen fijos a los valores encontrados con los picos de referencia, mientras que los parámetros específicos ($\log k_{0,i}$, relacionados con la hidrofobicidad de

los solutos) describen los tiempos de retención experimentales ($t_{R,i}$) para todos los solutos (compuestos de referencia o cualquier otro del cromatograma), cuando eluyen con el gradiente asociado con el cromatograma base.

- (v) Con esta información (b , c y $\log k_{0,i}$), se puede predecir el cromatograma para cualquier otro gradiente arbitrario.

Siguiendo este protocolo, el efecto del modificador se evalúa con varios gradientes que poseen perfiles muy distintos, utilizando solutos representativos, mientras que el efecto de la hidrofobicidad del soluto (que idealmente debería ser independiente de la fase móvil) se obtiene solamente con el gradiente del diseño que muestra un número máximo de picos. Se consigue así un vector que recoge los parámetros del modelo global [b , c , $\log k_{0,1}$, $\log k_{0,2}$...]. Este vector se puede reorganizar en una colección de vectores [b , c , $\log k_{0,i}$] de menor tamaño, cada uno de ellos asociado al modelo de retención individual para el soluto i .

Para acelerar y favorecer la convergencia del modelo global ampliado, se ensayaron varias opciones. La mejor consistió en realizar un ajuste secuencial, en el que los parámetros específicos de los solutos se determinan soluto a soluto, en orden decreciente de hidrofobicidad, de modo que el valor $\log k_0$ encontrado para el soluto i se emplee como una estimación inicial para el soluto $i - 1$. Este modo de operar acelera considerablemente el proceso de regresión e incrementa las posibilidades de obtener un buen ajuste en un solo intento. Otras opciones que se probaron con menor éxito fueron la realización de:

- (i) ajustes independientes usando la misma estimación inicial ($\log k_0$) para todos los solutos, y
- (ii) ajustes secuenciales, en los que la solución encontrada para el soluto i se aplicó en orden creciente de hidrofobicidad.

7.3.3. Modelos globales de anchura de pico para los compuestos de referencia

Para ser realista y práctica, la simulación de cromatogramas requiere, no sólo la predicción de la ubicación de los picos para cada componente de la muestra, a medida que cambian las condiciones de elución, sino también el valor de las anchuras de banda de los picos. Aunque algunos picos presentan anchuras de banda anómalas, a menudo debido a la co-elución parcial de compuestos químicos, u otros fenómenos, lo que realmente importa es que la mayoría de los picos en las huellas dactilares cromatográficas estén bien predichos.

En este trabajo, se simularon los perfiles de los picos cromatográficos utilizando un modelo gaussiano modificado (PMG1), donde la desviación estándar depende de la distancia al tiempo de retención [20,21]:

$$h(t) = h_0 \exp \left[-\frac{1}{2} \left(\frac{t - t_R}{s_0 + s_1 (t - t_R)} \right)^2 \right] \quad (7.3)$$

donde $h(t)$ es la altura en el tiempo t y h_0 la altura máxima del pico; s_0 es una medida de la anchura de pico establecida sobre una base gaussiana, y s_1 representa la distorsión del pico. Los coeficientes de la función lineal pueden calcularse fácilmente de la siguiente manera, utilizando los valores predichos de las semianchuras:

$$s_0 = 0.932 \frac{A_{0.1} B_{0.1}}{A_{0.1} + B_{0.1}} \quad (7.4)$$

$$s_1 = 0.466 \frac{B_{0.1} - A_{0.1}}{A_{0.1} + B_{0.1}} \quad (7.5)$$

El crecimiento anormal del pico obtenido con el modelo PMG1 se resuelve usando una función híbrida que combina la función PMG1 (Ecuación (7.3)), con decaimientos exponenciales a ambos lados del pico:

$$h = k_{1,izq} \exp\{k_{2,izq}(t - t_R)\} \quad \text{for } t < t_R - A_{0.1} \quad (7.6)$$

$$h = k_{1,der} \exp\{k_{2,der}(t - t_R)\} \quad \text{for } t > t_R + B_{0.1} \quad (7.7)$$

sujeto a la restricción de que las pendientes de las funciones gaussiana y exponencial, en los respectivos puntos de conexión, deben coincidir.

Los parámetros del modelo se calculan de la siguiente manera:

$$k_{1,izq} = 0.1 h_0 \exp(k_{2,izq} A_{0.1}) ; \quad k_{2,izq} = \frac{s_0 A_{0.1}}{(s_0 - c_1 A_{0.1})^3} \quad (7.8)$$

$$k_{1,der} = 0.1 h_0 \exp(k_{2,der} B_{0.1}) ; \quad k_{2,der} = -\frac{s_0 B_{0.1}}{(s_0 + c_1 B_{0.1})^3} \quad (7.9)$$

Los parámetros del modelo gaussiano pueden relacionarse con el tiempo de retención, el área y las anchuras (o semianchuras). A su vez, las anchuras de pico se pueden correlacionar con los tiempos de retención, dando lugar a una familia de modelos globales basados en la generalización del concepto de eficacia cromatográfica (N) [22–24]. Los modelos de anchura de pico describen la tendencia al ensanchamiento de los picos cromatográficos, a medida que aumenta el tiempo de retención. En este capítulo, la medida de las anchuras de pico se realizó cuando la señal alcanzó un 10% de la altura máxima del pico.

Si los datos iniciales son isocráticos, las anchuras de pico experimentales se correlacionan directamente con los respectivos tiempos de retención. Se suelen obtener tendencias parabólicas [23]:

$$w = \omega_0 + \omega_1 t_{iso} + \omega_2 t_{iso}^2 \quad (7.10)$$

pero puede considerarse a menudo un comportamiento lineal. En la Ecuación (7.10), w puede ser la anchura de pico o las semianchuras izquierda o derecha, y t_{iso} es el tiempo de retención isocrático.

En elución en gradiente, la relación entre las anchuras de pico y el tiempo de retención no es directa. Sin embargo, se puede conseguir suficiente precisión aplicando la aproximación de Jandera [25], aunque sólo es estrictamente válida para gradientes lineales. Esta aproximación postula que, utilizando elución en gradiente, la anchura de pico para un soluto i es la misma que la que se obtendría si migrara isocráticamente empleando una fase móvil de composición j , que se correspondería con la composición en el instante alcanzado por el gradiente j cuando el soluto abandona la columna. Aunque los datos primarios provienen de experimentos en gradiente, la predicción de los tiempos de retención en gradiente proporciona de manera colateral la composición instantánea cuando el soluto abandona la columna y, por lo tanto, es posible calcular los tiempos de retención equivalentes si la elución fuera isocrática. El tiempo isocrático correspondiente al valor j se denominará: “tiempo isocrático equivalente”.

La secuencia de operaciones necesarias para obtener los parámetros de los modelos globales de anchura de pico (ω_0 , ω_1 y ω_2 en la Ecuación (7.10)) es la siguiente:

- (i) Los datos de retención para cada soluto y gradiente se calculan resolviendo la ecuación fundamental para elución en gradiente [26–28], ya sea utilizando integración analítica o numérica. Una vez encontrado el tiempo a lo largo del gradiente que hace que la suma de las integrales coincida con el tiempo muerto, se obtiene colateralmente la composición instantánea j a la que el soluto abandona la columna.

- (ii) El tiempo de retención isocrático equivalente (al que cada soluto dejaría la columna si migrara a la composición j) se puede determinar sustituyendo la composición en el modelo de retención (por ejemplo, en la Ecuación (7.2)).
- (iii) La anchura de pico para el soluto i bajo el gradiente j se obtiene directamente, introduciendo t_{iso} en la Ecuación (7.10).

Finalmente, el modelo global de anchura de banda se ajusta modulando los parámetros en la Ecuación (7.10), tratando de obtener la mayor coincidencia entre las anchuras observadas y las predicciones correspondientes, utilizando los compuestos de referencia y todos los gradientes de entrenamiento.

7.4. Parte experimental

7.4.1. Preparación de extractos de plantas medicinales

Se estudió la separación mediante cromatografía líquida de fase inversa (RPLC, *reversed-phase liquid chromatography*) de extractos de tres plantas medicinales (té verde, melisa y tila). La melisa y la tila se compraron a granel en una tienda local, mientras que el té verde se adquirió en bolsas individuales en un supermercado. Los extractos de las tres plantas se procesaron siguiendo las recomendaciones de Álvarez-Segura *et al.* [16]. Debido a la heterogeneidad de las muestras, se trituraron porciones secas de cada planta. Se pesó un gramo del polvo resultante y se transfirió a un tubo Falcon, al que se añadió 15 mL de una disolución preparada con agua nanopura (Adrona B30 Trace, Burladingen, Alemania), y metanol al 70% (v/v) (Scharlau, Barcelona). El contenido del tubo Falcon se sonicó durante 60 min a 80 °C. Finalmente, la disolución se centrifugó a 3000 rpm durante 5 min.

7.4.2. Separación cromatográfica

Se tomó el sobrenadante del tubo Falcon con una jeringa y se filtró a través de una membrana de Nylon, con un tamaño de poro de 0.45 μm (Micron Separations, Westboro, MA, EE.UU.), transfiriéndose a un vial antes de la inyección. La separación se realizó mediante elución en gradiente con mezclas hidro-orgánicas, preparadas con agua nanopura y acetonitrilo grado HPLC (Scharlau), ambos conteniendo 0.1 % (v/v) de ácido fórmico (Acros Organics, Fair Lawn, NJ). La monitorización de los picos se llevó a cabo entre 210 y 280 nm, con incrementos de 10 nm. Otros detalles se proporcionan en el Capítulo 6.

Para establecer el intervalo de trabajo de acetonitrilo en el diseño experimental para cada planta medicinal, se utilizó un gradiente de exploración preliminar donde la concentración del modificador se incrementó linealmente del 5 al 100 % (v/v) en 60 min [13]. Se propusieron conjuntos de gradientes de entrenamiento atendiendo a la distribución de los picos en los cromatogramas obtenidos con el gradiente de exploración. Todos los gradientes incluyeron etapas adicionales necesarias para la limpieza de la columna para eliminar los componentes más hidrofóbicos y efectuar el reequilibrado antes de la siguiente inyección.

Para cada planta medicinal, se utilizó un diseño experimental de entrenamiento que constaba de 6 o 7 gradientes multi-lineales, con un nodo intermedio de posición variable (Figura 7.1). Estos diseños permitieron explorar un amplio intervalo de composiciones, sin dar lugar a tiempos de retención excesivos para los componentes más hidrofóbicos, ni demasiado cortos para los más hidrofílicos.

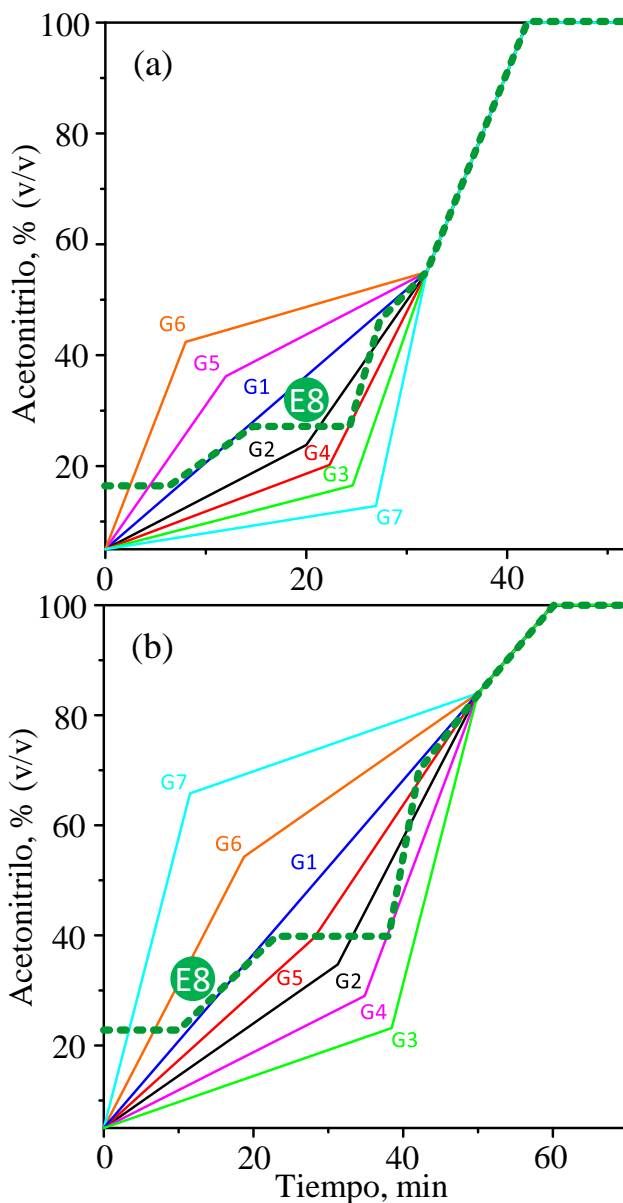


Figura 7.1. Gradientes de entrenamiento (G) y validación (E), utilizados para obtener los modelos globales y evaluar la exactitud de las predicciones de las huellas dactilares cromatográficas, respectivamente, para: (a) té verde, (b) melisa y (c) tila.

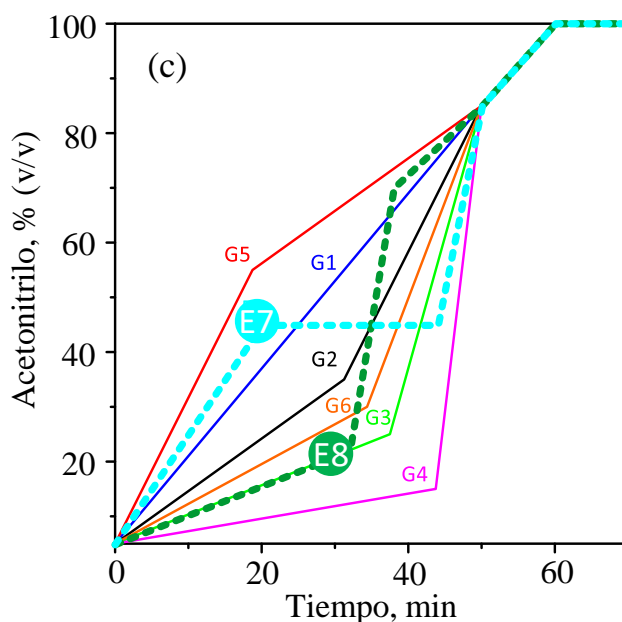


Figura 7.1 (continuación).

Una última ventaja es que este tipo de diseños facilita la trazabilidad de los picos de los compuestos de referencia cuando se varían las condiciones de elución. En la Figura 7.1, puede observarse que los intervalos de concentración del modificador, cubiertos por los gradientes de cada planta medicinal, son bastante distintos, reflejando las diferencias en la naturaleza de los componentes de cada muestra, y en consecuencia, en la distribución de los picos cromatográficos.

La construcción del diseño experimental de entrenamiento para cada tipo de muestra, así como otros detalles relacionados con la separación cromatográfica, aparecen en el Capítulo 6. Para verificar el rendimiento de predicción de los modelos globales, se utilizaron varios gradientes no incluidos en el diseño experimental (gradientes de validación etiquetados con la letra E en la Figura 7.1).

7.4.3. Software

Todo el tratamiento de datos se realizó con Matlab 2020a (The MathWorks Inc., Natick, MA, EE.UU.). La sustracción de la línea base en los cromatogramas experimentales se realizó con el algoritmo BEADS (*Baseline Estimation and Denoising using Sparsity*) [15]. La detección y medición automática de picos se llevó a cabo utilizando funciones de Matlab desarrolladas previamente por el grupo de investigación [16]. Estas funciones analizan automáticamente las señales, previa sustracción de la línea base, para localizar los picos y obtener los valores de tiempos de retención, semianchuras y áreas de pico, junto con otra información adicional.

7.5. Resultados y discusión

7.5.1. Medida de las señales cromatográficas

Como se ha indicado en la Sección 7.4.2, la monitorización de los picos se llevó a cabo en el intervalo de longitudes de onda entre 210 y 280 nm (utilizando ocho canales de adquisición separados entre sí por 10 nm). La longitud de onda de detección se seleccionó de acuerdo a dos enfoques. El primero hizo uso del “cromatograma total”, donde la absorbancia máxima en un determinado intervalo de longitud de onda se mide frente al tiempo de retención. Este cromatograma se puede procesar y utilizar como un cromatograma convencional. En el segundo enfoque, se seleccionó una longitud de onda de compromiso que equilibraba la detectabilidad y el ruido. Se prefirió este segundo enfoque, encontrándose que la longitud de onda más adecuada era 230 nm. A valores más elevados, los cromatogramas mostraban menos picos (la absorción era más selectiva), mientras que por debajo de 230

nm el ruido de fondo fue demasiado elevado, dificultando así el seguimiento de los picos.

Antes de procesar los cromatogramas, se eliminó la línea base mediante una función de Matlab, desarrollada previamente por el grupo de investigación, que automatiza y aplica el algoritmo BEADS [15]. BEADS realiza una descomposición de la señal basada en la frecuencia para obtener tres contribuciones: línea base, ruido y señal neta. El software construido en el laboratorio aplica el algoritmo de una manera muy flexible, permitiendo así un tratamiento muy satisfactorio para cromatogramas altamente complejos.

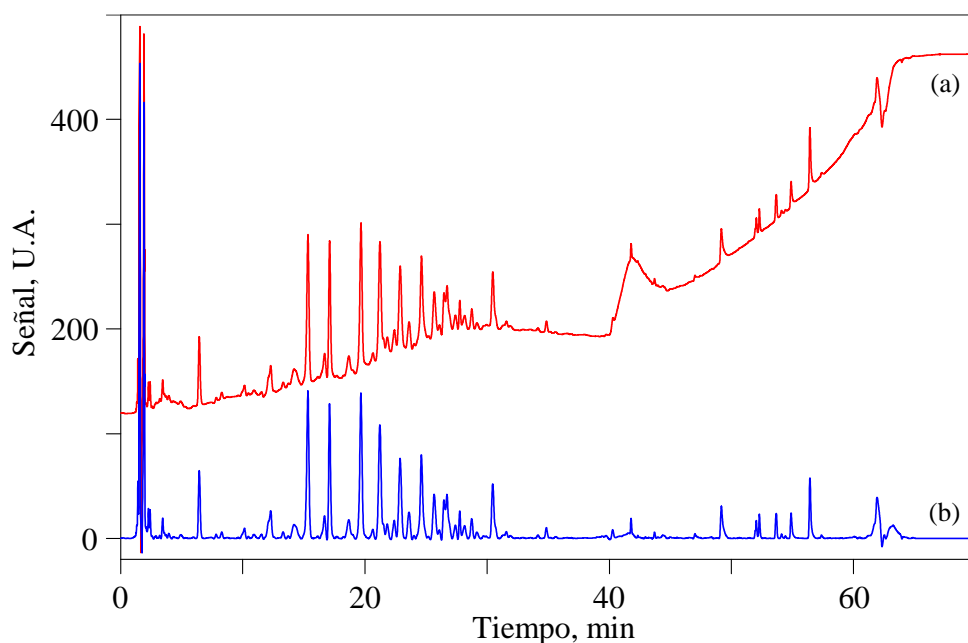


Figura 7.2. Cromatograma obtenido para el extracto de tila, utilizando el gradiente G3 (véase la Figura 7.1), antes (a) y después (b) de la sustracción de la línea base con el algoritmo BEADS asistido.

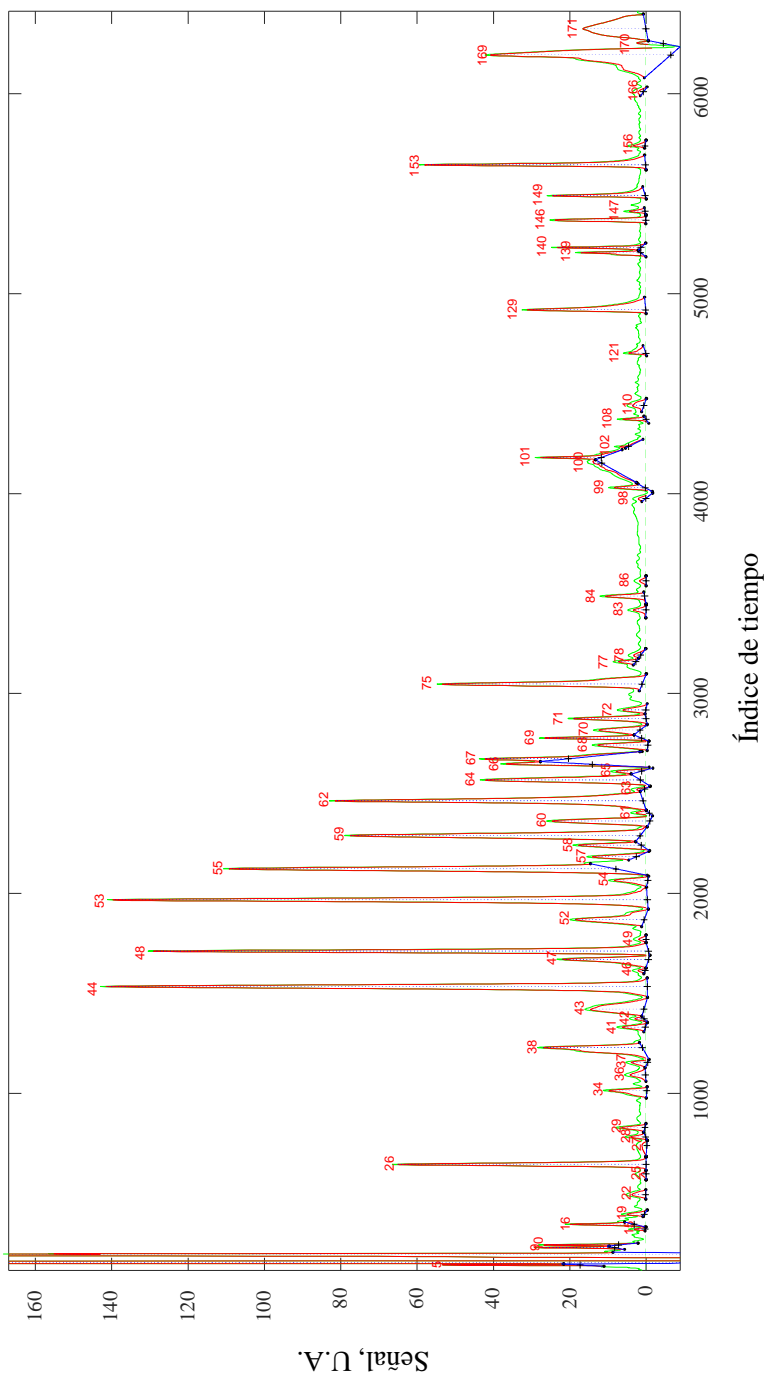


Figura 7.3. Análisis de detección de picos realizado con el algoritmo automático desarrollado por el grupo de investigación, para una de las réplicas de las huellas dactilares obtenidas con el gradiente G3, para la tía, previa sustracción de la línea base. El eje de abscisas corresponde a los índices del vector tiempo (frecuencia de adquisición de datos de cinco puntos por segundo).

La Figura 7.2 muestra un cromatograma representativo del extracto de tila, obtenido con el gradiente G3 (véase la Figura 7.1c). Como puede observarse, el algoritmo BEADS consigue una supresión adecuada de la línea base, eliminando casi por completo la perturbación asociada con el aumento repentino de la pendiente del gradiente a los 40 min.

La Figura 7.3 muestra el cromatograma del extracto de tila, una vez procesado por el algoritmo de detección automática, tras eliminar la línea base. Las señales simuladas tienen en cuenta el tamaño real del pico, que se midió automáticamente con la función de Matlab desarrollada para el análisis de señales.

7.5.2. Construcción de modelos globales de anchura de pico para simular cromatogramas

Como se ha comentado, la simulación de cromatogramas requiere, además de la disponibilidad de modelos de retención (consúltese la Sección 7.3.2), la construcción de modelos de anchura de pico para describir los perfiles de pico de los componentes de la muestra. En este capítulo, las anchuras de pico se predicen de acuerdo a su correlación con los tiempos de retención isocráticos (Sección 7.3.3). Sin embargo, no existe una correspondencia directa entre las anchuras de pico y los tiempos de retención obtenidos en elución en gradiente; por lo tanto, debe establecerse una relación interna con los tiempos que experimentaría el soluto si migrara isocráticamente, a la composición del modificador cuando éste abandona la columna utilizando un determinado gradiente (a los que nos referimos como tiempos isocráticos equivalentes).

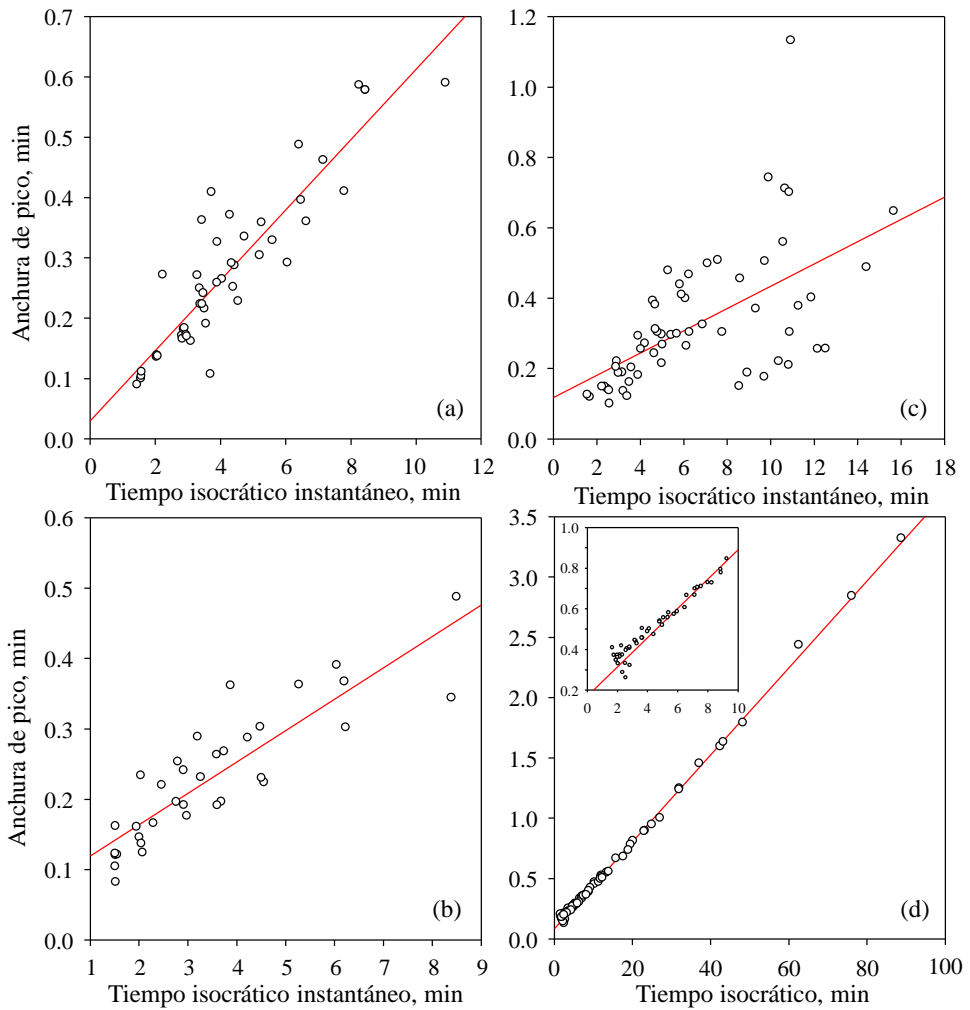


Figura 7.4. Gráficas de anchura de pico para: (a) té verde, (b) melisa, (c) tila, y (d) un conjunto de sulfonamidas. Véase el texto para más detalles.

La Sección 7.3.3 describe el protocolo para obtener los parámetros ω_0 , ω_1 y ω_2 del modelo global de anchura de pico (Ecuación (7.10)), basado en datos de gradiente. De manera similar a los datos isocráticos, las anchuras de pico de un conjunto de compuestos eluidos utilizando varios gradientes ofrecen una tendencia parabólica, cuando se representan frente a los tiempos de retención isocráticos equivalentes. Las Figuras 7.4a-7.4c muestran las tendencias de anchura para los picos de los compuestos de referencia, en los cromatogramas de los extractos de las tres plantas medicinales. Los datos representados en dichas figuras corresponden al conjunto total de compuestos de referencia, eluidos empleando todos los gradientes de los correspondientes diseños de entrenamiento. Con fines comparativos, los datos de anchura de pico para algunos compuestos relacionados estructuralmente (un conjunto de sulfonamidas), eluidos bajo elución isocrática, se han representado en la Figura 7.4d. Como se mostrará, los gráficos construidos para los compuestos de referencia muestran tendencias, que pueden ser útiles para la predicción de perfiles de pico para las huellas dactilares cromatográficas, a pesar de la dispersión intrínsecamente mayor, en comparación a la observada con compuestos estructuralmente relacionados.

Las plantas medicinales contienen compuestos de una gran diversidad en su naturaleza química, lo que da lugar a diversas cinéticas de interacción con la columna cromatográfica. Ésta es una de las razones de la mayor dispersión observada en las correlaciones de anchura de banda, en comparación con las sulfonamidas. La segunda razón que explica la mayor dispersión es que, en elución en gradiente, los tiempos de retención isocráticos corresponden al instante en que los solutos abandonan la columna. Cabe señalar que esto ocurre al inicio del gradiente en tiempos cortos para solutos de baja hidrofobicidad, y al final del gradiente para solutos de elevada hidrofobicidad donde la fuerza de

elución es mayor, dando lugar a una reducción de los tiempos de retención. Por lo tanto, los tiempos de retención más cortos, característicos de la elución en gradiente, hacen que la dispersión sea más evidente. Téngase en cuenta, sin embargo, que las simulaciones muestran un buen acuerdo con los picos experimentales (véanse las Figuras 7.5 a 7.7).

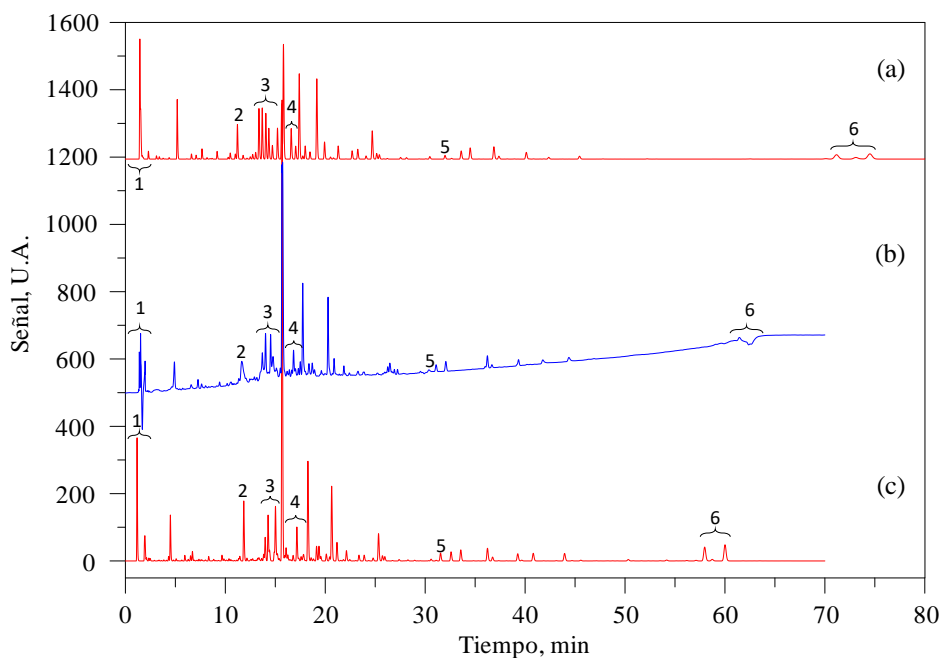


Figura 7.5. Comparación de una huella dactilar cromatográfica experimental para la melisa, correspondiente al gradiente G1 (b), con los cromatogramas predichos utilizando dos cromatogramas base distintos: (a) gradiente G7, y (c) gradiente G3, que incluyen una etapa inicial más rápida y más lenta, respectivamente. Consúltese la Figura 7.1 para conocer la identidad de los perfiles de gradiente.

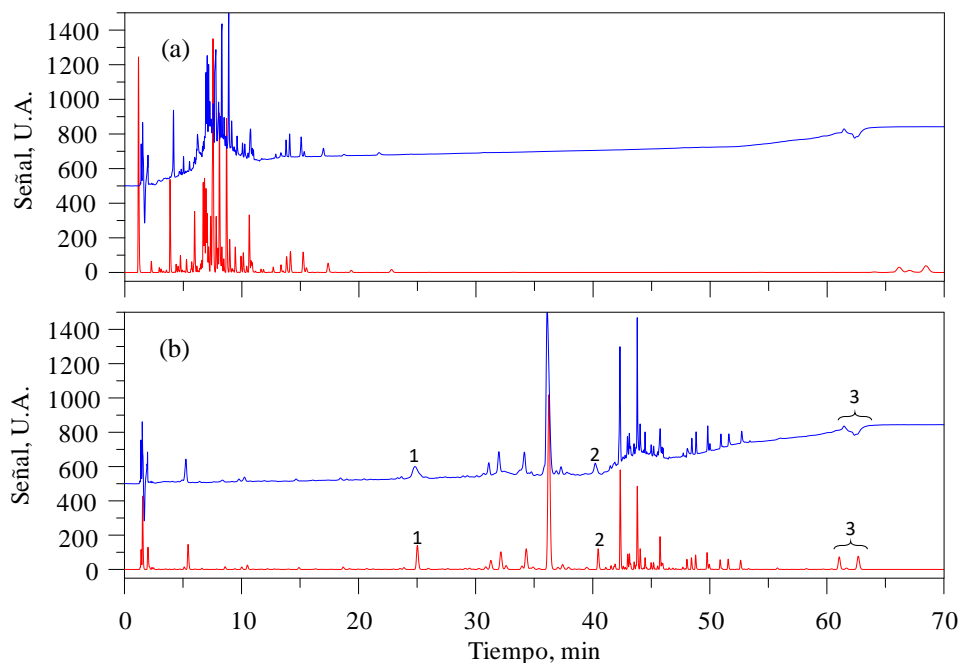


Figura 7.6. Comparación entre los cromatogramas experimentales (azules, parte superior) y predichos (rojos, parte inferior) para la melisa, obtenidos con los gradientes: (a) G7 y (b) G3 (véase la Figura 7.1). Se utilizaron estos mismos gradientes como cromatogramas base.

Cabe señalar que los modelos globales de anchura de pico para los picos de referencia son válidos para cualquier pico del cromatograma (los picos de referencia o cualquier otro). Éste no es el caso para los modelos globales de retención, que inicialmente se obtienen a partir de la información de los picos de referencia y deben adaptarse para predecir la retención de cualquier otro componente en la muestra, tal y como se ha explicado en la Sección 7.3.2.

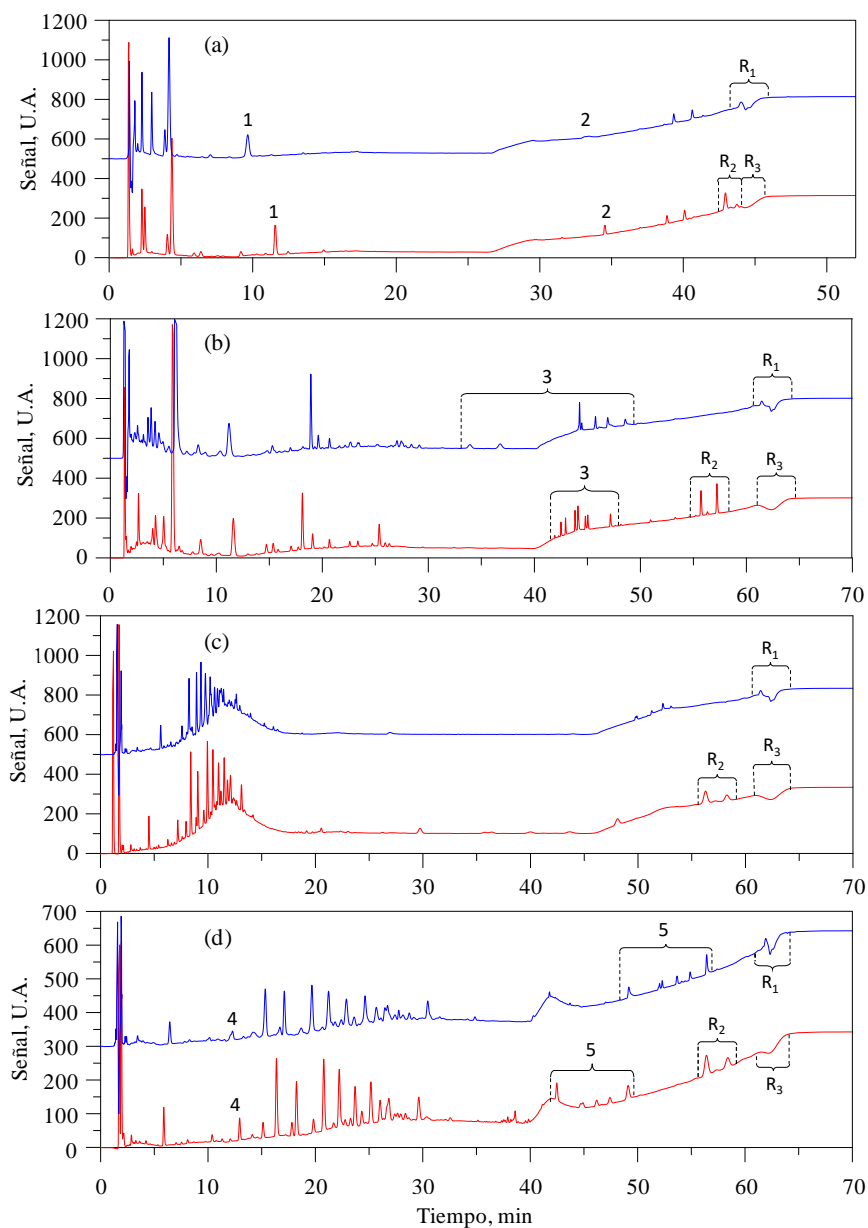


Figura 7.7. Comparación entre los cromatogramas experimentales (azules, parte superior) y predichos (rojos, parte inferior) obtenidos para las tres plantas medicinales, correspondientes a los gradientes de validación: (a) té verde con el gradiente E8 (véase la Figura 7.1), (b) melisa con el gradiente E8, (c) tila con el gradiente E7, y (d) tila con el gradiente E8.

7.5.3. *Algunos factores que afectan a la simulación de cromatogramas basados en modelos globales*

La calidad de las predicciones, llevadas a cabo con los modelos globales, se ha comprobado mediante la comparación de cromatogramas experimentales y predichos para:

- (i) Programas de gradientes multi-lineales, pertenecientes al diseño de entrenamiento experimental (Figura 7.1, gradientes G).
- (ii) Gradientes de validación externa, que contienen composiciones que exceden el intervalo cubierto por el diseño de entrenamiento (Figura 7.1, gradientes E). Estos gradientes también son multi-lineales, pero con perfiles muy diferentes a los del diseño de entrenamiento. En algunos casos, se incluyeron segmentos isocráticos.

Se utilizaron gradientes de validación para comprobar el rendimiento de predicción de los modelos globales, en condiciones desfavorables. Éste es el caso de aquellos gradientes en los que el programa comienza con concentraciones de modificador superiores a las utilizadas en el diseño de entrenamiento, o gradientes que incluyen segmentos isocráticos, más propensos a errores de predicción.

7.5.3.1. *Influencia de la elección del cromatograma base en las predicciones*

La construcción de un modelo global de retención, válido para realizar predicciones para todos los componentes de una muestra, requiere la selección arbitraria de un cromatograma experimental con el máximo número de picos (el cromatograma base, véase la Sección 7.3.3). La elección del cromatograma base es un factor que afecta de forma muy crítica a la calidad de las predicciones. Si el cromatograma seleccionado estuviera asociado al gradiente

con mayor pendiente inicial en el diseño experimental (por ejemplo, el gradiente G7 para la melisa, Figura 7.1b), las señales más pequeñas del cromatograma serían mayores debido al efecto de compresión del gradiente. Sin embargo, esto también favorecería la co-elución indeseable de picos vecinos. Por el contrario, si se utilizara el cromatograma con el gradiente más lento, los picos se resolverían mejor, pero el mayor tiempo de análisis puede empeorar la detectabilidad de las señales de menor tamaño. Sin embargo, si la rampa lenta se ve precedida de un segmento lineal más pronunciado (como en el gradiente G3), no se producirá la pérdida de perceptibilidad para los componentes más hidrofóbicos del cromatograma. Hay otras consideraciones a tener en cuenta al elegir el cromatograma base, como las diferencias en la incertidumbre de predicción de los picos que eluyen cerca de los nodos del gradiente, con fuertes cambios de pendiente.

Los parámetros $\log k_{0,i}$ específicos de los solutos, en los modelos globales utilizados para predecir las huellas dactilares cromatográficas, se calcularon a partir de los valores de los tiempos de retención para todos los picos encontrados en el cromatograma base, utilizando la función automática de detección de picos y análisis de señales. El conjunto de los parámetros $\log k_{0,i}$ y los parámetros asociados a la columna y al modificador (comunes para todos los solutos) se pueden emplear para predecir los cromatogramas bajo cualquier otro gradiente incluido dentro de la región experimental cubierta por el diseño de entrenamiento. Es interesante hacer notar que, en total, se detectaron 162, 205 y 203 picos para té verde, melisa y tila, respectivamente, en los respectivos cromatogramas base (obtenidos con los gradientes más lentos de sus diseños experimentales).

La Figura 7.5 muestra el cromatograma experimental para el extracto de melisa eluido con el gradiente G1, junto con dos cromatogramas predichos

(también para G1) obtenidos con el modelo global, pero utilizando dos cromatogramas base diferentes: G7 y G3 (consulte la identidad de los gradientes en la Figura 7.1b).

Las Figuras 7.5a y 7.5c muestran las predicciones realizadas considerando ambos gradientes: el más rápido (G7) y el más lento (G3) en el diseño experimental. En términos generales, las predicciones fueron más exactas con el modelo global desarrollado con G3. Como puede observarse, la concordancia entre los cromatogramas experimental y predicho es realmente satisfactoria.

Cabe indicar que la adquisición de cromatogramas se realizó a lo largo de un periodo de dos meses. En todos los experimentos, se utilizó un vial que contenía el mismo extracto, por lo que cualquier cambio químico en la muestra producido por degradación o formación de nuevos compuestos durante este periodo, estaría fuera del modelo ajustado. Otro factor a considerar es que el número de picos en el cromatograma predicho depende de los picos detectados en el cromatograma base. Así, en el cromatograma experimental obtenido con el gradiente G7 (donde los picos se hallan más próximos), sólo se muestran dos picos intermedios en la región 4 (Figura 7.5a). En consecuencia, si este cromatograma se utilizara como cromatograma base, cualquier predicción incluiría sólo dos picos dentro de esta región. Sin embargo, el cromatograma experimental con el gradiente G1 muestra al menos siete picos en la región 4 (Figura 7.5b). Si el cromatograma base hubiera sido el obtenido con el gradiente G3 (Figura 7.5c), habría sido posible predecir los siete picos para el gradiente G1.

7.5.3.2. Predicción de señales no asociadas a solutos retenidos

Naturalmente, la función automática para el análisis de señales no distingue entre picos genuinos y algunas otras señales no asociadas a solutos retenidos, como las siguientes:

- (i) *Señales cercanas al tiempo muerto*: Presentes al inicio del cromatograma como fluctuaciones refractivas, o señales que aparecen antes de la región del tiempo muerto, asociadas a fenómenos de arrastre o estabilización incompleta de la columna, debido a una inyección anterior. Si estas señales no se descartan, se procesarán como correspondientes a un soluto ficticio. Dado que no siguen el modelo global de retención, la predicción incidental fallará (véase la región 1 en la Figura 7.5).
- (ii) *Señales asociadas a la parada repentina de la rampa al final del gradiente*: La estabilización repentina de la pendiente al final del gradiente (por ejemplo, la región 6 en la Figura 7.5) también produce fluctuaciones refractivas, que aparecen a una posición fija. Estas señales no se corresponden con la elución de ningún soluto, sino con la parada brusca del aumento de la concentración del modificador en la etapa final del gradiente. Por lo tanto, son insensibles a los cambios producidos en el gradiente, siempre que el tiempo de gradiente t_G permanezca constante. Sin embargo, cuando los picos en esta región se asocian incorrectamente a solutos ficticios, su posición se vuelve susceptible a cambios cuando se utiliza un gradiente distinto al cromatograma base. Por lo tanto, estas señales deben ignorarse o eliminarse de la simulación. Análogamente, los cambios repentinos de pendiente en los gradientes multi-lineales pueden dar lugar a picos falsos que deben eliminarse.

7.5.3.3. Picos con anchuras anormales

Algunos picos, cuyas anchuras de banda son más amplias de las esperadas de acuerdo a la retención, se pueden encontrar frecuentemente asociados a la co-elución de dos o más componentes no resueltos, aunque estos picos pueden tener otro origen. Dado que el modelo de anchura de pico se establece con la información de picos para compuestos aislados, se predecirá un pico anormalmente estrecho, de acuerdo con la tendencia común de anchura para un compuesto aislado que eluye en esa posición. En consecuencia, cuando se aplican modelos globales de anchura, a fin de mantener la misma área, los picos simulados aparecen con una altura mayor que sus equivalentes experimentales (compárense los picos experimentales y simulados en la Figura 7.5).

Para evaluar la calidad de las predicciones de las anchuras de pico, eliminando las consecuencias de eventuales sesgos en la predicción de los tiempos de retención, se predijo el cromatograma para un gradiente seleccionado utilizándolo también como cromatograma base. Por lo tanto, las posiciones de los picos en realidad no se predijeron, sino tan sólo los perfiles de los picos. De acuerdo con esta idea, los cromatogramas asociados a los gradientes G3 y G7 se predijeron con los modelos globales de retención incluyendo todos los picos presentes en el cromatograma experimental.

Los cromatogramas experimentales y predichos se compararon para ambos gradientes G3 y G7 en las Figuras 7.6a y b, respectivamente. Como era de esperar, los picos anormalmente anchos se predicen más delgados e intensos. Éste es el caso de las regiones 2 y 5 en la Figura 7.5, y los picos 1 y 2 en la Figura 7.6b. Por otro lado, las señales refractivas que aparecen al final del gradiente (región 3 en la Figura 7.6b) se desplazan cuando cambia la composición del gradiente, ya que se procesan como componentes genuinos en la muestra. En consecuencia, se asigna un valor ficticio de $\log k_{0,i}$ a estas

señales y los cambios en la composición afectan su ubicación. En el ejemplo, la simulación sólo incluye áreas positivas y, por lo tanto, ambos picos de refracción son positivos. Estas señales se pueden identificar y eliminar fácilmente si se desea.

7.5.4. Validación de cromatogramas obtenidos con gradientes multi-lineales externos

También se simularon cromatogramas experimentales para gradientes multi-lineales no pertenecientes al diseño de entrenamiento y, por lo tanto, no utilizados para construir los modelos globales, con el objetivo de verificar el rendimiento predictivo en condiciones menos favorables. Estos gradientes de validación se muestran en la Figura 7.1 para las muestras de té verde (gradiente E8), melisa (E8) y tila (E7 y E8). Los gradientes de validación externos se obtuvieron tras las etapas de adquisición y modelización, generalmente dos semanas después de que se completara el diseño experimental. Para una comparación más realista, la contribución de la línea base, que es restada inicialmente por el algoritmo BEADS, se añadió a los cromatogramas predichos (Figura 7.7).

En el cromatograma obtenido para el té verde, se observan algunos picos experimentales, cuya predicción es anormalmente más estrecha (como es el caso de los picos 1 y 2 en la Figura 7.7a), ya que se procesan como picos genuinos asociados a un solo componente, cuando se predicen con el modelo global de anchura de pico. Obsérvese que las anchuras de pico de estas señales experimentales muestran diferencias con la tendencia observada para los picos vecinos. Por lo tanto, los picos anormalmente más anchos pueden ser el resultado de la co-elución de dos o más componentes. Otras plantas medicinales y gradientes también mostraron picos esporádicos más anchos (por ejemplo, el

pico 4 en el gradiente E8 para la tila, en la Figura 7.7d). El desplazamiento hacia tiempos más cortos de los picos asociados a las señales refractivas, existentes al final del gradiente, es igualmente perceptible en los cromatogramas. El perfil y la posición de la perturbación refractiva experimental R1, para las tres plantas, deben compararse con las señales R2 + R3 en los cromatogramas predichos. Estos cromatogramas se obtuvieron sumando la línea base encontrada por BEADS y los picos ficticios que modelan la perturbación refractiva.

Algunas diferencias observadas entre los cromatogramas experimentales y predichos pueden atribuirse a una lenta degradación de las muestras a lo largo de semanas, lo que se habría solucionado con la renovación periódica de las disoluciones de los extractos. Cabe señalar que los cromatogramas base se adquirieron varios días antes de realizar los experimentos de validación. Por lo tanto, ciertos picos están presentes en algunos cromatogramas experimentales, pero no en otros. Sin embargo, la mayoría de los picos conservan su presencia e intensidad originales.

Hay que tener en cuenta, además, que los gradientes de validación incluyen tramos isocráticos, seguidos de segmentos con fuertes incrementos de pendiente. Este tipo de configuración hace más incierta la posición de las señales, siendo los efectos acumulativos a lo largo del gradiente. La región 5 en el cromatograma de la tila, obtenido con el gradiente E8 (Figura 7.7d), ilustra este comportamiento como un cambio en la secuencia de picos. La magnitud y el signo del cambio dependen de la configuración particular del gradiente.

Se observa un efecto similar (región 3 en la Figura 7.7b), pero amplificado, en la melisa alrededor del nodo, cerca de un tiempo de análisis de 40 min, lo que es debido a una pendiente de gradiente más pronunciada (gradiente E8, véase la Figura 7.1b). Esta fuerte variación en la composición del eluyente,

junto con las incertidumbres progresivamente mayores en la posición del pico (típica de los solutos más lentos) da como resultado anchuras de pico diferentes para picos relativamente cercanos. Se puede observar que los dos primeros picos en la región 3 del cromatograma experimental (Figura 7.7b), que eluyen en el segmento isocrático del programa de gradiente (antes del cambio de pendiente), dan lugar a anchuras de banda más amplias. De acuerdo al modelo global, los compuestos asociados a estos picos son ligeramente más hidrofóbicos respecto a los experimentales; por lo tanto, se predicen con una retención más prolongada. Sin embargo, dado que estos picos se hallan ubicados cerca de un cambio pronunciado en la pendiente del gradiente, el valor ligeramente más alto para $\log k_{0,i}$ predicho (relacionado con la hidrofobicidad del soluto) implica que el soluto sea alcanzado por el siguiente segmento de pendiente más pronunciada del gradiente, cuando éste abandona la columna. Esto acelera la elución de estos picos que, en consecuencia, se comprimen. Por lo tanto, los cinco picos en la región 3 para el gradiente E8 se predicen correctamente, teniendo en cuenta su anchura de banda, pero experimentan sesgos graduales en su posición.

Finalmente, cabe señalar que para el té verde y la melisa, el intervalo de composición explorado por el conjunto de validación al comienzo del gradiente se halla fuera del dominio cubierto por el diseño de entrenamiento (16.4 % de acetonitrilo para el té verde y 23 % para la melisa, véase el gradiente E8 en las Figuras 7.1a y 7.1b). Esto significa que, para los compuestos menos retenidos, los gradientes no alcanzarán concentraciones tan altas en los primeros minutos y, por lo tanto, la predicción de la retención para estos compuestos se basará en extrapolaciones.

Los componentes más polares de las muestras, que eluyen al inicio del gradiente, son más sensibles a la falta de información, viéndose así afectados

por mayores incertidumbres. Dado que los gradientes de validación para el té verde y la melisa comienzan con elución isocrática, este problema se magnifica. Sin embargo, a pesar de esta limitación, los cromatogramas predichos y experimentales muestran una buena concordancia.

7.6. Conclusiones

Este capítulo trata sobre la idoneidad de los modelos globales para simular cromatogramas que contienen cientos de componentes, que pueden ser útiles con fines de optimización. Se han ampliado los modelos globales de retención obtenidos en el Capítulo 6 para compuestos seleccionados en las huellas dactilares cromatográficas de tres plantas medicinales, a fin de incluir todos los componentes de la muestra. Para ello, se incluyeron en el modelo de retención los datos de todos los picos detectados, que se encuentran en el cromatograma asociado al gradiente ensayado con la pendiente inicial más baja. Se ha demostrado que los modelos globales permiten la predicción de cromatogramas de alta complejidad, bajo diferentes condiciones de gradiente, con un notable nivel de aproximación a la realidad. La metodología propuesta ha sido comprobada con excelentes resultados, utilizando los extractos de tres plantas medicinales con cromatogramas afectados por problemas específicos. Para obtener una detección más segura de los picos de menor tamaño, se aplicó un algoritmo de corrección de la línea base, seguido de una función de Matlab no supervisada, construida en el laboratorio para la detección de picos.

En la construcción de modelos de retención individuales convencionales, todos los parámetros obtenidos al ajustar los datos de retención son específicos del soluto eluido, que se ajusta de forma independiente. Como consecuencia, cuando se comparan los parámetros específicos de los solutos ($\log k_{0,i}$), se

observa que éstos se ven afectados de manera desigual por su naturaleza química. Por el contrario, en los modelos globales, el proceso de regresión aísla los efectos comunes de la columna y el modificador de los específicos de cada soluto. Esto hace que la estimación de la hidrofobicidad del soluto dependa menos de las interacciones particulares de los analitos con la columna y el modificador. En consecuencia, la contribución de cada soluto a la retención sigue un orden más acorde a su hidrofobicidad [13].

Aunque la predicción del comportamiento de retención mediante un modelo global implica perder cierta especificidad para los solutos, distintiva de los modelos individuales, la pérdida en el rendimiento de la predicción es aceptable. La principal limitación de la propuesta presentada (y en general de los modelos globales en su estado actual) es que los cambios en el orden de elución de los componentes de la muestra, con la composición, requeriría identificar todos los picos presentes en un segundo cromatograma base, para poder relacionarlos con el primer cromatograma base.

Es también necesario indicar que el procesamiento de cromatogramas no asistido considera cualquier señal detectada como un componente genuino de la muestra. Así, en las regiones inicial y final de los cromatogramas, a menudo se observan picos positivos y negativos de naturaleza refractiva. En consecuencia, la predicción de estas señales se verá afectada por variaciones en el programa de gradiente y, si no se eliminan de las simulaciones, los picos asociados se predecirán con cambios proporcionales a su hidrofobicidad aparente. Lo mismo puede ocurrir con señales residuales asociadas con: (i) una corrección imperfecta de la línea base, (ii) artefactos de cálculo producidos por el algoritmo de corrección de línea base BEADS, o (iii) la presencia de picos que co-eluyen con anchuras de banda anormalmente más amplias. En esta etapa del trabajo, estas señales anormales se han conservado para mostrar sus efectos.

El objetivo de los Capítulos 6 y 7 ha sido estudiar exhaustivamente todos los aspectos relevantes y las limitaciones de los modelos globales para realizar la simulación de cromatogramas. La utilidad de los modelos globales va más allá del campo de las huellas dactilares cromatográficas: existen muchos problemas de separación donde no se dispone de estándares, e incluso se desconoce la identidad de la mayoría de los componentes de la muestra. Los modelos globales permitirían incluir compuestos desconocidos de cualquier muestra en las simulaciones. Finalmente, este trabajo abre la posibilidad de optimizar la separación de huellas dactilares cromatográficas utilizando métodos interpretativos.

7.7. Referencias

- [1] H. Sun, X. Chen, A. Zhang, T. Sakurai, J. Jiang, X. Wang, Chromatographic fingerprinting analysis of *Zhizhu Wan* preparation by high-performance liquid chromatography coupled with photodiode array detector, *Pharmacogn. Mag.* 10 (2014) 470–476.
- [2] P.K. Mukherjee, Quality Control and Evaluation of Herbal Drugs: Evaluating Natural Products and Traditional Medicine, Elsevier, Amsterdam, 2019.
- [3] H. Siddique, M. Sarwat (editors), Herbal Medicines: A Boon for Healthy Human Life, Academic Press, Cambridge, MA, EE.UU., 2022.
- [4] N. Cui, H. Hao, G. Wang, W. Wang, Y. Wang, Orthogonal design-directed optimization of an LC method for fingerprinting Mai-Luo-Ning injection, and validation of the method, *Chromatographia* 68 (2008) 33–39.

- [5] P. Wang, L. Li, H. Yang, S. Cheng, Y. Zeng, L. Nie, H. Zang, Chromatographic fingerprinting and quantitative analysis for the quality evaluation of *Xinkeshu* tablet, *J. Pharm. Anal.* 2 (2012) 422–430.
- [6] G. Alaerts, S. Pieters, H. Logie, M. Merino-Arévalo, B. Dejaegher, J. Smeyers-Verbeke, Y. Vander Heyden, Exploration and classification of chromatographic fingerprints as additional tool for identification and quality control of several *Artemisia* species, *J. Pharm. Biomed. Anal.* 95 (2014) 34–46.
- [7] O.A. Souza, R.L. Carneiro, T.H.M. Vieira, C.S. Funari, D. Rinaldo, Fingerprinting *Cynara scolymus* L. (Artichoke) by means of a green statistically developed HPLC PAD method, *Food Anal. Methods* 11 (2018) 1977–1985.
- [8] M.C. García Álvarez-Coque, J.R. Torres Lapasió, J.J. Baeza Baeza, Models and objective functions for the optimisation of selectivity in reversed-phase liquid chromatography, *Anal. Chim. Acta* 579 (2006) 125–145.
- [9] G. Jin, X. Xue, F. Zhang, X. Zhang, Q. Xu, Y. Jin, X. Liang, Prediction of retention times and peak shape parameters of unknown compounds in traditional chinese medicine under gradient conditions by ultra-performance liquid chromatography, *Anal. Chim. Acta* 628 (2008) 95–103.
- [10] T. Álvarez Segura, A. Gómez Díaz, C. Ortiz Bolsico, J.R. Torres Lapasió, M.C. García Álvarez-Coque, A chromatographic objective function to characterise chromatograms with unknown compounds or without standards available, *J. Chromatogr. A* 1409 (2015) 79–88.

- [11] B. Yan, X. Bai, Y. Sheng, F. Li, Statistical model based HPLC analytical method adjustment strategy to adapt to different sets of analytes in complicated samples, *Phytochem. Anal.* 28 (2017) 424–432.
- [12] A. Gisbert Alonso, J.A. Navarro Huerta, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Global retention models and their application to the prediction of chromatographic fingerprints, *J. Chromatogr. A* 1637 (2021) 461845.
- [13] A. Gisbert Alonso, S. López Ureña, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Chromatographic fingerprint-based analysis of extracts of green tea, lemon balm and linden: I. Development of global retention models without the use of standards, *J. Chromatogr. A* 1672 (2022) 463060.
- [14] A. Gisbert Alonso J.A. Navarro Huerta, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Testing experimental designs in liquid chromatography (II): Influence of the design geometry on the prediction performance of retention models, *J. Chromatogr. A* 1654 (2021) 462458.
- [15] J.A. Navarro Huerta, J.R. Torres Lapasió, S. López Ureña, M.C. García Álvarez-Coque, Assisted baseline subtraction in complex chromatograms using the BEADS algorithm, *J. Chromatogr. A* 1507 (2017) 1–10.
- [16] T. Álvarez Segura, E. Cabo Calvet, J.R. Torres Lapasió, M.C. García Álvarez-Coque, An approach to evaluate the information in chromatographic fingerprints: Application to the optimisation of the extraction and conservation conditions of medicinal herbs, *J. Chromatogr. A* 1422 (2015) 178–185.

- [17] L.R. Snyder, J.J. Kirkland, J.L. Glajch, Practical HPLC Method Development, 2^a ed., John Wiley & Sons, Nueva York, 1997.
- [18] P.J. Schoenmakers, H.A.H. Billiet, R. Tussen, L. de Galan, Gradient selection in reversed-phase liquid chromatography, *J. Chromatogr. A* 149 (1978) 519–537.
- [19] U.D. Neue, H.J. Kuss, Improved reversed-phase gradient retention modeling, *J. Chromatogr. A* 1217 (2010) 3794–3803.
- [20] J.R. Torres Lapasió, J.J. Baeza Baeza, M.C. García Álvarez-Coque, A model for the description, simulation and deconvolution of skewed chromatographic peaks, *Anal. Chem.* 69 (1997) 3822–3831.
- [21] G. Vivó Truyols, J.R. Torres Lapasió, A.M. van Nederkaassel, Y. Vander Heyden, D.L. Massart, Automatic program for peak detection and deconvolution of multi overlapped chromatographic signals: Part II: Peak model and deconvolution algorithms, *J. Chromatogr. A* 1096 (2005) 146–155.
- [22] J.J. Baeza Baeza, S. Pous Torres, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Approaches to characterise chromatographic column performance based on global parameters accounting for peak broadening and skewness, *J. Chromatogr. A* 1217 (2010) 2147–2157.
- [23] J.J. Baeza Baeza, M.J. Ruiz Ángel, M.C. García Álvarez-Coque, S. Carda Broch, Half-width plots, a simple tool to predict peak shape, reveal column kinetics and characterise chromatographic columns in liquid chromatography: State of the art and new results, *J. Chromatogr. A* 1314 (2013) 142–153.
- [24] J.R. Torres Lapasió, J.J. Baeza Baeza, M.C. García Álvarez-Coque, Modeling of peak shape and asymmetry, in Chemometrics in Chromatography (editores, L. Komsta, Y. Vander Heyden, J. Sherma),

- Chromatographic Science Series, CRC Press, Taylor and Francis Group, Boca Raton, FL, 2018, págs. 217–238.
- [25] P. Jandera, Predictive calculation methods for optimization of gradient elution using binary and ternary solvent gradients, *J. Chromatogr. A* 485 (1989) 113–141.
- [26] P. Nikitas, A. Pappa-Louisi, Expressions of the fundamental equation of gradient elution and a numerical solution of these equations under any gradient profile, *Anal. Chem.* 77 (2005) 5670–5677.
- [27] P. Nikitas, A. Pappa-Louisi, New approaches to linear gradient elution used for optimization in reversed-phase liquid chromatography, *J. Liq. Chromatogr. Relat. Technol.* 32 (2009) 1527–1576.
- [28] S. López Ureña, J.R. Torres Lapasió, M.C. García Álvarez-Coque, Enhancement in the computation of gradient retention times in liquid chromatography using root-finding, *J. Chromatogr. A* 1600 (2019) 137–147.

RESUMEN Y CONCLUSIONES

La cromatografía líquida (LC, *liquid chromatography*) y en su vertiente más utilizada, la cromatografía líquida de fase inversa (RPLC, *reversed-phase liquid chromatography*), es actualmente una de las técnicas más potentes de que se dispone en los laboratorios de análisis clínico, farmacéutico, alimentario, etc., para resolver problemas analíticos, que en algunos casos no son asumibles mediante otras técnicas instrumentales más simples. Esto se debe a que puede ser aplicada al análisis de compuestos no volátiles, desde pequeñas moléculas hasta grandes macromoléculas biológicas, abarcando una amplia gama de polaridades, siendo una técnica versátil, robusta y sensible. Sin embargo, la selectividad y el tiempo de análisis están condicionados por multitud de factores experimentales, como son la naturaleza de la columna (fase estacionaria), la elección de la naturaleza y la concentración del modificador orgánico en la fase móvil, su pH y la temperatura a la cual se lleva a cabo la separación. Debido a la multitud de variables que deben controlarse para lograr el éxito de la separación, las metodologías basadas en ensayos de prueba y error no son capaces de proporcionar las condiciones óptimas de resolución, excepto en el caso de las muestras sencillas.

Cuando el analista se enfrenta a muestras de cierta complejidad, lo que es habitual en la práctica cromatográfica, las condiciones de separación que den lugar a la resolución de los picos del cromatograma de la muestra deben obtenerse de forma preferente haciendo uso de las denominadas optimizaciones interpretativas o basadas en modelos, en las que se hace uso de la información extraída de un conjunto de experimentos cuidadosamente planificado y lo más reducido posible, al que denominaremos diseño experimental de entrenamiento. Éste debe cubrir todo el dominio de los factores experimentales, en el cual se obtendrán los datos con los que se ajuste el modelo matemático que permita simular la separación, bajo nuevas condiciones no contempladas en el diseño y,

por lo tanto, en segundo término, establecer cuáles son las condiciones óptimas con las que resolver el problema.

La Tesis Doctoral, cuya Memoria aquí se presenta, se ha enfocado a la realización de estudios fundamentales para mejorar las metodologías de optimización interpretativa, centrándose en la primera etapa, la modelización. Para ello se ha hecho uso de:

- (i) un conjunto de compuestos estructuralmente relacionados (una mezcla de 14 sulfonamidas),
- (ii) bases de datos de retención para multitud de compuestos de diversa funcionalidad química, y
- (iv) muestras de productos naturales (específicamente extractos de té verde, melisa y tila), como ejemplo de un caso extremadamente complejo en términos de separación.

Los análisis se han realizado utilizando como fase estacionaria una columna cromatográfica C18, y como eluyentes, tanto fases móviles isocráticas como gradientes, en todos los casos constituidos por mezclas hidro-orgánicas que hacen uso de acetonitrilo como modificador orgánico de la polaridad, en presencia de un tampón para controlar el pH. El trabajo se encuentra estructurado en dos partes:

- Parte I: Evaluación del rendimiento predictivo de diversos diseños experimentales.
- Parte II: Desarrollo de modelos globales de retención para su aplicación a muestras complejas.

A continuación, se resumen las conclusiones generales de cada uno de los capítulos considerados en la Memoria de Tesis Doctoral.

C.1. Evaluación del rendimiento predictivo de diversos diseños experimentales

El éxito de las estrategias interpretativas utilizadas para optimizar la resolución de una mezcla, mediante LC, depende críticamente tanto de la disponibilidad de ecuaciones que sean capaces de describir el comportamiento (retención y perfil) de los picos cromatográficos de los componentes de la muestra bajo estudio, como de la calidad de los datos cromatográficos con los que se llevará a cabo el ajuste de los modelos descriptores, lo que depende en gran medida de la idoneidad del diseño de entrenamiento experimental con el que se obtendrán, aspecto que se aborda en la Parte I de esta Memoria. Se han estudiado nuevos diseños como alternativa a los diseños isocráticos, que son ampliamente aceptados por proporcionar datos de mayor calidad. En los nuevos diseños, se busca recortar los largos tiempos de adquisición para los solutos más hidrofóbicos, sin renunciar a la disponibilidad de medidas de retención a baja fuerza eluyente. También se desarrolla una metodología basada en la teoría de propagación de errores, que permite categorizar cualquier tipo de diseño experimental de entrenamiento, en función de las incertidumbres asociadas a las predicciones realizadas tras el ajuste de los datos obtenidos con dicho diseño y, por lo tanto, de su calidad informativa. A continuación, se describen los aspectos más relevantes de cada propuesta.

C.1.1. Capítulo 2: Beneficios del uso de pulsos de concentración de modificador orgánico en la modelización del tiempo de retención en cromatografía líquida

- Para el desarrollo de métodos de análisis en LC, es crucial la obtención de modelos de gran exactitud para predecir la retención. Una de las etapas críticas es la elección del diseño experimental con el que modelizarla. Entre las diferentes opciones, son los diseños experimentales isocráticos los que brindan la información más rica sobre el comportamiento de los solutos, generando tras la modelización parámetros del modelo más exactos, e intervalos de confianza más estrechos que los diseños basados en gradientes, donde se aumenta gradualmente el contenido de modificador en función del tiempo. Sin embargo, el uso de diseños isocráticos se ve obstaculizado por los largos tiempos de retención de los solutos más hidrofóbicos, especialmente cuando se eluyen a bajos contenidos de modificador orgánico.
- La solución habitual es el uso de diseños experimentales formados con gradientes de modificador orgánico. El objetivo principal es obtener una resolución adecuada para todos los componentes de la muestra, aumentando la retención de los solutos poco retenidos y reduciéndola para los fuertemente hidrofóbicos. Los diseños experimentales de gradiente son capaces de reducir sustancialmente los largos tiempos de adquisición, pero proporcionan información más pobre sobre el comportamiento de retención cuando se utilizan composiciones de baja fuerza eluyente, que pueden ser decisivas para la separación. En consecuencia, estos diseños dan lugar a modelos menos exactos, por lo que poseen una capacidad predictiva peor.
- Como alternativa, se ha explorado el uso de diseños experimentales isocráticos que incluyen incrementos repentinos (es decir, pulsos) de

modificador orgánico en las fases móviles de menor fuerza eluyente. Los experimentos que contienen pulsos se pueden considerar como un tipo de gradientes multi-isocráticos que permiten obtener información cromatográfica para solutos altamente hidrofóbicos, manteniendo una baja concentración de modificador durante la mayor parte de la elución. La presencia transitoria de una alta concentración de modificador, durante un corto tiempo, mueve en bloque los tiempos de retención de los compuestos de elución tardía a tiempos más tempranos, para los que las concentraciones más bajas de modificador en el pulso tienen un peso significativo.

- La construcción de diseños mixtos con pulsos puede llevarse a cabo fácilmente insertando en las fases móviles isocráticas de menor fuerza eluyente uno o dos pulsos de corta duración, a tiempos intermedios. La ubicación del pulso puede establecerse arbitrariamente, siendo la mejor opción situarlo en una región intermedia vacía del cromatograma. De esta forma, los solutos más rápidos eluyen antes del pulso y los solutos más retenidos después del pulso, a tiempos aceptables. Dado que la aplicación de pulsos en fases isocráticas tiene efectos importantes en la selectividad y retención de los solutos que eluyen después del pulso, su posición, duración y aumento repentino del contenido de modificador orgánico deben adaptarse a cada muestra.
- La aplicación de pulsos a la modelización de la retención tiene como objetivo reducir el tiempo de elución de los solutos lentos, cuando se recopilan los datos de entrenamiento para el ajuste de modelos de retención. Sin embargo, la inclusión de pulsos no es práctica con fines de optimización, debido a la mayor superposición de los picos, especialmente en la región del

pulso, y porque los solutos sufren caídas importantes de eficacia tras el pulso.

- Para la predicción de condiciones de elución en diseños que contienen pulsos se ha hecho uso de la ecuación fundamental para la elución en gradiente. Se han evidenciado desviaciones notables en el cálculo de los tiempos de retención mediante integración numérica, para los solutos que eluyen en regiones cercanas al pulso, incluso haciendo uso de un modelo de retención con errores de predicción bajos. La corrección de tales desviaciones implica considerar el retraso intra-columnar, esto es, el tiempo que tarda el frente de fase móvil de una cierta composición, marcada por el programa de gradiente, en afectar a un soluto que se encuentra migrando a través de la columna. Tras considerar dicho retraso, las predicciones de los tiempos de retención observados en los cromatogramas experimentales fueron satisfactorias.
- Cuando las predicciones con diseños que contenían pulsos o gradientes se realizaron dentro del dominio experimental, la discrepancia entre los tiempos predichos y experimentales fue inferior a 0.01 min. Los diseños con pulsos proporcionaron parámetros para los modelos de retención similares a los obtenidos con diseños isocráticos, que como se ha comentado, son considerados los más exactos en términos predictivos. Para las predicciones fuera del dominio experimental, la capacidad predictiva de los diseños que contenían pulsos también fue similar al rendimiento obtenido con los diseños con experimentos puramente isocráticos.
- En general, los diseños que contienen pulsos han demostrado ser muy competitivos respecto a los diseños de gradiente, en términos de tiempo de análisis y consumo de modificador orgánico. Entre los diferentes diseños, el

construido con gradientes de tiempo de gradiente variable mostró el mejor rendimiento en términos de tiempo de análisis y consumo de modificador. Sin embargo, con ese diseño se obtuvieron parámetros del modelo con mayor error y mayores desviaciones en las predicciones extrapoladas.

C.1.2. Capítulo 3: Diseños experimentales en cromatografía líquida: Desarrollo y validación de un método para la inspección integral de diseños experimentales

- La fiabilidad de la predicción de la resolución depende críticamente de la exactitud de los modelos de retención obtenidos. Este hecho exalta la importancia de los diseños experimentales, que son los que determinan la calidad de la información obtenida necesaria para la realización del ajuste de los modelos. Es ampliamente aceptado que los diseños isocráticos de entrenamiento son los más informativos, pero son también los que necesitan mayores tiempos de adquisición de datos. Debido a ello, muchos analistas prefieren hacer uso de diseños de entrenamiento formados por gradientes, para llevar a cabo la modelización de la retención. Sin embargo, encontrar la distribución óptima de gradientes para la construcción del diseño no es sencillo.
- Con el fin de encontrar los mejores diseños experimentales, se muestra el desarrollo, la validación y la aplicación de una metodología universal que evalúa la calidad de diseños experimentales de entrenamiento, y que es capaz de revelar el mejor diseño. La metodología puede aplicarse con cualquier configuración arbitraria de gradientes y no se ve limitada por la disponibilidad de expresiones algebraicas para la predicción de tiempos de retención en este modo de elución.

- Se pretende estudiar los efectos asociados a cambios regulares en los diseños más usuales en HPLC isocráticos y de gradiente de modificador orgánico, y con ello deducir los principios generales que favorecen la calidad de la información en los diseños, relacionándolos con las propiedades de los solutos, particularmente con su hidrofobicidad.
- El enfoque utiliza el Principio de Optimalidad G, que se basa en la teoría de propagación de errores y permite la evaluación sistemática de las incertidumbres asociadas a la estimación de los tiempos de retención, cuando se predice una distribución regular de condiciones experimentales. El cálculo se realiza utilizando diseños de muestreo, en ambos modos de elución, a partir de los datos del diseño de entrenamiento utilizado para modelizar la retención.
- El tratamiento requiere el cálculo de matrices jacobianas asociadas a los experimentos de los diseños de entrenamiento ($\mathbf{J}_{\text{train}}$) y de muestreo (\mathbf{J}_{pred}) (Ecuación (4.19)). Las matrices jacobianas implican el cálculo de derivadas parciales de los modelos de retención, para un amplio conjunto de condiciones. En elución en gradiente, el tiempo de cálculo puede ser particularmente largo, ya que requiere la predicción del tiempo de retención por integración de la ecuación fundamental, lo que puede implicar cálculos masivos. Como solución se emplearon procedimientos numéricos, con los que el tiempo de cálculo se redujo considerablemente.
- La metodología propuesta fue validada comprobando la capacidad de cinco diseños de entrenamiento (ISO1, ISO2, G1, G2 y G3), muy utilizados en RPLC para la construcción de modelos, utilizando los tiempos de retención de 14 sulfonamidas de polaridad diversa. Para ello, se hizo uso de diseños de muestreo para ambas, elución isocrática y en gradiente, considerando dos

modelos de retención: el modelo de Snyder o de la fuerza lineal del disolvente (LSS) y el modelo de Neue-Kuss (NK).

- La ecuación propuesta por NK para describir la retención ofreció un mayor rendimiento predictivo que el modelo LSS, con errores relativos para las predicciones por debajo del 0.7 %. Por otro lado, se encontró que el modelo LSS, utilizado ampliamente en la elución en gradiente, presenta un importante sesgo. Por lo tanto, los resultados obtenidos con este modelo deberían analizarse con cautela, siendo sólo posible derivar conclusiones generales.
- Los diseños de entrenamiento se compararon asistidos por mapas de incertidumbre, en los que se representaron las incertidumbres relativas en las predicciones de cada uno de los compuestos bajo estudio, en función de las diferentes condiciones de elución recogidas en los diseños de muestreo. El diseño de muestreo isocrático consistió en una sucesión de fases móviles, cuya composición se varió en incrementos del 1 % de acetonitrilo, mientras que el diseño de muestreo de gradiente consistió en rampas con incrementos angulares constantes de 3°.
- Las incertidumbres relativas proporcionaron resultados más significativos e interpretables que las incertidumbres absolutas, que eran muy variables y dependían de la retención de los solutos. Los cambios sistemáticos observados en los gráficos de incertidumbre, a medida que disminuía la polaridad de los solutos, respaldan la fiabilidad de los cálculos, así como de los propios resultados. El nivel de precisión en el cálculo del tiempo de gradiente fue un factor crítico en el cálculo de las derivadas en las matrices jacobianas. Con un nivel de precisión insuficiente, los mapas de incertidumbre contenían curvas afectadas de un ruido significativo. Cuando

se alcanzó un nivel de precisión de alrededor de 10^{-15} , se obtuvieron curvas suaves y, en la mayoría de los casos, se observó un patrón en U característico para la elución en gradiente, con incrementos en ambos extremos y errores más bajos en el centro de la curva obtenida.

- La interpretación de las diferentes curvas de incertidumbre asociadas a cada diseño y soluto se ha basado en tres principios fundamentales: (i) cada composición a lo largo de un diseño de gradiente debe participar significativamente en la retención de los solutos, (ii) el intervalo de composiciones muestreadas a lo largo de la elución debe ser amplio, y (iii) los tiempos de retención encontrados a lo largo de un cierto diseño deben reflejar esa diversidad.
- Entre los diferentes diseños de entrenamiento estudiados, el Diseño ISO1 consistente en un conjunto de experimentos isocráticos, gradualmente concentrados a un bajo contenido de modificador orgánico, fue confirmado como el mejor, tanto para las predicciones isocráticas como de gradiente. Los diseños de gradiente a tiempo de gradiente fijo y contenido final de modificador orgánico variable (G1), y a contenido de modificador orgánico final fijo y tiempo de gradiente variable (G2), exhibieron un rendimiento insuficiente en la mayoría de las situaciones, siendo tan sólo aceptables para los eluyentes más lentos y los solutos más rápidos. El diseño G3, que combina algunas características de los diseños G1 y G2, proporcionó un rendimiento predictivo razonable para todos los compuestos de la muestra en estudio, sólo superado por el diseño ISO1.

*C.1.3. Capítulo 4: Diseños experimentales en cromatografía líquida:
Influencia de la geometría del diseño en la capacidad predictiva de los
modelos de retención*

- Se aplicó la metodología desarrollada a la evaluación de un número elevado de diseños, incluidas familias de diseños relacionados y mixtos. El estudio expuesto en el Capítulo 4 complementa al realizado en el capítulo anterior, donde se ha presentado el desarrollo de una metodología basada en la teoría de propagación de errores, utilizando el Principio de Optimalidad G, que permite evaluar las incertidumbres asociadas a los diferentes diseños experimentales de entrenamiento.
- El propósito es comparar, utilizando la metodología desarrollada, la capacidad de predicción de una amplia variedad de diseños relacionados geoméricamente, así como revelar el efecto de diferentes factores sobre el rendimiento predictivo cuando los diseños varían sistemáticamente el número y la distribución de los experimentos, la presencia de pulsos, el contenido de modificador inicial y final en los gradientes, así como sus pendientes y la ubicación de sus nodos.
- A partir de la observación de los resultados, procedentes de 38 diseños de diferentes tipos, construidos para conseguir una retención adecuada, se ofrecen algunas recomendaciones para construir diseños con capacidad predictiva satisfactoria, con lo que se dé repuesta a algunas preguntas generalmente resueltas de forma intuitiva, como son la geometría adecuada de un diseño y el número de experimentos necesarios. La metodología puede utilizarse en cualquier aplicación que requiera la modelización de la retención, como es la optimización de la selectividad, el desarrollo de métodos o la caracterización de las propiedades de los solutos.

- La introducción en la evaluación de un número considerable de diseños obliga a replantear nuevas formas para visualizar los resultados, respecto a las mostradas en el Capítulo 3, ya que conforme aumenta el número de diseños a comparar, su interpretación se vuelve inviable. Es por ello que, en el Capítulo 4 se han utilizado diagramas de caja y bigotes, así como diagramas triples que incorporan códigos de colores, permitiendo así una evaluación simultánea de la capacidad predictiva de una gran cantidad de diseños, atendiendo a las incertidumbres relativas esperadas.
- Los diagramas de caja y bigotes brindan una perspectiva rápida de la capacidad de predicción de los diseños. En ellos se muestra la distribución de las incertidumbres relativas, para todos los solutos y todos los experimentos de un cierto diseño. Por otro lado, los gráficos triples resumen los resultados obtenidos mediante un código de colores en función de la incertidumbre relativa, para cada soluto y fase móvil (o programa de gradiente). La combinación de ambos tipos de gráficos permite realizar múltiples comparaciones para un gran número de diseños y simplifica el análisis de los resultados, aunque sacrificando algunos detalles con respecto a los mapas de incertidumbre.
- En cuanto a las predicciones realizadas, se ha encontrado que los gradientes se predicen con incertidumbres intrínsecamente más bajas, independientemente del diseño experimental utilizado para modelizar la retención, siendo menos sensibles al tipo de diseño de entrenamiento que las predicciones isocráticas.
- Tal y como se ha visto en el estudio anterior, los diseños isocráticos (Diseños ISO1–ISO5) ofrecen la información más valiosa sobre la retención de los solutos. El número y la distribución de los experimentos isocráticos

(disponiendo al menos de un grado de libertad y siempre que el dominio de composiciones del eluyente esté bien cubierto) no resultó crítico en términos de capacidad predictiva. A pesar de ello, cuando la muestra contiene solutos altamente hidrofóbicos, este tipo de diseños poseen el inconveniente de originar largos tiempos para la adquisición de los cromatogramas. Tal y como se ha introducido en el Capítulo 2, en estas situaciones pueden acortarse los largos tiempos de adquisición al implementar un pulso de modificador orgánico, en la fase isocrática de menor fuerza eluyente (Diseños P1–P7). Se ha comprobado que, para los solutos lentos eluidos con fases móviles con bajo contenido de modificador, no se deteriora la capacidad predictiva tras la implementación del pulso, siempre y cuando la ubicación, la duración y la concentración máxima de modificador sean las adecuadas.

- Otra alternativa para reducir los largos tiempos de adquisición, cuando se muestrean regiones de baja fuerza eluyente, es la aplicación de diseños de gradiente. Sin embargo, éstos poseen el inconveniente de que, cuando el diseño experimental de entrenamiento está formado exclusivamente por gradientes, cada gradiente sólo examina intervalos estrechos del dominio de modificador. Este hecho da como resultado que la realización de predicciones isocráticas con bajas incertidumbres, a partir de diseños de entrenamiento compuestos por gradientes, sólo sea viable si la elución del soluto bajo una determinada fase móvil a predecir se encuentra bien muestreada por el conjunto de gradientes. Por el contrario, cuando el diseño de entrenamiento es exclusivamente isocrático, todas las fases móviles muestreadas participan al 100 % en la migración de los solutos.

- Tras la evaluación de las diferentes familias de diseños de entrenamiento, se ha comprobado que los diseños que presentan niveles comunes de concentración inicial (Diseños SL1–SL6) o final (Diseños ML1–ML6), en los gradientes, tienen una capacidad de muestreo intrínsecamente peor. Además, se ha encontrado que diseños de gradiente normalmente utilizados para modelizaciones rutinarias (Diseños SL1 y SL3) no proporcionan predicciones con baja incertidumbre en ciertas regiones del dominio. Mientras tanto, cuando se implementan tres o más niveles de modificador orgánico al principio y al final de los gradientes (Diseños ML7–ML12), la capacidad predictiva mejora. Prueba de ello es el Diseño ML12, que ofrece incertidumbres comparables a las encontradas con los diseños isocráticos, pero con la ventaja de un tiempo de adquisición de los datos significativamente más corto.
- El último grupo de diseños (N1–N4) se planificó para conseguir la separación de solutos en un intervalo de polaridad extremadamente amplio. Estos diseños tienen un perfil romboédrico, progresivamente ensanchado debido a la mayor diferencia de pendientes entre los dos segmentos lineales consecutivos unidos mediante un nodo. En el Capítulo 5, se hace uso, con excelentes resultados, de un diseño similar a N3 para obtener un modelo global que permita describir la retención de un gran número de componentes en extractos de manzanilla. Por otro lado, N5–N7 mantienen el carácter multi-lineal de los gradientes, pero diferenciando los niveles iniciales y finales de los gradientes. En base a los resultados obtenidos, dicha separación de los niveles de concentración inicial y final de los gradientes parece altamente recomendable.

- La posibilidad de muestrear composiciones tan diversas como sea posible, y a cualquier tiempo de gradiente, considerando todos los gradientes de un diseño, también es compartida por el Diseño multi-lineal N7, que es el siguiente en términos de capacidad predictiva tras el Diseño ML12 y con un tiempo de adquisición similar. En el estudio realizado, se ha comprobado que el uso de gradientes multi-lineales mejora la calidad de las predicciones, respecto a los diseños de gradientes lineales que poseen una única rampa. Sin embargo, los diseños de gradientes multi-lineales son más propensos a depender de los requisitos de información de cada soluto.

C.2. Desarrollo de modelos globales de retención para su aplicación a muestras complejas

En LC, el desarrollo de nuevos métodos de separación se sustenta en dos pilares fundamentales: la modelización del sistema y la optimización de la selectividad. Para ello, el primer paso consiste en la recopilación de información experimental con la cual poder construir un modelo que permita la predicción del comportamiento cromatográfico. El objetivo es, en segunda instancia, poder pronosticar cuáles son las condiciones óptimas de elución con las que resolver la mezcla. Habitualmente se hace uso de estándares, que pueden ser inyectados separadamente o en mezclas, a fin de obtener la información requerida para lograr el desarrollo de un nuevo método de separación. Sin embargo, en la práctica, los analistas pueden enfrentarse a mezclas en las que su complejidad, por el número de componentes presentes, y/o el desconocimiento de la identidad de algunos o la mayor parte de los componentes, imposibilite el uso de estándares para llevar a cabo la modelización. La Parte II se centra en este tipo de muestras, tomando como

caso de estudio cromatogramas de productos naturales, a los que se ha llamado huellas dactilares cromatográficas, que presentan una complejidad extrema en términos separativos. Para estas muestras, se ha desarrollado una metodología basada en la obtención de modelos globales de retención, que permiten la simulación de cromatogramas complejos.

C.2.1. Capítulo 5: Desarrollo de metodologías para obtener modelos globales de retención que describan grupos de compuestos

- En RPLC, el analista dispone de varios factores experimentales que puede modificar a voluntad, con el fin de lograr una separación selectiva, siendo los relacionados con la composición de la fase móvil los más útiles para lograr la separación. Este hecho subraya la necesidad de disponer de modelos fiables que describan la retención en función del contenido de modificador orgánico en la fase móvil. Normalmente, los datos proporcionados por el diseño experimental se ajustan a los modelos de retención que, una vez construidos, pueden aplicarse para predecir la calidad de la separación bajo nuevas condiciones experimentales arbitrarias.
- El trabajo presentado en el Capítulo 5 aborda el desarrollo y ajuste de modelos globales de retención, que representan una alternativa al uso de los modelos individuales para describir la retención en LC cuando, debido a las características de las muestras, el uso de estos últimos resulta inviable. En los modelos globales propuestos, algunos parámetros son específicos de cada soluto, mientras que otros describen los efectos generales de los cambios de modificador sobre la retención en la columna. Cabe mencionar que los modelos individuales son capaces de predecir el comportamiento de retención de un único compuesto, con gran exactitud, ya que se particulariza

la descripción al soluto, mientras que un modelo global obtiene una tendencia general, siendo capaz de predecir la retención de varios compuestos, pero pagando el precio de sacrificar la exactitud.

- Uno de los objetivos del Capítulo 5 fue desarrollar un procedimiento de regresión alterna, que permite el ajuste simultáneo de los datos de varios solutos a una ecuación de retención común, haciendo posible la descripción del comportamiento de conjuntos de compuestos. Se construyeron modelos globales de retención basados en varias ecuaciones, propuestas por Snyder, Schoenmakers, Neue y Kuss, Jandera, y Bosch y Rosés. Como casos de estudio se consideraron, por un lado, huellas dactilares cromatográficas de extractos de plantas medicinales, eluidos con gradientes multi-lineales y, por otro lado, un conjunto de datos isocráticos para 145 compuestos, con amplia variabilidad en sus estructuras químicas y funcionalidades.
- Las huellas dactilares cromatográficas representan un caso extremo de cromatogramas complejos con componentes de identidad desconocida. Cabe señalar que los modelos individuales son inviables para muestras con varios cientos de componentes, y un modelo basado en la información obtenida para cada compuesto, a partir de estándares, no es viable. Es por ello que, para llevar a cabo la modelización fuera necesaria la selección de picos de referencia, que deben ser trazables en las diferentes condiciones de elución del diseño experimental y estar distribuidos a lo largo del cromatograma. El procedimiento se inicia con el ajuste de los picos de referencia a modelos individuales, siendo el rendimiento de ajuste excelente, especialmente para las ecuaciones de NK y Schoenmakers, con $R^2 > 0.9999$, siendo peor para la ecuación de Snyder ($R^2 = 0.9983$).

- El procedimiento de regresión alterna ajusta, en primer lugar, un vector construido con las medianas de los parámetros que dependen de la columna y del modificador (S y T en la ecuación de Schoenmakers), calculadas a partir de los parámetros de los modelos individuales, y los parámetros específicos para cada soluto y relacionados con su hidrofobicidad ($\ln k_{w,i}$ en la ecuación de Schoenmakers). Mediante este procedimiento, los parámetros se ajustan, por etapas, atendiendo a todos los solutos y condiciones simultáneamente. En ciertas etapas, los valores de S y T son comunes y se mantienen constantes para todos los solutos, mientras que el proceso de regresión ajusta los valores de $\ln k_{w,i}$ para cada soluto. En otras etapas, los valores de S y T se modifican para que sean comunes para todos los solutos, mientras que los valores de $\ln k_{w,i}$ son constantes. Los cinco modelos globales de retención ensayados muestran menores diferencias de rendimiento que los modelos individuales, y una capacidad de predicción similar o mejor que la ofrecida por la ecuación de Snyder, cuando se ajusta soluto a soluto individualmente.
- Con el fin de validar los modelos globales de retención desarrollados con las huellas dactilares cromatográficas, se aplicó la metodología de regresión propuesta a un conjunto de solutos conocidos de amplia diversidad química, cuyos datos de retención fueron obtenidos en elución isocrática. De esta forma, el modelo global tenía que afrontar el reto de dar respuesta a un conjunto muy amplio de solutos, muy distintos entre sí, eluidos a intervalos extremos de concentración efectiva de modificador orgánico. Las experiencias isocráticas condicionan a una participación completa de cada fase móvil en el ajuste de los modelos de retención, mientras que los gradientes se benefician de intervalos de modificador más estrechos. Debido a ello, las predicciones son más exactas en elución en gradiente y las

ecuaciones necesarias son más simples. Estas ecuaciones suelen mostrar un rendimiento peor en predicciones isocráticas, como es el caso de la ecuación de Snyder, con una capacidad de predicción en elución en gradiente muy superior.

- El objetivo final del trabajo realizado ha sido la predicción de un gran número de picos en las huellas dactilares cromatográficas. Tras establecer los parámetros columna-modificador, mediante la obtención del modelo global de retención para los picos de referencia, se obtienen los parámetros específicos de los solutos para aquellos compuestos contenidos en el cromatograma que no han participado en la etapa anterior. Para ello, se utilizan los tiempos de retención experimentales de todos los picos detectados en el cromatograma más lento del diseño experimental. Finalmente, con esta información se obtiene un modelo global final que permite predecir con éxito el cromatograma en su totalidad, para cualquier gradiente arbitrario. La simulación de huellas dactilares cromatográficas se aborda en su totalidad en el Capítulo 7.
- Todos estos resultados indican que los modelos globales pueden constituir una vía interesante como base para el desarrollo de una función objetivo que oriente la optimización de la resolución cromatográfica, cuando la naturaleza de la muestra hace imposible atender a la selectividad específica de cada soluto. Otra aplicación interesante de los modelos globales es la predicción de propiedades moleculares, como el coeficiente de reparto octanol-agua ($\log P_{o/w}$), a partir de los parámetros específicos de los solutos, utilizando el modelo de Bosch-Rosés.

C.2.2. Capítulo 6: Análisis de huellas dactilares cromatográficas de extractos de té verde, melisa y tila: Desarrollo de modelos globales de retención sin el uso de estándares

- En los estudios de modelización de la retención, el primer paso consiste en la recopilación de información sobre el comportamiento de los constituyentes de la muestra (tiempos de retención y propiedades de pico), de acuerdo a un diseño experimental pre-establecido. Normalmente, las situaciones más favorables corresponden a muestras sencillas, de composición conocida y con disponibilidad de estándares, siendo posible el desarrollo de modelos de retención de calidad para cada constituyente de la muestra. En otras situaciones, se dispone de muestras con un elevado nivel de complejidad, cuyo número de componentes, algunos de ellos desconocidos, hace prohibitivo modelizar individualmente el comportamiento de retención para cada uno de los componentes presentes, siendo normal no disponer de estándares para todos ellos. En estos casos, la modelización de los diferentes constituyentes debe realizarse directamente con la información recopilada de los cromatogramas de las muestras.
- Como respuesta al problema de la modelización de muestras de extrema complejidad, se propone el uso de modelos globales y un tipo de diseño de gradiente de entrenamiento que facilita la trazabilidad de picos en los cromatogramas, en distintas condiciones de elución. En los modelos globales, los parámetros comunes para todos los componentes describen los efectos generales de la columna y el modificador orgánico sobre la retención, mientras que otros parámetros son específicos para cada componente. De este modo, los modelos globales de retención ofrecen ecuaciones válidas para un conjunto de componentes, a costa de una cierta

pérdida en la capacidad predictiva, respecto a los modelos individuales, lo cual no impide su uso en aplicaciones prácticas, como la optimización de cromatogramas con una alta densidad de picos.

- El estudio desarrollado en el Capítulo 6 tiene como objetivo validar el uso de modelos globales de retención en la modelización de huellas dactilares cromatográficas. Para ello se hace uso del procedimiento de regresión alterna desarrollado en el Capítulo 5, donde los parámetros del modelo global se ajustan en etapas consecutivas, atendiendo simultáneamente a todas las condiciones experimentales y componentes de la muestra. El procedimiento se aplica utilizando las ecuaciones de Snyder, Schoenmackers y NK, mientras que como muestras de estudio se han considerado tres plantas medicinales que ofrecen características de separación muy diversas: *Camellia sinensis* (té verde), *Melissa officinalis* (melisa) y *Tilia platyphyllos* (tila).
- La caracterización de las denominadas huellas dactilares cromatográficas se basa en la modelización de la retención para los picos de unos pocos componentes de las muestras, a los que se ha denominado picos de referencia. Con el fin de obtener un modelo satisfactorio, los picos seleccionados deben distribuirse a lo largo de todo el cromatograma. Además, debe ser posible su trazabilidad observando la similitud de patrones de distribución de picos bajo diferentes condiciones de gradiente. Por último, debe tenerse en cuenta su intensidad, debiendo ser lo suficiente para permitir su detección, incluso a bajas fuerzas eluyentes. El número de picos de referencia seleccionados fue 12, 13 y 10 para los extractos de té verde, melisa y tila, respectivamente.

- Una etapa relevante en la modelización, tal y como se ha visto en la Parte I, es la adecuación del diseño experimental al problema en estudio. En una investigación preliminar, y utilizando un extracto de té verde, se utilizó un diseño experimental basado en gradientes simples no relacionados. El modelo de retención obtenido no cumplió las expectativas en términos de capacidad predictiva, lo que fue debido no sólo a las ambigüedades en la identificación de los picos, que en ocasiones podían detectarse y corregirse parcialmente, sino también a la falta de disponibilidad de la información necesaria para modelizar componentes con una extrema diversidad química.
- Dicho estudio sugirió que el diseño de gradientes utilizado para ajustar los modelos de retención debía incluir variaciones ordenadas en la composición del eluyente, con el fin de facilitar la trazabilidad de los picos de referencia bajo diferentes gradientes de elución. Para ello, la secuencia de gradientes se debe construir paso a paso, de acuerdo a los resultados del experimento anterior. Los gradientes en estos diseños son multi-lineales, con dos segmentos lineales consecutivos que comparten un nodo intermedio de posición variable. El nodo se ubica siguiendo una diagonal, lo que da lugar a un perfil característico en forma de diamante. Dicha geometría permite explorar composiciones extremas, sin que los tiempos de retención resulten inaceptablemente elevados para los componentes más hidrofóbicos, o demasiado cortos para los más hidrofílicos. Previamente, este tipo de diseño ha sido evaluado en términos de capacidad predictiva arrojando buenos resultados (Capítulo 4). En el Capítulo 5, se ha utilizado por primera vez para modelizar la retención de compuestos de referencia seleccionados para extractos de manzanilla.

- El uso de diseños de gradiente en forma de diamante, para modelizar la retención, requiere su adaptación a cada tipo de muestra, con el fin de obtener una elución adecuada en la región más significativa de los cromatogramas de los extractos. Para establecer esta región se hace uso de un gradiente de exploración, que consiste en una rampa lineal con una duración de 60 min, en el que se barre todo el dominio de composiciones de acetonitrilo. Tanto este gradiente, que también pertenece al diseño experimental y que es común a todas las muestras, como el resto de gradientes del diseño, incluyen etapas finales de limpieza, así como un tiempo de reequilibrado de la columna antes de la siguiente inyección.
- Idealmente, el análisis de muestras altamente complejas debe llevarse a cabo a un valor de pH constante y controlado. En las separaciones isocráticas, la elución a pH constante no presenta ningún problema. Sin embargo, en elución en gradiente, los cambios de pH debidos a la variación de la concentración del modificador orgánico pueden afectar a la retención, ya que solutos con valores de pK_a cercanos al pH del tampón pueden sufrir desviaciones en sus predicciones respecto al cromatograma experimental. Sin embargo, en las tres muestras estudiadas, este problema no fue destacable. Además, siendo el objetivo mejorar la separación general, el hecho de que algunos picos se aparten de las posiciones esperadas entre cientos bien pronosticados, no condiciona la aplicabilidad de la metodología.
- A pesar de la gran dispersión en la hidrofobicidad de los solutos, en los extractos de las plantas medicinales, y el amplio intervalo de concentración de modificador cubierto por los gradientes ensayados, el rendimiento predictivo de los modelos globales es comparable al de los modelos individuales. Respecto a los resultados ofrecidos por las ecuaciones de

Snyder y Schoenmakers, se muestra que el modelo de retención propuesto por NK arroja las mejores predicciones, ofreciendo excelentes correlaciones de los tiempos de retención predichos frente a los experimentales, utilizando para ello datos provenientes de todos los gradientes y compuestos ensayados.

C.2.3. Capítulo 7: Análisis de huellas dactilares cromatográficas de extractos de té verde, melisa y tila: Simulación de cromatogramas haciendo uso de modelos globales de retención

- Debido a la eficacia y baja toxicidad de los preparados derivados de plantas medicinales, tanto en la prevención como en el tratamiento de enfermedades, su uso se ha ido extendiendo en gran medida en los últimos años. Para estos productos, una diversidad de factores pueden comprometer su seguridad y eficacia farmacológica, por lo que deben ser evaluados mediante un control de calidad que asegure que cumplen los estándares requeridos para su uso. Sin embargo, debido a la alta complejidad química de este tipo de muestras, el control de calidad es extremadamente complejo. Por ello, diversas organizaciones con reconocimiento internacional han admitido el uso de huellas dactilares cromatográficas para evaluar la calidad de estos productos.
- Como se ha comentado a lo largo de esta Memoria, para llevar a cabo una optimización que revele cuáles son las condiciones de elución óptimas para una muestra, es necesario modelizar previamente el comportamiento de retención de los solutos que la componen, haciendo uso de un diseño experimental de entrenamiento pre-establecido. Las metodologías convencionales que hacen uso de modelos de retención individuales no son

viales cuando se dispone de muestras complejas, como los productos naturales, debido principalmente a la gran cantidad de picos existentes en los cromatogramas, que pueden ser del orden de cientos, así como al desconocimiento de la identidad de la mayor parte de los componentes de la muestra inyectada. Por ello, la Parte II se ha enfocado en desarrollar una metodología para describir el comportamiento de retención de compuestos desconocidos en un cromatograma, mediante el uso de modelos globales, donde algunos parámetros son específicos de cada soluto, mientras que otros describen los efectos combinados de la columna y el modificador, comunes para todos los solutos.

- En el Capítulo 6, se han construido modelos globales a partir de un número limitado de señales que cumplen una serie de criterios que permiten su trazabilidad, a lo largo de los diferentes gradientes del diseño de entrenamiento, los denominados picos de referencia. La metodología se probó considerando tres plantas medicinales (té verde, melisa y tila) que exhiben diferentes escenarios en términos de separación cromatográfica.
- El objetivo final de las investigaciones descritas es proponer una metodología fiable, para obtener modelos globales de retención adecuados para predecir cromatogramas completos de huellas dactilares, considerando tanto los compuestos de referencia como cualquier otro que dé lugar a picos detectables. Para ello, en el Capítulo 7 se amplía el desarrollo de los modelos globales de retención basados en la ecuación de NK, y construidos para los compuestos de referencia, para incluir todos los componentes de las plantas medicinales, que junto con la posibilidad de desarrollar modelos globales que describan la anchura de banda, permiten la predicción de cromatogramas completos con cientos de componentes.

- Para extender el modelo global de retención a la totalidad de las señales presentes en la huella dactilar, en primer lugar, se selecciona el cromatograma obtenido con el gradiente de menor fuerza eluyente del diseño de entrenamiento ya que, generalmente, será aquél con el mayor número de picos detectables. Previamente a su procesamiento, y con el fin de obtener una detección más segura de los picos de menor tamaño, la línea base se sustrae del cromatograma experimental mediante un algoritmo adecuado, siendo el cromatograma resultante el “cromatograma base”. Seguidamente, y mediante una función automática de detección de picos, se mide la posición de todas aquellas señales que superen el umbral de detección, junto con otras medidas que definen el perfil de pico. Finalmente se realiza un ajuste por mínimos cuadrados, donde los parámetros asociados a la columna y al modificador se mantienen fijos a los valores encontrados con los picos de referencia, mientras que se ajustan los parámetros específicos para todos los solutos. Con esta información (b , c y $\log k_{0,i}$), para el modelo NK, se puede predecir el cromatograma para cualquier otro gradiente arbitrario.
- Es importante resaltar que, para que la simulación de cromatogramas sea realista, debe considerarse además de la ubicación de los picos para cada componente, el valor de las anchuras de banda de los picos. Para experimentos isocráticos, las anchuras de pico experimentales se correlacionan directamente con los respectivos tiempos de retención, obteniéndose tendencias parabólicas que, en la mayoría de los casos, pueden aproximarse a relaciones lineales. Sin embargo, en elución en gradiente, la relación entre las anchuras de pico y el tiempo de retención no es directa. En estos casos se puede aplicar la aproximación de Jandera, que establece que la anchura de pico en elución en gradiente es equivalente a la que se

obtendría si el soluto eluyera isocráticamente a la composición instantánea de fase móvil del gradiente a la que el soluto abandona la columna. Una vez obtenidos los datos, el modelo global de anchura de banda se ajusta modulando sus parámetros para obtener la mayor concordancia entre las anchuras de pico experimentales y las predichas, utilizando los compuestos de referencia y todos los gradientes de entrenamiento.

- Entre los diferentes factores que afectan a la calidad de las simulaciones de cromatogramas, la elección arbitraria del cromatograma base, como es de esperar, influye de forma crítica, debiéndose llegar a una solución de compromiso que maximice el número de picos visibles y su detectabilidad, y minimice la coelución entre picos. Entre las diferentes muestras estudiadas, se detectaron 162, 205 y 203 picos para té verde, melisa y tila, respectivamente, con los respectivos cromatogramas base.
- Otra limitación, intrínseca al uso de una función automática de detección de picos es que ésta considera cualquier señal detectada como un componente real de la muestra. Por ejemplo, picos de origen refractométrico que habitualmente se observan en las regiones inicial y final de los cromatogramas, si no se eliminan de las simulaciones, serán predichos con cambios proporcionales a su hidrofobicidad aparente. Lo mismo puede ocurrir con señales residuales asociadas con una corrección imperfecta de la línea base, o la presencia de picos que co-eluyen con anchuras de banda anormalmente más amplias.

- Se demuestra que los modelos globales permiten la predicción de cromatogramas de alta complejidad, bajo diferentes condiciones de gradiente, con un notable nivel de aproximación a la realidad. La utilidad de los modelos globales va más allá del campo de las huellas dactilares cromatográficas: existen muchos problemas de separación donde no se dispone de estándares, e incluso se desconoce la identidad de la mayoría de los componentes de la muestra. Los modelos globales permitirían incluir compuestos desconocidos de cualquier muestra en las simulaciones. Finalmente, se abre la posibilidad de optimizar la separación de huellas dactilares cromatográficas utilizando métodos interpretativos.

**CONTRIBUCIÓN DEL TRABAJO DE
DOCTORADO A LAS PUBLICACIONES**

Los capítulos contenidos en esta Tesis Doctoral corresponden a las siguientes publicaciones, ordenadas según su fecha de publicación. El porcentaje de aportación de Adrián Gisbert Alonso como estudiante de Doctorado se indica con cada artículo.

- José Antonio Navarro Huerta, Adrián Gisbert Alonso, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
Benefits of solvent concentration pulses in retention time modelling of liquid chromatography.
Journal of Chromatography A 1597 (2019) 76–88 (Capítulo 2).
Contribución: 50% Adrián Gisbert Alonso, 50% José Antonio Navarro Huerta
- José Antonio Navarro Huerta, Adrián Gisbert Alonso, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
Testing experimental designs in liquid chromatography (I): Development and validation of a method for the comprehensive inspection of experimental designs.
Journal of Chromatography A 1624 (2020) 461180 (Capítulo 3).
Contribución: 50% Adrián Gisbert Alonso, 50% José Antonio Navarro Huerta
- Adrián Gisbert Alonso, José Antonio Navarro Huerta, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
Testing experimental designs in liquid chromatography (II): Influence of the design geometry on the prediction performance of retention models.
Journal of Chromatography A 1654 (2021) 462458 (Capítulo 4).
Contribución: 100%

- Adrián Gisbert Alonso, José Antonio Navarro Huerta, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
Global retention models and their application to the prediction of chromatographic fingerprints.
Journal of Chromatography A 1637 (2021) 461845 (Capítulo 5).
Contribución: 100%
- Adrián Gisbert Alonso, Sergio López Ureña, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
Chromatographic fingerprint-based analysis of extracts of green tea, lemon balm and linden: I. Development of global models without the use of standards.
Journal of Chromatography A 1672 (2022) 463060 (Capítulo 6).
Contribución: 100%
- Adrián Gisbert Alonso, Álvaro Navarro Martínez, José Antonio Navarro Huerta, José Ramón Torres Lapasió, María Celia García Álvarez-Coque
Chromatographic fingerprint-based analysis of extracts of green tea, lemon balm and linden: II. Simulation of chromatograms using global models.
Journal of Chromatography A 1684 (2022) 463561 (Capítulo 7).
Contribución: 100%

Ambos directores de esta Tesis Doctoral (José Ramón Torres Lapasió y María Celia García Álvarez-Coque) aparecen como coautores de todos los artículos incluidos en la Memoria. También son investigadores principales de los proyectos que han financiado la investigación. José Ramón Torres Lapasió ha propuesto las líneas maestras de los trabajos llevados a cabo y ha

desarrollado los programas informáticos utilizados en el tratamiento de datos incluidos en los diferentes artículos. María Celia García Álvarez-Coque ha dado apoyo a las investigaciones y colaborado en la redacción de cada artículo y de esta Memoria.

José Antonio Navarro Huerta participó en el período inicial de la Tesis Doctoral, mediante la formación y supervisión del trabajo experimental (preparación de disoluciones, técnica experimental y manejo de instrumentos). Sergio López Ureña, del Departamento de Matemáticas de la Universitat de València, colaboró en el tratamiento matemático y procesamiento de datos. Además, algunos trabajos de investigación contribuyeron a la formación del estudiante del Grado de Química Álvaro Navarro Martínez, y de la estudiante del Máster de “Técnicas Experimentales en Química” Carmen Tébar Soler.

