



UNIVERSIDAD NACIONAL DEL LITORAL

FACULTAD DE BIOQUÍMICA Y CIENCIAS BIOLÓGICAS
CÁTEDRA DE QUÍMICA ANALÍTICA I
LABORATORIO DE DESARROLLO ANALÍTICO Y QUIMIOMETRÍA

*Modelado quimiométrico con señales instrumentales de segundo orden.
Aplicación a la determinación de fármacos y aditivos alimentarios en
muestras complejas.*

Lic. María Julia Culzoni

*Director de Tesis
Dr. Héctor C. Goicoechea*

2008

A mis queridos padres

*Dime y lo olvido,
enséñame y lo recuerdo,
involúcrame y lo aprendo.*

Benjamín Franklin (1706-1790)

Gracias...

Héctor, por darme esta valiosa oportunidad, y por todas las demás. Tu apoyo constante, tu entusiasmo y tu impulso de progreso son verdaderos estímulos que me alientan a asumir nuevos desafíos con alegría.

Gabriel, por tu gran solidaridad. Tus aportes resultaron fundamentales para la construcción de esta tesis, y te lo agradezco profundamente.

Alejandro Olivieri, por los conocimientos compartidos. Ellos constituyeron herramientas claves para el desarrollo y la concreción de este trabajo.

Lu, por tu amistad sincera. Por estar siempre dispuesta a escucharme y a brindarme tu ayuda.

Gonza, Merce, Víctor, Caco, Marisi, Silvia F., Silvia H., Nilda, Yami, Vane, Pablo. Es un privilegio compartir todos los días con personas como ustedes.

Familia: Franco, mamá, papá, Eugenia, Hermas y Joaquín. La vida no podría haberme dado un regalo mejor.

La presente tesis dio lugar a las siguientes publicaciones y presentaciones a reuniones científicas:

PUBLICACIONES

Evaluation of partial least-squares with second-order advantage for the multi-way spectroscopic analysis of complex biological samples in the presence of analyte-background interactions. M. J. Culzoni, H. C. Goicoechea, A. P. Pagani, M. A. Cabezón, A. C. Olivieri. **Analyst** 131 (2006) 718 - 723.

Experimental study of non-linear second-order analytical data with focus on the second-order advantage. M. J. Culzoni, P. C. Damiani, A. García-Reiriz, H. C. Goicoechea, A. C. Olivieri. **Analyst** 132 (2007) 654 - 663.

Determination of loratadine and pseudoephedrine sulfate in pharmaceuticals based on non-linear second-order spectrophotometric data generated by a pH gradient flow injection technique and artificial neural networks. M. J. Culzoni, H. C. Goicoechea. **Anal. Bioanal. Chem.** 389 (2007) 2217 - 2225.

Solving matrix-effects exploiting the second order advantage in the resolution and determination of eight tetracycline antibiotics in effluent wastewater by modelling liquid chromatography data with multivariate curve resolution-alternating least squares and unfolded-partial least squares followed by residual bilinearization algorithms. I. Effect of signal pre-treatment. M. M. De Zan, M. D. Gil García, M. J. Culzoni, G. G. Siano, H. C. Goicoechea, M. Martinez Galera. **J. Chromatogr. A** 1179 (2008) 106 - 114.

Solving matrix-effects exploiting the second order advantage in the resolution and determination of eight tetracycline antibiotics in effluent wastewater by modelling liquid chromatography data with multivariate curve resolution-alternating least squares and unfolded-partial least squares followed by residual bilinearization algorithms. II. Prediction and figures of merit. M. D. Gil García, M. J. Culzoni, M. M. De Zan, R. Santiago Valverde, M. Martinez Galera, H. C. Goicoechea. **J. Chromatogr. A** 1179 (2008) 115 - 124.

A versatile strategy for achieving the second-order advantage when applying different artificial neural networks to non-linear second-order data: unfolded principal component analysis/residual bilinearization. A. García-Reiriz, P. C. Damiani, M. J. Culzoni, H. C. Goicoechea, A. C. Olivieri. *Chemom. Intell. Lab. Syst.* 92 (2008) 61 - 70.

Second-order advantage from kinetic-spectroscopic data matrices in the presence of extreme spectral overlapping. A multivariate curve resolution - alternating least-squares approach. M. J. Culzoni, H. C. Goicoechea, G. A. Ibáñez, V. A. Lozano, N. R. Marsili, A. C. Olivieri, A. P. Pagani. *Anal. Chim. Acta* 614 (2008) 46 - 57.

Chemometric resolution of fully overlapped CE peaks: Quantitation of carbamazepine in human serum in the presence of several interferences. L. Vera-Candioti, M. J. Culzoni, A. C. Olivieri, H. C. Goicoechea. *Electrophoresis* 29 (2008) 1 - 12.

PRESENTACIONES ORALES Y EN PÓSTER A REUNIONES CIENTÍFICAS:

Un nuevo algoritmo quimiométrico para datos de segundo orden: regresión en cuadrados mínimos parciales con ventaja de segundo orden. Aplicaciones bioanalíticas. M. J. Culzoni, H. C. Goicoechea, A. P. Pagani, M. A. Cabezón, A. C. Olivieri. III Congreso Argentino de Química Analítica. Merlo, San Luis. 1 al 4 de octubre de 2005. Póster.

Diferentes estrategias analíticas aplicadas al control de calidad de fármacos. Uso de quimiometría en espectroscopia UV. M. J. Culzoni, L. Vera Candioti, M. M. De Zan, J. C. Robles, V. E. Mantovani, H. C. Goicoechea. 13° Encontro Nacional de Química Analítica/1° Congresso Ibero-Americano de Química Analítica. Universidade Federal Fluminense, Niteroi, RJ, Brasil. 12 al 16 de setiembre de 2005. Presentación oral.

Residual bilinearization. A useful tool providing PLS and ANN with the second order advantage. M. J. Culzoni, M. A. Cabezón, G. M. Escandar, A. García-

Reiriz, H. C. Goicoechea, N. Marsili, A. C. Olivieri, A. P. Pagani, G. Piccirilli. 10th International Conference on Chemometrics in Analytical Chemistry. Águas de Lindóia, SP, Brasil. 10 al 15 de setiembre de 2006. Presentación oral.

Regresión por cuadrados mínimos parciales con ventaja de segundo orden. Aplicación bioanalítica. M. J. Culzoni, J. C. Robles, H. C. Goicoechea. 10° Encuentro de Jóvenes Investigadores de la Universidad Nacional del Litoral y 1° Encuentro de Jóvenes Investigadores de Universidades de Santa Fe. Santa Fe. 25 y 26 de octubre de 2006. Póster.

Application of chemometric appealing to the second order advantage to solve interference problems in the determination of emerging contaminants in wastewater. M. Martínez Galera, M. M. De Zan, M. D. Gil García, M. J. Culzoni, R. Santiago Valverde, H. C. Goicoechea. 2nd EMCO workshop on Emerging Contaminants in Wastewaters: Monitoring Tools and Treatment Technologies. Belgrado, Serbia. 26 y 27 de abril de 2007. Póster.

Redes neuronales artificiales con ventaja de segundo orden aplicadas a la determinación de loratadina y pseudoefedrina sulfato en solución oral. M. J. Culzoni y H. C. Goicoechea. II Congreso Iberoamericano y IV Congreso Argentino de Química Analítica. Buenos Aires. 27 al 30 de agosto de 2007. Póster.

Determinación de tetraciclinas como contaminantes en agua por HPLC-DAD y modelado quimiométrico con ventaja de segundo orden. M. M. De Zan, M. D. Gil García, M. J. Culzoni, R. Santiago Valverde, H. C. Goicoechea, M. Martínez Galera. II Congreso Iberoamericano y IV Congreso Argentino de Química Analítica. Buenos Aires. 27 al 30 de agosto de 2007. Póster.

Análisis de enalapril en comprimidos usando datos de segundo orden generados por doble gradiente de pH y espectroscopia UV. M. S. Cámara, G. Santiago, M. J. Culzoni, J. C. Robles, H. C. Goicoechea. II Congreso

Iberoamericano y IV Congreso Argentino de Química Analítica. Buenos Aires. 27 al 30 de agosto de 2007. Póster.

Effect of PDS, baseline correction and mean centering as signal pretreatments in the resolution of eight tetracyclines by application of MCR-ALS to HPLC-DAD data from effluent wastewater samples. M. D. Gil García, M. M. De Zan, M. J. Culzoni, R. Santiago Valverde, G. G. Siano, H. G. Goicoechea, M. Martínez Galera. IX Internacional Symposium on Analytical Methodology in The Environmental Field. Pollensa, España. 1 al 3 de octubre de 2007. Póster.

Second order advantage when applying different artificial neural networks after unfolded principal component analysis/residual bilinearization. H. C. Goicoechea, A. García-Reiriz, P. Damiani, M. J. Culzoni, A. C. Olivieri. 11th International Conference on Chemometrics for Analytical Chemistry. Montpellier, Francia. 30 de junio al 4 de julio de 2008. Póster.

Chemometric resolution of fully overlapped CE peaks: Quantitation of carbamazepine in human serum in the presence of several interferences. H. C. Goicoechea, L. Vera-Candioti, M. J. Culzoni, A. C. Olivieri. Annual Meeting of the Federation for Analytical Chemistry and Spectroscopy Societies (FACSS 2008). Reno, NV, Estados Unidos de América. 28 de septiembre al 2 de octubre de 2008. Presentación oral.

LISTA DE SÍMBOLOS Y ABREVIATURAS

+	Matriz pseudoinversa (como supraíndice)
$\ \cdot \ $	Norma Euclidea de un vector
\otimes	Producto de Kronecker
γ_n	Sensibilidad analítica
ACE	Expectativas condicionales alternantes
ALS	Cuadrados mínimos alternantes
ANN	Redes neuronales artificiales
ATLD	Descomposición trilineal alternante
AParP	Gráfico de los residuos parciales aumentados de Mallows
APTLD	Descomposición trilineal alternante penalizada
A	Número de factores latentes
BLLS	Cuadrados mínimos bilineales
BLLS/RBL	Cuadrados mínimos bilineales seguidos de bilinealización residual
CANDECOMP	Descomposición canónica
CART	Árboles de clasificación y regresión
CLAR	Cromatografía líquida de alta resolución
CV	Coefficiente de variación
Df	Grados de libertad
DTLD	Descomposición trilineal directa
EEM	Matriz de excitación-emisión
EJCR	Intervalo de confianza elíptico conjunto
FIA	Análisis por inyección en flujo
GC	Cromatografía gaseosa
GCV	Error cuadrático penalizado
GN	Gauss-Newton
GRAM	Método de aniquilación de rango generalizado
I	Número de muestras de calibración
IR	Infrarrojo
LOD	Límite de detección
LOQ	Límite de cuantificación

LS-SVM	Máquinas de vectores soporte con ajuste por cuadrados mínimos
LV	A factores de PLS
MARS	<i>Splines</i> de regresión multivariada adaptativa
MCR-ALS	Resolución multivariada de curvas mediante cuadrados mínimos alternantes
MS	Espectrometría de masa
MNR	Resonancia magnética nuclear
MLP	Perceptrón multicapa
NAS	Señal neta del analito
N_{noesp}	Número de componentes no esperados
NIR	Infrarrojo cercano
NLPLS	Cuadrados mínimos parciales no lineales
N-PLS	Cuadrados mínimos parciales multivía
PARAFAC	Análisis paralelo de factores
PCA	Análisis en componentes principales
PCR	Regresión en componentes principales
PLS	Cuadrados mínimos parciales
RBF	Funciones de base radial
RBL	Bilinealización residual
RES	Variación residual
REP(%)	Error relativo de predicción
RMSEM	Error cuadrático medio de monitorización
RMSEP	Error cuadrático medio de predicción
RMSET	Error cuadrático medio de entrenamiento
RSD	Desviación estándar relativa
RTL	Trilinealización residual
SEL	Selectividad
SEN	Sensibilidad
SMART	Regresión aditiva múltiple suavizadora
<i>Spline</i> -PLS	<i>Spline</i> -Cuadrados mínimos parciales
SSE	Suma de los errores cuadráticos
SVD	Descomposición en valores singulares

SVM	Máquinas de vectores soporte
SWATLD	Descomposición trilineal alternante autoponderada
T	Matriz transpuesta (como supraíndice)
TLLS/RTL	Cuadrados mínimos trilineales seguidos de trilinealización residual
U-PCR	Regresión en componentes principales desdoblados
U-PLS	Cuadrados mínimos parciales desdoblados
U-PCA/RBL	Componentes principales desdoblados seguidos de bilinealización residual
U-PCR/RBL	Regresión en componentes principales desdoblados seguidos de bilinealización residual
U-PLS/RBL	Cuadrados mínimos parciales desdoblados seguidos de bilinealización residual
U-PLS/RTL	Cuadrados mínimos parciales desdoblados seguidos de trilinealización residual
vec	Operación de vectorización
<u>X</u>	Arreglo tridimensional formado por diferentes matrices

ÍNDICE DE CONTENIDOS

DESCRIPCIÓN	PÁGINA
CAPÍTULO 1: Introducción	1
1.1. Calibración multivariada	2
1.1.1. Clasificación de los datos instrumentales	2
1.1.2. Terminología	5
1.1.2.1. Tipos de componentes presentes en las muestras	5
1.1.2.2. Estructura de los datos instrumentales de segundo orden	6
1.1.3. Algoritmos quimiométricos	7
1.1.3.1. Algoritmos quimiométricos para datos de orden cero, uno y dos	7
1.1.3.2. Algoritmos quimiométricos para datos de N -dimensiones	10
1.1.4. Presente y futuro del análisis multivariado	10
1.2. Objetivos	12
CAPÍTULO 2: Evaluación del algoritmo U-PLS/RBL mediante su aplicación al análisis de muestras biológicas complejas en presencia de efecto matriz.	13
2.1. Introducción	14
2.2. Objetivos	15
2.3. Teoría	16
2.3.1. PARAFAC	16
2.3.2. U-PLS	19
2.3.2.1. Ventaja de segundo orden: U-PLS/RBL	20
2.4. Programas	22
2.5. Materiales y métodos	22
2.5.1. Equipamiento	22
2.5.2. Reactivos	23
2.5.3. Diseño del juego de calibración	23
2.5.4. Diseño del juego de validación	23
2.5.5. Procedimiento experimental	24
2.6. Resultados y discusión	24
2.6.1. Selección de regiones	24

2.6.2. Determinación del número de factores	24
2.6.3. Resultados	25
2.7. Conclusiones	29
CAPÍTULO 3: Estudio de diferentes sistemas experimentales con datos instrumentales de segundo orden en presencia de solapamiento extremo. Resolución basada en la obtención de la ventaja de segundo orden a través del algoritmo MCR-ALS.	30
3.1. Introducción	31
3.2. Objetivos	34
3.3. Teoría	34
3.3.1. MCR-ALS	34
3.4. Programas	40
3.5. Sistema 1: Determinación de Fe(II) presente como catalizador de la reacción de oxidación de naranja de metilo con bromato de potasio	41
3.5.1. Materiales y métodos	41
3.5.1.1. Equipamiento	41
3.5.1.2. Reactivos	41
3.5.1.3. Diseño de los juegos de calibración y validación	42
3.5.2. Resultados y discusión	42
3.6. Sistema 2: Determinación de tartrazina por oxidación con bromato de potasio catalizada por Fe(II), en presencia de una interferencia causada por azul brillante	45
3.6.1. Materiales y métodos	45
3.6.1.1. Equipamiento	45
3.6.1.2. Reactivos	46
3.6.1.3. Diseño de los juegos de calibración y validación	46
3.6.1.4. Muestras reales	47
3.6.1.5. Procedimiento cromatográfico para muestras reales	47
3.6.2. Resultados y discusión	47
3.6.2.1. Muestras de validación	47
3.6.2.2. Muestras reales	51
3.7. Sistema 3: Determinación de carbamazepina en suero en	

presencia de las interferencias carbamazepina epóxido y lamotrigina	53
3.7.1. Preprocesamiento de los datos	53
3.7.1.1. Corrección de la línea de base	53
3.7.1.2. Corrección del corrimiento del tiempo de migración de los picos	55
3.7.2. Materiales y métodos	55
3.7.2.1. Equipamiento	55
3.7.2.2. Reactivos	56
3.7.2.3. Procedimiento de extracción	56
3.7.2.4. Diseño de los juegos de calibración y validación	57
3.7.3. Resultados y discusión	58
3.7.3.1. Corrección de la línea de base y el corrimiento temporal	58
3.7.3.2. Aplicación de MCR-ALS a mezclas en SCR y muestras de suero adicionadas	63
3.7.3.2.1. Análisis exploratorio del rango de los juegos de datos	64
3.7.3.2.2. Resultados de la calibración multivariada	66
3.8. Conclusiones	71
CAPÍTULO 4: Estudio comparativo del rendimiento de U-PLS/RBL frente a MCR-ALS en la determinación de ocho tetraciclinas presentes en aguas residuales por cromatografía líquida de alta resolución con detección por arreglo de diodos, previa preconcentración en fase sólida.	72
4.1. Introducción	73
4.2. Objetivos	76
4.3. Teoría	76
4.3.1. U-PLS/RBL y MCR-ALS	76
4.3.1.1. Cifras de mérito analíticas en U-PLS/RBL	76
4.3.1.2. Cifras de mérito analíticas en MCR-ALS	76
4.3.2. Corrección de la línea de base	77
4.3.3. Estandarización directa sucesiva	77
4.4. Programas	78
4.5. Materiales y métodos	79

4.5.1. Equipamiento	79
4.5.2. Reactivos	79
4.5.3. Recolección y preparación de la muestra	80
4.5.4. Preparación de estándares y muestras adicionadas	80
4.5.5. Procedimientos	81
4.5.5.1. Procedimiento para preconcentrar tetraciclinas presentes en agua residual	81
4.5.5.2. Procedimiento cromatográfico	82
4.6. Resultados y discusión	83
4.6.1. Análisis cromatográfico	83
4.6.2. Pretratamiento de los datos	85
4.6.3. Resolución	88
4.6.3.1. Cuantificación de tetraciclinas en agua residual por MCR-ALS y U-PLS/RBL	88
4.6.3.2. Cifras de mérito analíticas multivariadas	99
4.7. Conclusiones	100
CAPÍTULO 5: U-PCA/RBL: una estrategia versátil para obtener la ventaja de segundo orden al aplicar diferentes redes neuronales artificiales a datos de segundo orden no lineales.	102
5.1. Introducción	103
5.2. Objetivos	107
5.3. Teoría	108
5.3.1. U-PCA/RBL	108
5.3.2. Perceptrones multicapa	109
5.3.3. Funciones de base radial	111
5.3.4. Máquinas de vectores soportes	113
5.4. Programas	115
5.5. Sistema Fe (II)	115
5.5.1. Materiales y métodos	116
5.5.1.1. Equipamiento	116
5.5.1.2. Reactivos	117
5.5.1.3. Diseño del juego de entrenamiento	117
5.5.1.4. Diseño del juego de validación	117

5.5.1.5. Procedimiento experimental	117
5.5.2. Resultados y discusión	118
5.6. Sistema loratadina-pseudoefedrina sulfato	122
5.6.1. Materiales y métodos	124
5.6.1.1. Equipamiento	124
5.6.1.2. Reactivos	125
5.6.1.3. Diseño del juego de entrenamiento	126
5.6.1.4. Diseño del juego de monitorización	126
5.6.1.5. Muestra real	126
5.6.1.6. Procedimiento cromatográfico para muestras reales	126
5.6.2. Resultados y discusión	128
5.6.2.1. Detección de no linealidades	128
5.6.2.2. Determinación de las arquitecturas de las redes MLPs	131
5.6.2.3. Predicción en muestras sintéticas y reales mediante U-PCA/RBL-MLP	133
5.6.2.3.1. Muestras de validación	133
5.6.2.3.2. Muestra real	136
5.6.2.4. Predicción de muestras sintéticas mediante U-PCA/RBL seguido de MLP, RBF y SVM	137
5.6.2.4.1. Consideraciones generales	137
5.6.2.4.2. Predicción de las muestras de validación	139
5.7. Conclusiones	143
CAPÍTULO 6: Conclusiones	144
CAPÍTULO 7: Resumen	148
CAPÍTULO 8: Bibliografía	153

ÍNDICE DE TABLAS

	DESCRIPCIÓN DE CONTENIDOS	PÁGINA
Tabla 2.1.	Resultados de las muestras de validación empleando los métodos PARAFAC y U-PLS/RBL para predecir la concentración de tetraciclina.	26
Tabla 3.1	Concentraciones nominales y predicciones de las muestras de validación del sistema 1.	45
Tabla 3.2	Concentraciones nominales y predicciones de las muestras de validación del sistema 2.	50
Tabla 3.3	Resultados obtenidos aplicando el test de comparación de medias entre métodos para los resultados obtenidos mediante MCR-ALS y CLAR.	52
Tabla 3.4	Composición de las muestras de validación.	58
Tabla 3.5	Rango estimado calculado a través de la descomposición en valores singulares de los juegos de datos individuales y aumentados.	65
Tabla 3.6	Concentraciones de las muestras de validación y resultados de la predicción mediante MCR-ALS.	68
Tabla 4.1	Composición de las muestras de validación obtenidas adicionando cantidades diferentes de tetraciclinas a un efluente doméstico de agua residual.	81
Tabla 4.2	Programa de gradiente usado para separar las ocho tetraciclinas.	82
Tabla 4.3	Regiones en las cuales los datos cromatográficos totales se dividieron para construir modelos quimiométricos individuales para el análisis mediante MCR-ALS y U-PLS/RBL con los correspondientes factores y componentes no esperados empleados.	90
Tabla 4.4	Predicciones mediante MCR-ALS para las muestras de validación obtenidas mediante adición de diferentes cantidades de tetraciclinas a un efluente doméstico de agua residual.	96

Tabla 4.5	Predicciones mediante U-PLS/RBL para las muestras de validación obtenidas mediante adición de diferentes cantidades de tetraciclinas a un efluente doméstico de agua residual.	97
Tabla 4.6	Cifras de mérito para el modelado mediante MCR-ALS y U-PLS/RBL.	100
Tabla 5.1	Resultados obtenidos de la aplicación de U-PLS/RBL y U-PCA/RBL-MLP a las muestras de validación para el sistema Fe(II).	120
Tabla 5.2	Resultados del entrenamiento de la red MLP para el sistema Fe(II).	121
Tabla 5.3	Composición del juego de calibración siguiendo un diseño experimental factorial completo en cinco niveles.	127
Tabla 5.4	Composición del juego de validación siguiendo un diseño experimental central compuesto.	128
Tabla 5.5	Resultados de las pruebas Runs y Durbin-Watson aplicadas a los gráficos APaRP para evaluar el grado de no linealidad.	131
Tabla 5.6	Parámetros de las redes MLPs	133
Tabla 5.7	Concentraciones de las muestras de validación y resultados de la predicción empleando U-PLS/RBL, MCR-ALS y U-PCA/RBL-MLP para LOR.	135
Tabla 5.8	Concentraciones de las muestras de validación y resultados de la predicción empleando PLS/RBL, MCR-ALS y U-PCA/RBL-MLP para PES.	136
Tabla 5.9	Predicciones y análisis estadístico para loratadina y pseudoefedrina sulfato.	142

ÍNDICE DE FIGURAS

DESCRIPCIÓN DE CONTENIDOS	PÁGINA
Figura 1.1. Arreglo tridimensional típico en donde se muestra la intensidad de fluorescencia de una muestra de tetraciclina en suero en función de las longitudes de onda de excitación y emisión.	4
Figura 1.2. Representación de los datos de complejidad creciente con el orden asociado en paréntesis. El término multivía hace referencia a datos con dos o más dimensiones por muestra. (Extraída de Escandar y col 2007).	5
Figura 1.3. Tres modos básicos de obtener la ventaja de segundo orden a partir de datos de orden superior: A) combinando los datos de la calibración y las muestras de validación antes de calcular los coeficientes de regresión, B) estimando las cargas (<i>loadings</i>) empleando solo los datos de calibración y calculando los coeficientes de regresión luego de introducir los datos correspondientes a las muestras de validación y C) modificando las puntuaciones (<i>scores</i>) de las muestras de validación antes de realizar la predicción con las ANNs, sin alterar los parámetros de las redes .	10
Figura 2.1. Arreglo de tres vías \underline{X} formado por el apilamiento de las I matrices de calibración y la matriz de la muestra incógnita, de dimensiones $(I + 1) \times J \times K$.	17
Figura 2.2. Representación gráfica de un modelo PARAFAC de dos componentes.	17
Figura 2.3. Gráfico que indica cómo se realiza la predicción de una muestra de validación que contiene componentes no esperados mediante U-PLS/RBL.	22
Figura 2.4. Figura tridimensional de las EEMs de fluorescencia para una muestra típica de suero (A) y para el analito	

	tetraciclina en una concentración de 3,00 mg L ⁻¹ (B), lo cual muestra la baja sensibilidad del sistema con respecto al analito.	28
Figura 3.1	Matriz aumentada D constituida por el apilamiento de las matrices de calibración y la matriz de la muestra incógnita.	35
Figura 3.2	Gráfico tridimensional de absorbancia en función de longitud de onda y tiempo de reacción para una muestra de calibración que contiene 5,00 mg L ⁻¹ de Fe(II), durante la reacción con bromato de potasio (los datos fueron convertidos de la manera descrita en el texto, es decir, se sustrajo el espectro a tiempo cero y se tomó el módulo).	43
Figura 3.3	Gráfico de los perfiles cinéticos y espectrales luego del procesamiento por MCR-ALS de una muestra de validación típica junto con las muestras de calibración del sistema 1. A) Perfiles cinéticos normalizados a la unidad y B) perfiles espectrales para matrices de muestras sucesivas aumentadas en el modo espectral (la primera es la muestra de validación y las restantes son las cuatro primeras muestras de calibración; las líneas punteadas verticales separan los sensores de cada muestra). Se muestran solo cuatro muestras de calibración por simplicidad. La línea sólida corresponde al analito y la línea de trazos a la interferencia.	44
Figura 3.4	Figura 3.4. Gráfico tridimensional de absorbancia en función de longitud de onda y tiempo de reacción para: A) una muestra de calibración que contiene 16,00 mg L ⁻¹ de tartrazina y B) una muestra de validación que contiene 15,00 mg L ⁻¹ tanto de tartrazina como de azul brillante, durante la reacción con bromato de potasio.	48

-
- Figura 3.5** Gráfico de los perfiles espectrales y cinéticos luego del procesamiento por MCR-ALS de una muestra de validación típica junto con las muestras de calibración del sistema 2. A) Perfiles espectrales normalizados a la unidad y B) perfiles cinéticos para matrices de muestras sucesivas aumentadas en el modo cinético (la primera es la muestra de validación y las restantes son las cuatro primeras muestras de calibración; las líneas punteadas verticales separan los sensores de cada muestra). Se muestran solo cuatro muestras de calibración por simplicidad. La línea sólida corresponde al analito y la línea de trazos a la interferencia. 51
- Figura 3.6** A) Superficie correspondiente a una muestra de suero (SC1) adicionada con CBZ (4,96 mg L⁻¹), CBZ-EP (4,99 mg L⁻¹) y LAMO (5,66 mg L⁻¹). B) Matriz *background* correspondiente a la superficie de la Figura 3.6 A. C) Línea sólida: electroferograma correspondiente a una muestra de suero (SC1); línea de trazos y puntos: su correspondiente *background*; línea de trazos: electroferograma corregido (todos registrados a 250 nm). 60
- Figura 3.7** A) Electroferograma registrado a 250 nm para cinco estándares de CBZ (líneas sólidas) y una muestra de suero adicionada (SC1, línea de rayas). B) Variación RES(%) [ver Ecuación (3.14)] cuando la matriz estudiada se modifica con respecto a la matriz de referencia. C) Electroferogramas con corrimientos corregidos registrados a 250 nm. 62
- Figura 3.8** A) Electroferograma de una muestra adicionada preparada en SRC, en la cual los picos corresponden a: 1) FEN (20,00 mg L⁻¹); 2) CBZ, CBZ-EP (5,00 mg L⁻¹), CAF y LAMO (10,00 mg L⁻¹); 3) PAR (10,00 mg L⁻¹);
-

	4) PHT (10,00 mg L ⁻¹); 5) IBU y TEO (10,00 mg L ⁻¹); 6) PHB (10,00 mg L ⁻¹) y 7) AAS (10,00 mg L ⁻¹). B) Electroferograma de la misma muestra luego del procedimiento de extracción: 1) CBZ, CBZ-EP y LAMO; 2) PHT; 3) IBU y TEO y 4) PHB. Ambos electroferogramas se registraron a 250 nm.	63
Figura 3.9	Electroferograma (registrado a 250 nm) correspondiente a una muestra de suero SB que contiene CBZ y CBZ-EP solapadas en el pico 2 y componentes no esperados en el pico 1 y 3.	66
Figura 3.10	Perfiles temporales obtenidos del procesamiento de una muestra SC mediante MCR-ALS correspondientes a: A) los tres componentes encontrados en la matriz SC; B) un estándar de CBZ-EP (5,00 mg L ⁻¹); C) un estándar de LAMO (5,00 mg L ⁻¹) y D) un estándar de CBZ (5,00 mg L ⁻¹).	69
Figura 3.11	Comparación de los perfiles espectrales obtenidos del procesamiento de la muestra SC analizada en la Figura 3.10. En los tres casos, los espectros de los estándares puros se representan en líneas de trazos mientras que los perfiles extraídos se muestran como líneas sólidas. A) CBZ, B) CBZ-EP y C) LAMO.	70
Figura 4.1	Cromatograma correspondiente a una solución estándar mezcla con 2,00 mg L ⁻¹ de (1) MIC, (2) OTC, (3) TC, (4) DMC, 5) CTC, 6) MTC, 7) DOC y 8) MEC. La detección se realizó a 360 nm.	84
Figura 4.2	Cromatogramas ($\lambda = 360$ nm) correspondientes a: (a) una solución estándar conteniendo 0,50 mg L ⁻¹ de las ocho tetraciclinas, (b) un extracto de una muestra real y (c) un extracto de la misma muestra obtenido luego de adicionarla con 2,00 μ g L ⁻¹ de las ocho tetraciclinas. Los números están asociados a los nombres en la Tabla 4.1.	84

Figura 4.3	Cromatogramas ($\lambda = 360 \text{ nm}$) correspondientes a (a) una solución estándar conteniendo $0,50 \text{ mg L}^{-1}$ de las ocho tetraciclinas y (b) una solución estándar conteniendo $2,00 \mu\text{g L}^{-1}$ de las ocho tetraciclinas luego del procedimiento de SPE.	85
Figura 4.4	Cromatogramas ($\lambda = 360 \text{ nm}$) restringidos al primer pico correspondientes a (a) un extracto SPE de $250,00 \text{ mL}$ de una muestra de agua milliQ adicionada con $2,00 \mu\text{g L}^{-1}$ de minociclina, (b) una solución estándar conteniendo $0,50 \text{ mg L}^{-1}$ de minociclina y (c) la última señal luego de la estandarización.	87
Figura 4.5	Cromatogramas originales ($\lambda = 360 \text{ nm}$) en líneas sólidas correspondiente a (a) un extracto de la muestra real y (b) este extracto luego de ser adicionado con $0,50 \text{ mg L}^{-1}$ de minociclina (pico N° 1). (c y d) son los correspondientes cromatogramas corregidos en líneas de puntos y (e) el perfil temporal correspondiente a un estándar de minociclina $0,50 \text{ mg L}^{-1}$.	88
Figura 4.6	Cromatograma registrado a múltiples longitudes de ondas correspondiente a una solución estándar mezcla conteniendo $0,50 \text{ mg L}^{-1}$ de cada tetraciclina.	89
Figura 4.7	Datos de segundo orden correspondientes al pico N° 8 (meclociclina) registrado para (A) una solución estándar con $0,50 \text{ mg L}^{-1}$, (B) una muestra real sin el analito antes del procedimiento de pretratamiento, (C) un extracto de muestra real sin el analito luego del paso de pretratamiento y (D) este extracto de muestra con $0,50 \text{ mg L}^{-1}$ de meclociclina.	91
Figura 4.8	Variación de los residuos de predicción en función de los valores de prueba para la muestra replicado R-1 cuando se analiza el pico N°2 (oxitetraciclina, tiempo de retención = $6,03 \text{ min}$).	92

Figura 4.9	Perfiles temporales de MCR-ALS para el pico N°2: (a) perfil de concentración para oxitetraciclina y (b) perfil de tiempo para la interferencia. (c) Perfil temporal de U-PLS/RBL para la interferencia presente en el pico N°2. (d) Cromatograma del componente puro normalizado para oxitetraciclina.	94
Figura 4.10	Perfiles espectrales de MCR-ALS para el pico N°2: (a) perfil espectral para oxitetraciclina y (b) perfil espectral para la interferencia. (c) Perfil espectral de U-PLS/RBL para la interferencia presente en el pico N°2. (d) Espectro normalizado del componente puro para oxitetraciclina.	94
Figura 4.11	(A) Perfiles de tiempo sucesivos correspondientes a cuatro componentes (el analito y las interferencias) encontrados mediante el análisis por MCR-ALS para el pico N°8 (meclociclina) en la muestra R-1. La línea roja sólida corresponde al perfil de meclociclina, mientras que la línea roja de círculos representa el componente puro normalizado del componente a 404 nm. (B) Los restantes cinco perfiles corresponden a cinco estándares de muestras con 0,10 - 0,20 - 0,40 - 0,60 y 1,00 mg L ⁻¹ respectivamente de meclociclina.	95
Figura 4.12	Gráficos de barras tridimensionales en donde se representan las recuperaciones obtenidas para cada tetraciclina, en las muestras de validación M-1 a M-6, mediante A) MCR-ALS y B) U-PLS/RBL.	98
Figura 4.13	Gráfico de barras tridimensional en donde se representan las recuperaciones obtenidas para las muestras R1-4 mediante MCR-ALS y U-PLS/RBL.	99
Figura 5.1	Gráfico que indica como se realiza la predicción de una muestra de validación que contiene componentes no esperados mediante U-PCA/RBL.	109
Figura 5.2	Espectro de absorción de una muestra de calibración	

	de Fe(II) (1,00 mg L ⁻¹) en naranja de metilo (16,00 mg L ⁻¹) (celeste), una solución de púrpura de bromocresol (12,00 mg L ⁻¹) (azul) y una muestra de Fe(II) en naranja de metilo con adición de púrpura de bromocresol (12,00 mg L ⁻¹) como interferencia (violeta).	116
Figura 5.3	Gráfica tridimensional de absorbancia en función de longitud de onda y tiempo para la reacción entre naranja de metilo (16,00 mg L ⁻¹) y bromato de potasio (2×10 ⁻³ mol L ⁻¹), catalizada por 1,00 mg L ⁻¹ de Fe(II).	118
Figura 5.4	Región elíptica de confianza conjunta para la pendiente y para la ordenada al origen, usando el ajuste cuadrados mínimos ordinarios entre concentración nominal versus la predicha, para U-PCA/RBL-MLP (azul) y para U-PLS/RBL (violeta). La marca en forma de cruz indica el punto teórico (1,0).	121
Figura 5.5	Perfiles espectrales (A) y temporales (B) para el componente no esperado en el sistema Fe(II). Los perfiles experimentales (líneas sólidas) y provistos por RBL (líneas de rayas) se normalizaron a la unidad. El procedimiento de RBL se aplicó a una muestra conteniendo naranja de metilo (16,00 mg L ⁻¹), bromato de potasio (2×10 ⁻³ mol L ⁻¹), Fe(II) (1,00 mg L ⁻¹) y púrpura de bromocresol como componente no esperado (12,00 mg L ⁻¹).	122
Figura 5.6	Espectro UV-Vis para LOR (línea violeta; 8,33 mg L ⁻¹), PES (línea anaranjada; 100,00 mg L ⁻¹) y una mezcla de los excipientes (línea verde) a valores de pH de 2,0 (líneas sólidas) y 7,0 (líneas de rayas).	124
Figura 5.7	Sistema de inyección en flujo.	125
Figura 5.8	Gráfico de los residuos parciales aumentados de Mallows (APaRP) para LOR (a) y PES (b) cuando se incluyen en el modelo, para ambos, los seis primeros	

	LVs.	129
Figura 5.9	Medias e intervalos de Tukey HSD versus redes para A) LOR, con arquitecturas A=4-6-1, B=4-7-1, C=4-8-1, D=5-8-1 y E=5-9-1 y B) PES, con arquitecturas A=8-9-1, B=8-8-1, C=8-7-1, D=7-7-1 y E=7-5-1.	132
Figura 5.10	Superficie obtenida recogiendo los espectros UV a diferentes puntos de tiempo FIA para una de las muestras de calibración que contenía 7,50 mg L ⁻¹ de LOR y 80,00 mg L ⁻¹ de PES.	140

CAPÍTULO 1

Introducción

1.1. CALIBRACIÓN MULTIVARIADA

Se denomina calibración multivariada a un conjunto de métodos quimiométricos que tienen por objetivo la cuantificación de uno o más componentes en mezclas incógnitas de multicomponentes (Martens y Næs 1989).

La quimiometría es una disciplina científica que se inició en 1969 y que ha crecido de manera sorprendente en los últimos años. Tal es así, que se cuenta con revistas especializadas de investigación en quimiometría, como el Chemometrics and Intelligent Laboratory System y el Journal of Chemometrics (Ramis Ramos y García Álvarez Coque 2001).

La quimiometría trata, específicamente, de todos aquellos procesos que transforman señales analíticas y datos más o menos complejos en información, para lo cual se utilizan métodos de origen matemático, estadístico y otros procedentes del campo de la lógica formal (Massart y col 1997). Esta disciplina se puede dividir en tres áreas generales que incluyen: la optimización de los procedimientos experimentales y las mediciones químicas, la extracción de la máxima información química de los datos analíticos, y la calibración, la validación y la significancia de las mediciones analíticas (Brereton 1990).

Los modelos matemáticos involucrados en la calibración multivariada relacionan múltiples señales instrumentales, no selectivas, con las concentraciones de analitos (Martens y Næs 1989, Massart y col 1997). Esto implica que, a diferencia de la calibración univariada, la calibración multivariada comprende el estudio de mezclas de varios componentes, tanto durante la etapa de calibración como la de predicción (Haaland y Thomas 1988).

1.1.1. CLASIFICACIÓN DE LOS DATOS INSTRUMENTALES

La calibración multivariada está estrechamente relacionada con el tipo de datos instrumentales de los cuales se dispone para cuantificar el o los componentes de interés en una muestra. Es el tipo de datos obtenido el que condiciona el tipo de calibración que debe realizarse. Dentro de este

contexto, los datos se clasifican según su orden, es decir, el número de dimensiones que presentan para una única muestra, del siguiente modo (Booksh y Kowalski 1994, Escandar y col 2007):

- Datos de orden cero: son generados por instrumentos que producen una única respuesta por muestra (un tensor de orden cero) como, por ejemplo, la absorbancia medida a una única longitud de onda o la lectura de un electrodo ión selectivo.
- Datos de primer orden: los datos de primer orden para una dada muestra están ordenados en forma de vector o tensor de primer orden. Son ejemplo de este tipo de datos los espectros obtenidos mediante espectrofotometría UV-Vis, espectrofluorimetría, espectroscopia de infrarrojo (IR) o infrarrojo cercano (NIR) y resonancia magnética nuclear (MNR); o los barridos electroquímicos obtenidos empleando voltametría o cronoamperometría.
- Datos de segundo orden: se obtiene este tipo de datos cuando por cada muestra analizada se genera una matriz, la cual se puede visualizar como una superficie (ver Figura 1.1). Se pueden obtener de dos maneras: empleando un único instrumento como, por ejemplo, un espectrofluorímetro registrando matrices de excitación-emisión (EEMs) o un espectrofotómetro de arreglo de diodos siguiendo la cinética de una reacción química; o mediante el acoplamiento de dos instrumentos de primer orden como, por ejemplo, cromatografía de gases acoplada a espectrometría de masas (GC-MS), GC-GC, y MS(MS)² en tándem, etc. Cuando los datos de segundo orden para un grupo de muestras se juntan en un único arreglo tridimensional, el objeto resultante se conoce como arreglo de tres vías. Con el término vía se indica el número de dimensiones que tiene un arreglo de datos formado por la agrupación de un cierto número de muestras. Por lo tanto, los datos de segundo orden también se conocen como datos de tres vías.
- Datos de tercer orden: la introducción de una dimensión extra en los datos conduce a datos de orden superior. Por ejemplo, si se agrupan datos de

tercer orden correspondientes a varias muestras, se obtiene un arreglo de cuatro vías. Un ejemplo de este tipo de arreglos es el obtenido al estudiar la cinética de una reacción con EEMs.

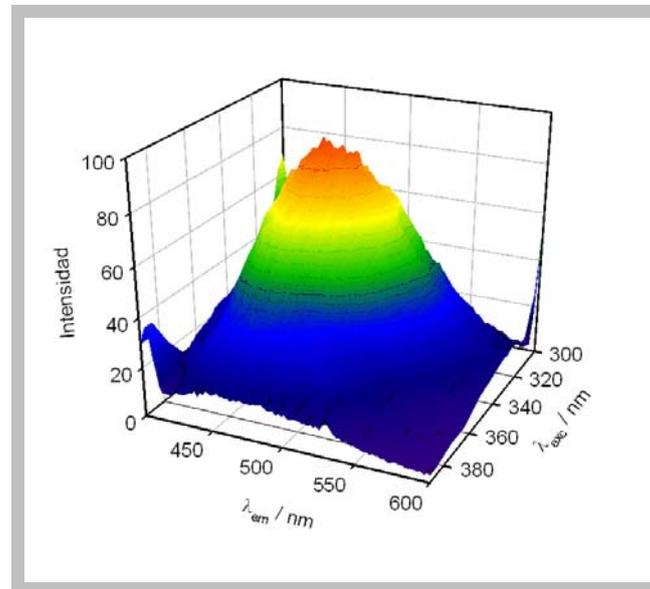


Figura 1.1. Arreglo tridimensional típico en donde se muestra la intensidad de fluorescencia de una muestra de tetraciclina en suero en función de las longitudes de onda de excitación y emisión.

En la Figura 1.2 se representa la evolución de los datos desde los más simples, de orden cero, hasta los multivariados. Los datos que no son univariados se conocen como multivariados, mientras que los datos de segundo orden y órdenes superiores se conocen como datos multivía.

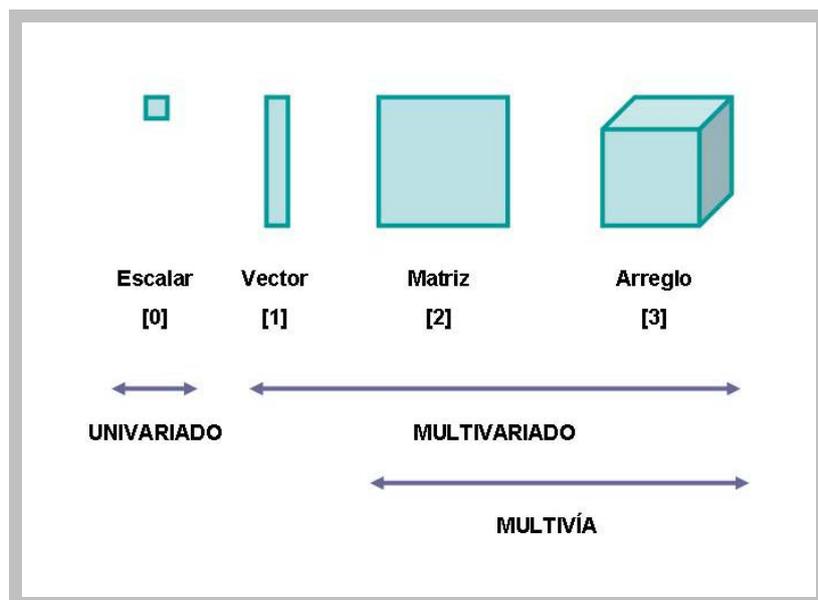


Figura 1.2. Representación de los datos de complejidad creciente con el orden asociado en paréntesis. El término multivía hace referencia a datos con dos o más dimensiones por muestra. (Figura extraída de Escandar y col 2007).

1.1.2. TERMINOLOGÍA

Es preciso definir cierta terminología que será utilizada en el transcurso de la presente tesis, con el objetivo de interpretar correctamente lo que se expone.

1.1.2.1. Tipos de componentes presentes en las muestras

Los componentes presentes tanto en las muestras de calibración como en las de validación se denotan como 'esperados', debido a que se espera que estén presentes en la mayoría de las muestras de validación y se los emplea para construir un juego de muestras de entrenamiento lo suficientemente representativo. Los componentes esperados pueden a su vez diferenciarse en 'calibrados' y 'no calibrados', haciendo referencia a si las concentraciones de cada uno están o no disponibles. Por otra parte, las muestras pueden contener componentes adicionales realmente desconocidos denominados 'componentes no esperados' (Escandar y col 2007, Olivieri 2008).

Cabe mencionar que las potenciales interferencias no siempre van a generar un error sistemático en la determinación del analito de interés (van der Linden 1989), ya que esto dependerá del tipo de señales instrumentales

generadas y de la metodología de calibración empleada. Para datos instrumentales de primer orden los componentes no esperados muy probablemente constituirán una interferencia. Sin embargo, como se verá más adelante, este inconveniente puede ser superado si se cuenta con datos de segundo orden y algoritmos para procesarlos que gocen de la ventaja de segundo orden (Escandar y col 2007, Olivieri 2008).

1.1.2.2. Estructura de los datos instrumentales de segundo orden

En general, la matriz de datos Z ($J \times K$) para un componente puro en ausencia de ruido, se puede expresar como el producto interno de dos vectores:

$$Z = \mathbf{xy}^T \quad (1.1)$$

en donde \mathbf{x} ($J \times 1$) e \mathbf{y} ($K \times 1$) denotan los perfiles en ambas dimensiones. Idealmente, los perfiles de \mathbf{x} e \mathbf{y} son independientes uno de otro. Este tipo de datos se denomina 'bilineal', es decir, lineal en una variable cuando la otra está fija y viceversa.

Cuando una matriz de datos se agrupa en un arreglo de tres vías, la resultante se puede expresar como:

$$M_{ijk} = x_i y_j z_k \quad (1.2)$$

donde M_{ijk} es un elemento del arreglo \mathbf{M} de tres vías (de tamaño $I \times J \times K$) y x_i , y_j y z_k son los elementos de los perfiles de una dimensión \mathbf{x} , \mathbf{y} y \mathbf{z} que describen el comportamiento de la señal medida en cada una de las dimensiones de los datos. En este caso el arreglo de datos se conoce como 'trilineal', indicando que existe una relación interna entre cada una de las matrices y que, en general, la descomposición de estos arreglos es única (Escandar y col 2007, Bro 1998).

1.1.3. ALGORITMOS QUIMIOMÉTRICOS

1.1.3.1. Algoritmos quimiométricos para datos de orden cero, uno y dos

Se puede decir que la razón por la cual se emplean métodos multivariados no es la obtención de un mejor ajuste, sino modelos más adecuados, robustos e interpretables. Los métodos multivariados son menos sensibles al ruido y generan información específica que puede relacionarse directamente con los diferentes modos de un arreglo multivariado (Bro 1997).

Para cada tipo de datos existen una variedad de algoritmos quimiométricos que permiten procesarlos con propósitos analíticos.

Con los datos de orden cero se realiza una calibración univariada, ajustando los datos por cuadrados mínimos, que requiere de selectividad completa para el analito de interés (Danzer y Currie 1998).

Los datos de primer orden se procesan mediante algoritmos de primer orden adecuados, siendo los más populares la regresión en componentes principales (PCR) y los cuadrados mínimos parciales (PLS) (Martens y Naes 1989). El objetivo principal de estos algoritmos es compensar la falta de selectividad, extrayendo de la señal total la porción que se puede emplear para predecir la concentración de un componente particular de la muestra. Por otra parte, cuando los datos se comportan de un modo no lineal con respecto al analito, se pueden emplear distintos tipos de redes neuronales artificiales (ANNs), como por ejemplo: el perceptrón multicapa (MLP) (Zupan y Gasteiger 1999), las funciones de base radial (RBF) (Derks y col 1995) o las máquinas de vectores soportes con ajuste por cuadrados mínimos (LS-SVM) (Suykens y col 2002). Cuando se aplican los algoritmos de primer orden, en la mayoría de los casos, la calibración se puede realizar ignorando las concentraciones de todos los componentes excepto las del analito de interés. Si las muestras de predicción contienen componentes no esperados, la determinación será errónea. Afortunadamente, esta muestra podrá ser identificada como anómala debido a que los residuos espectrales serán anormalmente grandes en comparación con el ruido instrumental. Esta propiedad se conoce como la ventaja de primer orden (Booksh y Kowalski 1994).

El analista puede recurrir a un número grande de algoritmos para manejar datos de segundo orden. Dentro de los más empleados se pueden citar: el análisis paralelo de factores (PARAFAC) (Bro 1997), el método de aniquilación de rango generalizado (GRAM) (Sanchez y Kowalski 1986), la descomposición trilineal directa (DTLD) (Sanchez y Kowalski 1990), la resolución multivariada de curvas mediante cuadrados mínimos alternantes (MCR-ALS) (De Juan y col 2000), los cuadrados mínimos bilineales (BLLS) (Linder y Sundberg 1998, Linder y Sundberg 2002) y la descomposición trilineal alternante (ATLD) (Wu y col 1998) y sus variantes descomposición trilineal alternante autoponderada (SWATLD) (Chen y col 2000) y descomposición trilineal alternante penalizada (APTLD) (Hu y col 2006, Xia y col 2005). Estos algoritmos cobran mayor importancia en el análisis de mezclas complejas debido a que cuentan con la ventaja de segundo orden. Esta propiedad permite no solo señalar la muestra que tiene componentes no esperados como anómala sino, también, modelar la presencia de las potenciales interferencias causadas por componentes no esperados en la muestra y cuantificar de manera exacta el o los analitos de interés (Booksh y Kowalski 1994). Además, los analitos pueden ser calibrados empleando un número pequeño de muestras en comparación con la cantidad de muestras que se necesitan para realizar una calibración univariada. De los algoritmos mencionados, en el presente trabajo de tesis se estudiaron las capacidades del algoritmo MCR-ALS para resolver sistemas experimentales con datos de segundo orden, en presencia de solapamiento extremo.

Una alternativa para trabajar con datos de segundo orden es reordenar los datos en vectores y aplicar metodologías de primer orden, por ejemplo, a través de componentes principales desdoblados (U-PCR) o cuadrados mínimos parciales desdoblados (U-PLS) (Wold y col 1987). Estos algoritmos se usaron antes de que aparezcan los métodos genuinos de segundo orden como el algoritmo cuadrados mínimos parciales multivía (N-PLS) (Bro 1996). Sin embargo, ninguno de ellos presenta la ventaja de segundo orden, a menos que se los acople a un procedimiento separado de bilinealización residual (Öhman y col 1990a). En la presente tesis se evaluará el rendimiento del algoritmo cuadrados mínimos parciales desdoblados seguidos de bilinealización residual (U-PLS/RBL) (Olivieri 2005a). El mismo es un método promisorio que presenta

las capacidades de las metodologías basadas en factores latentes, preservando la ventaja de segundo orden.

Sin embargo, el analista estará condicionado a la estructura de los datos que posee. Los algoritmos mencionados no se pueden implementar cuando las señales registradas presentan una relación no lineal con la concentración de analito. En la presente tesis se propuso la obtención de la ventaja de segundo orden, a partir de información de segundo orden no lineal, mediante la aplicación de un nuevo algoritmo: componentes principales desdoblados seguidos de bilinealización residual (U-PCA/RBL), como un modo de preprocesar datos instrumentales de segundo orden para poder ser tratados luego con diferentes tipos de redes neuronales artificiales.

En general, existen tres maneras de obtener la ventaja de segundo orden. En la primera, los datos de las muestras de validación tienen influencia en el cálculo de los coeficientes de regresión que conducen a la predicción (ver Figura 1.3 A), mientras que en la segunda se realiza primero la calibración usando los datos de entrenamiento, con las muestras de validación conduciendo a coeficientes de regresión específicos de la muestra, en un paso posterior (ver Figura 1.3 B). La última estrategia consiste en modificar las puntuaciones (*scores*) de las muestras de validación antes de realizar la predicción con ANNs, sin alterar los parámetros de las redes (ver Figura 1.3 C). Todos los casos implican que la muestra de validación es parte del proceso global de calibración (Escandar y col 2007).

Los algoritmos PARAFAC, SWATLD, GRAM, DLTD y MCR-ALS obtienen la ventaja de segundo orden empleando el esquema expuesto en la Figura 1.3 A, mientras que BLLS/RBL y U-PLS/RBL emplean el procedimiento presentado en la Figura 1.3 B. Ambos modos de obtener la ventaja de segundo orden tienen sus ventajas y desventajas. Mientras que la descomposición trilineal provee perfiles de los componentes con significado físico (perfiles espectrales, temporales, de pH, etc), esta información se pierde parcialmente dentro de la estructura de variable latente de la bilinealización residual. Sin embargo, este último algoritmo es más flexible en lo que respecta a los datos multivía que no cumplen con el modelo requerido para la descomposición trilineal

(Olivieri 2008). En el Capítulo 5 se presentará por primera vez el algoritmo U-PCA/RBL, el cual emplea la forma de calibrar mostrada en la Figura 1.3 C.

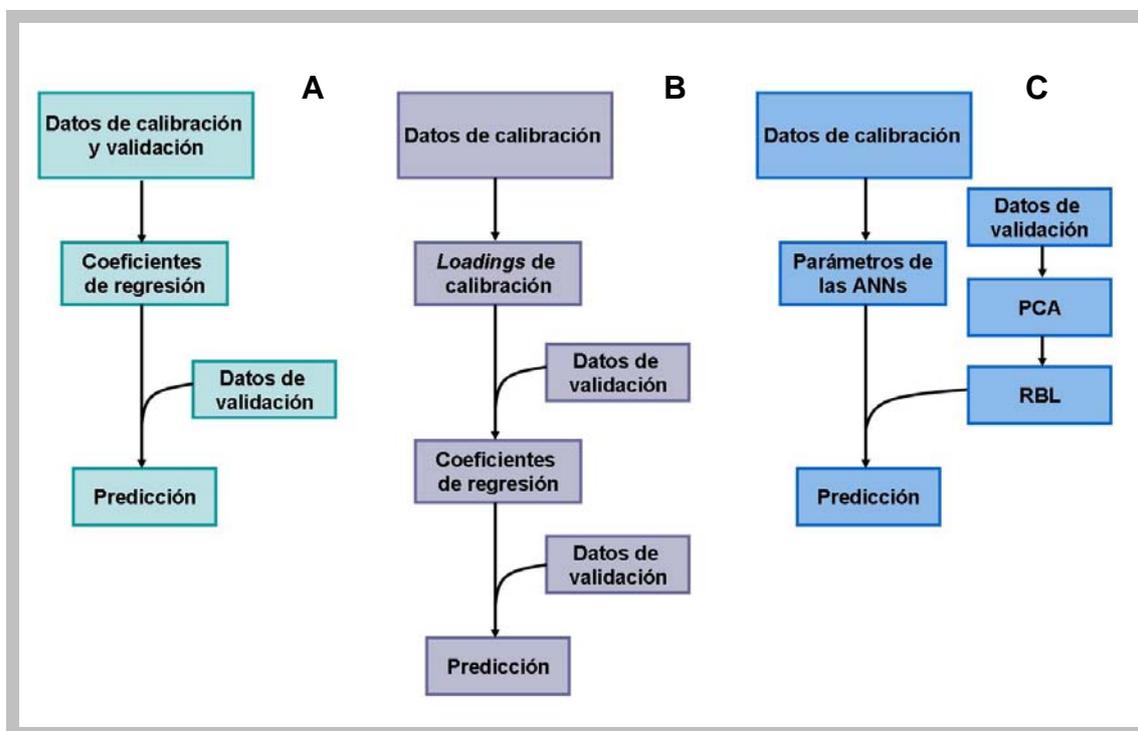


Figura 1.3. Tres modos básicos de obtener la ventaja de segundo orden a partir de datos de orden superior: A) combinando los datos de la calibración y las muestras de validación antes de calcular los coeficientes de regresión, B) estimando las cargas (*loadings*) empleando solo los datos de calibración y calculando los coeficientes de regresión luego de introducir los datos correspondientes a las muestras de validación y C) modificando las puntuaciones (*scores*) de las muestras de validación antes de realizar la predicción con las ANNs, sin alterar los parámetros de las redes.

1.1.3.2. Algoritmos quimiométricos para datos de N -dimensiones

Una característica importante de los algoritmos de segundo orden es que se pueden extender a datos de órdenes superiores. En este sentido, U-PCR, U-PLS, N-PLS y PARAFAC se pueden usar para datos de N -dimensiones. Sin embargo, para BLLS/RBL y U-PLS/RBL se encuentran disponibles extensiones para el tercer orden que dan lugar a TLLS/RTL (Arancibia y col 2006) y a U-PLS/RTL (Arancibia y col 2006), respectivamente.

1.1.4. PRESENTE Y FUTURO DEL ANÁLISIS MULTIVARIADO

La variedad de datos instrumentales multidimensionales que se están generando hoy en día, gracias al avance de la tecnología en lo que respecta a instrumentación moderna cada vez más sofisticada, hace que el análisis multivariado de segundo y tercer orden cobre cada vez mayor relevancia y aceptación dentro de la comunidad científica. A esto se le suma la multiplicidad de algoritmos quimiométricos disponibles para analizar eficientemente esta información y poder resolver muestras complejas provenientes de sistemas medioambientales, biológicos y alimenticios (Escandar y col 2007).

La investigación científica en el contexto del análisis multivariado es constante e intensiva, y pone de manifiesto que las aplicaciones y las potencialidades de la ventaja de segundo orden para resolver problemas analíticos son inmensas. Si bien se sugirió una ventaja teórica para los datos N -dimensionales (Booksh y Kowalski 1994), hasta el momento no se evidencian ventajas adicionales a las ya descritas (Olivieri 2008).

Como se mencionó, la ventaja de primer orden permite marcar una muestra que contiene componentes no esperados como anómala e impedir su cuantificación. Por otra parte, la ventaja de segundo orden permite la cuantificación exacta del o los analitos de interés, aún en presencia de componentes no esperados en la muestra. Siguiendo con este análisis, cabría preguntarse si es posible que los datos de órdenes mayores a dos provean ventajas adicionales. Algunos autores proponen alternativas para ser consideradas como posibles ventajas de tercer orden: 1) cuando se dispone de datos trilineales para una dada muestra, la posibilidad de obtener los tres perfiles contribuyentes, de manera única, para cada especie presente en la muestra (Booksh y Kowalski 1994, Sihna y col 2004a, Sihna y col 2004b), 2) la resolución algorítmica mejorada de los datos de tercer orden altamente colineales (Xia y col 2007), o 3) el aumento en la sensibilidad y la selectividad que acompaña a los datos multivía de mayores dimensiones (van Mispelaar 2005). Aunque todavía no existe consenso sobre cuál es o si realmente existe una ventaja adicional al procesar este tipo de datos, será la incesante investigación en el tema la cual se encargará de corroborar, refutar o generar una nueva respuesta a este interrogante.

1.2. OBJETIVOS

Los objetivos de la presente tesis son:

- 1) Desarrollar y evaluar las potencialidades de nuevos algoritmos quimiométricos para el tratamiento de datos multidimensionales.

- 2) Desarrollar métodos de análisis espectroscópicos, sencillos y económicos, que permitan la determinación de especies químicas que suelen formar parte de muestras complejas: fármacos y sus metabolitos en materiales biológicos (sangre, orina, líquidos de punción, líquido amniótico, lágrimas, leche, etc.), en preparaciones farmacéuticas y en muestras ambientales y aditivos en muestras alimenticias.

- 3) Validar las metodologías analíticas desarrolladas y sugerir su uso en estudios de farmacocinética y biodisponibilidad de medicamentos, en el control de calidad de preparaciones farmacéuticas y productos alimenticios, y en la determinación de fármacos como contaminantes ambientales.

CAPÍTULO 2

Evaluación del algoritmo U-PLS/RBL mediante su aplicación al análisis de muestras biológicas complejas en presencia de efecto matriz

2.1. INTRODUCCIÓN

El uso de matrices de excitación-emisión para la determinación de analitos en muestras biológicas complejas (Arancibia y Escandar 2003, Damiani y col 2004, Arancibia y col 2002, Muñoz de la Peña y col 2003, Escandar y col 2004, Herget y Escandar 2003, Cao y col 2002, Esteves Da Silva y col 2002, Saurina y col 2000, Saurina y Tauler 2000, Beltran y col 1998, Trevisan y Poppi 2003, Rodríguez-Cuesta y col 2003, Moberg y col 2001, Mahedero y col 2005, Christensen y col 2005), generalmente implica al algoritmo PARAFAC (Bro 1997, Andersen y Bro 2003, Faber y col 2003), el cual realiza una descomposición única en arreglos de tres dimensiones, permitiendo la extracción directa de los perfiles de concentración y espectrales de los componentes de la muestra, constituyendo la base de la ventaja de segundo orden (Booksh y Kowalsky 1994). Por otra parte, como se comentó en el capítulo anterior, la combinación de cuadrados mínimos bilineales con bilinealización residual (BLLS/RBL) mostró ser útil para el análisis de muestras complejas (Damiani y col 2004, Linder y Sundberg 1998, Linder y Sundberg 2002, Goicoechea y Olivieri 2005, Marsili y col 2005). Otras metodologías para procesar este tipo de datos, basadas en el uso de variables latentes, como N-PLS, U-PLS y U-PCR, (Wold y col 1987) no exhiben la citada ventaja (Bro 1996), aunque esta puede ser obtenida mediante la aplicación de un paso de bilinealización residual (RBL) (Öhman y col 1990a, Öhman y col 1990b, Olivieri 2005a).

La combinación de U-PLS con RBL resultó ser una buena alternativa para procesar datos instrumentales de segundo orden (extensible a órdenes superiores) que presentan dependencia lineal debida al pH o a fenómenos cinéticos, conservando la ventaja de segundo orden (Olivieri 2005a). Este algoritmo está basado en la conocida regresión en cuadrados mínimos parciales (PLS), aplicada una vez que los datos originales se han desdoblado para transformarlos en vectores, seguida de un paso posterior llamado bilinealización residual (RBL). Esta última etapa permite al nuevo algoritmo obtener la ventaja de segundo orden, esencial para el análisis de mezclas complejas conteniendo interferencias no calibradas. El mismo goza de las conocidas bondades de los algoritmos quimiométricos basados en factores

espectrales, que les otorgan mayor flexibilidad y robustez en comparación con métodos clásicos de cuadrados mínimos, además de permitir la presencia de compuestos químicos desconocidos en el conjunto de calibrado. Estas características lo hacen especialmente útil para casos en que las señales del analito se modifican en presencia de la matriz, que tradicionalmente deben resolverse por la técnica de adición estándar, más engorrosa desde el punto de vista experimental y más proclive a errores.

Las características de U-PLS/RBL se han probado sobre datos simulados digitalmente en una variedad de sistemas, en los que ha mostrado igual o mejor capacidad predictiva que sus competidores, particularmente el popular PARAFAC (Olivieri 2005a).

En el presente estudio se realizó el procesamiento de datos experimentales de excitación-emisión de fluorescencia correspondientes a un sistema biomédico que contiene interferencias no modeladas. Las muestras estaban compuestas por diferentes sueros humanos adicionados con el antibiótico tetraciclina. La interacción con el suero modifica de manera significativa las propiedades espectrales del analito (Goicoechea y Olivieri 1999, Sena y col 2005), por lo que la calibración debió incluir a la matriz biológica. Debido a esto, las interacciones entre el analito y la matriz, así como los cambios espectrales asociados, pueden diferir de muestra a muestra, generando un problema para PARAFAC que necesita que los perfiles sean reproducibles, o dicho de otra manera, que los datos sean trilineales (Bro 1997, Andersen y Bro 2003). Esto no es necesariamente cierto para U-PLS/RBL, el cual requiere que los datos de calibración se modelen con el número adecuado de *loadings* y los posibles componentes no esperados mediante descomposición en valores singulares (SVD) (Olivieri 2005a).

2.2. OBJETIVOS

- Evaluar el algoritmo U-PLS/RBL para el procesamiento de datos instrumentales de segundo orden.
- Comparar los resultados de la predicción por U-PLS/RBL de una variedad de muestras de suero adicionadas con tetraciclina, con los obtenidos por el clásico algoritmo PARAFAC. Un problema adicional es la baja sensibilidad del

sistema con respecto a la tetraciclina, lo cual se presenta como un desafío para ambos métodos.

2.3. TEORÍA

2.3.1. PARAFAC

PARAFAC es un algoritmo para datos multivariados de segundo orden y órdenes superiores que descompone los arreglos de datos en juegos de *loadings* y puntuaciones (*scores*) para describirlos de una forma más condensada que la original (Bro 1997). Fue propuesto simultánea e independientemente por Harshman bajo el nombre de PARAFAC (Harshman 1970) y por Carroll y Chang (Carroll y Chang 1970) los cuales lo llamaron CANDECOMP (del inglés, *Canonical decomposition*). De aquí en adelante nos referiremos al mismo a través de su forma más usual: PARAFAC.

Puede ser considerado como una generalización del análisis en componentes principales (PCA) a sistemas de tres vías y es ampliamente utilizado en quimiometría debido a que presenta una estructura similar a muchos modelos físicos de datos obtenidos a partir de instrumentos de uso común y a su propiedad de unicidad, la cual implica que los datos en concordancia con este modelo pueden ser descompuestos en sus contribuciones individuales de una manera única (Harshman 1972, Sidiropoulos y Bro 2000). De esta manera, PARAFAC es capaz de resolver matemáticamente, entre otros, problemas de solapamiento espectral, proporcionando además los espectros verdaderos de los analitos involucrados.

Si para una dada muestra se obtienen datos experimentales en forma de matrices de $J \times K$, estamos en presencia de un arreglo de dos vías, en donde J y K denotan el número de datos en la primera y la segunda dimensión, respectivamente. Si las I matrices de calibración y la matriz de la muestra incógnita se apilan, se genera un arreglo de tres vías $\underline{\mathbf{X}}$, con dimensiones $[(I + 1) \times J \times K]$ (ver Figura 2.1). De esta manera se agrega al sistema información sobre las posibles interferencias presentes en la muestra, aunque sin necesidad de identificarlas o conocer su concentración. Si $\underline{\mathbf{X}}$ sigue un comportamiento trilineal, el modelo PARAFAC estará dado por tres matrices *loadings* \mathbf{A} , \mathbf{B} y \mathbf{C} con elementos a_n , b_n y c_n que incluyen, respectivamente, la

concentración relativa o *scores* $[(I + 1) \times 1]$, los perfiles en un sentido $(J \times 1)$ y los perfiles en el otro sentido $(K \times 1)$, para cada uno de los n componentes que la conforman. En fluorescencia, J corresponde a los perfiles de emisión mientras que K a los de excitación. La expresión específica para el modelo es:

$$\underline{\mathbf{X}} = \sum_{n=1}^N \mathbf{a}_n \otimes \mathbf{b}_n \otimes \mathbf{c}_n + \underline{\mathbf{E}} \quad (2.1)$$

en donde \otimes indica el producto de Kronecker, N es el número total de componentes y $\underline{\mathbf{E}}$ es una hipermatriz que contiene los residuos no ajustados por el modelo. Esta ecuación se muestra gráficamente en la Figura 2.2 para un sistema con dos componentes.

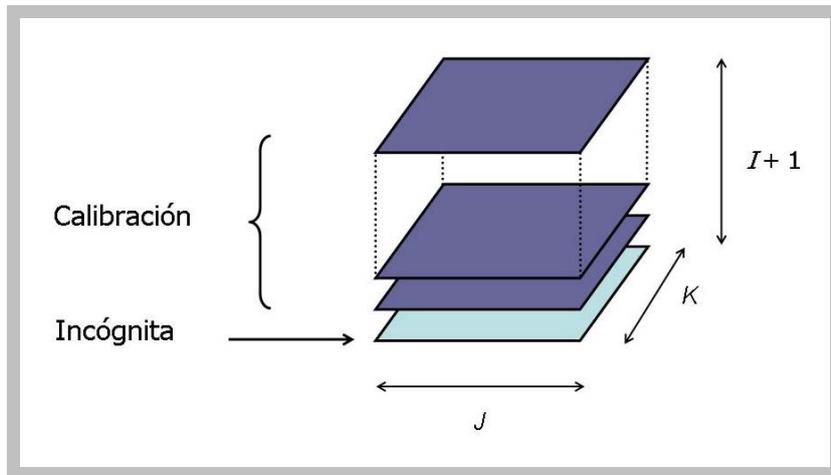


Figura 2.1. Arreglo de tres vías $\underline{\mathbf{X}}$ formado por el apilamiento de las I matrices de calibración y la matriz de la muestra incógnita, de dimensiones $(I + 1) \times J \times K$.

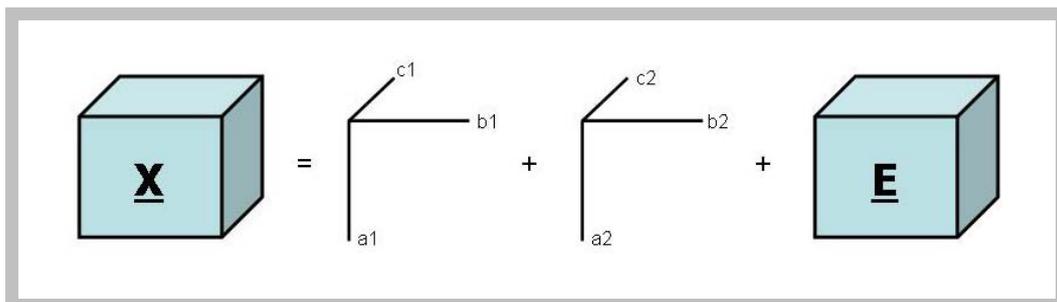


Figura 2.2. Representación gráfica de un modelo PARAFAC de dos componentes.

El modelo descrito por la Ecuación (2.1) permite obtener los perfiles espectrales (**B** y **C**) y de concentración (**A**) de los componentes individuales de las $(I + 1)$ muestras, a partir de la descomposición de \underline{X} . Esto es posible aún si los componentes son químicamente desconocidos, constituyendo la base de la ventaja de segundo orden.

La solución al modelo se puede alcanzar mediante la aplicación de diferentes algoritmos de descomposición (Bro 2002, Faber y col 2003, Kiers 1997), aunque el más frecuentemente utilizado, y aplicado en la presente tesis, es cuadrados mínimos alternantes (ALS) (Harshman 1970, Carroll y Chang 1970). El ajuste se realiza asumiendo como conocidos los *loadings* en dos sentidos y estimando los parámetros desconocidos del tercer *loading*, y así sucesivamente hasta alcanzar la convergencia. Cabe aclarar que una de las desventajas de PARAFAC es el largo tiempo requerido para calcular los modelos.

Con el objetivo de obtener las concentraciones absolutas de analito se realiza una calibración adecuada (preparando un juego de estándares de composición conocida) a partir de la cual se obtienen valores relativos de **A** mediante la descomposición del arreglo de datos de tres vías. Una vez realizado el ajuste, se realiza una regresión entre los primeros I elementos de la columna a_n en función de las concentraciones conocidas de los estándares de calibración y para el analito n :

$$k = \mathbf{y}^+ \times [a_{1n} | \dots | a_{In}] \quad (2.2)$$

en donde $+$ implica realizar la pseudoinversa.

Finalmente, la conversión de concentraciones relativas a absolutas de la muestra desconocida n se realiza a partir del último elemento de la columna, $a_n [a_{(I+1)n}]$, y la pendiente de la curva de calibración k :

$$y_u = a_{(I+1)n} / k \quad (2.3)$$

Es importante resaltar que PARAFAC incluye tanto las muestras de calibración como la muestra incógnita en el ajuste del modelo, y que la

introducción de la información de las concentraciones de los patrones se realiza posteriormente a través de un paso de calibración pseudounivariada [ver Ecuación (2.2)].

2.3.2. U-PLS

El algoritmo U-PLS opera en dos pasos (Wold 1987). Primero, se introduce la información de la concentración en el paso de calibración (sin incluir datos de la muestra incógnita) para obtener dos clases de variables latentes: los *loadings* contenidos en la matriz \mathbf{P} y los pesos de las cargas (*weight loadings*) contenidos en \mathbf{W} . Ambos son estimados a partir de las muestras de calibración $\mathbf{X}_{c,i}$ (matrices de $J \times K$) y las concentraciones de calibración \mathbf{y} ($N_c \times 1$, en donde N_c es el número de muestras de calibración). Las matrices $\mathbf{X}_{c,i}$ se vectorizan y se agrupan dentro de una matriz \mathbf{Z}_x :

$$\mathbf{Z}_x = [\text{vec}(\mathbf{X}_{c,1}) \mid \text{vec}(\mathbf{X}_{c,2}) \mid \dots \mid \text{vec}(\mathbf{X}_{c,i})] \quad (2.4)$$

donde 'vec' denota la operación de vectorización, la cual convierte las matrices de $J \times K$ en vectores de $JK \times 1$.

Con esta matriz \mathbf{Z}_x ($JK \times 1$) se calibra un modelo U-PLS convencional, el cual provee un juego de *loadings* \mathbf{P} y de *weight loadings* \mathbf{W} , ambos de tamaño $JK \times A$ en donde A es el número de factores latentes, y de vectores de regresión \mathbf{v} de $A \times 1$. El parámetro A se selecciona usualmente mediante técnicas de validación cruzada (Haaland y Thomas 1988). Si la calibración resulta conveniente, \mathbf{v} se puede emplear para estimar las concentraciones de analito en una muestra incógnita:

$$y_u = \mathbf{t}_u^T \mathbf{v} \quad (2.5)$$

donde \mathbf{t}_u es el *score* de la muestra incógnita, obtenido por proyección de los datos vectorizados para la muestra incógnita \mathbf{X}_u dentro del espacio de los factores latentes.

$$\mathbf{t}_u = (\mathbf{W}^T \mathbf{P})^{-1} \mathbf{W}^T \text{vec}(\mathbf{X}_u) \quad (2.6)$$

2.3.2.1. Ventaja de segundo orden: U-PLS/RBL

Cuando existen componentes no calibrados en X_u , los *loadings* de la muestra obtenidos con la Ecuación (2.6) no son adecuados para realizar la predicción del analito a través de la Ecuación (2.5). En este caso, los residuos del paso de predicción mediante U-PLS (s_p) serán anormalmente grandes en comparación con el ruido instrumental típico:

$$\begin{aligned} s_p &= || e_p || / (JK-A)^{1/2} = || \text{vec}(X_u) - P (W^T P)^{-1} W^T \text{vec}(X_u) || / (JK-A)^{1/2} = \\ &= || \text{vec}(X_u) - P t_u || / (JK-A)^{1/2} \end{aligned} \quad (2.7)$$

donde $|| \cdot ||$ indica la norma Euclídea.

Si hay interferencias en las muestras de validación, la situación puede ser solucionada mediante un procedimiento aparte llamado **bilinealización residual**:

Dada una muestra de validación que contiene componentes no calibrados, su señal puede ser descompuesta en dos partes: una que puede ser modelada usando las variables latentes de la calibración (X_{mod}) y otra que no puede ser modelada con estas variables (X_{nomod}):

$$X_u = X_{\text{mod}} + X_{\text{nomod}} \quad (2.8)$$

X_{mod} se puede expresar en función de las variables latentes de la calibración P y de los *loadings* de la muestra asociada con t_u , de la siguiente manera:

$$\text{vec}(X_u) = P t_u + e_{\text{mod}} + \text{vec}(X_{\text{nomod}}) \quad (2.9)$$

donde e_{mod} es el vector de los residuos no ajustados en X_{mod} mediante el modelo PLS de A factores y tiene elementos del orden del ruido instrumental. Si en X_{nomod} existe algo que tenga una estructura bilineal con valores mayores que el ruido experimental, puede ser modelado por SVD, el cual permite la estimación de los perfiles de las interferencias b_{int} y c_{int} mediante la

minimización de la norma del vector de los residuos e_u (calculado mientras se ajustaron los datos de la muestra a la suma de las contribuciones relevantes):

$$\text{vec}(X_u) = P t_u + \text{vec}[g_{\text{int}} b_{\text{int}} (c_{\text{int}})^T] + e_u \quad (2.10)$$

Durante este procedimiento, P se mantiene constante a los valores de calibración, t_u se varía hasta minimizar $\| e_u \|$, y los perfiles de las interferencias se estiman por SVD de una matriz residual obtenida luego de reorganizar e_p en una matriz de $J \times K$.

$$(g_{\text{int}}, b_{\text{int}}, c_{\text{int}}) = \text{SVD}_1(E_p) \quad (2.11)$$

donde E_p es una matriz de $J \times K$ obtenida luego de reordenar el vector de $JK \times 1$.

La minimización se realiza mediante el procedimiento de Gauss-Newton (GN) comenzando por t_u de la Ecuación (2.6). Una vez que se minimiza $\| e_u \|$ en la Ecuación (2.10), las concentraciones de los analitos se calculan usando la Ecuación (2.5), mediante la introducción del vector final encontrado por el procedimiento de GN, t_u .

El número de interferencias M_i se puede estimar comparando los residuos finales s_u con el nivel de ruido experimental:

$$s_u = \| e_u \| / [JK - (N_{\text{cal}} + M_i)]^{1/2} \quad (2.12)$$

donde e_u viene de la Ecuación (2.10) y N_{cal} es el número de analitos calibrados.

En general, un gráfico de s_u para un dado número de componentes mostrará valores decrecientes, comenzando en s_p cuando el número de componentes es igual a A , hasta que se estabiliza en un valor compatible con el ruido experimental, permitiendo localizar el número correcto de componentes (Olivieri 2005a).

En la Figura 2.3 se resumen gráficamente los pasos involucrados al aplicar el algoritmo U-PLS/RBL a una muestra que contiene componentes no esperados.

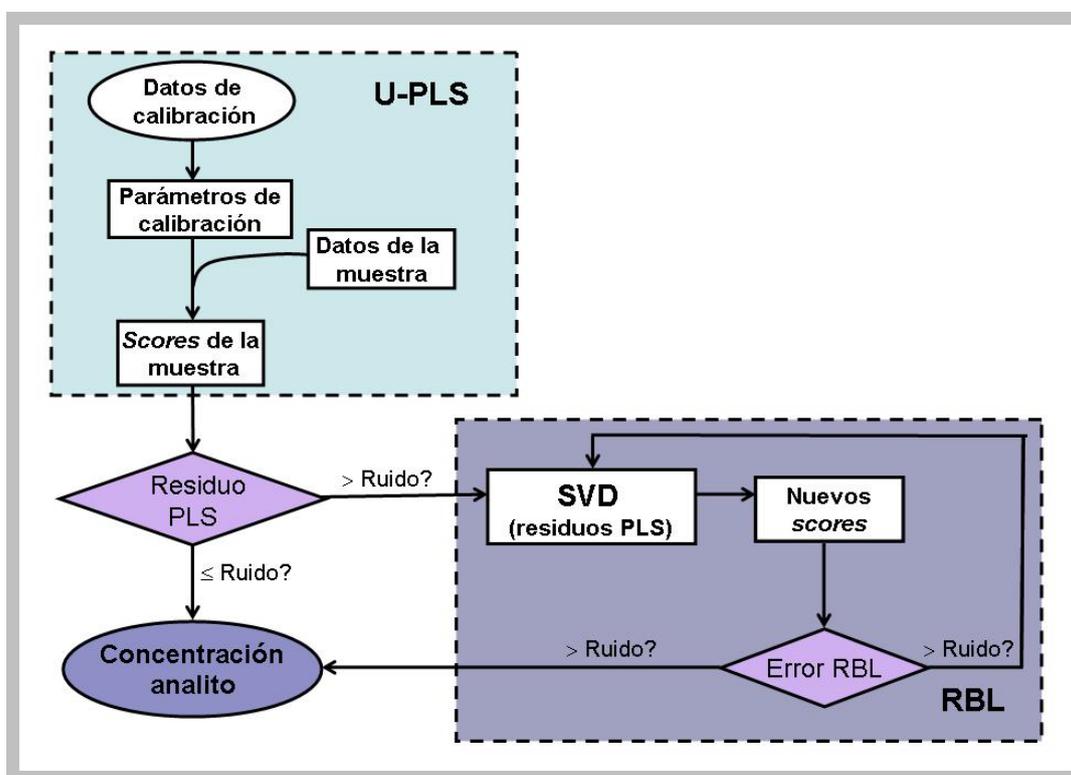


Figura 2.3. Gráfico que indica el modo en que se realiza la predicción de una muestra de validación que contiene componentes no esperados mediante U-PLS/RBL.

2.4. PROGRAMAS

Todas las rutinas empleadas fueron escritas en MATLAB (Matlab 6.0, 2000). Las necesarias para la aplicación de PARAFAC están disponibles en INTERNET gracias a Bro (<http://www.models.kvl.dk/source/>). U-PLS/RBL se implementó a través rutinas escritas por el Dr. Olivieri incluidas en la interfaz gráfica MVC2, similar a la descrita para calibración multivariada de primer orden MVC1 (Olivieri y col 2004).

2.5. MATERIALES Y MÉTODOS

2.5.1. EQUIPAMIENTO

Las medidas fluorescentes se realizaron en un espectrofluorímetro Varian Cary-Eclipse equipado con lámpara flash de xenón, empleando cubetas de cuarzo de 1,00 cm y ancho de ranura de emisión y de excitación de 5,00 nm. Los espectros de excitación-emisión de fluorescencia se registraron en los rangos 400 - 600 nm cada 2 nm para la emisión y 300 - 450 cada 5 nm para la excitación, generándose matrices de 101×31 datos por muestra.

2.5.2. REACTIVOS

Todas las experiencias se realizaron con drogas calidad para análisis y agua milliQ. Se preparó una solución stock de tetraciclina de $1,000 \text{ g L}^{-1}$ mediante disolución de la droga sólida en agua. Las muestras de suero se prepararon adicionando cantidades apropiadas de esta solución a sueros blanco. Por otra parte se preparó una solución de amoníaco $0,1 \text{ mol L}^{-1}$ y una de acetato de magnesio $5 \times 10^{-3} \text{ mol L}^{-1}$ en hidróxido de amonio 3 mol L^{-1} .

2.5.3. DISEÑO DEL JUEGO DE CALIBRACIÓN

Debido a interacciones entre las proteínas del suero y la tetraciclina, las cuales modifican las propiedades fluorescentes del analito, se prepararon cinco muestras de calibración adicionando estándar puro a sueros normales para alcanzar las siguientes concentraciones: 0,50 - 1,00 - 2,00 - 3,00 y 4,00 mg L^{-1} , cubriéndose así el rango terapéutico. A continuación se realizó una dilución 1:3 de las muestras para obtener valores de concentración comprendidos en el rango lineal de fluorescencia-concentración: 0,167 - 1,333 mg L^{-1} .

2.5.4. DISEÑO DEL JUEGO DE VALIDACIÓN

Se prepararon dieciocho muestras de validación compuestas por suero normal (distinto al empleado para calibrar) y tetraciclina adicionadas, en algunos casos, con una de las siguientes cuatro drogas que pueden ser ocasionalmente coadministradas con tetraciclina: salicilato, paracetamol, ibuprofeno y doxorubicina (ver Tabla 2.1).

2.5.5. PROCEDIMIENTO EXPERIMENTAL

Se agregaron cantidades adecuadas de solución stock de tetraciclina a 1,00 mL de suero para obtener las concentraciones deseadas. Luego de homogeneizar, se agregaron 2,00 mL de acetato de magnesio 5×10^{-3} mol L⁻¹ en hidróxido de amonio 3 mol L⁻¹ (Poiger y Schlatter 1976). Las EEMs se recogieron dentro de un período de 30 minutos a partir de la preparación de la muestra, con el objetivo de evitar cambios en la intensidad de la señal.

2.6. RESULTADOS Y DISCUSIÓN

2.6.1. SELECCIÓN DE REGIONES

Con el objetivo de eliminar las dispersiones de Rayleigh y Raman, y de focalizar en aquellas longitudes de onda en las cuales la señal debida al analito de interés es mayor, se seleccionaron regiones de trabajo de excitación entre 300 y 390 nm y de emisión entre 400 y 600 nm.

2.6.2. DETERMINACIÓN DEL NÚMERO DE FACTORES

El primer paso para la aplicación de PARAFAC es la determinación del número correcto de componentes que explican la variabilidad de los datos. En este caso se realizó mediante la consideración del parámetro interno de PARAFAC conocido como *core consistency* (Bro y Kiers 2003). Para explicar la variabilidad de las muestras de calibración se necesitaron dos componentes: uno debido a la tetraciclina y otro al suero como un todo. Por otra parte, para las muestras de validación el número de componentes resultó ser tres o cuatro, indicando la presencia de variabilidad adicional. Esto es coherente con el hecho de que fueron preparadas en sueros diferentes a los empleados para calibrar y con la adición, en algunos casos, de drogas que son ocasionalmente coadministradas con tetraciclina (ver Tabla 2.1).

En el caso de U-PLS, se estimaron dos variables latentes mediante validación cruzada (Haaland y Thomas 1988) y uno o dos componentes adicionales a través del procedimiento de RBL (ver Tabla 2.1), sugiriendo nuevamente que N_{noesp} depende tanto del tipo de suero como del componente adicionado. El número de componentes no esperados se evaluó mediante la

comparación de los residuos finales (s_u) con el ruido instrumental de la manera descrita en el apartado 2.3.2.1.

2.6.3. RESULTADOS

Las predicciones realizadas mediante PARAFAC y U-PLS/RBL se muestran en la Tabla 2.1. Para evaluar la calidad de los resultados se calcularon las siguientes cifras de mérito:

- Error cuadrático medio de la predicción (RMSEP): $\left[\frac{1}{l} \sum_1^m (c_{\text{nom}} - c_{\text{pred}})^2 \right]^{1/2}$ (2.13)

- Error relativo de la predicción (REP %): $\frac{100}{\bar{C}} \left[\frac{1}{l} \sum_1^m (c_{\text{nom}} - c_{\text{pred}})^2 \right]^{1/2}$ (2.14)

donde c_{nom} es la concentración nominal del componente en las l muestras de validación, c_{pred} es la concentración predicha por el método y \bar{C} es el valor medio de concentración de las muestras de calibración.

De la comparación visual de las cifras de mérito presentadas en la Tabla 2.1 se evidencia que U-PLS/RBL proporciona mejores resultados que PARAFAC.

Tabla 2.1. Resultados de las muestras de validación empleando los métodos PARAFAC y U-PLS/RBL para predecir la concentración de tetraciclina.

Muestra ^a	Componentes no esperados ^b / mg L ⁻¹	Concentración de tetraciclina / mg L ⁻¹				
		Nominal	Predicha			
			PARAFAC ^c	Recup. / %	U-PLS/RBL ^d	Recup. / %
TA-1	-	1,30	1,33 (3)	102,3	1,36 (1)	104,6
TA-2	-	1,30	1,37 (3)	105,4	1,29 (1)	99,2
TA-3	-	1,30	1,33 (3)	102,3	1,20 (1)	92,3
TB-1	-	1,80	2,70 (4)	150,0	1,78 (2)	98,9
TB-2	-	1,80	2,88 (4)	160,0	1,76 (2)	97,8
TB-3	-	1,80	2,52 (4)	140,0	1,82 (2)	101,1
TB-4	-	1,80	2,82 (4)	156,7	1,78 (2)	98,9
TB-5	-	1,80	2,86 (4)	158,9	1,83 (2)	101,7
TC-1	-	2,20	2,53 (3)	115,0	1,78 (1)	80,9
TC-2	Salicilato (12)	2,20	2,57 (4)	116,8	1,77 (2)	80,5
TC-3	Paracetamol(12)	2,20	2,06 (4)	93,6	1,86 (2)	84,5
TC-4	Ibuprofeno (20)	2,20	1,86 (4)	84,5	1,92 (2)	87,3
TC-5	Doxorubicina(12)	2,20	2,03 (4)	92,3	2,30 (2)	104,5
TD-1	-	2,80	3,35 (3)	119,6	3,23 (1)	115,4
TD-2	Salicilato (12)	2,80	3,56 (4)	127,1	3,24 (2)	115,7
TD-3	Paracetamol(12)	2,80	3,46 (4)	123,6	3,23 (2)	115,4
TD-4	Ibuprofeno (20)	2,80	3,32 (4)	118,6	3,28 (2)	117,1
TD-5	Doxorubicina(12)	2,80	3,53 (4)	126,1	3,02 (2)	107,9
RMSEP			0,63		0,28	
REP%			30		13	

^a La segunda letra mayúscula corresponde al tipo de suero empleado. Todas las muestras se prepararon empleando sueros diferentes a los usados en la calibración.

^b Concentración de los componentes no esperados en paréntesis.

^c Número de componentes de PARAFAC en paréntesis.

^d Número de componentes no esperados (N_{noesp} , ver el texto) en paréntesis.

Una manera más adecuada de comparar los rendimientos de diferentes métodos frente a juegos de datos distintos, es a través del cálculo de la sensibilidad (Olivieri 2005a, Olivieri 2004). Recientemente, Olivieri proporcionó las siguientes ecuaciones para su cálculo (Olivieri 2005a):

$$SEN_n = 1 / || (\mathbf{P}_{\text{eff}}^+)^T \mathbf{v} || \quad (2.15)$$

donde \mathbf{v} es el vector latente ($A \times 1$) de coeficientes de regresión para el modelo PLS [ver Ecuación (2.5)] y \mathbf{P}_{eff} es una matriz dada por:

$$\mathbf{P}_{\text{eff}} = (\mathbf{P}_{\text{c,noesp}} \otimes \mathbf{P}_{\text{b,noesp}})^T \mathbf{P} \quad (2.16)$$

donde \mathbf{P} es la matriz de los *loadings* ($JA \times A$) provista por el modelo PLS desdoblado, y $\mathbf{P}_{\text{c,noesp}}$ y $\mathbf{P}_{\text{b,noesp}}$ son las matrices de proyección provistas por las siguientes ecuaciones:

$$\mathbf{P}_{\text{c,noesp}} = \mathbf{I} - \mathbf{C}_{\text{noesp}} \mathbf{C}_{\text{noesp}}^+ \quad (2.17)$$

$$\mathbf{P}_{\text{b,noesp}} = \mathbf{I} - \mathbf{B}_{\text{noesp}} \mathbf{B}_{\text{noesp}}^+ \quad (2.18)$$

donde $\mathbf{B}_{\text{noesp}}$ y $\mathbf{C}_{\text{noesp}}$ contienen los perfiles para los componentes no esperados como columnas.

Si se emplea la ventaja de segundo orden, la Ecuación (2.15) implica que la SEN_n es específica de la muestra y no puede ser definida para el método multivariado como un todo. Por lo tanto, se puede estimar e informar un valor promedio para un juego de muestras.

Por otra parte, se puede calcular la sensibilidad analítica (γ_n) como el cociente entre la SEN_n y el ruido instrumental. Su inversa establece la mínima diferencia de concentración que puede apreciarse en el rango lineal y es independiente del instrumento o la escala (Cuadros Rodríguez y col 1993).

Con el objetivo de determinar la eficacia de U-PLS/RBL frente a PARAFAC en sistemas que presentan baja sensibilidad con respecto al analito de interés, los datos del presente trabajo se compararon con los obtenidos por integrantes del Departamento de Química Analítica de la Facultad de Ciencias Bioquímicas y Farmacéuticas, perteneciente a la Universidad Nacional de Rosario. Los mismos realizaron un estudio similar al ya expuesto con el objetivo de determinar salicilato en suero, en presencia de interacciones entre ambos. En algunos casos las muestras se adicionaron, además, con el

antiinflamatorio naproxeno y, en otros, los sueros eran lipémicos. Las predicciones para este sistema aplicando PARAFAC y U-PLS/RBL resultaron comparables, con RMSEP y REP% aceptablemente bajos, teniendo en cuenta la complejidad de las muestras analizadas. La sensibilidad y la sensibilidad analítica, dadas como un promedio sobre un juego de muestras conteniendo suero normal y salicilato, resultaron $3,2 \times 10^3$ UF L mg^{-1} (UF = unidades de fluorescencia) y $1,6 \times 10^{-4}$ mg L $^{-1}$, respectivamente, ambas referidas a las concentraciones en la celda de medición. Por otro lado, la sensibilidad y la sensibilidad analítica para el sistema de tetraciclina resultaron 66 UF L mg^{-1} y $2,5 \times 10^{-2}$ mg L $^{-1}$, respectivamente, siguiendo el criterio mencionado para salicilato. Estas cifras de mérito indican que el sistema tetraciclina-suero es alrededor de dos órdenes de magnitud menos sensible con respecto al analito que el sistema salicilato-suero. Este hecho se puede observar gráficamente en la Figura 2.4 a través de la comparación de las EEMs para suero (Figura 2.4 A) y tetraciclina (Figura 2.4 B), ambas graficadas en escalas verticales iguales.

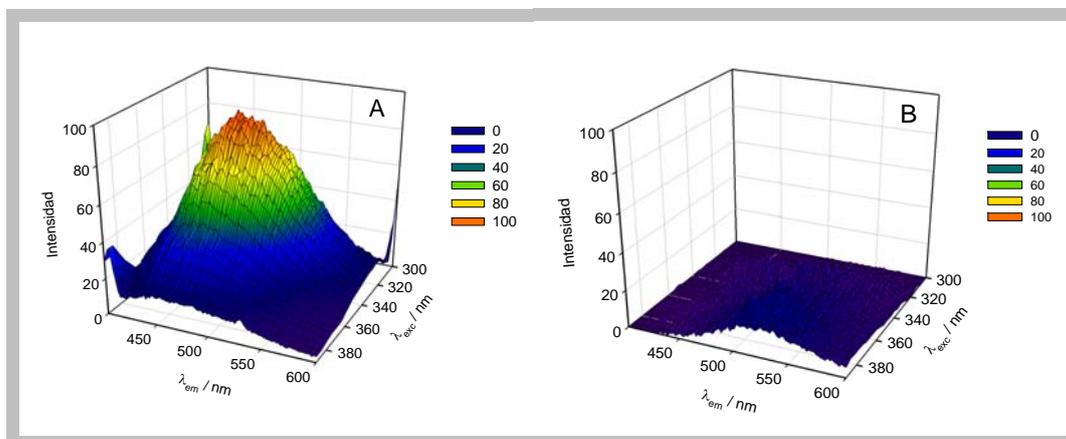


Figura 2.4. Figura tridimensional de las EEMs de fluorescencia para una muestra típica de suero (A) y para el analito tetraciclina en una concentración de $3,00 \text{ mg L}^{-1}$ (B), lo cual muestra la baja sensibilidad del sistema con respecto al analito.

Considerando lo expuesto, se pueden postular dos posibilidades por las cuales los resultados obtenidos mediante U-PLS/RBL son mejores que los obtenidos mediante PARAFAC: 1) la habilidad predictiva del primero es mejor en condiciones de baja sensibilidad con respecto del analito y/o 2) U-PLS/RBL

permite contemplar de una manera más eficiente las interacciones analito-matriz debido a sus propiedades de variabilidad latente.

2.7. CONCLUSIONES

La aplicación del algoritmo cuadrados mínimos parciales a datos multivariados desdoblados, en combinación con un paso de bilinealización residual, mostró ser una buena alternativa para extraer información útil de un sistema complejo que requiere estricta adherencia a la ventaja de segundo orden. U-PLS/RBL se puede aplicar exitosamente a una variedad de problemas multivariados, principalmente en aquellos en donde existen interacciones entre los componentes de la muestra que modifican las características espectrales del analito causando variaciones espectrales de muestra a muestra, aún en presencia de gran superposición espectral, baja sensibilidad, interacción entre componentes y presencia de componentes no esperados.

CAPÍTULO 3

Estudio de diferentes sistemas experimentales con datos instrumentales de segundo orden en presencia de solapamiento extremo. Resolución basada en la obtención de la ventaja de segundo orden a través del algoritmo MCR-ALS

3.1. INTRODUCCIÓN

El seguimiento de la evolución espectral de una dada reacción cinética o corrida electroforética, entre otros, genera datos instrumentales de segundo orden los cuales, en principio, pueden ser procesados de manera satisfactoria a través de la aplicación de calibración multivariada de segundo orden (Escandar y col 2007). Como se mencionó en los capítulos anteriores, los datos de segundo orden pueden mostrar la propiedad intrínseca de la ventaja de segundo orden (Booksh y Kowalski 1994), la cual permite la cuantificación del analito de interés en muestras que contienen componentes no esperados.

Los sistemas espectroscópicos-cinéticos pueden resultar en una fuente de información valiosa para el analista, la que puede ser empleada con múltiples propósitos analíticos (Mottola 1988, Perez-Bendito y Silva 1988). Los datos cinéticos se utilizaron tradicionalmente para la determinación de uno o varios analitos en base a ecuaciones cinéticas diferenciales o integradas, con metodologías que varían de aplicación en aplicación (Mottola 1988, Perez-Bendito y Silva 1988, Wu y col 2001). Por ejemplo, el método de las ecuaciones proporcionales se aplicó a muestras binarias bajo condiciones de primer o pseudoprimer orden (Mottola 1988). Sin embargo, en este trabajo se utilizó solo una pequeña porción de los datos cinéticos, lo que condujo a una baja precisión. También se propusieron filtros de Kalman (Wentzell y col 1989, Xiong 1991) y ajustes por cuadrados mínimos no lineales (Cladera y col 1993a). Por otra parte se realizó análisis de multicomponentes mediante la aplicación de PCR, PLS o métodos basados en la señal neta del analito (NAS), los cuales pueden corregir efectos de interacciones entre componentes (Blanco y col 1993, Cladera y col 1993b, Havel y col 1993, Blanco y col 1994, Blanco y col 1995, García y col 1995, Xie y col 1996, Durán-Merás y col 1995, Bouhsain y col 1996, de la Guardia y col 1996, Muñoz de la Peña y col 2002).

Es de esperar que la resolución de estos sistemas empleando datos de segundo orden proporcione mejores resultados analíticos, ya que los datos cinéticos a una sola longitud de onda pueden contener menor información que los datos obtenidos a múltiples longitudes de onda (Esteves da Silva y Oliveira 1999, Bro 1996, Azubel y col 1999, Pettersson y Karlberg 1997, Coello y col 2000). Algunos ejemplos de la literatura en donde la información

espectroscópica-cinética se emplea para obtener la ventaja de segundo orden incluyen las determinaciones de creatinina en presencia de componentes del suero (albúmina, bilirubina y glucosa) (Guterres y col 2004), pesticidas carbarilo y cloropirifo en muestras comerciales conteniendo sustancias no esperadas (Espinosa Mansilla y col 2004), nitrito en muestras de agua y carne (Nazi y col 2005) y amoxicilina en muestras de orina (García-Reiriz y col 2007).

Un aspecto importante relativo al análisis de datos cinéticos-espectrales de segundo orden es el hecho de que los analitos y las potenciales interferencias pueden tener diferentes velocidades de reacción, pero espectros idénticos (ya sea propios o a través de reactivos y productos). Por ejemplo, si los componentes de la muestra reaccionan con un reactivo para dar un producto común que absorbe en la región UV-Visible, entonces los datos de segundo orden mostrarán perfiles idénticos en la región espectral. La misma situación se generará si un reactivo que proporciona una dada respuesta es consumido por reacción con los componentes de la muestra. Este fenómeno se puede considerar como un caso específico de dependencia lineal entre los perfiles de los componentes. Una forma de hacer frente a este problema es incluir las potenciales interferencias en el juego de calibración, con la desventaja de tener que conocer todas aquellas posibles que pueden estar presentes en nuevas muestras, conduciendo a un incremento considerable del tamaño del juego de calibración.

Por otra parte, la electroforesis capilar (EC) se ha vuelto una técnica de elección para una amplia variedad de aplicaciones analíticas, entre otras, medioambientales, bioanalíticas, farmacéuticas y alimenticias (Kraly y col 2006). Cuando se desarrolla un método mediante EC, es necesario optimizar las condiciones experimentales con el objetivo de garantizar la separación completa de todos los componentes de la muestra (Hanrahan y col 2008). Sin embargo, si este requerimiento no se cumple, la aplicación de quimiometría puede ser un recurso útil para lograr la cuantificación exacta del analito (Sentellas y Saurina 2003), siempre que se registren datos de segundo orden. En este contexto, se encuentran en la literatura un número pequeño de aplicaciones en las cuales se resolvieron picos parcialmente solapados a través

de MCR-ALS (Sentellas y col 2001, Li y col 2003, Zhang y Li 2006, Li y col 2006), aunque también se aplicaron de manera satisfactoria metodologías tales como redes neuronales artificiales (Zhang y col 2006) y transformada de *wavelet* (Jiao y col 2008).

Con frecuencia aparecen diferencias en la difusión axial que pueden originar disimilaridades en la forma de los picos, generando datos no trilineales y haciendo necesario el uso de MCR-ALS para poder distinguir perfiles temporales en cada muestra experimental (Bro 1998). Por otra parte, es importante tener en cuenta que en el análisis de los datos espectrales-electroforéticos de segundo orden, los analitos y las potenciales interferencias pueden mostrar espectros diferentes pero tiempos de migración idénticos, por ejemplo, cuando los componentes de una muestra presentan propiedades fisicoquímicas similares (pKas o masas moleculares, entre otras) que dificultan la separación.

La estructura de la mayoría de los algoritmos de segundo orden no permite resolver sistemas que presentan solapamiento total entre el analito y las potenciales interferencias en la dimensión espectral o temporal. MCR-ALS es un algoritmo que puede resolver el problema de la presencia de dependencia lineal a través del recurso matemático de generación de una matriz aumentada. Este procedimiento consiste en construir una matriz ensamblando matrices de datos instrumentales de diferentes muestras en la dirección de las columnas, en la dirección de las filas o en ambas direcciones simultáneamente. Tanto en los sistemas espectroscópicos-cinéticos como en los espectroscópicos-electroforéticos, es común aumentar las matrices en la dirección temporal, es decir, acomodarlas una arriba de la otra cuando las filas representan los espectros y las columnas los perfiles temporales, debido a que esto disminuye los problemas en esta dimensión asociados a las diferencias entre muestras. Sin embargo, en el primer sistema existen casos en los cuales es preferible aumentar en la dirección espectral ya que esto permite la diferenciación entre analitos e interferentes. Se debe aclarar que PARAFAC2, una variante de PARAFAC, es también capaz de modelar datos que presenten dependencias lineales y puede ser empleado para la resolución de los presentes casos de análisis (Bro 1998).

En el presente capítulo se estudian dos sistemas espectroscópicos-cinéticos y uno espectroscópico-electroforético: 1) la oxidación de naranja de metilo con bromato de potasio catalizada por Fe(II) (el analito) en presencia de Zn(II) como interferencia, el cual también cataliza la reacción pero con diferente cinética (Safavi y col 2001), 2) la reacción de tartrazina (el analito) con bromato de potasio catalizada por Fe(II) en presencia de azul brillante como interferencia, el cual reacciona produciendo un compuesto con un espectro de absorción muy similar al de la tartrazina, y 3) un sistema en donde el analito de interés (carbamazepina) se encuentra solapado en la dimensión temporal con algunos de los componentes de la muestra. Estos sistemas experimentales ponen de manifiesto la importancia de seleccionar MCR-ALS como algoritmo de procesamiento, empleando el modo de aumento correcto dependiendo del grado de solapamiento espectral.

3.2. OBJETIVOS

- Evaluar el algoritmo MCR-ALS para el procesamiento de datos instrumentales de segundo orden espectroscópicos-cinéticos y espectroscópicos-electroforéticos con solapamiento espectral severo o total de los componentes presentes en el sistema, a través de la obtención de la ventaja de segundo orden.
- Testear las diferentes opciones de construcción de la matriz de partida en la aplicación de MCR-ALS, por aumento temporal o espectral, y su repercusión en la resolución de diferentes sistemas experimentales.

3.3. TEORÍA

3.3.1. MCR-ALS

El algoritmo MCR-ALS es ampliamente utilizado para la resolución de componentes múltiples en muestras desconocidas que den como resultado tablas o matrices de datos que puedan ser descritas por un modelo bilineal (Jaumot y col 2005). Los datos en forma de tabla o matriz, en donde una dirección está relacionada con la variación de la composición del sistema y la

otra con la variación de la respuesta, ayudan a la diferenciación entre los componentes.

MCR-ALS posee la capacidad de procesar múltiples matrices de datos simultáneamente reduciendo ambigüedades intrínsecas del análisis de factores y/o deficiencias de rango presentes en los datos. Se considera que una matriz es deficiente de rango cuando el número de contribuciones químicas detectadas por los métodos de análisis de factores, como por ejemplo SVD, es menor que el número de componentes químicos presentes en el sistema, lo cual puede deberse a similitudes muy marcadas entre los espectros o los perfiles de concentración de las especies que intervienen (Ruckebusch y col 2006). Por otra parte, este algoritmo permite la aplicación flexible y diversa de restricciones que ayudan a la resolución, mejorando los resultados (Jaumot y col 2005). En la Figura 3.1 se observa una matriz de datos aumentada a partir de las matrices de calibración y validación.

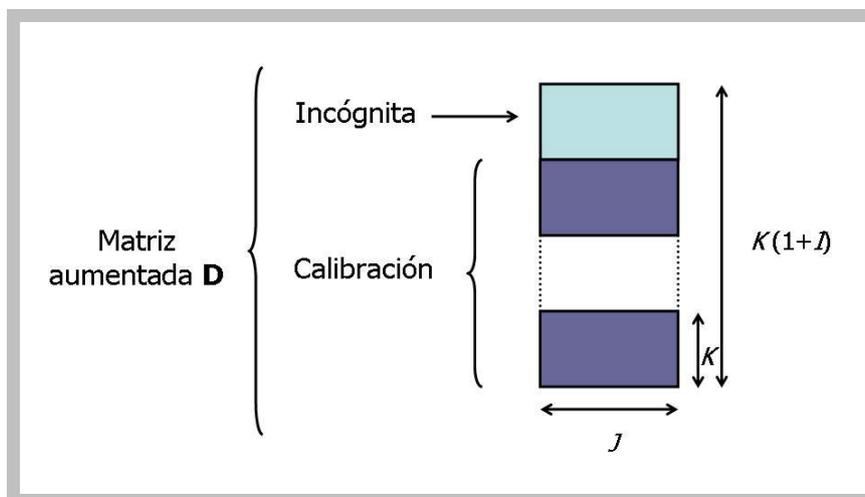


Figura 3.1. Matriz aumentada D constituida por el apilamiento de las matrices de calibración y la matriz de la muestra incógnita.

Si las matrices son de tamaño $J \times K$ (J es el número de longitudes de onda y K el número de tiempos de reacción), las columnas representan el modo espectral y las filas el modo temporal. La forma usual consiste en aumentar las matrices en la dirección temporal, con el objetivo de hacer frente a posibles variaciones de los perfiles temporales entre muestra y muestra. En

otras palabras, MCR-ALS es capaz de hacer frente a juegos de datos que se desvían de la trilinealidad. Esto se logra aumentando las matrices en el modo sospechado de romper la estructura trilineal. Es decir, si ocurre una variación de los perfiles temporales de matriz a matriz, se crea una matriz aumentada en este sentido. En el modo de aumento temporal las matrices de calibración y la matriz incógnita se acomodan una al lado de la otra (De Juan y col 2000). Luego, la descomposición bilineal de la matriz aumentada se realiza de acuerdo a la siguiente expresión:

$$D_{ta} = S_{ta} G_{ta}^T + E_{ta} \quad (3.1)$$

donde las columnas de D_{ta} contienen los espectros de absorción medidos para diferentes muestras a diferentes tiempos de reacción, las columnas de G_{ta} contienen los perfiles temporales de las especies que intervienen, las columnas de S_{ta} sus espectros relacionados y E_{ta} es una matriz de residuos no ajustados por el modelo (el subíndice 'ta' indica aumento en el modo temporal). Las dimensiones apropiadas de D_{ta} , G_{ta} , S_{ta} y E_{ta} son $J \times [(I+1)K]$, $[(I+1)K] \times N$, $J \times N$ y $J \times [(I+1)K]$, respectivamente (I es el número de muestras de calibración, N es el número de componentes que dan respuesta y se asume una sola muestra de validación). La descomposición de D_{ta} se logra a través de la minimización por cuadrados mínimos alternantes (ALS) de la norma de E_{ta} . La minimización comienza con la provisión de los espectros estimados de los componentes, los cuales se emplean para estimar \hat{G}_{ta} (el 'sombrero' indica una matriz estimada) a partir de la Ecuación (3.1):

$$\hat{G}_{ta} = D_{ta}^T (S_{ta}^+)^T \quad (3.2)$$

en donde '+' indica la inversa generalizada. Con la matriz \hat{G}_{ta} de la Ecuación (3.2) y la matriz de datos original D_{ta} se vuelve a estimar la matriz S_{ta} mediante cuadrados mínimos:

$$\hat{S}_{ta} = D_{ta} (\hat{G}_{ta}^T)^+ \quad (3.3)$$

y finalmente se calcula E_{ta} a partir de la Ecuación (3.1) usando D_{ta} y las matrices estimadas \hat{G}_{ta} y \hat{S}_{ta} . Estos pasos se repiten hasta alcanzar la convergencia imponiendo ciertas restricciones adecuadas al sistema en estudio durante el proceso de ALS. Entre las restricciones más aplicadas se encuentran (Muñoz y De Juan 2007, De Juan y Tauler 2003, Jaumot y col 2005):

- No negatividad: previene la presencia de valores negativos en los perfiles. Se aplica a los perfiles de concentración y a espectros UV-Vis y de fluorescencia, debido a que sus valores son siempre positivos.
- Unimodalidad: fuerza la presencia de un único máximo en el perfil de una especie. Se aplica a procesos en donde los perfiles de concentración muestran una forma de aparición-decaimiento (perfiles de reacción, picos cromatográficos y electroforéticos, voltamogramas).
- Clausura (cierre o balance de masa): se relaciona con ecuaciones de balance de masa en sistemas cerrados. Se aplica cuando la concentración total de las especies detectables en cada etapa del proceso es constante, pudiéndose fijar a un valor único o variable. Cuando hay especies no-absorbentes en el sistema no se puede aplicar debido a que las especies que contribuyen a la señal medida no forman un sistema cerrado.
- Rango local/selectividad: impone la ausencia de algunas especies en un rango específico de perfiles de concentración, por ejemplo, en una cierta ventana de tiempo o pH. Este conocimiento puede tener una base química o ser adquirido mediante la aplicación de métodos de análisis de rango local.
- Igualdad: se puede introducir en la optimización información química asociada al conocimiento de los perfiles puros de concentración o espectros. Los perfiles conocidos pueden ser establecidos como invariantes durante el proceso iterativo. No es necesario conocer completamente un perfil; si hay algunos elementos conocidos se pueden fijar.
- Modelado rígido: se aplica cuando se conoce la ley química (cinética o termodinámica) asociada a un proceso particular. Los perfiles de

concentración asociados a componentes en el proceso se fuerzan a presentar la forma descrita por la ley química.

- Trilinealidad: es la restricción de tres vías más esencial ya que, cuando es aplicable, asegura que los resultados son únicos; en otras palabras, que los perfiles resueltos no tienen ambigüedad.

En el presente trabajo se aplicaron las restricciones de no negatividad en perfiles espectrales, y unimodalidad (solo en el caso de matrices electroforéticas) y no negatividad en perfiles temporales. Se espera que los espectros puros de los compuestos sean los mismos en todos los experimentos, mientras que los perfiles temporales en las diferentes submatrices G_{ta} no necesitan tener la misma forma. Esta es la razón por la cual los experimentos cinéticos realizados en condiciones diferentes (por ejemplo, de temperatura o pH) o las corridas electroforéticas, las cuales se ven afectadas muchas veces por variaciones instrumentales (ver Sistema 3), se pueden analizar juntos, siempre que los espectros permanezcan invariantes.

En este contexto, es importante mencionar que MCR-ALS requiere ser inicializado con parámetros del sistema lo más cercanos posibles al resultado final. Se pueden proveer, por ejemplo, espectros de las especies obtenidos a partir de sus estándares puros o del análisis de los llamados espectros ‘más puros’ a través de la metodología SIMPLISMA (del inglés, *simple interactive self-modelling mixture analysis*), un algoritmo de resolución multivariada de curvas que extrae los espectros puros de los componentes a partir de una mezcla de espectros de composición variada (Windig y Guilment 1991). En el presente trabajo se empleó la primera alternativa para los sistemas cinéticos. Para el análisis de muestras realmente desconocidas en presencia de posibles interferencias del sistema electroforético, la aplicación de MCR-ALS necesitó de la combinación de ambos procedimientos, es decir, espectros puros de los analitos conocidos y análisis de los espectros más puros para la interferencia. Finalmente, la resolución satisfactoria de MCR se logra incluyendo información de la correspondencia entre especies en cada matriz, a través de una matriz binaria que indica si un dado componente existe o no en una dada muestra (Ruckebush y col 2006).

La condición para realizar lo descrito es que los espectros de las especies que intervienen sean lo suficientemente diferentes. En el caso de solapamiento espectral extremo la inversa generalizada requerida en la Ecuación (3.2) se vuelve inestable. Una alternativa para resolver este problema es realizar el aumento en el modo espectral. En este caso, las matrices se ubican una arriba de la otra y el proceso de cuadrados mínimos se representa por el siguiente juego de ecuaciones:

$$D_{sa} = S_{sa} G_{sa}^T + E_{sa} \quad (3.4)$$

$$\hat{S}_{sa} = D_{sa} (G_{sa}^T)^+ \quad (3.5)$$

$$\hat{G}_{sa} = (\hat{S}_{sa}^+ D_{sa})^T \quad (3.6)$$

en donde ahora los tamaños de las matrices involucradas son: D_{sa} , $[(I + 1)J] \times K$, S_{sa} , $[(I + 1)J] \times N$, G_{sa} , $K \times N$ ('sa' indica aumento en el modo espectral). En este modo particular de aumento, la inversa generalizada de G_{sa} en la Ecuación (3.5) puede ser obtenida siempre que los perfiles temporales de los componentes de la muestra sean diferentes (siendo esta una de las principales asunciones del presente modelo). Más importante aún, la inversa generalizada de S_{sa} puede ser obtenida de manera confiable debido a que el aumento de la matriz rompe la dependencia lineal de los perfiles espectrales individuales. Esto se debe al hecho de que la matriz espectral S_{sa} ahora tiene $[(I + 1)J] \times N$ elementos, en vez de los $J \times N$ correspondientes al aumento en el modo temporal. Cada columna de S_{sa} consiste en vectores sucesivos de tamaño $J \times 1$, cada uno de ellos teniendo la forma de un espectro de un componente particular, con una intensidad proporcional a su concentración en una dada muestra. Por lo tanto, las columnas de S_{sa} no son más linealmente dependientes.

Luego de la descomposición de D_{sa} mediante MCR-ALS, la información de concentración contenida en S_{sa} puede ser utilizada para realizar predicciones cuantitativas. Esto se realiza construyendo una curva de calibración

pseudounivariada con las concentraciones relativas de las muestras de calibración contenidas en la matriz optimizada S_{sa} . Dado un componente de interés particular, se seleccionan los elementos de S_{sa} correspondientes a las I muestras de calibración, los cuales van de $S_{sa}(1, n)$ a $S_{sa}(IK, n)$. Cada serie de K valores corresponde al perfil en la dimensión temporal para una muestra en particular. Si el área bajo el perfil se considera proporcional a la concentración del componente, entonces se construye la curva pseudounivariada de la siguiente forma:

$$\left[\sum_{k=1}^K S_{sa}(k, n) \mid \sum_{k=1}^K S_{sa}[(k+K), n] \mid \dots \mid \sum_{k=1}^K S_{sa} \{[k+Kl], n\} \right] = k \mathbf{y} \quad (3.7)$$

donde cada sumatoria agrupa los K elementos de S_{sa} correspondientes a un componente particular y la muestra estándar \mathbf{y} es el vector de las concentraciones de calibración. Una vez que se obtiene k de la Ecuación (3.7), la interpolación de los valores de las muestras incógnitas provee la concentración del componente en la incógnita y_u , es decir:

$$y_u = \frac{1}{K} \sum_{k=1}^K S_{sa} \{[k+(I+1)K], n\} \quad (3.8)$$

Cuando el aumento se realiza en el modo temporal también se emplean las ecuaciones (3.7) y (3.8), con los elementos de $S_{sa}(k, n)$ reemplazados por las series análogas de J elementos de $G_{ta}(j, n)$ tomadas a partir de la matriz G_{ta} .

3.4. PROGRAMAS

MCR-ALS se implementó a través de la interfaz gráfica provista por R. Tauler en su página <http://www.ub.edu/mcr/welcome.html> (Jaumot y col 2005).

Las muestras correspondientes al sistema espectroscópico-electroforético se sometieron a corrección de la línea de base a través de una rutina de MATLAB escrita por Eilers (Eilers y col 2006) y adaptada en nuestro

laboratorio. Además, los datos fueron sometidos a corrección del desplazamiento de tiempo empleando una rutina de MATLAB escrita de acuerdo a lo publicado por Prazen y col (Prazen y col 1998).

Todas las aplicaciones se ejecutaron utilizando MATLAB 7.0 (MATLAB 7.0, 2005).

3.5. SISTEMA 1: Determinación de Fe(II) presente como catalizador de la reacción de oxidación de naranja de metilo con bromato de potasio

3.5.1. MATERIALES Y MÉTODOS

3.5.1.1. Equipamiento

Las medidas de absorción electrónica se llevaron a cabo en un espectrofotómetro Perkin Elmer Lambda 20 (Perkin Elmer, Waltham, Massachusetts, USA) usando cubetas de cuarzo de 1,00 cm y ancho de ranura de 2 nm. Para cada muestra se registraron veinte espectros consecutivos cada 10 s en el rango de longitudes de onda de 450 - 550 nm cada 2 nm, con una velocidad de barrido de 2880 nm min⁻¹. Por lo tanto, el tamaño de cada matriz de datos resultó de 51 × 20.

Los espectros se registraron de manera aleatoria con respecto a las concentraciones de analito y los correspondientes al juego de calibración de obtuvieron en días diferentes a los del juego de validación. Los datos se guardaron en formato ASCII y se transfirieron a una PC para su posterior manipulación mediante el programa MCR-ALS.

3.5.1.2. Reactivos

Todas las experiencias se realizaron con drogas calidad para análisis y agua milliQ. Se preparó una solución stock de Fe(II) de 1000,0 mg L⁻¹ mediante la disolución de 3,5833 g de Fe(NH₄)(SO₄)₂·6H₂O (Cicarelli, San Lorenzo, Argentina) en agua, el agregado de 5,00 mL de H₂SO₄ y la dilución hasta la marca, en un matraz volumétrico de 500,00 mL. Por otra parte se prepararon una solución de naranja de metilo (Carlo Erba Reagenti, Milan, Italia) 1240,0 mg L⁻¹ y otra de Zn(II) 92,7 mg L⁻¹ a partir de ZnSO₄·7H₂O (Merck, Darmstadt, Germany), disolviendo cantidades adecuadas de cada compuesto en agua.

Además, se prepararon soluciones de bromato de potasio (Cicarelli, San Lorenzo, Argentina) $0,1 \text{ mol L}^{-1}$ y ácido fosfórico (Cicarelli, San Lorenzo, Argentina) $3,0 \text{ mol L}^{-1}$.

3.5.1.3. Diseño de los juegos de calibración y validación

Se preparó un juego de calibración conteniendo diez muestras con concentraciones de Fe(II) comprendidas en el rango de $1,00$ a $10,00 \text{ mg L}^{-1}$ cada $1,00 \text{ mg L}^{-1}$. Por otra parte se prepararon por duplicado seis muestras de validación conteniendo Fe(II) en diferentes concentraciones (ver Tabla 3.1) y $3,00 \text{ mg L}^{-1}$ de Zn(II).

El procedimiento cinético se adaptó del propuesto por Safavi y col (Safavi y col 2001). En matraces de $5,00 \text{ mL}$ conteniendo cantidades adecuadas de solución stock de Fe(II) se agregaron $1,00 \text{ mL}$ de solución de naranja de metilo $80,0 \text{ mg L}^{-1}$ y $1,00 \text{ mL}$ de solución de ácido fosfórico $3,0 \text{ mol L}^{-1}$. Las muestras se llevaron a enrase con agua milliQ y se homogeneizaron. Posteriormente, se transfirieron $2,00 \text{ mL}$ de cada solución a una cubeta de cuarzo de $1,00 \text{ cm}$, y pasados 10 s de la adición de $40,0 \mu\text{L}$ de solución de bromato de potasio $0,1 \text{ mol L}^{-1}$, se registraron veinte espectros consecutivos cada 10 s .

3.5.2. RESULTADOS Y DISCUSIÓN

En este sistema experimental, se observa un decaimiento de la intensidad de la señal correspondiente al naranja de metilo a medida que transcurre la reacción de oxidación con bromato de potasio, catalizada por la presencia de iones metálicos, en este caso Fe(II) (el analito) y Zn(II) (la potencial interferencia). Con el objetivo de resaltar la información provista por los datos experimentales obtenidos durante la reacción, se aplicó el siguiente preprocesamiento: se sustrajo el espectro a tiempo cero de todos los espectros contenidos en cada matriz de datos, a todos los tiempos. De este modo se resaltaron las diferencias entre los perfiles cinéticos para las diferentes muestras. A continuación se calcularon los módulos de los nuevos datos y se emplearon para construir los modelos. La Figura 3.2 muestra la evolución temporal de la matriz de datos, preprocesada de la manera descrita arriba, para una muestra de calibración que contiene $5,00 \text{ mg L}^{-1}$ de Fe(II): las

diferencias en las absorbancias, las cuales inicialmente son cero a todas las longitudes de onda, parecen incrementarse a medida que la reacción progresa.

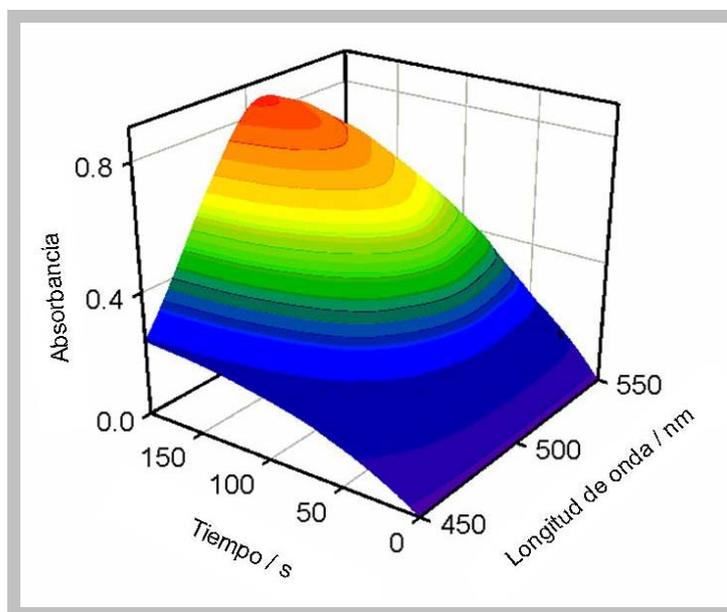


Figura 3.2. Gráfico tridimensional de absorbancia en función de longitud de onda y tiempo de reacción para una muestra de calibración que contiene $5,00 \text{ mg L}^{-1}$ de Fe(II), durante la reacción con bromato de potasio (los datos fueron convertidos de la manera descrita en el texto, es decir, se sustrajo el espectro a tiempo cero y se tomó el módulo).

Este sistema puede ser analizado solamente usando MCR-ALS aumentando las matrices en el modo espectral, debido a que existe un espectro común para ambas especies, mientras que los perfiles cinéticos difieren. Por lo tanto, las matrices de calibración y cada una de las matrices de validación se acomodaron una arriba de la otra. La descomposición se realizó imponiendo la restricción de no negatividad para los dos componentes en ambas dimensiones. Otra restricción importante impuesta al modelo durante la minimización por ALS fue la correspondencia entre especies, es decir, se suministró información de la ausencia del componente Zn(II) en todas las muestras de calibración, permitiéndole aparecer solo en las muestras de validación. Los perfiles generados por el programa se muestran en la Figuras 3.3 A y B en la dimensión cinética y espectral (aumentada), respectivamente.

Resulta interesante notar en la Figura 3.3 B que los perfiles muestran un incremento en el cambio de absorbancia, debido al preprocesamiento aplicado. Como se puede observar, las cinéticas se pueden distinguir muy bien, mientras que el espectro es común para ambos componentes.

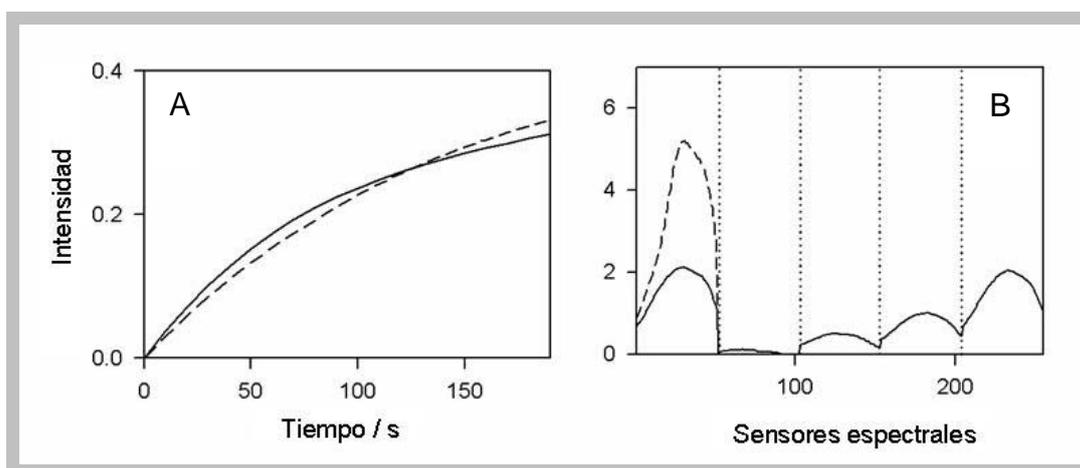


Figura 3.3. Gráfico de los perfiles cinéticos y espectrales luego del procesamiento por MCR-ALS de una muestra de validación típica junto con las muestras de calibración del sistema 1. A) Perfiles cinéticos normalizados a la unidad y B) perfiles espectrales para matrices de muestras sucesivas aumentadas en el modo espectral (la primera es la muestra de validación y las restantes son las cuatro primeras muestras de calibración; las líneas punteadas verticales separan los sensores de cada muestra). Se muestran solo cuatro muestras de calibración por simplicidad. La línea sólida corresponde al analito y la línea de trazos a la interferencia.

Los resultados de la predicción con la metodología pseudounivariada discutida anteriormente se presentan en la Tabla 3.1. Los valores nominales y predichos muestran una concordancia excelente, con un RMSEP de $0,16 \text{ mg L}^{-1}$ que implica un REP de 2,8% (calculado con respecto al valor medio de concentración de la calibración).

Tabla 3.1. Concentraciones nominales y predicciones de las muestras de validación del sistema 1.

Muestra de validación ^a	Fe (II) / mg L ⁻¹	
	Nominal	Predicha ^b
1	1,00	1,16
2	3,50	3,49
3	5,00	4,92
4	6,50	6,31
5	8,00	7,73
6	9,50	9,45
RMSE / mg L ⁻¹		0,16
REP%		2,8

^a RMSE = error cuadrático medio; REP% = error relativo de predicción calculado con respecto al valor medio de concentración de la calibración. Todas las muestras de validación contienen 3,00 mg L⁻¹ de Zn(II).

^b Los resultados son promedios de análisis por duplicados.

3.6. SISTEMA 2: Determinación de tartrazina por oxidación con bromato de potasio catalizada por Fe(II), en presencia de una interferencia causada por azul brillante

3.6.1. MATERIALES Y MÉTODOS

3.6.1.1. Equipamiento

El sistema 2 se implementó a través de un sistema de análisis por inyección en flujo (FIA) con flujo interrumpido, desarrollado usando cinco módulos (degasificador, bomba, válvula de inyección, inyector automático y detector) de un instrumento Agilent 1100 Series (Agilent Technologies, Waldbronn, Alemania). Se desarrolló un programa de inyección para mezclar de manera automática 95 μ L de cada muestra con 5 μ L de bromato 0,1 mol L⁻¹ e inyectar la mezcla dentro de un solvente transportador compuesto por agua milliQ fluyendo a 0,6 mL min⁻¹, a través de una cámara de mezcla de 200 cm \times 0,12 mm de diámetro interno. Transcurridos 12 s desde la inyección de la muestra, la bomba se paró y se monitorizó la reacción durante 108 s. Una vez que se alcanzó el tiempo establecido, el flujo se reanudó. Para cada pico FI se registraron espectros en el rango de 400 - 700 nm cada 1 nm, a intervalos regulares de 0,4 s por un tiempo total de 120 s. Por lo tanto, se generaron

matrices de tamaño 301×300 por muestra, aunque se seleccionaron regiones en ambas dimensiones antes de emplearlas en calibración multivariada.

Los espectros se registraron de manera aleatoria con respecto a las concentraciones de analito y los correspondientes al juego de calibración se obtuvieron en días diferentes a los del juego de validación. Los datos se guardaron en formato ASCII y se transfirieron a una PC para su posterior manipulación mediante el programa MCR-ALS.

Las mediciones cromatográficas se llevaron a cabo en un instrumento Agilent 1100 Series (Agilent Technologies, Waldbronn, Alemania) empleando una columna ($4,6 \times 150$ mm) Zorbax Eclipse XDB-C18 de tamaño de partícula $5 \mu\text{m}$. Los cromatogramas se realizaron a temperatura ambiente, utilizando como fase móvil metanol: acetato de amonio acuoso (45:55) a una velocidad de flujo de 1 mL min^{-1} , con detección ultravioleta a 454 nm para tartrazina y 632 nm para azul brillante (Pereira Alves y col 2008).

3.6.1.2. Reactivos

Todas las experiencias se realizaron con drogas calidad para análisis y agua milliQ. Se emplearon las soluciones de Fe(II), bromato de potasio y ácido fosfórico descritas para el sistema 1.

Se prepararon soluciones stock de tartrazina y azul brillante (Ardinet, Buenos Aires, Argentina) de $1000,0 \text{ mg L}^{-1}$ disolviendo cantidades apropiadas de cada compuesto en agua.

3.6.1.3. Diseño de los juegos de calibración y validación

Para calibrar se preparó un juego de seis muestras con concentraciones de tartrazina entre $4,00$ y $24,00 \text{ mg L}^{-1}$ cada $4,00 \text{ mg L}^{-1}$. Para la validación se preparó un juego adicional de cinco muestras con las concentraciones de tartrazina y azul brillante mostradas en la Tabla 3.2. Todas las muestras se prepararon por duplicado mediante el agregado de $20,0 \mu\text{L}$ de solución stock de Fe(II) a matraces volumétricos de $2,00 \text{ mL}$ conteniendo cantidades adecuadas de tartrazina y/o azul brillante y $40,0 \mu\text{L}$ de ácido fosfórico $3,0 \text{ mol L}^{-1}$. Las muestras se llevaron a enrase con agua milliQ y se homogeneizaron.

3.6.1.4. Muestras reales

Las muestras analizadas fueron las gelatinas Frutigelatin sabor manzana (Arcor, Argentina), denominada G1, y Godet sabor Kiwi (Arcor, Argentina), denominada G2. Ambas declaraban contener azúcar, gelatina, ácido cítrico, ácido fumárico, citrato de sodio, tartrazina, azul brillante y aromatizante.

Debido a la presencia de un efecto matriz observado al tratar de cuantificar las muestras por el método propuesto, la determinación de ambos colorantes se llevó a cabo mediante MCR-ALS combinado con adición estándar (Saurina y Tauler 2000). Las muestras se prepararon pesando 5,00 g de gelatina en polvo, disolviendo con agua milliQ a 60°C y llevando a volumen en matraz volumétrico de 20,00 mL. A continuación se procedió a realizar las adiciones estándares. Para esto, en matraces volumétricos de 20,00 mL se agregaron 5,00 mL de la solución anterior, 200 μL de solución stock de Fe(II), 400 μL de ácido fosfórico 3,0 mol L⁻¹ y se adicionaron 0,00 - 2,00 - 4,00 y 6,00 mg L⁻¹ de cada colorante. Las muestras se llevaron a enrase con agua milliQ y se sometieron al método cinético.

3.6.1.5. Procedimiento cromatográfico para muestras reales

Las soluciones de trabajo de tartrazina y azul brillante se prepararon tomando alícuotas adecuadas de cada solución stock y diluyéndolas en fase móvil para obtener concentraciones de 4,00 y 2,00 mg L⁻¹, respectivamente. Las muestras reales previamente disueltas en agua milliQ (5,00 g de gelatina en 20,00 mL de agua milliQ) se diluyeron 1:5 en fase móvil. Todas las soluciones se filtraron a través de membranas de 0,45 μm de tamaño de poro y se inyectaron 50 μL en el cromatógrafo.

3.6.2. RESULTADOS Y DISCUSIÓN

3.6.2.1. Muestras de validación

En este sistema experimental, la tartrazina reacciona con el bromato de potasio en presencia de Fe(II) y su absorbancia decrece significativamente debido a la reacción de oxidación (ver Figura 3.4 A). En la Figura 3.4 B se pone en evidencia como la potencial interferencia, azul brillante, reacciona a

su vez con el bromato de potasio generando un producto cuyo espectro es muy similar al de la tartrazina pero no idéntico. Una distinción espectral exitosa podría permitir retomar el aumento en el modo cinético usual, el cual podría proveer la ventaja adicional de hacer frente a cambios en los perfiles cinéticos de muestra a muestra.

La Figura 3.4 B muestra la evolución temporal de la absorbancia en la región espectral completa para una muestra de validación que contiene 15,00 mg L⁻¹ tanto de tartrazina como de azul brillante. Como se mencionó anteriormente, se seleccionaron regiones de sensores apropiados en ambas dimensiones antes de construir los modelos. Específicamente, las longitudes de onda se restringieron a 400 - 524 nm y los tiempos a 14 - 115,6 s; conduciendo a $125 \times 259 = 32375$ datos por muestra.

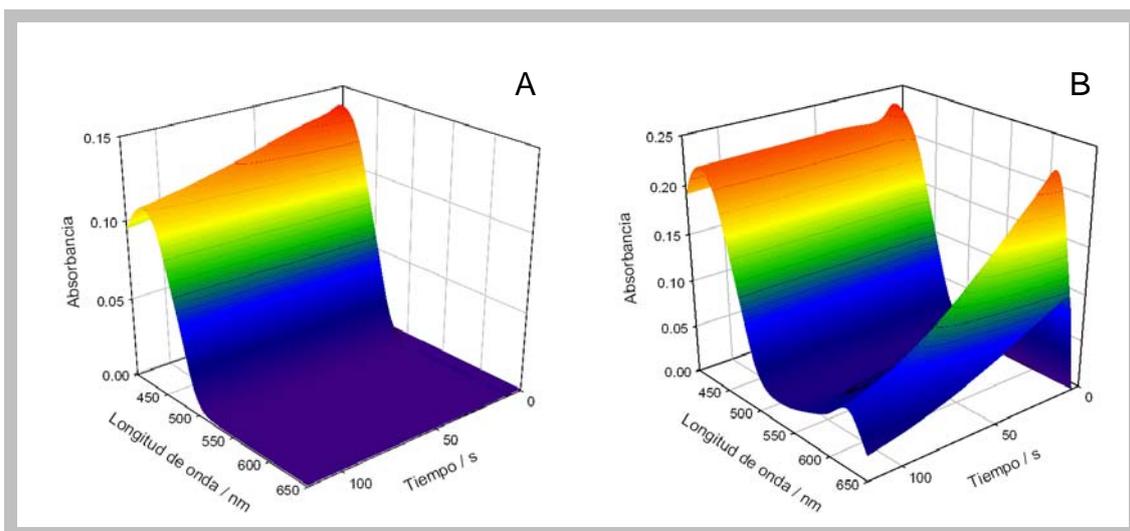


Figura 3.4. Gráfico tridimensional de absorbancia en función de longitud de onda y tiempo de reacción para: A) una muestra de calibración que contiene 16,00 mg L⁻¹ de tartrazina y B) una muestra de validación que contiene 15,00 mg L⁻¹ tanto de tartrazina como de azul brillante, durante la reacción con bromato de potasio.

Con el objetivo de medir el grado de solapamiento espectral entre los componentes 1 y 2 (S_{12}), en la presente tesis se propuso el empleo de la siguiente expresión:

$$S_{12} = \frac{||s_1^T s_2||}{||s_1|| ||s_2||} \quad (3.9)$$

en donde s_1 y s_2 son los espectros para los componentes 1 y 2, respectivamente. El valor de S_{12} está comprendido entre cero y uno, correspondiendo a situaciones extremas de no solapamiento y de solapamiento completo, respectivamente.

El grado de solapamiento espectral, estimado empleando la Ecuación (3.9), es de 0,9856. Los dos únicos componentes de las muestras de validación, que tienen tanto tartrazina como azul brillante, que presentan respuesta espectral son: la tartrazina y el producto de oxidación del azul brillante. MCR-ALS se aplicó a este sistema aumentado las matrices tanto en el modo temporal como en el espectral y se obtuvieron los resultados analíticos presentados en la Tabla 3.2. Como se puede ver, se obtienen mejores resultados cuando se aumenta en el modo temporal con un RMSE (REP%) de 0,54 (4,4%). Por otra parte, se obtiene un valor de RMSE (REP%) de 1,63 (13,1%) cuando se realiza el aumento en el modo espectral. Este hecho se puede adjudicar a las diferencias espectrales entre la tartrazina y el producto de oxidación del azul brillante, las cuales son suficientemente grandes como para estabilizar la inversa generalizada del componente puro de la matriz espectral S . De hecho, la Ecuación (3.9) se puede emplear para estimar el grado de solapamiento cinético entre los componentes a través del reemplazo de los espectros por los correspondientes perfiles cinéticos. El resultado es 0,9969; mostrando un fuerte solapamiento en la dimensión temporal en comparación con la dimensión espectral. Esta es la razón por la cual se obtienen mejores cifras de mérito al aumentar en el modo temporal.

Tabla 3.2. Concentraciones nominales y predicciones de las muestras de validación del sistema 2.

Muestra de validación ^a	Tartrazina / mg L ⁻¹			Azul brillante / mg L ⁻¹
	Nominal	Predicha ^b	Predicha ^c	Nominal
1	10,00	10,92	13,07	10,00
2	15,00	15,28	16,42	10,00
3	10,00	10,33	11,99	15,00
4	15,00	14,82	15,03	15,00
5	12,50	13,15	13,80	12,50
RMSE / mg L ⁻¹		0,54	1,63	
REP%		4,4	13,1	

^a RMSE = error cuadrático medio; REP% = error relativo de predicción calculado con respecto al valor medio de concentración de la calibración.

^b Modo de aumento temporal. Los resultados son promedios de análisis por duplicado.

^c Modo de aumento espectral. Los resultados son promedios de análisis por duplicado.

Las Figuras 3.5 A y B muestran los perfiles espectrales y cinéticos (aumentados) para una muestra de validación típica y algunas de las muestras de calibración luego del procesamiento de los datos por MCR-ALS. El éxito del modelo quimiométrico aplicado para descomponer la contribución de ambos constituyentes de la muestra es evidente. No solo los espectros proporcionados (Figura 3.5 A) se correlacionan de manera satisfactoria con los espectros reales de los componentes, sino que además los perfiles cinéticos (Figura 3.5 B) están de acuerdo con la concentración creciente del producto de oxidación del azul brillante en la muestra de validación. Además, tanto en las muestras de validación como en las de calibración, se evidencia la disminución de la concentración de la tartrazina con el tiempo (Figura 3.5 B).

En este sistema experimental en particular, las diferencias en los perfiles espectrales proveen suficiente información, por lo que el modo de aumento temporal usual funciona mejor que el aumento espectral en lo que se refiere a cifras de mérito analíticas. Sin embargo, en un sistema con solapamiento espectral más severo lo opuesto puede ser verdad. Los resultados aquí presentados indican que en estos casos se deben chequear ambas opciones antes de decidir cuál provee mejores resultados analíticos.

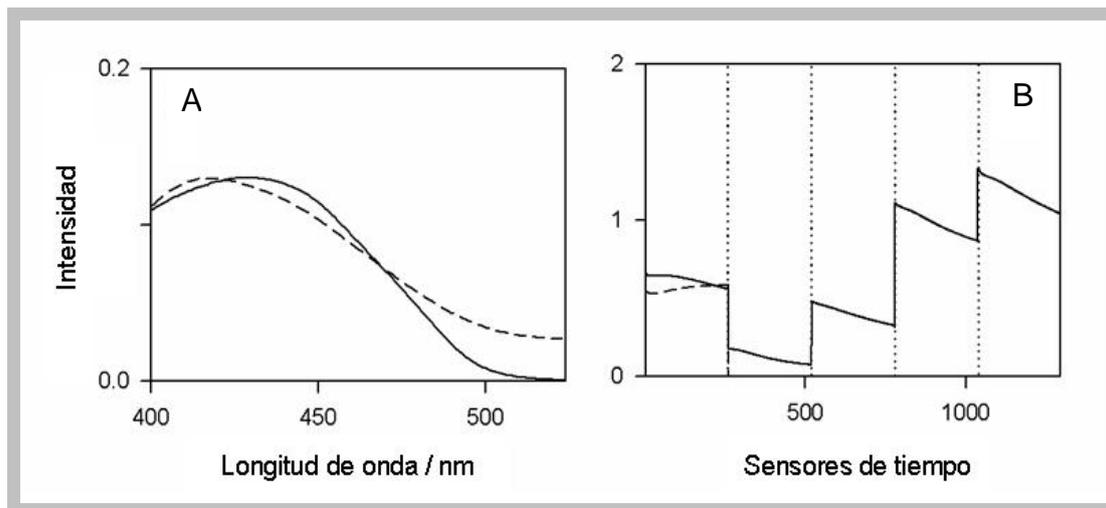


Figura 3.5. Gráfico de los perfiles espectrales y cinéticos luego del procesamiento por MCR-ALS de una muestra de validación típica junto con las muestras de calibración del sistema 2. A) Perfiles espectrales normalizados a la unidad y B) perfiles cinéticos para matrices de muestras sucesivas aumentadas en el modo cinético (la primera es la muestra de validación y las restantes son las cuatro primeras muestras de calibración; las líneas punteadas verticales separan los sensores de cada muestra). Se muestran solo cuatro muestras de calibración por simplicidad. La línea sólida corresponde al analito y la línea de trazos a la interferencia.

3.6.2.2. Muestras reales

Las matrices de segundo orden obtenidas por el método cinético para cada muestra de gelatina y sus correspondientes adiciones estándares se sometieron al análisis mediante MCR-ALS de la manera descrita para las muestras de validación. Cada muestra se analizó por quintuplicado. Posteriormente se construyó una gráfica de adición estándar con las áreas relativas obtenidas para cada muestra y sus correspondientes adiciones. La intersección de la línea de calibración con la abcisa permitió conocer la concentración de tartrazina en la muestra. El mismo procedimiento se empleó para cuantificar azul brillante, pero en este caso las longitudes de onda se restringieron a 525 - 700 nm y los tiempos a 14 - 115,6 s; conduciendo a 176×259 datos por muestra. En este último análisis fue necesario considerar un solo componente para la resolución mediante MCR-ALS, ya que en la región seleccionada la respuesta espectral corresponde solo a la generada por el azul brillante.

En la Tabla 3.3 se observan los resultados de la cuantificación de tartrazina y azul brillante obtenidos para cada muestra de gelatina, tanto por MCR-ALS como por el método CLAR publicado por Pereira Alves y col (Pereira Alves y col 2008). Con el objetivo de comparar estadísticamente los resultados obtenidos mediante MCR-ALS con los proporcionados por el método CLAR, se realizó un análisis estadístico de comparación de medias con variancias diferentes sobre los resultados de cinco replicados procesados con cada método. Se concluyó que no existen diferencias significativas entre los resultados obtenidos mediante ambos métodos ya que todos los $t_{\text{calculado}}$ fueron menores que el t_{tabulado} para $\alpha = 0,05$ y 4 grados de libertad. Los resultados del análisis estadístico se presentan en la Tabla 3.3.

Tabla 3.3. Resultados obtenidos aplicando el test de comparación de medias entre métodos para los resultados obtenidos mediante MCR-ALS y CLAR.

	Tartrazina (mg / 100g)			Azul brillante (mg / 100g)		
	Media	SD	$t_{\text{calculado}}^a$	Media	SD	$t_{\text{calculado}}^a$
	G1					
CLAR	9,20	0,03		2,47	0,04	
MCR-ALS	9,14	0,43	0,31	2,38	0,18	1,09
	G2					
CLAR	8,42	0,03		1,03	0,02	
MCR-ALS	8,36	0,39	0,34	1,13	0,10	2,19

^a Los grados de libertad fueron 4, el límite de confianza 95% y valor tabulado para $t = 2,78$.

En vista de los resultados obtenidos, la presente metodología puede ser recomendada para realizar análisis de rutina de tartrazina y azul brillante en gelatinas.

3.7. SISTEMA 3: Determinación de carbamazapina en suero en presencia de las interferencias carbamazapina epóxido y lamotrigina.

3.7.1. PREPROCESAMIENTO DE LOS DATOS

La instrumentación utilizada para realizar EC es conocida por presentar problemas que se relacionan con la línea de base y los corrimientos en el tiempo de migración entre corridas, que requieren del preprocesamiento de los datos (Daszykowski y Walczak 2006).

3.7.1.1. Corrección de la línea de base

La implementación de un preprocesamiento apropiado es de extrema importancia para garantizar la calidad de las señales instrumentales, lo cual a su vez determina la calidad de los resultados finales del análisis electroforético. Del mismo modo que cualquier otra señal instrumental, los electroferogramas están compuestos de tres componentes principales: señal, ruido y línea de base. En consecuencia, la eliminación del ruido o la línea de base de los componentes puede resultar en un aumento de la señal (Daszykowski y Walczak 2006).

Además, la eliminación de la línea de base del electroferograma es un paso crítico para reducir el número y la complejidad de los componentes no esperados. En la literatura existen diferentes estrategias para la corrección de la línea de base (Daszykowski y Walczak 2006). En el presente trabajo, se aplicó el método cuadrados mínimos asimétricos propuesto por Eilers (Eilers 2004), el cual recientemente fue adaptado para datos multidimensionales (Eilers y col 2006). Este método consiste en la estimación de una matriz de fondo (*background*) F ($J \times K$) a partir de la matriz de datos M ($J \times K$), en donde J es el número de longitudes de onda digitalizadas y K es el número de tiempos de migración. Para lograr este objetivo, se usó una matriz de base curva definida a tramos mediante polinomios (*spline*) B_1 ($L \times J$) a lo largo de las filas de la matriz M y una matriz de base *spline* B_2 ($M \times K$) a lo largo de las columnas de la matriz M . En base a la literatura (Kaczmarek y col 2005) se utilizó un compromiso de 10 funciones bases, es decir $L = M = 10$. F se puede representar como:

$$f_{j,k} = \sum_{L,M} b1_{Lj} b2_{Mk} a_{LM} \quad (3.10)$$

en donde a_{LM} es el elemento (L,M) de una matriz A que contiene los coeficientes de regresión, los cuales se pueden calcular a través de la minimización de la siguiente función costo:

$$Q = \sum_{JK} v_{JK} (y_{JK} - f_{JK})^2 + P \quad (3.11)$$

donde y es la señal experimental, f es una tendencia del ruido (la aproximación de la línea de base) y v son los pesos anteriores. Los elementos de v deberían tener valores grandes en las partes de la señal en donde está permitido afectar la estimación de la línea de base. Si se considera la siguiente elección de pesos asimétricos: $v_{JK} = p$ si $v_{JK} > f_{JK}$ y $v_{JK} = 1 - p$ si $v_{JK} \leq f_{JK}$ con $0 < p < 1$, se observa que desviaciones positivas de la tendencia resultarán en pesos diferentes de los residuos negativos. La experiencia muestra que comenzar a partir de $v \cong 1$, e iterar entre dos cálculos, conduce de manera rápida y confiable a una solución en aproximadamente 10 iteraciones. Finalmente, en la Ecuación (3.11) hay un término de penalización definido por:

$$P = \lambda \left[\sum_L (\Delta_1^d a_{\cdot L})^2 + \sum_M (\Delta_2^d a_{M \cdot})^2 \right] \quad (3.12)$$

en donde Δ_1 y Δ_2 son diferencias de orden d calculadas para cada columna de A ($a_{\cdot L}$) y cada fila de A ($a_{M \cdot}$), respectivamente. Como se puede observar en la Ecuación (3.12), si se usan valores diferentes para el parámetro de regularización λ , la penalización podría tener diferente influencia en las direcciones verticales y horizontales. En el presente trabajo se utilizó un único parámetro de regularización.

3.7.1.2. Corrección del corrimiento del tiempo de migración de los picos

Con el propósito de alinear los picos se empleó la metodología propuesta por Fraga y col (Fraga y col 2001), la cual tiene en cuenta la estructura de segundo orden de los datos. Este algoritmo se basa en la SVD de una matriz de N/M que contiene las matrices de datos N y M, en donde N se toma como referencia y M será corregido en relación a N:

$$(U, S, V) = \text{SVD}(N/M) \quad (3.13)$$

En la Ecuación (3.13), S es una matriz diagonal que contiene los valores singulares, mientras que U y V son matrices que contienen los vectores singulares izquierdo y derecho de N/M, respectivamente. La corrección se realiza calculando la variancia residual (RES) a medida que la matriz M se mueve en relación a N, usando un número preestablecido de puntos que pueden estimarse mediante inspección de los electroferogramas de N y M.

$$\text{RES} = 100 \frac{\sum_{a=A+1}^{\min(J,K)} S_a^2}{\sum_{a=1}^{\min(J,K)} S_a^2} \frac{JK}{(J-A)(K-A)} \quad (3.14)$$

donde A es el número de valores singulares significativos, idealmente igual al número de especies presentes en N/M, y J y K tienen el significado descrito anteriormente. Cuando las matrices N y M se alinean, los valores de RES deberían alcanzar un mínimo.

3.7.2. MATERIALES Y MÉTODOS

3.7.2.1. Equipamiento

Todos los experimentos se llevaron a cabo en un sistema de electroforesis capilar (Agilent Technologies, Waldbronn, Alemania) equipado con detector de arreglo de diodos. El instrumento se operó bajo condiciones de polaridad positiva. Las mediciones se realizaron en un capilar desnudo de sílice fundida (Microsolv, USA) de 53,5 cm de largo total (largo efectivo 45 cm) con 75 μm

de diámetro interno. El cartucho se mantuvo a 25°C. Se aplicó un voltaje constante de +20 kV con una corriente de aproximadamente 90 μ A. La inyección se realizó desde el cátodo del capilar aplicando una presión de 50 mbar durante 10 s.

Para cada electroferograma se registraron espectros en el rango de 189 - 351 nm cada 2 nm, a intervalos regulares de 0,003 s por un tiempo total de 5 min. Por lo tanto, se generaron matrices de tamaño 616 \times 82 por muestra, aunque se seleccionaron regiones antes de realizar la calibración multivariada.

En todos los casos los electroferogramas se registraron de manera aleatoria con respecto a las concentraciones de analito y los correspondientes a las muestras de calibración se obtuvieron en días diferentes a los del juego de validación.

3.7.2.2. Reactivos

Todas las experiencias se realizaron con drogas calidad para análisis (borato de sodio, fosfato de sodio, hidróxido de sodio, diclorometano y metanol) y agua milliQ. Carbamazepina (CBZ), su principal metabolito carbamazepina epóxido (CBZ-EP), fenobarbital (PHB), fenitoína (PHT), fenilefrina (FEN), ibuprofeno (IBU), acetaminofén (PAR), teofilina (TEO), cafeína (CAF) y ácido acetilsalicílico (AAS) fueron provistos por Sigma-Aldrich Inc. Lamotrigina (LAMO) fue provista por Química Alkano S.A.

Las soluciones stock se prepararon en los siguientes solventes: CBZ, CBZ-EP y LAMO en metanol; FEN, PHB, PHT, PAR, AAS, CAF y TEO en agua e IBU en etanol, a concentraciones de 2000,0 mg L⁻¹, y se almacenaron en oscuridad a -4°C. Las soluciones estándares de todos los analitos se prepararon diariamente mediante dilución de las soluciones stock en solución reguladora de corrida (SRC). Esta última solución consistió en una solución reguladora de borato de sodio 0,025 mol L⁻¹ de pH = 9,20.

3.7.2.3. Procedimiento de extracción

Para realizar la extracción se transfirieron 500 μ L de una dada muestra de suero a un tubo de fondo circular y se extrajeron los analitos con tres

alícuotas sucesivas de 1,00 mL de diclorometano. El tubo se tapó, se mezcló en vórtex y se centrifugó a 4000 rpm durante 5 min. La capa inferior orgánica se transfirió a un tubo limpio y se evaporó a sequedad a 45 °C en una manta calefactora. Finalmente el residuo se redisolvió en 500 µL de SRC y se filtró a través de una membrana de 0,45 µm de tamaño de poro, antes de ser inyectado en el instrumento.

3.7.2.4. Diseño de los juegos de calibración y validación

Para calibrar se preparó un juego de veintiuna muestras estándares por dilución de cantidades apropiadas de solución stock de CBZ en SRC. Se prepararon por triplicado siete niveles de concentración en el rango lineal comprendido entre 1,00 y 15,00 mg L⁻¹ y se analizaron en el sistema de EC. Este rango incluye las concentraciones terapéuticas medias de CBZ informadas en la literatura (de 4,00 a 12,00 mg L⁻¹) (Smith y Reynard 1993).

Por otra parte se prepararon varias muestras de validación conteniendo mezclas de CBZ y CBZ-EP a niveles de concentración diferentes a los empleados para calibrar, mediante dilución de las soluciones stock de estándares en SRC (muestras B1-B5, ver Tabla 3.4).

Además se prepararon quince muestras de validación compuestas por sueros blanco obtenidos de pacientes sanos, los cuales no estaban bajo tratamiento con antiepilépticos, adicionados con cantidades conocidas de solución estándar de CBZ y CBZ-EP: SA1-SA5, SB1-SB5 y SC1-SC5 (ver Tabla 3.4). A las cinco últimas muestras se les agregaron, además, concentraciones conocidas de LAMO. Las mezclas resultantes se sometieron al procedimiento de extracción. Varias muestras de suero que contenían la misma concentración de CBZ se emplearon para evaluar precisión.

Finalmente se preparó una muestra adicional para ser empleada en un estudio de interferencias. Consistió en una muestra de suero adicionada con las siguientes drogas: a) CBZ y CBZ-EP (5,00 mg L⁻¹), b) FEN (20,00 mg L⁻¹) y c) LAMO, PAR, IBU, PHT, PHB, TEO CAF y AAS (10,00 mg L⁻¹).

Tabla 3.4. Composición de las muestras de validación.

Muestra ^a	CBZ / mg L ⁻¹	CBZ-EP / mg L ⁻¹	LAMO / mg L ⁻¹
B1	2,50	2,50	-
B2	2,98	5,99	-
B3	3,97	7,99	-
B4	6,99	2,00	-
B5	7,50	7,38	-
SA1	3,02	5,99	-
SA2	3,02	7,03	-
SA3	3,97	2,00	-
SA4	3,97	7,99	-
SA5	6,03	5,99	-
SB1	2,00	8,00	-
SB2	3,00	5,00	-
SB3	4,00	6,00	-
SB4	6,00	4,00	-
SB5	8,00	2,00	-
SC1	4,96	4,99	5,66
SC2	4,96	4,99	7,55
SC3	4,96	4,99	9,44
SC4	4,96	4,99	11,33
SC5	4,96	4,99	13,21

^a La primera letra mayúscula corresponde al tipo de matriz en la cual se preparó cada muestra: B, SRC; S, suero. La segunda letra mayúscula identifica otros componentes presentes en cada muestra: A, CBZ-EP; B, CBZ-EP y componentes no esperados del suero; C, CBZ-EP y LAMO.

3.7.3. RESULTADOS Y DISCUSIÓN

3.7.3.1. Corrección de la línea de base y el corrimiento temporal

Con el objetivo de simplificar el análisis se realizó la eliminación de la línea de base y la corrección del corrimiento del tiempo de migración en los

datos de segundo orden electroforéticos-espectrales. Ambos procedimientos de preprocesamiento se implementaron a través de algoritmos previamente adaptados por sus autores a datos multivariados (Eilers y col 2006, Fraga y col 2001).

Para obtener la matriz de corrección del *background* F se usaron diez funciones de base *spline* con un sólo parámetro de regularización cuyo valor fue 1. La Figura 3.6 A muestra la superficie correspondiente a una muestra de suero (SC1 en la Tabla 3.4), la cual fue adicionada con CBZ, CBZ-EP y LAMO. Como se puede ver en esta figura, hay un solo pico en el dominio temporal, indicando el solapamiento completo de las tres sustancias. La Figura 3.6 B muestra la correspondiente matriz *background* (F). Se usó una escala diferente con el propósito de mostrar la variación de la línea de base a diferentes longitudes de onda. Como se puede observar, la línea de base electroforética se incrementa con las longitudes de onda hasta que se alcanza un máximo a aproximadamente 300 nm, hecho que justifica el uso de un algoritmo de corrección de línea de base que contemple ambas vías al mismo tiempo. La sustracción de esta matriz permite obtener una matriz de datos corregida, la cual, a continuación, se somete a la corrección del desplazamiento temporal. Finalmente, la Figura 3.6 C muestra el electroferograma original graficado a 250 nm y el correspondiente electroferograma con la línea de base y el corrimiento corregidos.

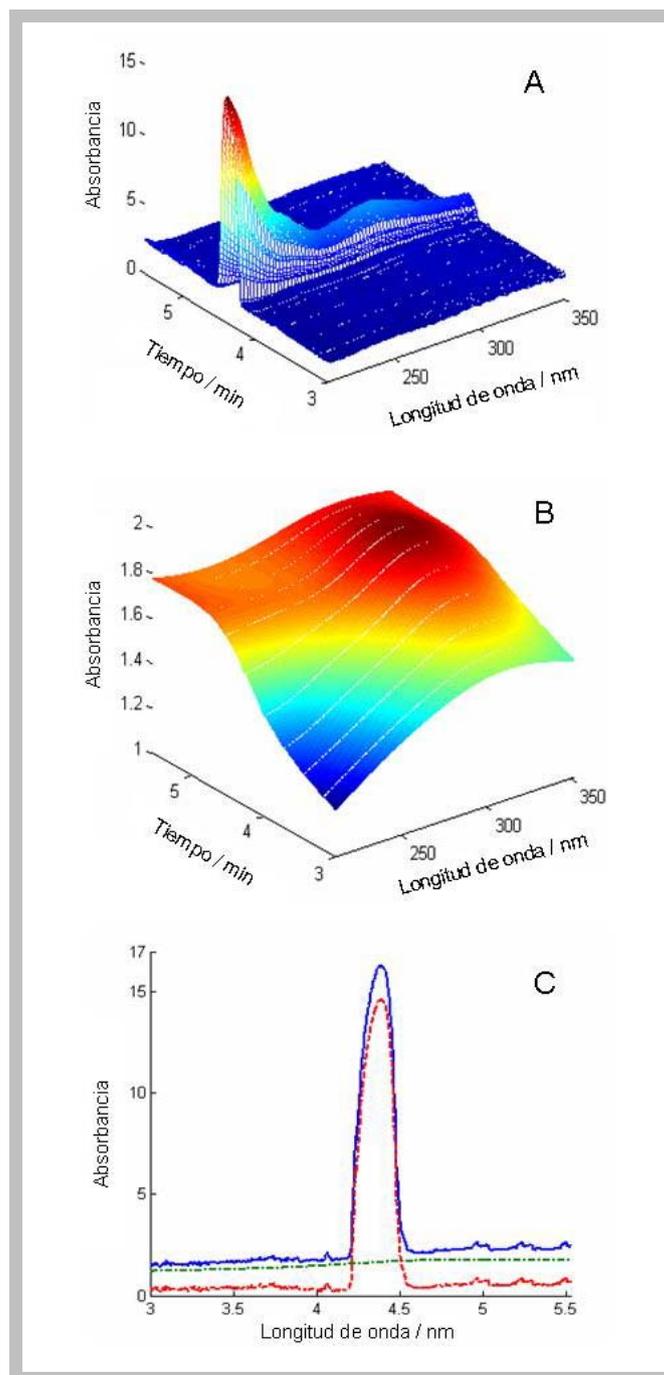


Figura 3.6. A) Superficie correspondiente a una muestra de suero (SC1) adicionada con CBZ ($4,96 \text{ mg L}^{-1}$), CBZ-EP ($4,99 \text{ mg L}^{-1}$) y LAMO ($5,66 \text{ mg L}^{-1}$). B) Matriz *background* correspondiente a la superficie de la Figura 3.6 A. C) Línea sólida: electroferograma correspondiente a una muestra de suero (SC1); línea de trazos y puntos: su correspondiente *background*; línea de trazos: electroferograma corregido (todos registrados a 250 nm).

La necesidad de la corrección temporal se puede apreciar en la Figura 3.7 A, en donde se graficaron los electroferogramas de cinco estándares de carbamazepina y una muestra de suero adicionada (la misma que se

seleccionó para ilustrar la corrección de la línea de base) registrados a 250 nm. El corrimiento temporal es evidente y, aunque MCR es capaz de hacer frente a datos no sincronizados, análisis posteriores mostraron que la corrección mejora los resultados. Como se mencionó anteriormente, la corrección de una matriz M (línea sólida en la Figura 3.7 A), tomando una matriz N como referencia (graficada en línea de trazos), consiste en realizar la descomposición en valores singulares de la matriz de datos conjunta N/M . El número de componentes necesarios fue de 2 o 3 dependiendo de la matriz analizada. Además, la ventana se varió con el objetivo de optimizar la aplicación. A modo de ejemplo, la Figura 3.7 B muestra la variación del RES(%) [ver Ecuación (3.14)] cuando la matriz estudiada se cambia con respecto a la matriz de referencia. El pico más remoto necesitó una ventana de corrección de 39 puntos temporales para alcanzar un mínimo. Finalmente, la Figura 3.7 C muestra los electroferogramas registrados a 250 nm para los seis picos alineados con respecto al de referencia. Resulta interesante ver que, aunque los picos parecen aproximadamente alineados, la pérdida de trilinealidad es evidente debido a que las formas de las bandas individuales son todas diferentes, haciendo necesaria la aplicación del algoritmo MCR-ALS para la resolución.

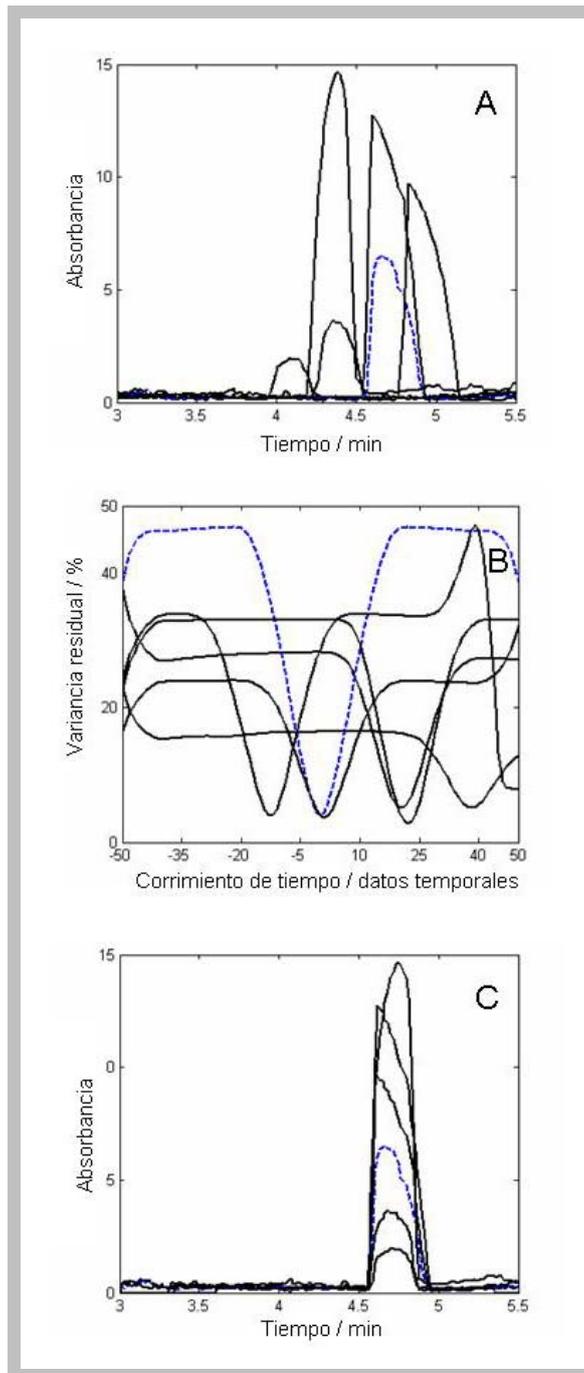


Figura 3.7. A) Electroferograma registrado a 250 nm para cinco estándares de CBZ (líneas sólidas) y una muestra de suero adicionada (SC1, línea de trazos). B) Variación RES(%) [ver Ecuación (3.14)] cuando la matriz estudiada se modifica con respecto a la matriz de referencia. C) Electroferogramas con corrimientos corregidos registrados a 250 nm.

3.7.3.2. Aplicación de MCR-ALS a mezclas en SRC y muestras de suero adicionadas

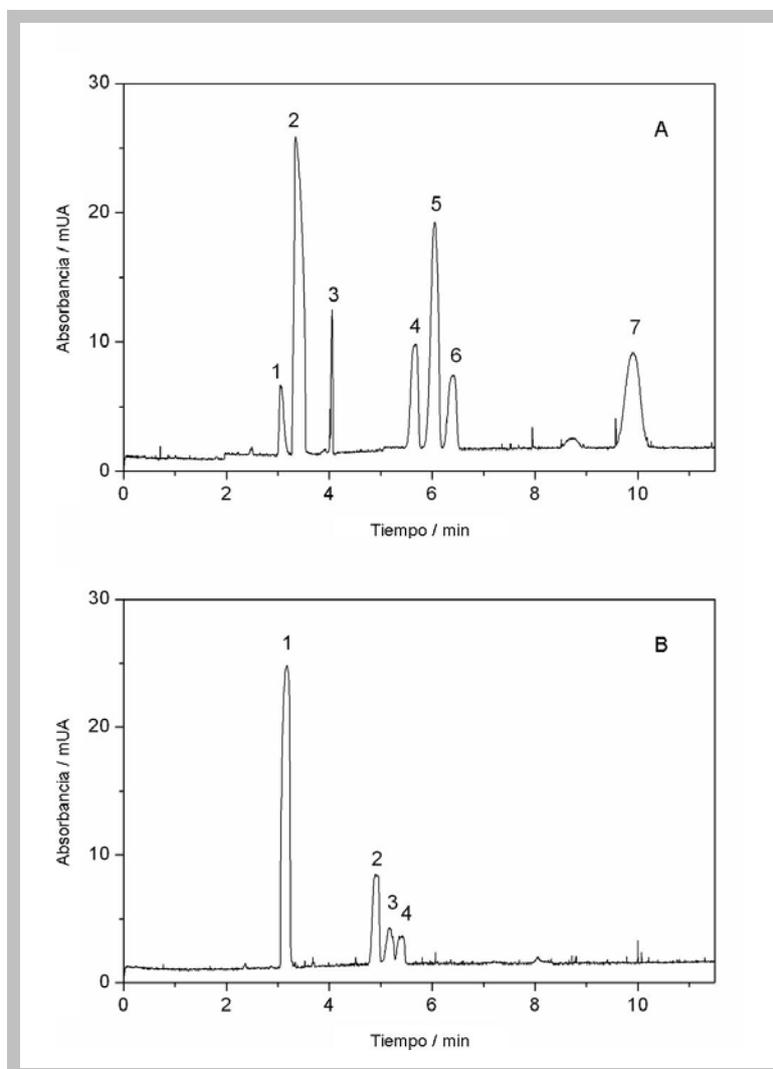


Figura 3.8. A) Electroferograma de una muestra adicionada preparada en SRC, en la cual los picos corresponden a: 1) FEN ($20,00 \text{ mg L}^{-1}$); 2) CBZ, CBZ-EP ($5,00 \text{ mg L}^{-1}$), CAF y LAMO ($10,00 \text{ mg L}^{-1}$); 3) PAR ($10,00 \text{ mg L}^{-1}$); 4) PHT ($10,00 \text{ mg L}^{-1}$); 5) IBU y TEO ($10,00 \text{ mg L}^{-1}$); 6) PHB ($10,00 \text{ mg L}^{-1}$) y 7) AAS ($10,00 \text{ mg L}^{-1}$). B) Electroferograma de la misma muestra luego del procedimiento de extracción: 1) CBZ, CBZ-EP y LAMO; 2) PHT; 3) IBU y TEO y 4) PHB. Ambos electroferogramas se registraron a 250 nm .

La Figura 3.8 A muestra el electroferograma completo de una muestra preparada en SRC que contiene no solo CBZ y su principal metabolito CBZ-EP, sino también otras drogas antiepilépticas como PHT, PHB y LAMO, las cuales probablemente pueden ser coadministradas con CBZ, además de una variedad

de fármacos de uso común: FEN, PAR, TEO, IBU, CAF y AAS. Como se puede observar, la separación electroforética no es completa conduciendo al solapamiento de CBZ, CBZ-EP, CAF y LAMO (pico número 2) y TEO e IBU (pico número 5). Además, el procedimiento de extracción aplicado a la muestra recupera tres de los cuatro analitos incluidos en el pico número 2 (CBZ, CBZ-EP y LAMO), haciendo imposible la determinación de CBZ por calibración univariada (ver Figura 3.8 B).

3.7.3.2.1. Análisis exploratorio del rango de los juegos de datos

En MCR-ALS, el rendimiento de la resolución depende fuertemente del conocimiento de las propiedades globales y locales de los datos, particularmente de aquellas relacionadas al rango matemático y químico (Ruckebusch y col 2006). En la práctica, una matriz de datos se considera de rango completo si el rango estimado es igual al número de especies químicas que absorben, asumiendo que ese número es conocido (Ruckebusch y col 2006).

El primer paso para la aplicación de MCR-ALS es la determinación del número correcto de componentes presentes en la matriz estudiada aplicando, por ejemplo, SVD (Muñoz y De Juan 2007). En el presente caso, cuando se analizaron matrices de muestras que contenían tres especies químicas con tiempos electroforéticos casi idénticos (CBZ, CBZ-EP y LAMO), SVD no fue capaz de detectar el número correcto de componentes que explican la variancia del sistema, ya que estimó un único componente principal. Esto se debe a que los perfiles temporales idénticos conducen a matrices deficientes de rango, a pesar de la existencia de cierta selectividad en el dominio espectral.

Para solucionar el problema de la deficiencia de rango se puede implementar el recurso matemático de aumentar la matriz. Esta es una manera simple e inteligente de romper la deficiencia de rango y facilitar la resolución (De Juan y col 2000). La estrategia empleada en este trabajo consistió en aumentar la matriz bajo estudio, deficiente de rango, con matrices de estándares de analitos puros para luego estimar el rango de la matriz aumentada mediante SVD. El aumento se realizó en el modo temporal,

único posible en este caso en donde el perfil de tiempo del analito de interés coincide con el de las interferencias. Un caso típico se muestra en la Tabla 3.5 para la muestra SC3. La aplicación de SVD solamente a la matriz de esta muestra que contiene CBZ, CBZ-EP y LAMO (Tabla 3.4) detecta un único valor singular significativo en lugar de tres, el cual es el número de componentes químicos presentes en la muestra (primera entrada en la Tabla 3.5). Solo cuando la matriz se aumenta mediante el agregado de la información de los estándares de los analitos CBZ-EP y LAMO, la aplicación de SVD a la matriz aumentada es capaz de estimar correctamente el rango, es decir tres (última entrada en la Tabla 3.5). Cabe destacar que este procedimiento colaboró además en la detección de componentes no esperados en las muestras de suero correspondientes al set SB (ver Tabla 3.4). Debido a que todas las muestras de suero eran diferentes, algunas contenían componentes que se vieron durante el análisis por MCR-ALS mientras que otros no generaron interferencias significativas.

Tabla 3.5. Rango estimado calculado a través de la descomposición en valores singulares de los juegos de datos individuales y aumentados.

Matriz aumentada ^{a,b}	Rango químico conocido	Rango estimado por SVD en la matriz aumentada	Condición
[SC3]	3	1	Deficiencia de rango
[SC3,C]	3	2	Deficiencia de rango
[SC3,E,C]	3	2	Deficiencia de rango
[SC3,L,C]	3	2	Deficiencia de rango
[SC3,E,L,C]	3	3	Rango completo

^a La nomenclatura $[X_1, X_2, \dots, X_M]$ implica que se crea una matriz aumentada D [ver Ecuación (3.3)] acomodando las matrices X_1, X_2, \dots, X_M una al lado de la otra. Se asume que la dirección de las columnas es la longitud de onda espectral mientras que la dirección de las filas es el tiempo electroforético.

^b SC3 es una muestra de suero que contiene CBZ, CBZ-EP y LAMO (ver Tabla 3.4), mientras que C, E y L son las matrices de los estándares puros de CBZ, CBZ-EP y LAMO, respectivamente.

3.7.3.2.2. Resultados de la calibración multivariada

Previamente a la resolución mediante MCR-ALS, se realizó en todas las matrices la corrección de la línea de base y el corrimiento de tiempos, excepto en el set SB. Como se puede observar en la Figura 3.9, el electroferograma de una muestra SB contiene un pico ancho que corresponde a componentes del suero que eluyen junto con CBZ y CBZ-EP. Este hecho permite realizar solo la corrección de la línea de base pero hace imposible la alineación del pico 2. Para poder realizar la cuantificación mediante MCR-ALS, aprovechando su capacidad de manejar datos no trilineales, se seleccionaron regiones de tiempo apropiadas en las matrices correspondientes a las muestras SB, de acuerdo con la posición del pico de interés en cada electroferograma.

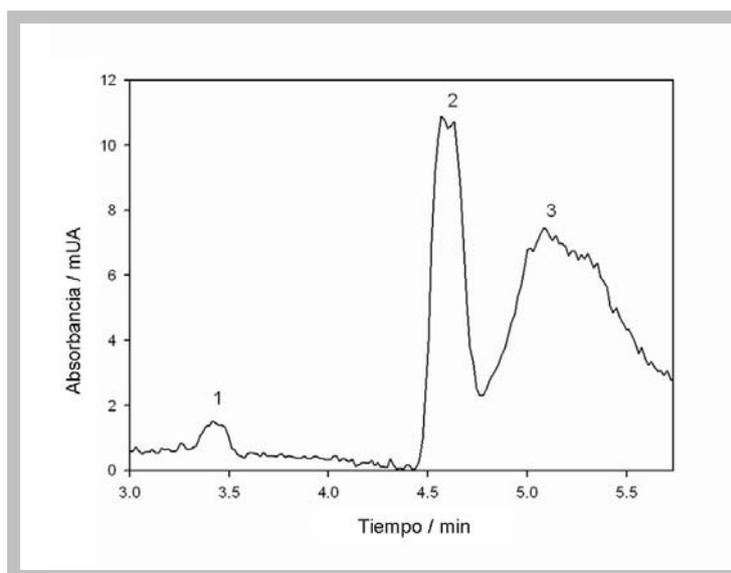


Figura 3.9. Electroferograma (registrado a 250 nm) correspondiente a una muestra de suero SB que contiene CBZ y CBZ-EP solapadas en el pico 2 y componentes no esperados en los picos 1 y 3.

Cada una de las muestras de validación se analizó como se explica a continuación. La matriz espectral-temporal para una dada muestra se aumentó en el modo temporal con las veintiuna matrices registradas para las muestras de calibración (que solo contenían CBZ). Una matriz adicional de un estándar de CBZ-EP se añadió en todos los casos a la matriz aumentada. Finalmente, en el caso de las muestras de suero SC se agregó una matriz de

un estándar de LAMO. Las matrices de CBZ-EP y LAMO se agregaron con el propósito de brindar a MCR-ALS suficiente información para la resolución de los componentes presentes en las muestras. La Tabla 3.6 muestra los resultados del análisis de las muestras de validación para CBZ. SVD detectó dos componentes para los sets B y SA, uno correspondiente a CBZ y el otro a CBZ-EP, mientras que se necesitó un componente adicional para los sets SB y SC, correspondiente a un componente no esperado del suero y a LAMO, respectivamente. Para los sets de muestras B, SA y SC, los cuales contienen CBZ y CBZ-EP (y también LAMO en el último caso), las estimaciones iniciales se construyeron a partir de los espectros de los estándares puros de cada analito. Aquellas para el set SB también incluyeron el perfil de un componente no esperado obtenido a través de la aplicación de SIMPLISMA (Windig y Guilment 1991) a una muestra SB. Las restricciones impuestas durante el procedimiento ALS fueron no negatividad en concentración y espectros, unimodalidad en concentración y correspondencia entre especies.

Tabla 3.6. Concentraciones de las muestras de validación y resultados de la predicción mediante MCR-ALS.

Muestra ^a	Concentración de CBZ / mg L ⁻¹		Recuperación / %
	Nominal	Predicha	
B1	2,50	2,42	96,8
B2	2,98	2,81	94,3
B3	3,97	4,47	112,6
B4	6,99	7,07	101,1
B5	7,50	8,13	108,4
SA1	3,02	2,30	76,2
SA2	3,02	2,91	96,4
SA3	3,97	3,58	90,2
SA4	3,97	4,23	106,5
SA5	6,03	3,86	64,0
SB1	2,00	1,92	95,8
SB2	3,00	2,44	81,4
SB3	4,00	4,42	110,4
SB4	6,00	5,52	92,0
SB5	8,00	9,15	114,4
SC1	4,96	4,71	95,0
SC2	4,96	4,98	100,4
SC3	4,96	4,64	93,5
SC4	4,96	4,53	91,3
SC5	4,96	4,53	91,3

^a La primera letra mayúscula corresponde al tipo de matriz en la cual se preparó cada muestra: B, SRC; S, suero. La segunda letra mayúscula identifica otros componentes presentes en cada muestra: A, CBZ-EP; B, CBZ-EP y componentes no esperados del suero; C, CBZ-EP y LAMO.

Las Figuras 3.10 (A-D) y 3.11 (A-C) muestran los perfiles temporales y espectrales correspondientes a los tres componentes encontrados mediante el análisis por MCR-ALS para una muestra SC. Es evidente que el solapamiento en ambas dimensiones es severo, sin embargo MCR-ALS es capaz de descomponer

correctamente los datos en sus contribuciones relevantes. Esto permite el aislamiento de la señal atribuible al analito en cada muestra de validación y su cuantificación exacta. Además, la Figura 3.11 (A-C) muestra la comparación entre los perfiles espectrales obtenidos mediante MCR-ALS y los registrados experimentalmente para los estándares de CBZ, CBZ-EP y LAMO. Como se puede observar, se encontró una concordancia razonablemente buena entre los espectros correspondientes al analito y a las interferencias (coeficientes de correlación iguales a 0,9857; 0,9843 y 0,9971; respectivamente). Se debe notar la gran similitud entre los espectros de CBZ y CBZ-EP, hecho que puede explicar los problemas encontrados al estimar los rangos de las matrices para los sistemas que contienen las tres sustancias.

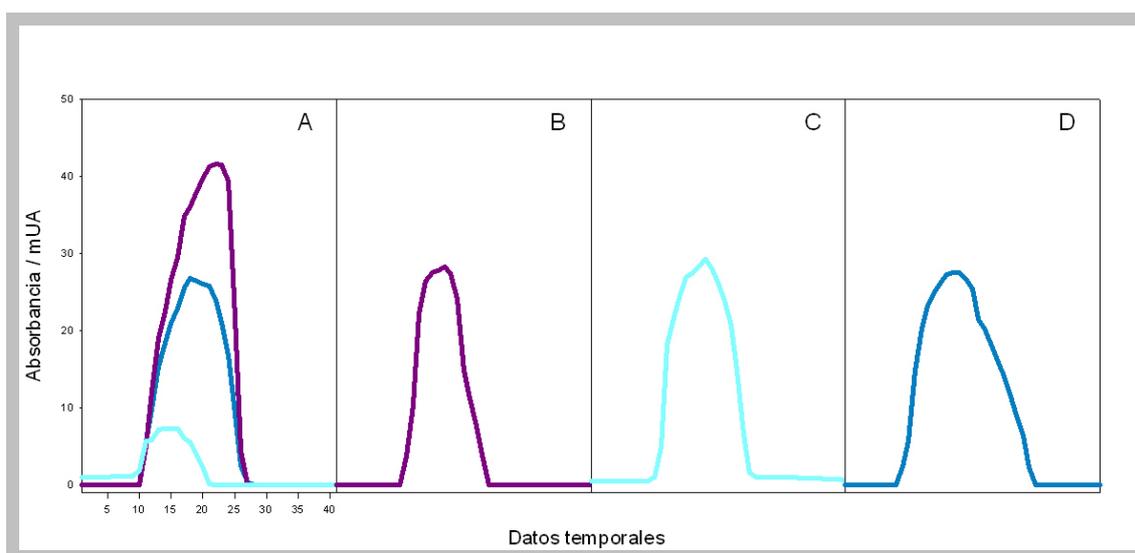


Figura 3.10. Perfiles temporales obtenidos del procesamiento de una muestra SC mediante MCR-ALS correspondientes a: A) los tres componentes encontrados en la matriz SC; B) un estándar de CBZ-EP ($5,00 \text{ mg L}^{-1}$); C) un estándar de LAMO ($5,00 \text{ mg L}^{-1}$) y D) un estándar de CBZ ($5,00 \text{ mg L}^{-1}$).

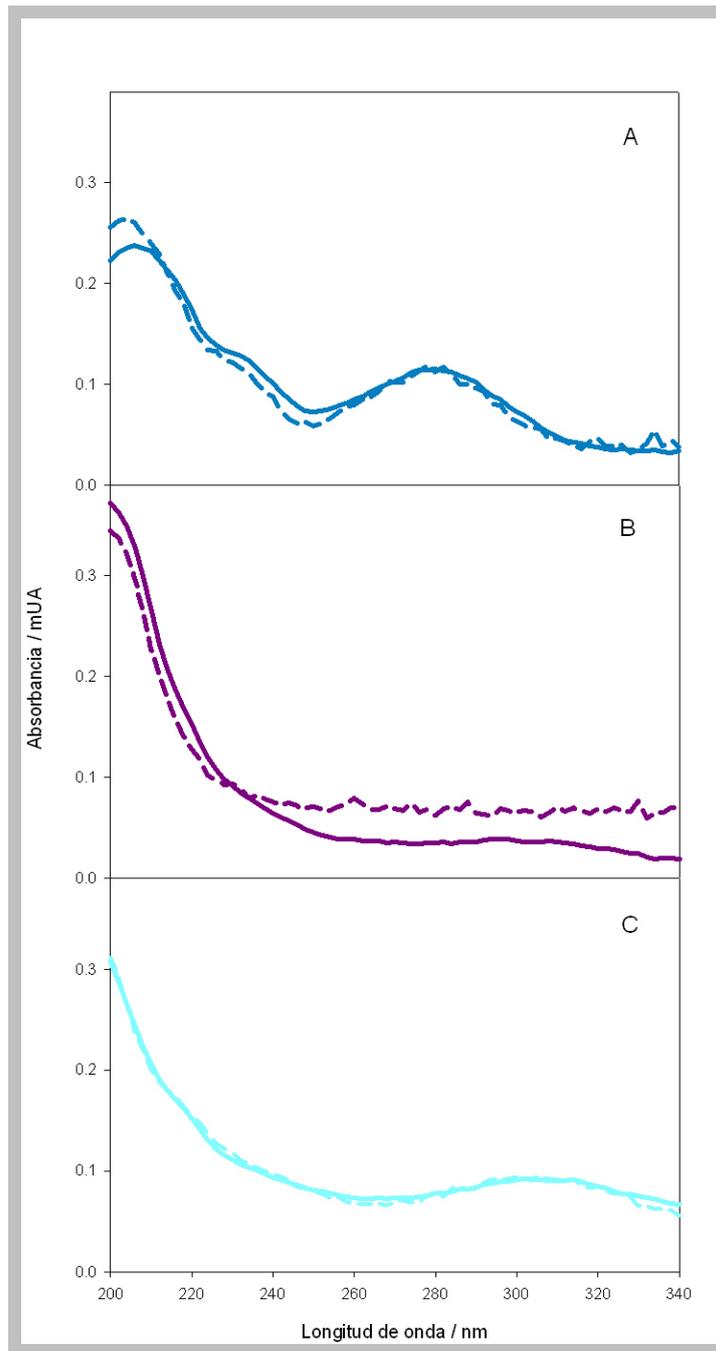


Figura 3.11. Comparación de los perfiles espectrales obtenidos del procesamiento de la muestra SC analizada en la Figura 3.10. En los tres casos, los espectros de los estándares puros se representan en líneas de trazos mientras que los perfiles extraídos se muestran como líneas sólidas. A) CBZ, B) CBZ-EP y C) LAMO.

Los resultados de la predicción pueden ser considerados razonables, y están incluidos en el rango aceptable sugerido por la ICH (Harmonised Tripartite Guideline 1994). Además, aquellos obtenidos para el set SC, compuesto de cinco muestras que tienen la misma concentración de CBZ, no

solo muestran excelentes recuperaciones sino además una muy buena reproducibilidad debido al bajo coeficiente de variación alcanzado [CV(%) = 4,0 %].

3.8. CONCLUSIONES

En el presente estudio se evidenció que MCR-ALS es un algoritmo que presenta un alto grado de adecuación para procesar datos de segundo orden con distintos grados de solapamiento en la dimensión espectral o temporal.

En presencia de solapamiento espectral severo, el aumento en la dimensión espectral permite una descomposición adecuada de la matriz conduciendo a la obtención de la ventaja de segundo orden.

Cuando la selectividad espectral es baja (pero no cero), las diferencias en los espectros de los componentes pueden permitir aumentar la matriz en el modo temporal y de esta manera hacer frente a variaciones en los perfiles temporales de muestra a muestra. No obstante, en este caso es recomendable testear ambas formas de aumento para emplear la que provea mejores cifras de mérito analíticas.

En presencia de solapamiento temporal severo, es el aumento en la dimensión temporal el cual permite alcanzar la ventaja de segundo orden, valiéndose de las diferencias espectrales entre el analito y las interferencias, y, por lo tanto, resolver el sistema de manera satisfactoria.

CAPÍTULO 4

Estudio comparativo del rendimiento de U-PLS/RBL frente a MCR-ALS en la determinación de ocho tetraciclinas presentes en aguas residuales por cromatografía líquida de alta resolución con detección por arreglo de diodos, previa preconcentración en fase sólida

4.1. INTRODUCCIÓN

Las tetraciclinas (TCs) son agentes bacteriostáticos de amplio espectro, activos contra bacterias Gram positivas y Gram negativas, que actúan inhibiendo la síntesis de proteínas. Estos antibióticos han sido usados ampliamente en medicina veterinaria y humana para la prevención y el tratamiento de enfermedades, y como aditivos alimentarios para promover el crecimiento de animales.

Un porcentaje considerable de los antibióticos consumidos por humanos y animales se excretan vía orina y heces sin sufrir modificaciones, y son descargados en las plantas de tratamiento de agua residual (PTARs), en donde son parcialmente eliminados. Por lo tanto, existe un riesgo potencial de que los residuos de antibióticos sean liberados en los efluentes de las PTARs al medio ambiente acuático. De hecho, las PTARs son consideradas como recursos puntuales de contaminación con antibióticos de aguas superficiales y subterráneas. Recientemente ha surgido preocupación con respecto a la aparición de antibióticos en el medio ambiente acuático y su influencia en la salud humana, ya que en los efluentes de las PTARs puede aparecer proliferación de bacterias resistentes (Yang y col 2005). En relación a estas consideraciones resulta necesaria la generación de métodos analíticos confiables y sensibles para determinar la concentración de estos compuestos en agua residual, con el objetivo de controlar su consumo en el medioambiente.

En la última década, la cromatografía líquida (LC) acoplada a espectrometría de masas (MS) o masas en tándem (MS-MS), vía interfaces de ionización a presión atmosférica (API) [ionización por electrospray (ESI) e ionización química por presión atmosférica (APCI)] ha resultado la técnica analítica de elección para la determinación de TCs en matrices medioambientales como agua natural y residual. LC-MS (Kolpin y col 2002, Lindsey y col 2001, Reverete y col 2003, Choi y col 2007, Ye y col 2007) y LC-MS-MS (Yang y col 2005, Hirsch y col 1999, Miao y col 2004, Zhu y col 2001, Hamscher y col 2002, Kim y Carlson 2007, De Ruyck y De Ridder 2007) fueron las técnicas de elección para la determinación de estas sustancias debido a su alta sensibilidad y habilidad de proveer confirmación sobre los compuestos.

Sin embargo, un aspecto crítico en el análisis cuantitativo con LC-MS es la aparición de un efecto matriz que puede conducir a diferencias significativas en la respuesta de un analito en una muestra en comparación con una solución de estándar puro. Para reducir o compensar este efecto se propusieron estrategias basadas en el procesamiento de la muestra (extracción selectiva, limpieza de la muestra luego de la extracción o mejoras en la separación cromatográfica) o en la realización de calibraciones adecuadas (uso de estándares internos, adición de los analitos dentro de cada muestra, estándares preparados en blanco de matriz o dilución de los extractos de las muestras), pero presentan las desventajas de que pueden consumir mucho tiempo (Stueber y Reemtsma 2004), requerir de estándares internos marcados isotópicamente (Choi y col 2001) o de matrices blanco reales que muestren el mismo comportamiento que las muestras. De todos modos, a pesar de que estas estrategias pueden compensar el efecto matriz, no evitan la pérdida de sensibilidad relacionada a la supresión de la señal en ESI y la posible falta de confianza en los resultados obtenidos de diferentes series de muestras (Kloepfer y col 2005).

Una alternativa interesante para cuantificar estos contaminantes de manera confiable consiste en el empleo de datos de segundo orden (y órdenes superiores) acoplados a algoritmos de calibración multivariada que presenten la ventaja de segundo orden, antes mencionada.

PARAFAC (Bro 1997) ha sido ampliamente empleado principalmente en la resolución de sistemas con EEMs (Muñoz de la Peña y col 2003, Olivieri y col 2004, Nahorniak y col 2005, Gimenez y col 2005, Bosco y col 2006), aunque se informaron varias aplicaciones relativas a la determinación de diferentes analitos mediante métodos cromatográficos como GC-TOF-MS (TOF, del inglés, *time of flight*) (Sihna y col 2004, Fraga y Corley 2005, Sihna y col 2004, Comas y col 2004, Wiberg y Jacobsson 2004), LC-DAD (Braga y col 2007) y LC-IR (IR, del inglés, *infrared*) (István y col 2006). Sin embargo, debido a que PARAFAC requiere de datos de segundo orden que sigan estructuras trilineales, MCR-ALS (De Juan y col 2000) es el algoritmo que se aplica cada vez con más frecuencia (De Juan y col 2000, Comas y col 2004, Mas y col 2007, García y col

2007, Zhang y Li 2006, De Juan y Tauler 2003) y parece ser la alternativa de elección para resolver sistemas de gran complejidad.

La combinación de cuadrados mínimos bilineales con bilinealización residual (BLLS/RBL) (Linder y Sundberg 1998, Linder y Sundberg 2002, Goicoechea y Olivieri 2005), presentando la ventaja de segundo orden, también mostró ser útil en el análisis de muestras con matrices complejas, incluyendo datos obtenidos mediante LC-DAD (Braga y col 2007).

U-PLS/RBL (Olivieri 2005a, Escandar y col 2007) se presentó en el Capítulo 2 como una alternativa apropiada para procesar datos de segundo orden de fluorescencia, conservando la ventaja de segundo orden. Por otra parte, aunque su rendimiento no fue estudiado en el modelado de datos más complejos, existen razones para plantear la hipótesis de que este algoritmo es lo suficientemente robusto como para manejar convenientemente el tipo de datos que se analizan en la presente sección.

En este capítulo se presenta la primera aplicación de U-PLS/RBL a datos cromatográficos, con el objetivo de estudiar las potencialidades de la estructura de variable latente del algoritmo, especialmente en casos en donde, al menos en principio, puede proveer resultados comparables con el ya establecido MCR-ALS.

Se debe considerar que, aunque se cuente con la ventaja de segundo orden, pueden aparecer problemas al analizar muestras medioambientales cuando existen compuestos que coeluyen con el analito de interés generando interferencias muy significativas. Debido a esto, se informaron limitaciones en la resolución y cuantificación con MCR-ALS de las mismas mezclas de analitos en muestras de sedimentos y aguas residuales por LC-DAD (Peré Trepát y col 2004) y LC-MS (Peré Trepát y col 2005).

Una manera de aumentar el rendimiento cromatográfico es a través del pretratamiento de los datos. Como se mencionó en el Capítulo 3, la eliminación de la línea de base del cromatograma resulta crucial para reducir tanto la complejidad como el número de componentes no esperados (Daszykowski y Walczak 2006). Por otra parte, el empleo de estandarización para lograr la compensación del factor de recuperación que sería aplicado cuando la muestra es preconcentrada, mejora la calidad de las señales

cromatográficas de segundo orden y, en consecuencia, el rendimiento de la resolución y la cuantificación de analitos presentes en matrices complejas mediante los métodos multivariados.

4.2. OBJETIVOS

- Evaluar el rendimiento del algoritmo U-PLS/RBL en el procesamiento de datos cromatográficos de segundo orden sometidos a pretratamiento, correspondientes a TCs presentes en muestras complejas, en relación con MCR-ALS.

4.3. TEORÍA

4.3.1. U-PLS/RBL y MCR-ALS

Los fundamentos teóricos de los algoritmos U-PLS/RBL y MCR-ALS se presentaron en los Capítulos 2 y 3, respectivamente.

4.3.1.1. Cifras de mérito analíticas en U-PLS/RBL

La estimación de la sensibilidad en U-PLS/RBL fue discutida por Olivieri (Olivieri 2005) y presentada en el Capítulo 2 de la presente tesis.

El límite de detección (LOD) se define en analogía con la calibración univariada y puede ser estimado como:

$$\text{LOD} = 3,3 s_x / \text{SEN}_n \quad (4.1)$$

siendo s_x el ruido instrumental y SEN la sensibilidad.

4.3.1.2. Cifras de mérito analíticas en MCR-ALS

En MCR-ALS se propuso un enfoque simple para determinar las cifras de mérito basado en la recuperación de los perfiles de respuesta puros de los analitos, luego del proceso de resolución por curvas (Saurina y col 2001). Por lo tanto, la sensibilidad se define como la pendiente de la curva de calibración obtenida graficando las respuestas relativas versus la

concentración de los estándares. Por otra parte, s_x [ver Ecuación (4.1)] se calcula para una muestra a concentración cero teniendo en cuenta los errores de la pendiente y la ordenada al origen de la ecuación de calibración (Saurina y col 2001, Currie 1999).

4.3.2. CORRECCIÓN DE LA LÍNEA DE BASE

La eliminación de la línea de base del cromatograma se realizó a través del método de cuadrados mínimos asimétricos propuesto por Eilers (Eilers 2004), el cual consiste en la minimización de la función:

$$Q = \sum_i v_i (y_i - f_i)^2 + \lambda \sum_i (\Delta^2 f_i)^2 \quad (4.2)$$

en donde y es la señal experimental, f una tendencia del ruido (la aproximación de la línea de base), v pesos anteriores, λ es un parámetro positivo que pesa el segundo término y Δ es la derivada de f .

4.3.3. ESTANDARIZACIÓN DIRECTA SUCESIVA

La estandarización directa sucesiva, conocida como PDS (del inglés, *Piecewise Direct Standardization*), es una estrategia para estandarizar espectros, es decir, cambiarles la forma. Esto permite corregir cambios en la respuesta instrumental debidos a diferentes causas, entre ellas el pretratamiento de la muestra (Feudale y col 2002, Santiago Valverde y col 2006).

PDS relaciona las respuestas de muestras obtenidas en una situación A (X_A) con las respuestas obtenidas en una situación B (X_B), siendo descrita la relación mediante una matriz de transformación F, de acuerdo con:

$$X_A = X_B F \quad (4.3)$$

En el presente caso, la situación A corresponde a las señales LC-DAD medidas luego de la extracción en fase sólida (SPE) de los estándares preparados en agua milliQ y la situación B a las señales LC-DAD de los estándares preparados en solvente. La matriz F está compuesta por vectores

de regresión del tipo PLS, cada uno de los cuales se obtuvo haciendo sucesivamente una regresión entre datos de la variable A y sus correspondientes en la variable B, flanqueada por una cierta cantidad de variables vecinas. En PDS, se llama ventana a la cantidad de variables secundarias tenidas en cuenta en cada uno de esos cálculos locales, y tolerancia al valor umbral que permite determinar el número óptimo de factores para cada modelo local. Los vectores individuales son luego ensamblados para formar una matriz diagonal F, de acuerdo con:

$$F = \text{diag} (b_1^T, b_2^T, \dots, b_j^T, \dots, b_k^T) \quad (4.4)$$

donde k es el número de variables. Finalmente, la respuesta de cada muestra desconocida se puede estandarizar de acuerdo con:

$$x_s^T = x_0^T F \quad (4.5)$$

donde x_0 y x_s corresponden a los datos de la muestra antes y después de la estandarización.

Como se puede observar, la metodología expuesta es adecuada para señales de primer orden (vectores). En consecuencia, los datos de segundo orden se desdoblaron en vectores y se trataron como una concatenación de señales de primer orden independientes para poder aplicar el método.

4.4. PROGRAMAS

Todas las rutinas utilizadas en este trabajo fueron escritas en MATLAB 6.0 (MATLAB 6.0, 2000).

Las rutinas para la corrección de la línea de base son una adaptación para datos de segundo orden de las publicadas por Eilers (Eilers 2004) y fueron provistas por J. Braga (Unicamp, Brazil). La aplicación de PDS realizó a través de una rutina incluida en el PLS toolbox 3.5 (Wise y col 2005).

MCR-ALS se implementó a través de la interfaz gráfica provista por R. Tauler en su página <http://www.ub.edu/mcr/welcome.html> (Jaumot y col

2005). U-PLS/RBL se implementó a través rutinas escritas por el Dr. Olivieri incluidas en la interfaz gráfica MVC2, similar a la descrita para calibración multivariada de primer orden MVC1 (Olivieri y col 2004).

4.5. MATERIALES Y MÉTODOS

4.5.1. EQUIPAMIENTO

Debido a que este trabajo se comenzó en Almería (España) y se finalizó en Santa Fe (Argentina), se emplearon dos instrumentos cromatográficos diferentes. El primero consistió en un instrumento Waters (Milford, MA, USA), compuesto por una bomba cuaternaria, una válvula de inyección de seis puertos y un detector DAD, mientras que el segundo fue un instrumento Agilent 1100 Series (Agilent Technologies, Waldbronn, Alemania) compuesto por un degasificador, una bomba cuaternaria, un inyector automático y un detector DAD. La detección fotométrica se realizó en el rango de 230 - 400 nm con resolución espectral de 1,5 nm y período de integración de 0,4 s por espectro. Por lo tanto, se generaron matrices de datos de 1950 filas (tiempos de migración entre 3 y 16 min) y 114 columnas (número de longitudes de onda).

Las separaciones mediante cromatografía líquida se realizaron en una columna (150 × 4,6mm) Aquasil C18 de tamaño de partícula 5 µm.

Como sistema de preconcentración se utilizó, en Almería, un sistema de vacío Büchi Vac V-500 y un controlador de vacío V-800 (Suiza) conectado a un dispositivo de extracción Waters (Miniford, USA) y, en Santa Fe, un dispositivo fabricado en el laboratorio.

4.5.2. REACTIVOS

Los estándares analíticos (calidad peptanal) de tetraciclina clorhidrato (TC; 97,3%), oxitetraciclina clorhidrato (OTC; 96,2%), clortetraciclina clorhidrato (CTC; 85,7%), demeclociclina clorhidrato hemidrato (DMC; 92,8%) y metaciclina clorhidrato (MTC; 99,6%) se obtuvieron de Riedel-de Haën (Alemania), doxicilina hclato (DOC, 98%) se obtuvo de Fluka (Suiza) y meclociclina sulfosalicilato (MEC, 98%) y minociclina clorhidrato (MIC, 90%) se

obtuvieron de Sigma-Aldrich (Alemania). Los solventes acetonitrilo y metanol de grado HPLC se obtuvieron de J.T. Baker (Holanda). Los ácidos oxálico 2-hidrato (99%) y fórmico (85%) se obtuvieron de Panreac (Barcelona, España).

Los componentes acuosos de la fase móvil se filtraron a través de membranas de acetato de celulosa de 0,45 μm y los orgánicos a través de membranas de politetrafluoroetileno (PTFE).

Los cartuchos Oasis MAX 6 cm^3 que contenían 0,5 g de fase estacionaria se obtuvieron de Waters (Milford, MA, USA).

4.5.3. RECOLECCIÓN Y PREPARACIÓN DE LA MUESTRA

Las muestras de efluente final de agua residual se recolectaron de una planta de tratamiento de agua residual en Almería (España). Antes del análisis se centrifugaron a $5000 \times g$, se filtraron a través de membranas de 0,45 μm y se almacenaron a 4°C en heladera. El tiempo máximo de almacenamiento fue de 2 días para minimizar la degradación microbiana.

4.5.4. PREPARACIÓN DE ESTÁNDARES Y MUESTRAS ADICIONADAS

Se prepararon soluciones individuales de estándares analíticos de tetraciclinas ($200,0 \text{ mg L}^{-1}$) pesando exactamente y disolviendo los compuestos correspondientes en metanol. Las soluciones estándares se protegieron de la luz y se almacenaron a 4 °C en heladera. En estas condiciones resultaron estables por al menos 3 meses.

Las soluciones estándares de trabajo se prepararon diariamente mediante dilución de una alícuota apropiada de cada solución estándar en ácido fórmico 1% (v/v) y se filtraron a través de membranas de PTFE de 0,45 μm antes de ser inyectadas en el sistema cromatográfico.

Para realizar la calibración se preparó un juego de siete muestras, cada una por triplicado, conteniendo 0,05 - 0,10 - 0,20 - 0,40 - 0,60 - 0,80 y 1,00 mg L^{-1} de las ocho tetraciclinas disueltas en solvente puro.

Para evaluar el efecto del volumen de rotura se prepararon dos soluciones estándares, una conteniendo 0,50 mg L^{-1} de las ocho tetraciclinas disueltas en solvente puro y la otra conteniendo 2,00 $\mu\text{g L}^{-1}$ de las ocho tetraciclinas

disueltas en agua milliQ. Para estandarizar, se prepararon tres muestras en agua milliQ conteniendo 0,80 - 2,40 y 4,00 $\mu\text{g L}^{-1}$ de las ocho tetraciclinas.

El juego de validación consistió en seis muestras (M1-M6) y cuatro replicados (R1/4) de una muestra diferente (ver Tabla 4.1). Las muestras se prepararon adicionando volúmenes apropiados de estándares a 250,00 ml de agua residual tratada de la manera descrita arriba. A continuación se prepararon para la extracción mediante del agregado de 1 mL de Na_2EDTA 5%, 285 mg de ácido cítrico y ajuste del pH a 2,8 con HCl inmediatamente antes de la extracción.

Tabla 4.1. Composición de las muestras de validación obtenidas adicionando cantidades diferentes de tetraciclinas a un efluente doméstico de agua residual.

Muestra	Componente / $\mu\text{g L}^{-1}$							
	MIC/1	OTC/2	TC/3	DMC/4	CTC/5	MTC/6	DOC/7	MEC/8
M-1	0,40	2,00	4,00	0,80	0,40	0,80	0,40	3,20
M-2	0,80	0,40	1,60	2,40	1,60	2,00	0,80	4,00
M-3	2,00	1,20	0,40	1,60	4,00	4,00	0,80	1,60
M-4	3,20	0,80	1,60	0,00	0,80	0,40	4,00	1,20
M-5	0,00	4,00	2,40	4,00	2,40	2,00	2,40	0,40
M-6	4,00	1,60	0,80	0,40	0,00	1,20	3,20	1,60
R-1/4	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00

4.5.5. PROCEDIMIENTOS

4.5.5.1. Procedimiento para preconcentrar tetraciclinas presentes en agua residual

Los cartuchos de extracción en fase sólida Oasis MAX se preconditionaron con 5 ml de metanol y 5 ml de agua MilliQ a pH = 2,8. A continuación se pasaron a través de ellos 250,00 ml de muestra acondicionada de la manera descrita, con una velocidad de flujo de 8 mL min^{-1} . Luego de la extracción, los cartuchos se lavaron con 15 mL de agua milliQ (pH = 2,8), 5 mL

de 5% NH₄OH y 10 mL de metanol, y se secaron mediante una corriente de aire durante 20 min. Los analitos retenidos en el cartucho de SPE se eluyeron con 1 mL de ácido fórmico 1% (v/v) seguido de 5 ml de metanol. Los extractos se concentraron a 1 mL bajo una corriente de N₂ y se filtraron a través de membranas de PTFE de 0,45 μm antes de ser inyectados en el sistema cromatográfico.

4.5.5.2. Procedimiento cromatográfico

Las ocho tetraciclinas se separaron en la columna cromatográfica siguiendo el método informado por Cinquina y col (Cinquina y col 2003) para la determinación de cuatro tetraciclinas (oxitetraciclina, tetraciclina, clortetraciclina y doxiciclina) por LC-DAD, el cual fue modificado para adecuarlo a un número mayor de analitos. Esta modificación consistió en implementar un programa de gradiente combinando velocidad de flujo y una fase móvil binaria (A:B, v/v), en lugar de utilizar una fase móvil isocrática a flujo constante. El solvente A fue ácido oxálico 0,01 mol L⁻¹ y el solvente B metanol:acetonitrilo (50:50, v/v). El programa de gradiente se detalla en la Tabla 4.2. La inyección consistió en 100 μL de estándares preparados en ácido fórmico 1% (v/v) o extractos de muestras obtenidos de la manera descrita arriba.

Tabla 4.2. Programa de gradiente usado para separar las ocho tetraciclinas.

Tiempo / min	Velocidad de flujo / mL min ⁻¹	A:B ^a
0	1,0	85:15
2	1,0	70:30
6	1,0	70:30
8	1,2	65:35
10	1,5	60:40
16	1,5	60:40
18	1,5	85:15
20	1,0	85:15

^a A: ácido oxálico 0,01 mol L⁻¹; B: metanol:acetonitrilo (50:50, v/v)

4.6. RESULTADOS Y DISCUSIÓN

4.6.1. ANÁLISIS CROMATOGRÁFICO

En la Figura 4.1 se observa el cromatograma correspondiente a una solución estándar mezcla conteniendo $2,00 \text{ mg L}^{-1}$ de las ocho tetraciclinas. En las condiciones cromatográficas seleccionadas cada corrida duró 20 minutos.

Los pasos de preconcentración no necesariamente son selectivos para los analitos de interés y pueden retener otros compuestos de características similares. En este caso, la gran complejidad del problema analítico bajo estudio se puede apreciar en la Figura 4.2, en donde se representan tres cromatogramas correspondientes a: a) una solución estándar mezcla compuesta por $0,50 \text{ mg L}^{-1}$ de las ocho tetraciclinas, b) un extracto de una muestra de agua residual y c) un extracto de la muestra de agua residual obtenido luego de adicionarle $2,00 \mu\text{g L}^{-1}$ de las ocho tetraciclinas, la cual es la concentración cercana a la esperada en agua residual (Yang y col 2005, Karthikeyan y Meyer 2006). Como se evidencia, en el paso de SPE se retienen un número grande de compuestos desconocidos, lo cual puede deberse a semejanzas entre sus polaridades y las de los analitos de interés. Este hecho causa importantes solapamientos entre los picos de los analitos y las interferencias, así como derivas severas en la línea de base. Debido a esto resulta imposible realizar la identificación de las tetraciclinas o la cuantificación cromatográfica mediante calibración univariada, aún estando convenientemente separadas.

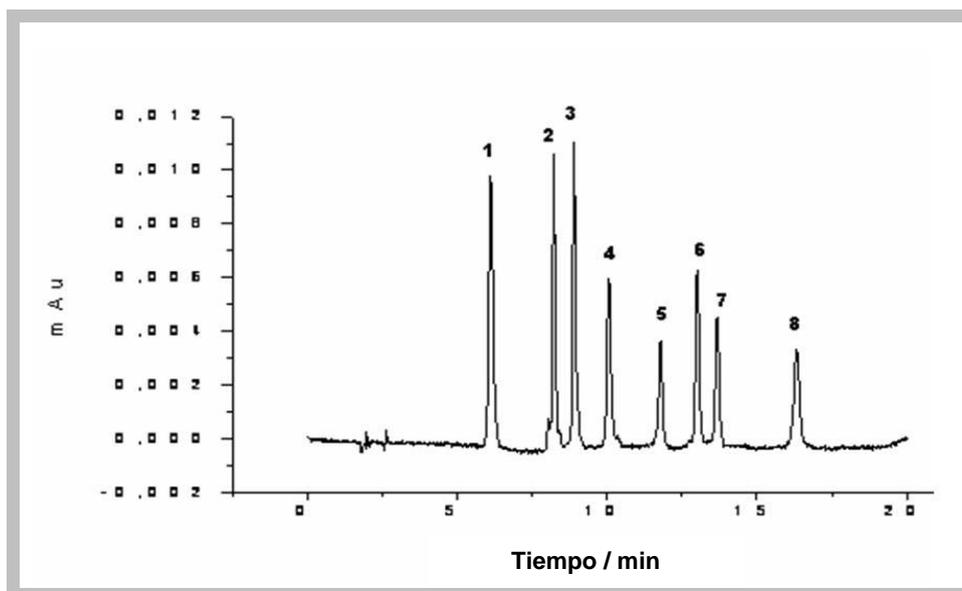


Figura 4.1. Cromatograma correspondiente a una solución estándar mezcla con $2,00 \text{ mg L}^{-1}$ de (1) MIC, (2) OTC, (3) TC, (4) DMC, (5) CTC, (6) MTC, (7) DOC y (8) MEC. La detección se realizó a 360 nm .

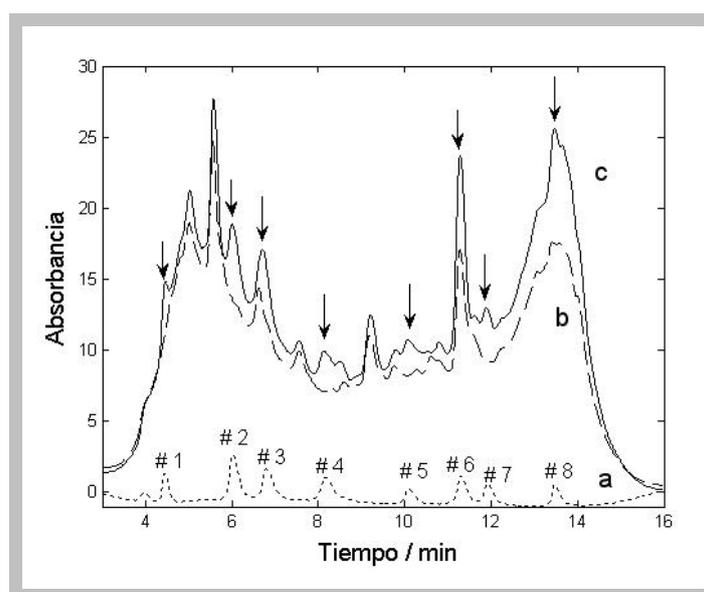


Figura 4.2. Cromatogramas ($\lambda = 360 \text{ nm}$) correspondientes a: (a) una solución estándar conteniendo $0,50 \text{ mg L}^{-1}$ de las ocho tetraciclinas, (b) un extracto de una muestra real y (c) un extracto de la misma muestra obtenido luego de adicionarla con $2,00 \mu\text{g L}^{-1}$ de las ocho tetraciclinas. Los números están asociados a los nombres en la Tabla 4.1.

4.6.2. PRETRATAMIENTO DE LOS DATOS

Las matrices de datos obtenidas en este trabajo fueron sometidas a dos pretratamientos antes de poder ser empleadas para predecir los analitos de interés de manera adecuada.

Como primera medida se realizó una estandarización para que las muestras normales, es decir, las soluciones estándares sin preconcentrar, adquirieran las características de las muestras preconcentradas. La razón de su aplicación fue la necesidad de corregir el efecto del volumen de rotura observado para los analitos más polares luego del paso de SPE (Feudale y col 2002, Santiago Valverde y col 2006). La Figura 4.3 muestra como los primeros cinco compuestos se pierden parcialmente durante el paso de preconcentración, apareciendo un nuevo pico a aproximadamente 4 min.

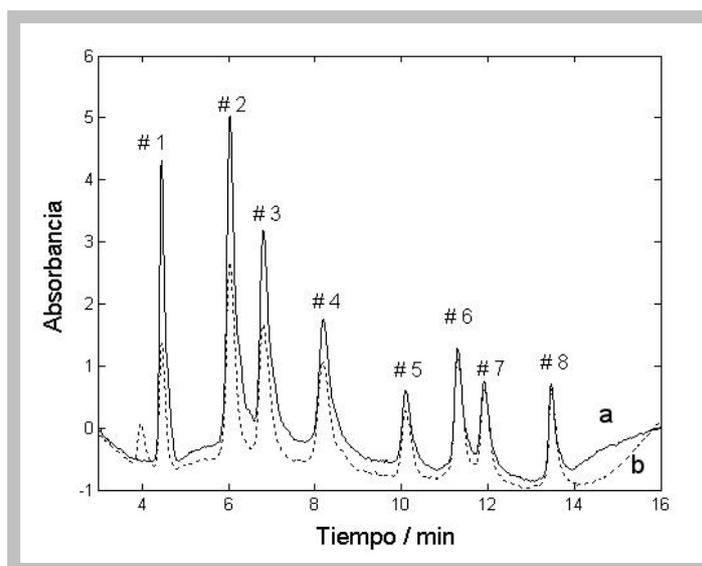


Figura 4.3. Cromatogramas ($\lambda = 360 \text{ nm}$) correspondientes a (a) una solución estándar conteniendo $0,50 \text{ mg L}^{-1}$ de las ocho tetraciclinas y (b) una solución estándar conteniendo $2,00 \text{ µg L}^{-1}$ de las ocho tetraciclinas luego del procedimiento de SPE.

Una estrategia para solucionar este problema podría haber sido someter los estándares analíticos al paso de SPE. Sin embargo, esta opción requiere de tiempo y costos adicionales. Para resolver este problema de manera eficiente se recurrió al uso de modelos de transferencia, en donde se necesitan solo un

número reducido del total de las muestras de calibración en la situación real, para construir un modelo que permita predecir nuevas muestras.

La estandarización se realizó mediante PDS (Yang y col 1991), utilizando los datos cromatográficos correspondientes a tres estándares preparados en solvente puro, con concentraciones que se consideraron representativas del rango de concentración, y tres estándares diluidos preparados en 250,00 mL de agua milliQ, que luego de ser sometidos a la preconcentración deberían tener las mismas concentraciones que los estándares preparados en solvente. Con estos datos se halló una matriz de transferencia que permitió convertir las señales originadas por las 21 muestras de estándares preparados en solvente hacia la nueva situación. Estas señales estandarizadas se utilizaron para predecir las muestras incógnitas con los modelos multivariados construidos. En otras palabras, las señales correspondientes a los estándares preparados en solvente puro se transfirieron a las señales obtenidas luego de la SPE, y las muestras reales se predijeron con los modelos construidos con los primeros luego de ser transferidos. Los parámetros óptimos empleados en la estandarización fueron: ventana igual a 3 y tolerancia igual a 0,01.

El impacto del pretratamiento se puede apreciar en la Figura 4.4. Esta figura muestra tres cromatogramas (registrados a 360 nm) restringidos al primer pico (3,80 - 5,00 min) correspondientes a: (a) un extracto de SPE de 250,00 mL de una muestra de agua milliQ adicionada con $2,00 \mu\text{g L}^{-1}$ de minociclina, (b) una solución estándar conteniendo $0,50 \text{ mg L}^{-1}$ de minociclina y (c) a la última señal luego de la estandarización.

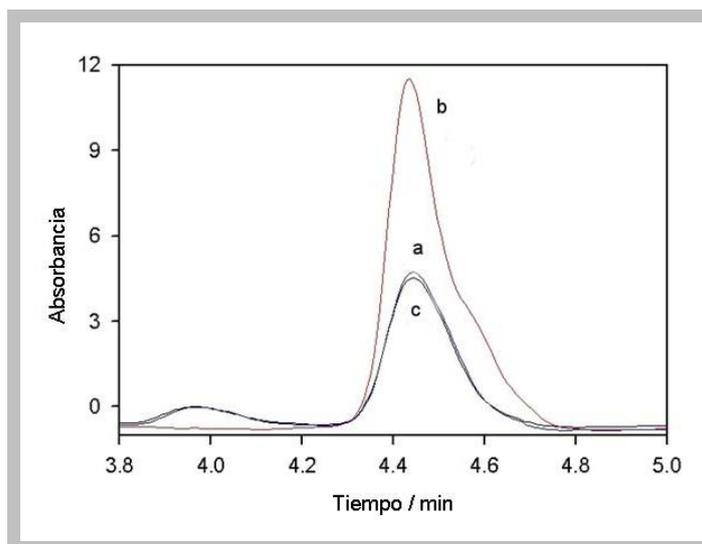


Figura 4.4. Cromatogramas ($\lambda = 360$ nm) restringidos al primer pico correspondientes a (a) un extracto SPE de 250,00 mL de una muestra de agua milliQ adicionada con $2,00 \mu\text{g L}^{-1}$ de minociclina, (b) una solución estándar conteniendo $0,50 \text{mg L}^{-1}$ de minociclina y (c) la última señal luego de la estandarización.

Por otra parte, dos problemas típicos en el análisis cromatográfico son la línea de base y el corrimiento de picos, siendo su eliminación de gran importancia para la calidad analítica de las predicciones (Daszykowski y Walczak 2006). Cabe destacar que para MCR-ALS el corrimiento de los picos no debería ser un problema debido a que este algoritmo no requiere que los datos sean perfectamente trilineales (Comas y col 2004).

La Figura 4.5 muestra el efecto de la corrección de la línea de base para minociclina (pico N°1). Como se puede observar, la eliminación de la línea de base permite trabajar en una región en donde las interferencias casi han desaparecido (entre 4,2 y 4,8 min) y solo se evidencia un cambio en el tiempo cuando se compara la señal correspondiente a un estándar con la de la muestra adicionada luego de la corrección. Es importante aclarar que no todos los picos fueron corregidos de esta manera tan categórica, sin embargo el impacto de las interferencias fue disminuido de manera significativa permitiendo la resolución de los ocho analitos de manera razonable.

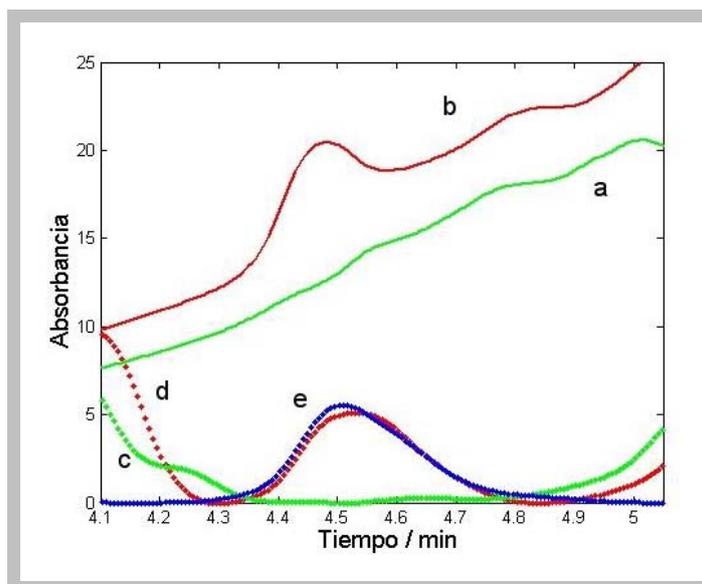


Figura 4.5. Cromatogramas originales ($\lambda = 360 \text{ nm}$) en líneas sólidas correspondiente a (a) un extracto de la muestra real y (b) este extracto luego de ser adicionado con $0,50 \text{ mg L}^{-1}$ de minociclina (pico N° 1). (c y d) son los correspondientes cromatogramas corregidos en líneas de puntos y (e) el perfil temporal correspondiente a un estándar de minociclina $0,50 \text{ mg L}^{-1}$.

4.6.3. RESOLUCIÓN

4.6.3.1. Cuantificación de tetraciclinas en agua residual por MCR-ALS y U-PLS/RBL

La Figura 4.6 muestra el cromatograma registrado a múltiples longitudes de onda correspondiente a una solución estándar mezcla conteniendo $0,50 \text{ mg L}^{-1}$ de cada tetraciclina.

Con el objetivo de simplificar los modelos y en consecuencia el análisis, los datos cromatográficos totales registrados para cada muestra se dividieron en ocho regiones, cada una de ellas correspondiente a cada analito. De esta manera se obtuvo un modelo para cada tetraciclina formado por un solo analito y las interferencias correspondientes. La Tabla 4.3 muestra las regiones espectrales y temporales empleadas.

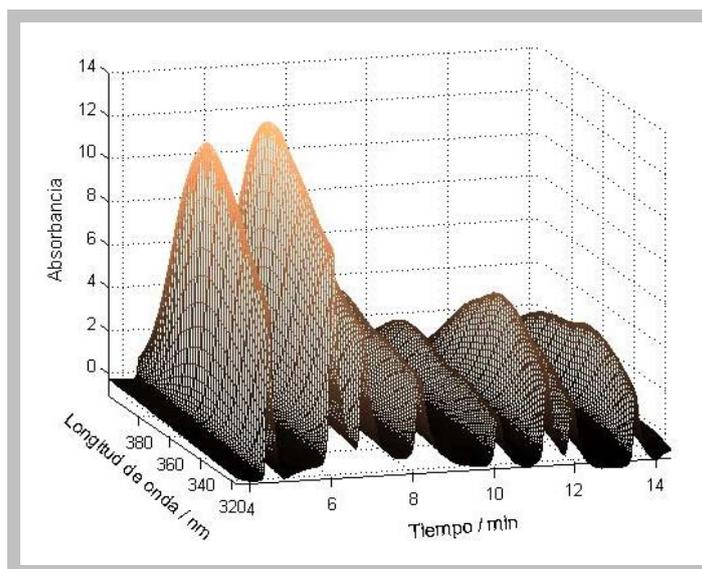


Figura 4.6. Cromatograma registrado a múltiples longitudes de ondas correspondiente a una solución estándar mezcla conteniendo $0,50 \text{ mg L}^{-1}$ de cada tetraciclina.

La complejidad del sistema bajo estudio se evidencia observando las Figuras 4.7 (A-D), las cuales muestran los datos de segundo orden correspondientes a meclociclina (analito N° 8 en orden de elusión) en diferentes situaciones. La Figura 4.7 A corresponde a una solución estándar de $0,50 \text{ mg L}^{-1}$, mientras que la Figura 4.7 B corresponde a un extracto de una muestra real sin pretratamiento de la señal. Una comparación visual de estas dos figuras permite evaluar la enorme complejidad del problema, debido a que un gran número de interferencias coeluyen con los analitos bajo estudio. En la Figura 4.7 C se observa el resultado de la corrección de la línea de base aplicada a los datos correspondientes a la muestra graficada en la Figura 4.7 B, en donde se evidencia una disminución de la señal debida a la matriz. Finalmente, la Figura 4.7 D corresponde a la muestra luego de la adición de $2,00 \mu\text{g L}^{-1}$ de meclociclina (luego del paso de preconcentración, la concentración del analito resulta $0,50 \text{ mg L}^{-1}$). Una inspección minuciosa de las Figuras 4.7 C y 4.7 D indica que, a pesar de la corrección de la línea de base, aparecen varios componentes no esperados en la muestra real que imposibilitan la predicción mediante calibración univariada.

Tabla 4.3. Regiones en las cuales los datos cromatográficos totales se dividieron para construir modelos quimiométricos individuales para el análisis mediante MCR-ALS y U-PLS/RBL, con los correspondientes factores y componentes no esperados empleados.

Analito/ N° de pico	Región temporal N° de sensor / min	Región espectral N° de sensor / nm	Factores de MCR-ALS ^a	M_{noesp} de U-PLS/RBL ^b
MIN / 1	201-280	1-56	3/1	3/1
	4,3562-4,8628	272-354		
OTC / 2	392-538	45-105	3/2	4/3
	5,6095-6,5828	285-354		
TC / 3	568-628	60-86	3/3	4/2
	6,7828-7,1828	361-400		
DMC / 4	752-812	1-86	2/1	3/1
	8,0095-8,4095	272-400		
CTC / 5	1037-1107	70-86	2/1	3/1
	9,9095-10,3762	375-400		
MTC / 6	1220-1280	15-86	4/4	3/2
	11,1295-11,5295	293-400		
DOC / 7	1321-1361	40-70	5/3	4/3
	11,8028-12,0695	330-376		
MEC / 8	1538-1658	1-86	5/4	5/5
	13,2495-14,0495	272-400		

^a Número de factores de MCR-ALS antes y después de pretratamiento de los datos.

^b Número de componentes no esperados de U-PLS/RBL antes y después de pretratamiento de los datos. El número de factores para el modelo U-PLS/RBL fue 1 para las ocho TCs.

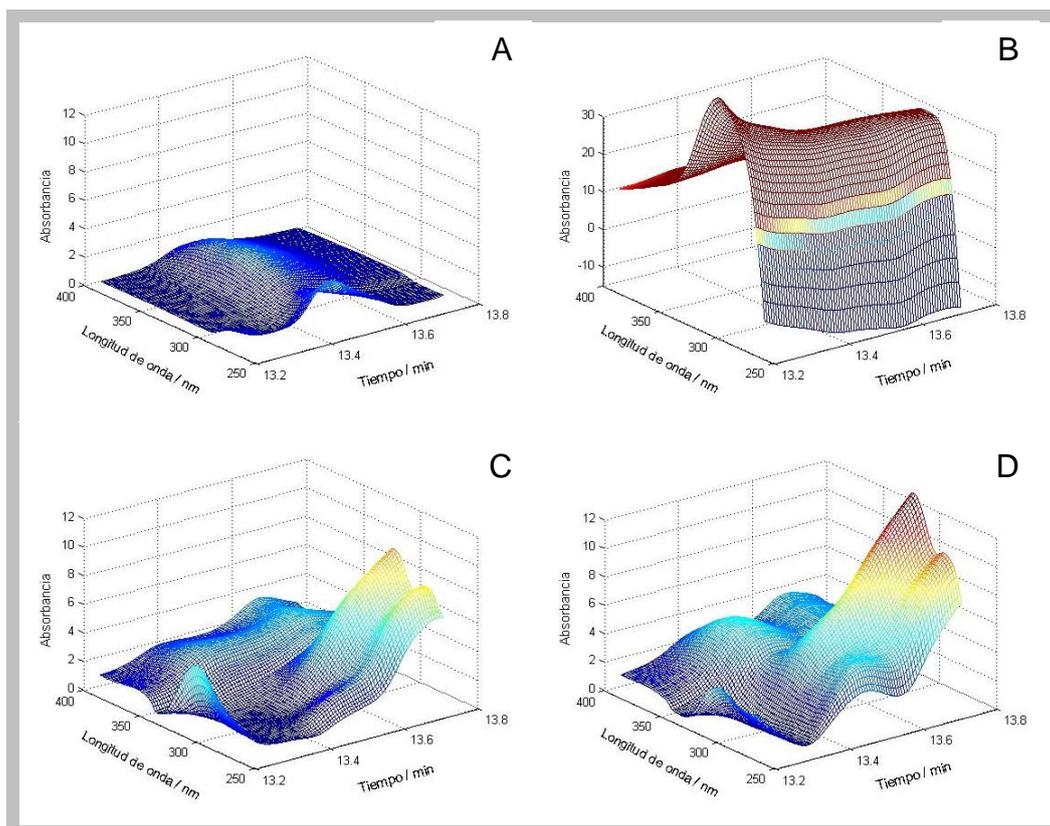


Figura 4.7. Datos de segundo orden correspondientes al pico N° 8 (meclociclina) registrado para (A) una solución estándar con $0,50 \text{ mg L}^{-1}$, (B) una muestra real sin el analito antes del procedimiento de pretratamiento, (C) un extracto de muestra real sin el analito luego del paso de pretratamiento y (D) este extracto de muestra con $0,50 \text{ mg L}^{-1}$ de meclociclina.

Para aplicar MCR-ALS se construyeron modelos individuales formados por matrices D aumentadas con las 21 muestras de calibración y la muestra incógnita, en el modo temporal. De esta manera se aprovechó la diferencia observada entre los espectros de los analitos y las interferencias para resolver las muestras de manera adecuada. Antes de empezar la resolución, se determinó el número de contribuciones presentes en cada matriz de datos aplicando SVD. Luego, se construyeron las estimaciones iniciales de los espectros mediante SIMPLISMA (Windig y Guilment 1991). Datos D y S^T , se aplicaron restricciones apropiadas (no-negatividad, unimodalidad y trilinealidad) con el objetivo de obtener la solución correcta.

En el caso de U-PLS/RBL, el número de variables latentes de calibración para cada analito se estimó mediante el método de validación cruzada (Haaland y Thomas 1988). Por otra parte, tuvo que ser considerada la presencia de componentes no esperados con el propósito de disminuir los

residuos obtenidos en la predicción de las muestras de validación, hasta estabilizarlos a un valor compatible con el ruido instrumental. La Figura 4.8 muestra la variación de los residuos de la predicción s_u en función de los valores de prueba de M_{noesp} para uno de los replicados de las muestras (R-1) cuando se analiza el pico N°2 (oxitetraciclina, tiempo de retención = 6,03 min). En comparación con el ruido instrumental (aproximadamente 0,16 unidades de absorbancia en este sistema), se evidencia que $M_{noesp} = 1$ es la elección correcta.

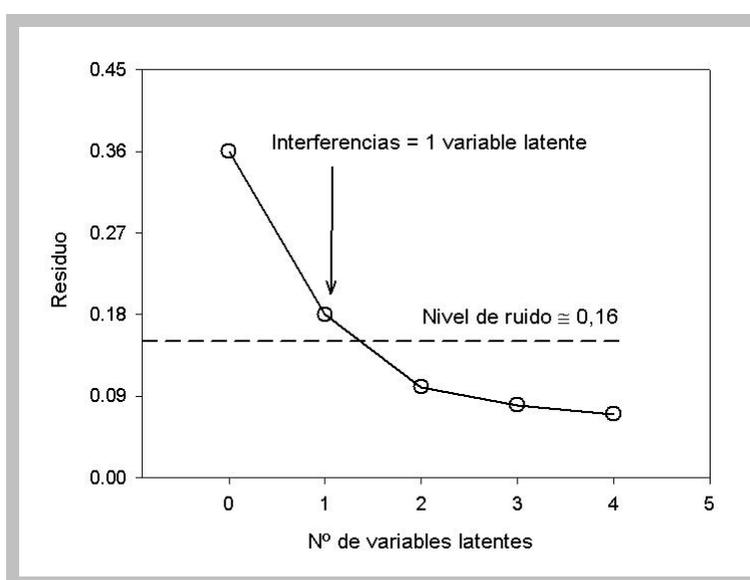


Figura 4.8. Variación de los residuos de predicción en función de los valores de prueba para la muestra replicado R-1 cuando se analiza el pico N°2 (oxitetraciclina, tiempo de retención = 6,03 min).

El número de componentes no esperados o de factores que contribuyen a cada matriz D depende tanto del número de interferencias retenidas en el procedimiento de extracción en fase sólida, como del preprocesamiento exitoso de corrección de la línea de base. La Tabla 4.3 muestra el número de factores y el número de componentes no esperados que debieron ser considerados en MCR-ALS y U-PLS/RBL, respectivamente, antes y después del pretratamiento de la señal. Como se observa, la complejidad de la mayoría de los picos se simplificó al mejorar las señales analíticas mediante los pretratamientos.

Para el pico N°2 se realizó un análisis comparativo del rendimiento de ambos algoritmos. La Figura 4.9 muestra los perfiles cromatográficos correspondientes al analito y a la interferencia extraídos mediante MCR-ALS y el perfil de la interferencia extraído mediante U-PLS/RBL, a partir de la muestra desconocida R-1. Esta figura muestra, además, la correspondiente señal instrumental pura normalizada. Como se puede observar, se consideraron dos componentes en el análisis mediante MCR-ALS (ambos graficados en líneas sólidas). La concordancia entre la línea sólida, la cual representa el perfil cromatográfico correspondiente al analito y la línea de círculos, la cual representa el componente puro normalizado a 360 nm, es un indicador de que el perfil extraído corresponde al analito de interés. Este análisis permite comparar el perfil extraído por MCR-ALS para la interferencia con el extraído mediante el procedimiento de RBL (señalados con las letras b y c en la Figura 4.9, respectivamente). La Tabla 4.3 muestra el número de componentes usados para construir todos los modelos MCR-ALS. Puede notarse que este número se encuentra comprendido en el rango de 1, el sistema más simple analizado en la Figura 4.5 (analito N°1), a 4. Se pueden realizar consideraciones similares analizando la Figura 4.10, en la cual se representan los perfiles espectrales extraídos mediante MCR-ALS y U-PLS/RBL y los espectros puros normalizados del analito N°2.

Mediante la inspección de las Figuras 4.9 y 4.10, se puede realizar una observación con respecto a la habilidad de RBL para extraer las contribuciones no esperadas en los perfiles temporales y espectrales, los cuales fueron marcados con la letra c. Estos son similares a los perfiles extraídos por MCR-ALS, los cuales se marcaron con la letra b en las figuras mencionadas, aunque se puede observar un cambio entre los perfiles de los componentes no esperados extraídos por ambos algoritmos. Este hecho confirma que, en este caso, la metodología U-PLS/RBL es capaz de obtener la ventaja de segundo orden a través del aislamiento correcto de la contribución del analito de la correspondiente al componente no esperado, esté o no presente en el juego de calibración.

Se debe remarcar que este análisis de comparación de perfiles extraídos por ambos algoritmos no se puede realizar para más de una interferencia.

Como fue mencionado anteriormente en el Capítulo 2, los siguientes perfiles obtenidos mediante U-PLS/RBL no reflejan los perfiles verdaderos para los componentes no esperados. Esta es una de las principales desventajas cuando se compara U-PLS/RBL con MCR-ALS.

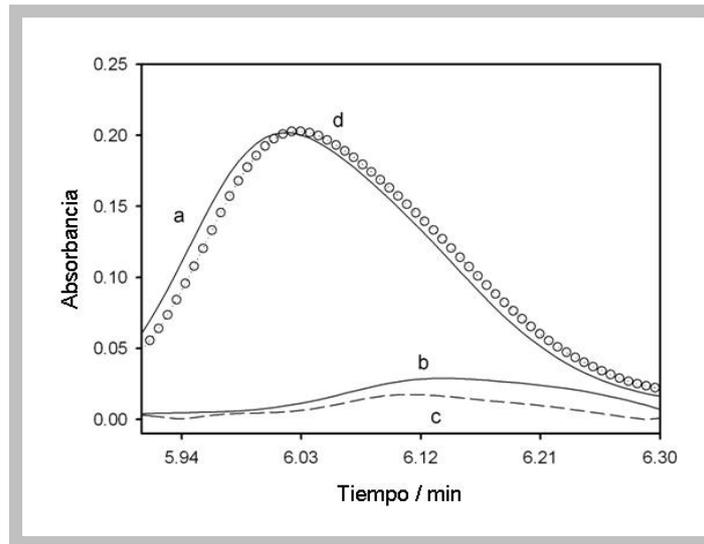


Figura 4.9. Perfiles temporales de MCR-ALS para el pico N°2: (a) perfil de concentración para oxitetraciclina y (b) perfil de tiempo para la interferencia. (c) Perfil temporal de U-PLS/RBL para la interferencia presente en el pico N°2. (d) Cromatograma del componente puro normalizado para oxitetraciclina.

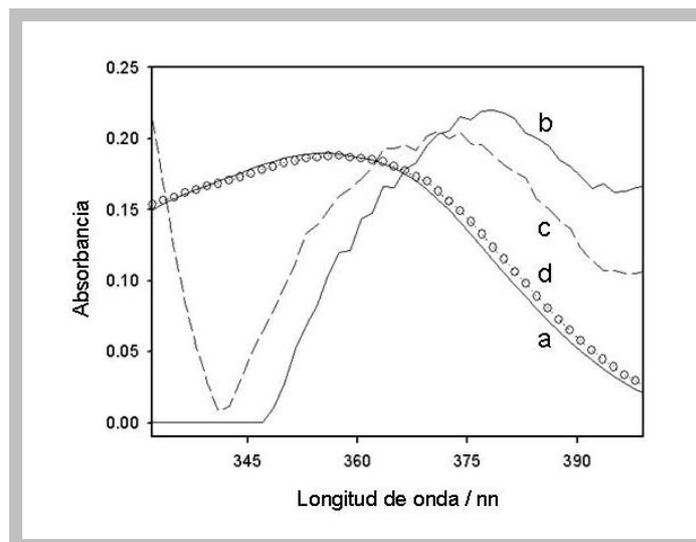


Figura 4.10. Perfiles espectrales de MCR-ALS para el pico N°2: (a) perfil espectral para oxitetraciclina y (b) perfil espectral para la interferencia. (c) Perfil espectral de U-PLS/RBL para la interferencia presente en el pico N°2. (d) Espectro normalizado del componente puro para oxitetraciclina.

La Figura 4.11 muestra los perfiles temporales sucesivos correspondientes al análisis por MCR-ALS del pico N°8 (meclociclina) de la muestra R-1. En el primer recuadro se representa la muestra desconocida para la cual se consideraron cuatro componentes (todos graficados en líneas sólidas). Otra vez, la relación entre la línea sólida roja y la línea de círculos roja, la cual representa el componente puro normalizado del cromatograma a 360 nm, indica que este perfil corresponde al analito de interés. Los restantes cinco recuadros corresponden a los perfiles extraídos de cinco muestras estándares con 0,10 - 0,20 - 0,40 - 0,60 y 1,00 mg L⁻¹, respectivamente. La calibración pseudounivariada empleando las áreas relativas permitió obtener la concentración de las muestras desconocidas.

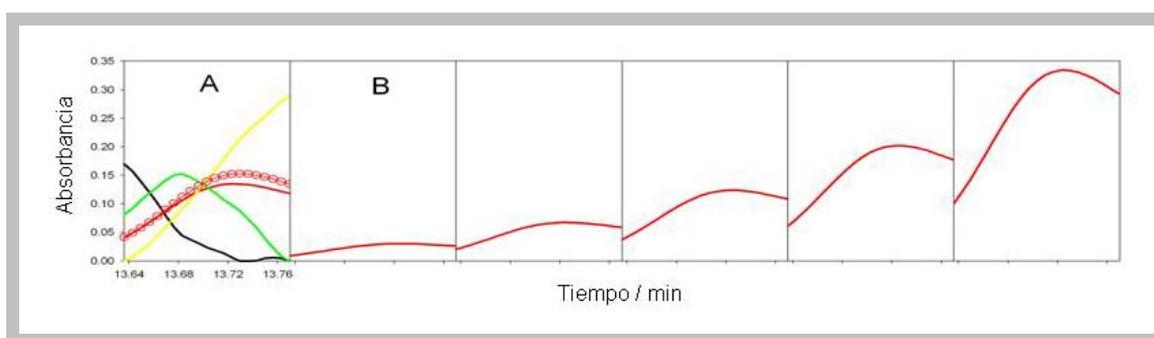


Figura 4.11. (A) Perfiles de tiempo sucesivos correspondientes a cuatro componentes (el analito y las interferencias) encontrados mediante el análisis por MCR-ALS para el pico N°8 (meclociclina) en la muestra R-1. La línea roja sólida corresponde al perfil de meclociclina, mientras que la línea roja de círculos representa el componente puro normalizado del componente a 404 nm. (B) Los restantes cinco perfiles corresponden a cinco estándares con 0,10 - 0,20 - 0,40 - 0,60 y 1,00 mg L⁻¹ de meclociclina, respectivamente.

Las predicciones obtenidas cuando se analiza la muestra real sin y con adición de antibióticos se presentan en la Tabla 4.4 para MCR-ALS y en la Tabla 4.5 para U-PLS/RBL. Como se puede observar, a pesar de la complejidad de la muestra analizada, se obtienen resultados aceptables para la mayoría de las muestras y los analitos. Se obtuvieron recuperaciones notablemente bajas solo para el analito N°1 (minociclina) por ambos métodos. El hecho de que para este analito ambas metodologías implementadas arrojen resultados comparables habla de la posibilidad de errores en la preparación

de la muestra, en su análisis o en la estabilidad del analito. Para las restantes TCs, los resultados obtenidos mediante U-PLS/RBL parecen ser ligeramente mejores que los obtenidos por MCR-ALS, siendo el número de predicciones excesivamente menores (< 75%) mayor en MCR-ALS. En la Figura 4.12 se presentan dos gráficos de barras tridimensionales en donde se pueden comparar visualmente las recuperaciones obtenidas para cada tetraciclina, en las muestras de validación M-1 a M-6, mediante MCR-ALS y U-PLS/RBL.

Tabla 4.4. Predicciones mediante MCR-ALS para las muestras de validación obtenidas mediante adición de diferentes cantidades de tetraciclinas a un efluente doméstico de agua residual^a.

Muestra	Componente / $\mu\text{g L}^{-1}$							
	MIC	OTC	TC	DMC	CTC	MTC	DOC	MEC
Basal	-0,04	-0,06	0,02	0,15	0,04	0,06	0,30	0,03
M-1	0,68 (170)	0,64 (32)	2,60 (65)	0,76 (95)	0,44 (110)	1,04 (130)	0,47 (117)	0,48 (15)
M-2	-0,08 (-)	-0,12 (-)	1,64 (102)	2,16 (90)	1,04 (65)	1,20 (60)	0,79 (99)	0,96 (24)
M-3	0,76 (38)	0,64 (53)	1,12 (93)	1,32 (82)	2,84 (71)	3,24 (81)	0,16 (80)	1,48 (92)
M-4	0,24 (7)	0,20 (23)	1,16 (72)	0,16 (-)	0,72 (90)	0,56 (140)	2,92 (73)	1,56 (130)
M-5	0,54 (-)	2,04 (51)	1,92 (80)	2,56 (64)	1,40 (58)	1,44 (72)	1,72 (72)	1,08 (90)
M-6	0,92 (23)	0,84 (52)	0,64 (80)	0,64 (160)	0,04 (-)	0,92 (77)	2,20 (69)	1,12 (70)
R-1/4	2,05 (103)	1,89 (95)	2,39 (120)	1,93 (97)	1,85 (93)	1,09 (55)	1,72 (86)	2,25 (112)
Desviación estándar ^b	0,38	0,38	0,44	0,13	0,10	0,11	0,24	0,24

^a Los valores entre paréntesis corresponden a % de recuperación de la cantidad adicionada.

^b Desviación estándar para los cuatro replicados (R-1 a R-4).

Tabla 4.5. Predicciones mediante U-PLS/RBL para las muestras de validación obtenidas mediante adición de diferentes cantidades de tetraciclinas a un efluente doméstico de agua residual^a.

Muestra	Componente / $\mu\text{g L}^{-1}$							
	MIC	OTC	TC	DMC	CTC	MTC	DOC	MEC
Basal	-0,04	005	0,08	0,03	0,04	0,08	0,02	-0,04
M-1	0,70 (175)	1,20 (60)	1,28 (32)	0,60 (75)	0,52 (130)	1,52 (190)	0,40 (100)	3,08 (96)
M-2	-0,05 (-)	0,68 (170)	2,12 (132)	2,12 (88)	1,24 (78)	2,08 (104)	0,60 (75)	3,36 (84)
M-3	0,74 (37)	1,36 (113)	1,24 (103)	1,24 (78)	3,16 (79)	3,28 (82)	1,04 (130)	0,76 (47)
M-4	0,22 (7)	0,96 (120)	1,60 (100)	0,04 (-)	0,88 (110)	0,32 (80)	3,60 (90)	0,72 (60)
M-5	0,64 (-)	3,80 (95)	2,48 (103)	3,48 (87)	1,68 (70)	2,20 (110)	2,52 (105)	1,20 (100)
M-6	1,80 (45)	1,40 (87)	0,40 (50)	0,40 (100)	0,02 (-)	1,44 (120)	2,56 (80)	1,24 (78)
R-1/4	2,12 (106)	1,92 (96)	2,26 (113)	1,66 (83)	1,92 (96)	1,43 (71)	1,84 (92)	2,09 (105)
Desviación estándar ^b	0,25	0,17	0,12	0,12	0,17	0,21	0,28	0,21

^a Los valores entre paréntesis corresponden a % de recuperación de la cantidad adicionada.

^b Desviación estándar para los cuatro replicados (R-1 a R-4).

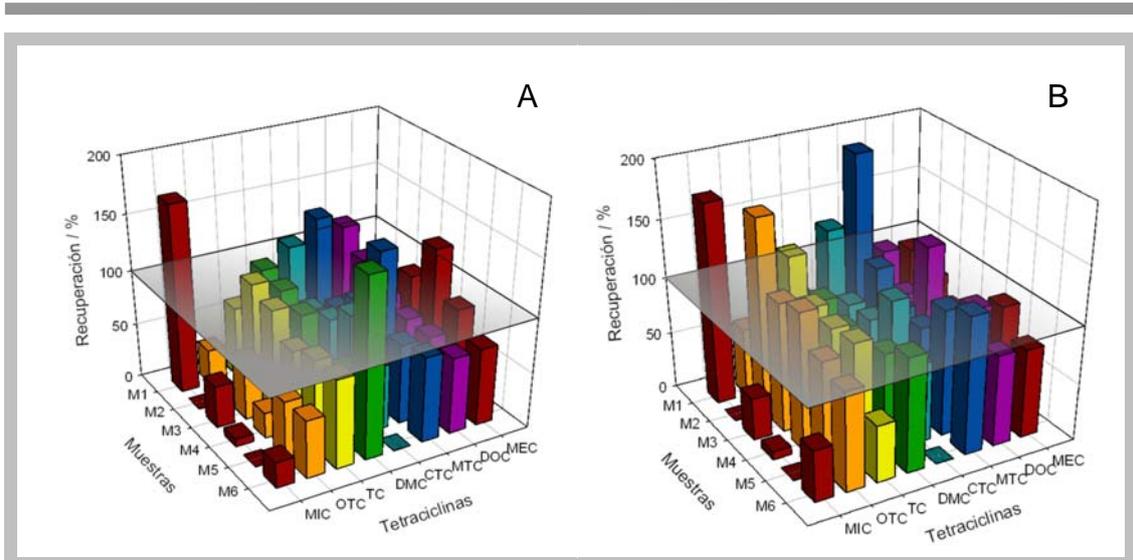


Figura 4.12. Gráficos de barras tridimensionales en donde se representan las recuperaciones obtenidas para cada tetraciclina, en las muestras de validación M-1 a M-6, mediante A) MCR-ALS y B) U-PLS/RBL.

Cuando se analizaron los cuatro replicados (muestras R-1/4) los resultados obtenidos por U-PLS/RBL resultaron más aceptables que los obtenidos por MCR-ALS para los analitos N°3, N°6, N°7 y N°8, mientras que las recuperaciones fueron comparables para los analitos N°1, N°2, N°4 y N°5 (ver Figura 4.13). Las desviaciones estándares fueron menores por U-PLS/RBL para los analitos N°1, N°2 y N°3 y ligeramente menores mediante MCR-ALS para los analitos N°5, N°6, N°7 y N°8.

Estos resultados muestran que U-PLS/RBL presenta un rendimiento comparable a MCR-ALS al resolver problemas analíticos altamente complejos. Sin embargo se necesitan estudios adicionales para dilucidar esta cuestión.

Los intentos de obtener predicciones sin pretratamientos de la señal produjeron resultados absolutamente inaceptables.

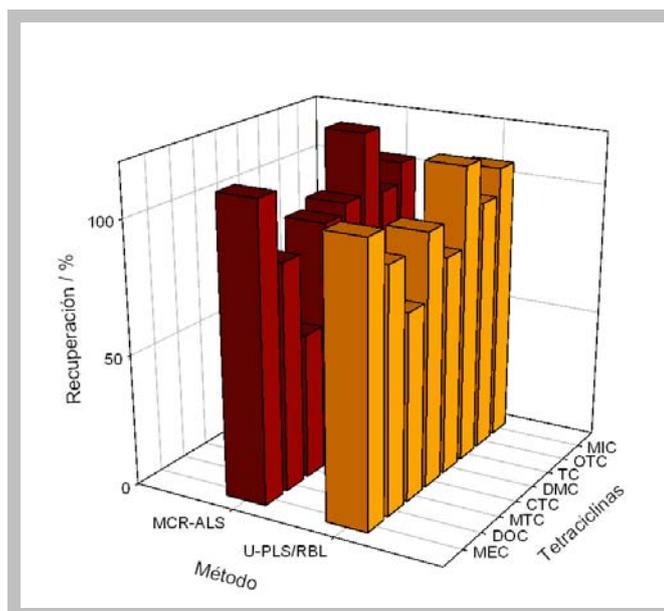


Figura 4.13. Gráfico de barras tridimensional en donde se representan las recuperaciones obtenidas para las muestras R1-4 mediante MCR-ALS y U-PLS/RBL.

4.6.3.2. Cifras de mérito analíticas multivariadas

La Tabla 4.6 muestra las cifras de mérito calculadas para los resultados obtenidos de la aplicación del algoritmo U-PLS/RBL al juego de datos. Esta última tabla muestra, además, las cifras de mérito correspondientes a la aplicación de MCR-ALS, calculadas a partir del gráfico de calibración pseudounivariado cuya explicación teórica se desarrolló en el Capítulo 3. Como se puede observar, debido a los diferentes esquemas para obtener tanto la ventaja de segundo orden como las cifras de mérito, se obtienen sensibilidades que en algunos casos son muy diferentes.

Tabla 4.6. Cifras de mérito para el modelado mediante MCR-ALS y U-PLS/RBL.

Analito /Nº de pico	Cifras de mérito					
	MCR-ALS			U-PLS/RBL		
	SEN	γ^{-1} / $\mu\text{g L}^{-1}$	LOD / $\mu\text{g L}^{-1}$	SEN	γ^{-1} / $\mu\text{g L}^{-1}$	LOD / $\mu\text{g L}^{-1}$
MIN / 1	3,5	0,02	0,23	12,1	0,03	0,09
OTC / 2	2,5	0,03	0,38	2,9	0,04	0,52
TC / 3	2,7	0,02	0,24	0,43	0,03	0,09
DMC / 4	2,4	0,03	0,39	0,56	0,05	0,14
CTC / 5	2,9	0,02	0,23	0,36	0,06	0,19
MTC / 6	2,5	0,03	0,36	0,13	0,06	0,18
DOC / 7	5,4	0,004	0,06	0,63	0,01	0,05
MEC / 8	4,0	0,005	0,10	56,2	0,01	0,06

4.7. CONCLUSIONES

Los datos cromatográficos con detección DAD se pueden analizar mediante U-PLS/RBL obteniendo la ventaja de segundo orden. Además, el uso de la estructura de variable latente permite manejar datos de gran complejidad, que presentan leve desviación a la trilinealidad, y generar resultados de predicción comparables a los obtenidos con MCR-ALS.

Una de las ventajas más importantes de MCR-ALS es que realiza el análisis generando información con significado físico. Sin embargo, involucra el manejo de un programa complejo y la introducción de las estimaciones iniciales de los espectros de los componentes o los perfiles temporales para alcanzar una descomposición satisfactoria. Por otra parte, U-PLS/RBL presentó un rendimiento analítico aceptable y una operación del programa rápida y fácil. A pesar de que no es posible una interpretación física del sistema, excepto cuando solo hay un componente no esperado, desde un punto de vista analítico y considerando otras posibilidades potencialidades, como implementación automatizada, U-PLS/RBL puede ser considerado como

un algoritmo alternativo para datos como los analizados en el presente estudio.

CAPÍTULO 5

U-PCA/RBL: una estrategia versátil para obtener la ventaja de segundo orden al aplicar diferentes redes neuronales artificiales a datos de segundo orden no lineales

5.1. INTRODUCCIÓN

La química analítica presenta una variedad de sistemas en donde los datos experimentales siguen una relación no lineal con respecto a la concentración de los componentes. Este fenómeno ocurre, por ejemplo, cuando existen interacciones significativas entre los componentes de la muestra (Blanco y col 1995, Goicoechea y col 2002). Además aparece en sistemas espectroscópicos-cinéticos, por ejemplo, cuando se monitoriza el producto de una reacción en donde el analito interviene como catalizador (Absalan y col 2001, Safavi y col 2001, Magni y col 2004, Boudart y col 1968), o cuando el producto de la reacción es el resultado de cinéticas de pseudoprimer orden con respecto al reactivo (Blanco y col 1996).

Intrínsecamente, los datos no lineales no pueden ser procesados con los métodos clásicos de calibración que hacen uso de modelos lineales. Es posible que en los casos anteriores se puedan encontrar regiones locales variables en las cuales la respuesta se comporta de manera lineal con respecto a la concentración de analito. Por otra parte, los algoritmos para datos lineales, que utilizan variables latentes, permiten estudiar componentes cuyos comportamientos se desvían ligeramente de la linealidad, introduciendo factores adicionales en el análisis (Despaigne y Massart 1998). Sin embargo, si se necesita explorar el conjunto de datos experimentales, hay que recurrir a alternativas adecuadas.

Muy recientemente, la atención se focalizó en la obtención de la ventaja de segundo orden a partir de información no lineal (Olivieri 2005b, García-Reiriz y col 2007). Para alcanzar este objetivo se necesita del acoplamiento de dos métodos separados que sean capaces de: 1) modelar los datos de calibración y validación de manera que la contribución de los componentes no esperados, que no están presentes en el juego de calibración, sea removida de las muestras de validación y 2) modelar la relación no lineal entre los datos de calibración y la concentración de analito, e interpolar los datos de validación preprocesados con el objetivo de realizar predicciones (Olivieri 2005a).

El primero de estos dos desafíos se puede alcanzar usando una combinación entre componentes principales desdoblados (U-PCA) y

bilinealización residual (RBL). El procedimiento de RBL se presentó en el Capítulo 2 como una herramienta útil para la obtención de la ventaja de segundo orden en U-PLS (Öhman y col 1990, Olivieri 2005a, Olivieri 2005b). Mientras que U-PCA modela los datos de calibración empleando un número limitado de componentes principales, RBL es útil para filtrar la contribución de los componentes no esperados en las muestras (Öhman y col 1990a). A pesar de que, en principio, otras técnicas de procesamiento pueden ser empleadas con este objetivo, se encontró que U-PCA es lo suficientemente flexible como para resolver casos en donde las señales de segundo orden no son bilineales.

Con el propósito de cumplir con el segundo objetivo se encuentran disponibles una variedad de técnicas de regresión no paramétricas. En general, se pueden considerar tres grupos de métodos para modelar satisfactoriamente los datos no lineales: 1) métodos basados en suavizadores (Friedman 1984), como expectativas condicionales alternantes (ACE), regresión aditiva múltiple suavizadora (SMART) y cuadrados mínimos parciales no lineales (NLPLS), 2) métodos basados en *splines* (Wold 1974, Wegman y Wright 1983), como árboles de clasificación y regresión (CART), *splines* de regresión multivariada adaptativa (MARS) y *Spline*-PLS y 3) redes neuronales artificiales (ANN) (Haykin 1999). Esta última es una metodología popular que tiene en cuenta no linealidades presentes en la relación entre los datos instrumentales y la concentración, y que puede ser subdividida en: 1) perceptrón multicapa (MLP) con retropropagación del error, 2) funciones de base radial (RBF) y 3) máquinas de vectores soportes (SVM). En la bibliografía se encuentran muy buenos trabajos de revisión sobre regresión no lineal (Frank 1995) y aplicación de redes MLP a problemas de calibración multivariada (Despaigne y Massart 1998).

La red MLP consiste en una serie de capas de neuronas o unidades de cómputo (Zupan y Gasteiger 1999, Despaigne y Massart 1998, Zupan y col 1997). Este algoritmo se basa en conceptos relacionados con el comportamiento del cerebro humano, estando la función matemática asociada a las conexiones intraneuronales. La capa de entrada recibe los datos y los transfiere a la capa siguiente luego de la aplicación de una función

de activación, generalmente de forma sigmoidea. Las neuronas en la capa oculta reciben estas salidas, las multiplican por pesos adecuados, las suman y le aplican la función sigmoidea antes de enviar los resultados a la capa siguiente. Este procedimiento continúa hasta que se alcanza la capa de salida. Entrenar una red MLP implica modificar los pesos a través del procedimiento de retropropagación del error, hasta que las salidas finales resulten comparables con valores de referencia de un dado juego de muestras de composición conocida. Existen algunos inconvenientes asociados con el proceso de entrenamiento: no hay garantía de convergencia y reproducibilidad en los resultados para un mismo juego de datos cuando son ejecutados varias veces empleando pesos iniciales diferentes. Por otra parte, MLP presenta la desventaja de que se requieren muchas muestras experimentales para ajustar los modelos, la falta de interpretación de los parámetros internos y la dificultad en generar nuevas hipótesis sobre los datos.

Recientemente, RBF y SVM se introdujeron como metodologías de redes neuronales promisorias, capaces de realizar estimación de funciones multivariadas no lineales y regresión no lineal (Haykin 1999). RBF se basa en una sola capa de neuronas ocultas que usan funciones de transferencia gaussianas y una capa de salida activada linealmente. En comparación con las redes MLP, RBF parece ofrecer algunas ventajas como robustez frente a datos ruidosos y procedimientos de entrenamiento más rápidos (Derks y col 1995). SVM, por otra parte, incluye la solución de un problema de programación cuadrática conduciendo a modelos generalmente únicos (Vapnik 1995). SVM se basa en el principio de minimización del riesgo estructural, el cual promete buena capacidad de generalización y excelente rendimiento en el caso de pocas muestras. Los cálculos computacionales de SVM fueron simplificados por Suykens y col a través de la implementación de la versión de cuadrados mínimos (LS-SVM) (Suykens y col 2002). Esta implica la resolución de un juego de ecuaciones lineales en vez del clásico problema cuadrático. Mientras que SVM estándar es más adecuada para problemas de clasificación, LS-SVM se emplea para regresión y estimación de funciones, y se relaciona mucho con las redes de regularización (Evgeniou y col 2000) y los procesos gaussianos

(Wahba 1990). Es relevante a la aplicación de la metodología LS-SVM, la función de activación entre nodos de redes; en este trabajo se utilizó la gaussiana de kernel o función de base radial.

Merecen ser citadas publicaciones recientes que describen la aplicación de RBF a datos espectrales multivariados de primer orden como, por ejemplo, el análisis mediante infrarrojo cercano de: materia orgánica en suelos (Fidêncio y col 2002), glucosa en sangre (Fischbacher y col 1997) y contenido de agua en productos de pescadería (Carlin y col 1994). En el caso de las aplicaciones de LS-SVM, se pueden citar el análisis mediante infrarrojo cercano de: adulterantes en leche en polvo (Borin y col 2006), acidez en uvas (Chauchard y col 2004) y propiedades físicas de madera (Cogdill y col 2004), y el análisis de Raman de masas de monómeros durante la polimerización (Thissen y col 2004a, Thissen y col 2004b).

En la literatura científica se han informado pocas aplicaciones de redes neuronales a datos de segundo orden no lineales: la determinación cinética-espectroscópica de pesticidas de carbamato (Ni y col 2004), la correlación de espectros magnéticos nucleares bidimensionales a la composición y las propiedades de muestras de combustibles (Vaananem y col 2002) y el monitoreo de procesos fermentativos (Lopes y Menezes 2004). Se sabe aún menos sobre la obtención de la ventaja de segundo orden a partir de información de órdenes mayores en presencia de componentes no esperados en la muestra.

En este trabajo se discuten dos sistemas experimentales en donde se producen desvíos de la relación lineal concentración-respuesta para los analitos calibrados. Además, las muestras están inmersas en matrices complejas que contienen componentes no contemplados en la calibración, por lo que se necesita de la ventaja de segundo orden para obtener resultados analíticos satisfactorios.

En el primer sistema, el Fe(II) (el analito) actúa como catalizador de la oxidación del colorante naranja de metilo con bromato de potasio siguiendo una cinética no lineal, en presencia del colorante púrpura de bromocresol como interferencia. Este sistema se resolvió empleando el algoritmo U-PCA/RBL y las redes MLP para modelar las no linealidades y, por otra parte, el

algoritmo U-PLS/RBL. Se realizó una comparación del rendimiento obtenido con ambos algoritmos.

El segundo sistema no lineal analizado consistió en una preparación farmacéutica conteniendo loratadina y pseudoefedrina sulfato. En una primera etapa este sistema se procesó mediante U-PCA/RBL seguido de las redes MLP (optimizando el número de neuronas de la capa de entrada) y se comparó el rendimiento con el obtenido mediante U-PLS/RBL y MCR-ALS. Si se considera que las redes MLP construyen modelos de aproximación global a relaciones de entrada-salida no lineales y las redes RBF y SVM, empleando funciones gaussianas, tienden a construir aproximaciones locales, resulta de interés investigar qué modelo (MLP, RBF o SVM) aproxima mejor las no linealidades presentes en este sistema. Por lo tanto, una vez filtradas las contribuciones de los componentes no esperados en las muestras mediante U-PCA/RBL, se modeló la relación *score*-concentración con las redes neuronales artificiales RBF y SVM. La comparación de las cifras de mérito de estos modelos, junto con las del modelo MLP, permite determinar cuál es la técnica de procesamiento no lineal más adecuada para este sistema.

5.2. OBJETIVOS

- Evaluar el algoritmo U-PCA/RBL para el procesamiento de datos instrumentales de segundo orden no lineales.
- Comparar los resultados de la predicción de U-PCA/RBL seguido de las redes MLP (U-PCA/RBL-MLP), con los obtenidos mediante la aplicación de algoritmos lineales (MCR-ALS y/o U-PLS/RBL), con el objetivo de corroborar que la primera estrategia ofrece mejor habilidad para modelar datos no lineales.
- Comparar los resultados de la predicción obtenidos mediante el empleo de las redes neuronales artificiales MLP, RBF y SVM, con el objetivo de establecer si existen diferencias en la habilidad de cada tipo de red para modelar sistemas no lineales.

5.3. TEORÍA

5.3.1. U-PCA/RBL

El análisis en componentes principales de datos desdoblados provee, en principio, un número de *scores* de entrenamiento y de validación que se pueden emplear para el entrenamiento y la predicción mediante ANN. Sin embargo, cuando aparecen componentes no esperados en las muestras de validación, sus *scores* no son adecuados para realizar la predicción empleando la ANN entrenada. En este caso, es necesario recurrir a una técnica que marque a la nueva muestra como anómala, indicando que se necesitan acciones adicionales antes de realizar la predicción mediante ANN, y luego aisle la contribución de los componentes no esperados de la de los analitos calibrados, con el objetivo de recalcular *scores* apropiados para las muestras de validación. En U-PCA la muestra será considerada como anómala si los residuos del análisis de los datos de validación desdoblados X_u [s_p , ver Ecuación (5.1)] son anormalmente grandes en comparación con el ruido instrumental típico:

$$\begin{aligned} s_p &= || e_p || / (JK-A)^{1/2} = || \text{vec}(X_u) - P P^T \text{vec}(X_u) || / (JK-A)^{1/2} = \\ &= || \text{vec}(X_u) - P t_u || / (JK-A)^{1/2} \end{aligned} \quad (5.1)$$

en donde $|| \cdot ||$ indica la norma euclídea, P es la matriz que contiene los primeros A *loadings* obtenidos de la aplicación de PCA a los datos de entrenamiento desdoblados, t_u es el vector de los *scores* de las muestras de validación y $\text{vec}()$ indica la operación de vectorización. Los tamaños de los arreglos involucrados en la Ecuación (5.1) son: e_p y $\text{vec}(X_u)$, $JK \times 1$; X_u , $J \times K$; P , $JK \times A$ y t_u , $A \times 1$, donde J y K son el número de sensores en cada dimensión. RBL se puede emplear para modelar la presencia de componentes no esperados en las muestras de validación mediante la descomposición sus señales en dos partes: una modelada usando las variables latentes de calibración (X_{mod}) y la restante que no puede ser modelada con estas variables (X_{nomod}). El tratamiento matemático es similar al descrito en el Capítulo 2, y comprende las ecuaciones (2.8) a (2.12)

Finalmente, una vez que se minimizó $\| e_u \|^2$ y se encontraron los *scores* t_u correctos de la muestra, el vector t_u final se introduce en las neuronas de entrada de la ANN entrenada y se obtiene como salida la concentración de analito. El proceso completo se muestra en la Figura 5.1.

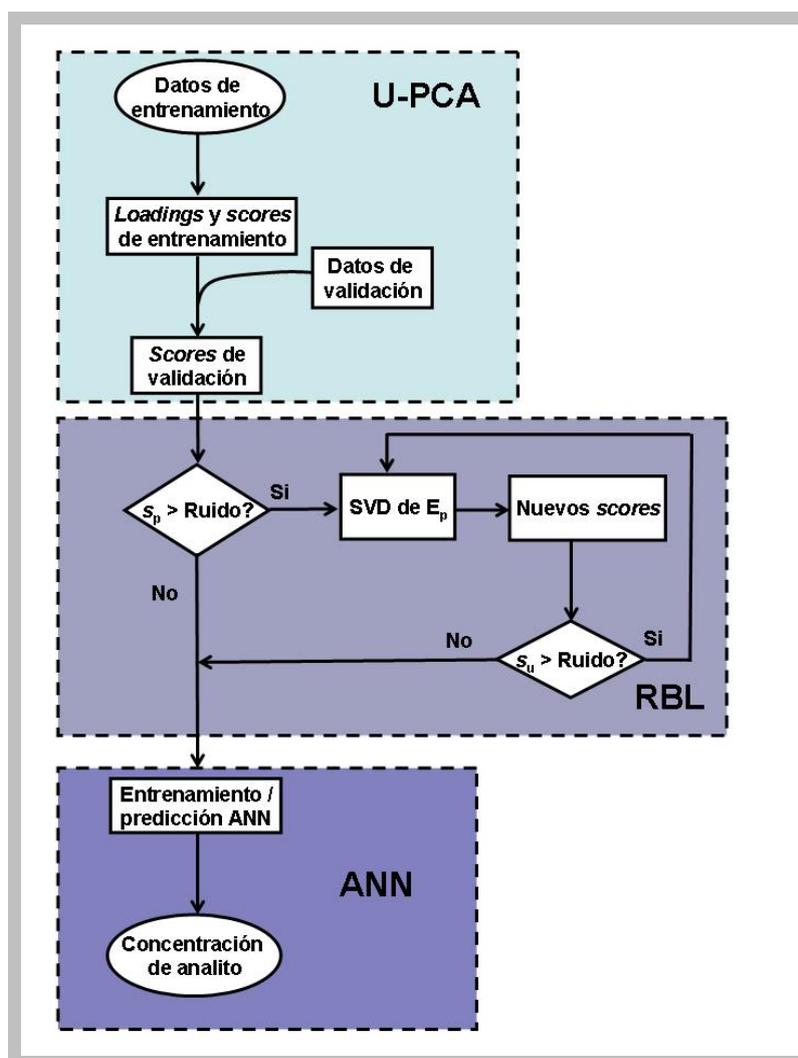


Figura 5.1. Gráfico que indica el modo en que se realiza la predicción de una muestra de validación que contiene componentes no esperados mediante U-PCA/RBL.

5.3.2. PERCEPTRONES MULTICAPA

Las redes MLP son las ANNs más comúnmente empleadas para calibración y su arquitectura consiste en tres capas con número variable de neuronas (Despaigne y Massart 1998). La primera, conocida como capa de entrada, se emplea para acomodar los datos de entrada. La segunda es una capa oculta y

su número se ajusta mediante prueba y error. La última capa esta formada por una sola neurona, la cual proporciona la concentración del analito de interés en cada muestra. Tanto la capa de entrada como la oculta incluyen además la llamada neurona umbral (*bias*), cuyos valores de entrada son iguales a 1. La práctica común es comprimir los datos de entrada en un número reducido (A) de componentes principales (PC) o *scores*. De este modo, el número de neuronas es igual a A , donde en general $A \ll JK$ (número de sensores luego del procedimiento de desdoblamiento, ver más adelante). El valor de A puede estimarse, por ejemplo, calculando el porcentaje de variancia explicada por los componentes principales de la matriz de datos de entrenamiento desdoblada (de tamaño $I \times JK$, I es el número de muestras de entrenamiento) y seleccionando los primeros A PCs que explican más de un cierto porcentaje de la variancia total (por ejemplo, 99%).

Una vez que los primeros A PCs se cargan dentro de las A neuronas de la capa de entrada, se calculan las salidas para cada muestra de entrenamiento empleando un juego de pesos iniciales seleccionados al azar que son transferidos de capa a capa a través de una función de transferencia adecuada (Zupan y Gasteiger 1999). En este caso se empleó la función de transferencia sigmoidea $f(x)=1/[1+\exp(-x)]$. Luego los pesos son modificados de acuerdo con la metodología de retropropagación del error, la cual compara las salidas de la red MLP con los valores nominales de la concentración de analito en las muestras de entrenamiento y monitorización. De esta comparación se obtienen dos cifras:

1. el error cuadrático medio de entrenamiento (RMSET), el cual se calcula en cada ciclo de entrenamiento y permite corregir los pesos de la red que conducen a disminuir el RMSET y
2. el RMSEM (M de monitorización), el cual simultáneamente monitoriza el rendimiento de la red MLP a través de los resultados provistos por un juego de monitorización independiente.

Ambas cifras se calculan como $\left[\frac{1}{I} \sum_1^m (c_{\text{nom}} - c_{\text{pred}})^2 \right]^{1/2}$, siendo c_{nom} la concentración nominal del componente en las I muestras de entrenamiento o

monitorización, según corresponda, y c_{pred} la concentración predicha por el método para estas mismas muestras.

En general, la red se entrena durante un número de épocas hasta que se alcanza un mínimo en el valor de RMSEM (compatible con el nivel de ruido presente en el sistema), con el objetivo de no generar sobreentrenamiento (Zupan y Gasteiger 1999). Dos parámetros importantes para el entrenamiento de la red son la velocidad de aprendizaje, la cual tiende hacia una convergencia rápida en pasos descendentes, y el momento, una función que previene a la solución de ser atrapada en un mínimo local. Estos parámetros usualmente se ajustan mediante prueba y error alrededor de un valor de 0,5. El juego final de pesos contiene los parámetros del modelo y se almacena para la predicción futura de nuevas muestras.

5.3.3. FUNCIONES DE BASE RADIAL

La estructura de las redes de base radial consiste en tres capas. La capa de entrada solo se emplea para distribuir las variables de entrada (en este caso los *A scores* obtenidos a través del análisis PCA de los datos desdoblados) a la capa oculta. Cada una de las M neuronas de la capa oculta usa una función de transferencia (en este caso fue la gaussiana) y, en contraste con las redes MLP, la neurona de salida emplea siempre una función de transferencia lineal. La implementación exitosa de estas redes implica encontrar parámetros adecuados para la función gaussiana de la capa oculta: los centros (contenidos en el vector c_m de $A \times 1$) y los anchos σ , los cuales usualmente se consideran iguales para todas las funciones. La salida de la m th. neurona oculta, para un dado objeto de entrada x_i , está dado por:

$$\text{out}_m = \exp\left(-\frac{1}{2\sigma^2} \|x_i - c_m\|^2\right) \quad (5.2)$$

en donde $\|x_i - c_m\|$ es la distancia Euclídea entre x_i y c_m . El valor de entrada para la neurona de salida es la suma ponderada de todas las salidas de los nodos ocultos, con la respuesta del nodo de salida relacionada linealmente a

su entrada. Por lo tanto, la salida de la red RBF (out_i) para un dado objeto de entrada x_i está dado por:

$$out_i = w_0 + \sum_{m=1}^M w_m \exp\left(-\frac{1}{2\sigma^2} \|x_i - c_m\|^2\right) \quad (5.3)$$

en donde w_0 es el *bias* y w_m es el peso asignado a la m th. salida oculta. Estos pesos se ajustan para minimizar el error cuadrático medio de la salida de la red. Por lo tanto se ajustan dos juegos de parámetros (los centros y los anchos) en la capa oculta y un juego de pesos en la capa de salida. RBF cuenta con un procedimiento garantizado para alcanzar la convergencia: dados los centros de las M funciones de base y un juego de I objetos de entrenamiento con valores de señal conocidos (x_i) y concentraciones conocidas (y_i), el menor error cuadrático en la predicción de y se alcanza cuando los pesos están dados por:

$$w = (H^T H)^{-1} H^T y \quad (5.4)$$

en donde w es un vector ($M \times 1$) que contiene los pesos, y ($I \times 1$) es el vector de las concentraciones conocidas y H ($I \times M$) es la llamada matriz de diseño, cuyos elementos se calculan como:

$$H(i, m) = \exp\left(-\frac{1}{2\sigma^2} \|x_i - c_m\|^2\right) \quad (5.5)$$

Existen varios procedimientos para eliminar la alta dimensionalidad de la capa oculta. Una posibilidad es controlar la complejidad de la red usando un subset de posibles centros. Un procedimiento ampliamente utilizado es la selección hacia delante, la cual comienza con un modelo vacío y agrega nuevas funciones de base hasta que se alcanza cierto criterio. Estas funciones de base poseen coordenadas en el espacio multivariado que se corresponden con los valores de los *scores* de las muestra de calibración y se agregan tantas como sea necesario hasta que la disminución del error cuadrático no resulte

significativa. Orr (Orr 1995) introdujo la regularización en conjunción con la selección hacia delante, la cual incluye el parámetro de regularización λ en la Ecuación (5.4), con el objetivo de penalizar pesos grandes:

$$\mathbf{w} = (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^T \mathbf{y} \quad (5.6)$$

en donde \mathbf{I} es una matriz identidad de dimensiones apropiadas.

Las siguientes características distinguen a las redes RBF de las MLP: 1) las redes RBF tienen una sola capa de neuronas ocultas mientras que las MLP pueden tener una o más capas ocultas, 2) la capa oculta (no lineal) de RBF tiene un propósito diferente a la capa de salida (lineal), mientras que en MLP hay un modelo neuronal común entre todas las capas y 3) el argumento de la función de transferencia es la distancia euclídea entre el vector de entrada y la función centro, mientras que MLP computa el producto interno del vector de entrada y del vector de pesos sinápticos.

5.3.4. MÁQUINAS DE VECTORES SOPORTES

SVM intenta modelar los valores objetivo usando una suma ponderada de funciones. En la versión de cuadrados mínimos del algoritmo SVM es necesario minimizar una función costo (C) que contiene un error de regresión penalizado (Suykens y Vanderwalle 1999):

$$C = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^l e_i^2 \quad (5.7)$$

sujeta a las siguientes restricciones (que en sí mismas definen la expresión empleada por SVM para modelar los datos y en función de las entradas \mathbf{x}):

$$y_i = \mathbf{w}^T \phi(\mathbf{x}_i) + b + e_i \quad (5.8)$$

donde \mathbf{w} es el vector de los pesos ($l \times 1$), ϕ es la función de modelado, $\phi(\mathbf{x}_i)$ es el vector de valores inducidos por ϕ en un objeto de entrada \mathbf{x}_i , b es el *bias* y e_i es el error de entrenamiento. La primera parte de la función costo es un

decaimiento del peso empleado para regularizar los tamaños de los pesos y penalizar los pesos grandes con el objetivo de converger a valores menores. Pesos grandes deterioran la capacidad de generalización de la SVM porque pueden causar una variancia excesiva. La segunda parte de la Ecuación (5.7) es el error de regresión para todos los datos de entrenamiento. El parámetro γ , el cual tiene que ser optimizado por el usuario, controla la importancia relativa de cada término en el lado derecho de la Ecuación (5.7).

El análisis de las ecuaciones (5.7) y (5.8) muestra que SVM constituye un problema de optimización típico que puede ser resuelto usando el método de los multiplicadores de Lagrange. Esto conduce a la siguiente expresión para los pesos:

$$\mathbf{w} = \sum_{i=1}^l \alpha_i \phi(\mathbf{x}_i) = \sum_{i=1}^l \gamma e_i \phi(\mathbf{x}_i) \quad (5.9)$$

Un resultado importante de este análisis es que los pesos (\mathbf{w}) se pueden escribir como combinaciones lineales de los multiplicadores de Lagrange con los datos de entrenamiento correspondientes (\mathbf{x}_i). Al mismo tiempo todos los valores de α_i se obtienen automáticamente. Poniendo el resultado de la Ecuación (5.9) dentro de la línea original de regresión [Ecuación (5.8)], se obtiene el siguiente resultado para un dado valor y :

$$y_i = \sum_{i=1}^l \alpha_i \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_i) + b \quad (5.10)$$

el cual expresa el valor de y para un dado objeto en función de sus valores de entrada y los del juego de entrenamiento. Como se puede observar, se emplea un kernel positivo definido de la siguiente manera:

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_{i'}) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_{i'}) \quad (5.11)$$

y por lo tanto:

$$y_i = \sum_{i=1}^l \alpha_i K(\mathbf{x}_i, \mathbf{x}_i) + b \quad (5.12)$$

Generalmente todos los multiplicadores de Lagrange (los vectores soporte) son distintos de cero, lo cual significa que todos los objetos de entrenamiento contribuyen a la solución.

La función de kernel más usada es la función de base radial $\exp[-(|\mathbf{x}_i - \mathbf{x}_j|^2)/2\sigma^2]$, en donde σ^2 es el ancho de la función Gaussiana, el cual debe ser optimizado por el usuario en combinación con la constante de regularización γ , con el objetivo de obtener un vector soporte.

5.4. PROGRAMAS

Los algoritmos multivariados se emplearon en MATLAB (Matlab 6.0, 2000). U-PCA/RBL y U-PLS/RBL se implementaron a través rutinas escritas por el Dr. Olivieri incluidas en la interfaz gráfica MVC2, similar a la ya descrita para calibración multivariada de primer orden MVC1 (Olivieri y col 2004). MCR-ALS está disponible en Internet a través de la página <http://www.ub.edu/mcr/welcome.html>.

Para entrenar las redes neuronales MLP se empleó una rutina escrita por el Dr. Olivieri sobre la base de algoritmos conocidos. Las redes RBF se implementaron utilizando el método de la selección hacia delante descrito por Orr (Orr 1999), disponible en Internet a través de la página <http://www.anc.ed.ac.uk/rbf/rbf.html>. Las redes SVM se emplearon mediante herramientas de laboratorio para LS-SVM (MATLAB/C Toolbox for Least-Squares Support Vector Machines), disponibles en <http://www.esat.kuleuven.ac.be/sista/lssvmlab/>.

Para realizar los análisis estadísticos se utilizó el programa Statgraphics Plus 5.1 (Statgraphics Plus 1994-2000).

5.5. SISTEMA Fe (II)

El Fe(II) actúa como catalizador de la oxidación del colorante naranja de metilo en presencia de bromato de potasio siguiendo una cinética no lineal.

Este comportamiento se utilizó con el fin de determinar Fe(II) en muestras de naranja de metilo, en presencia del colorante púrpura de bromocresol como interferencia. En el Capítulo 3 se implementó este mismo sistema, pero debido a que las concentraciones en las que se encontraba el Fe(II) eran mayores, el sistema resultó lineal y se determinó empleando algoritmos adecuados para este tipo de datos. En la Figura 5.2 se evidencia el solapamiento espectral existente entre ambos colorantes, lo que imposibilita la resolución del sistema sin recurrir a la ventaja de segundo orden.

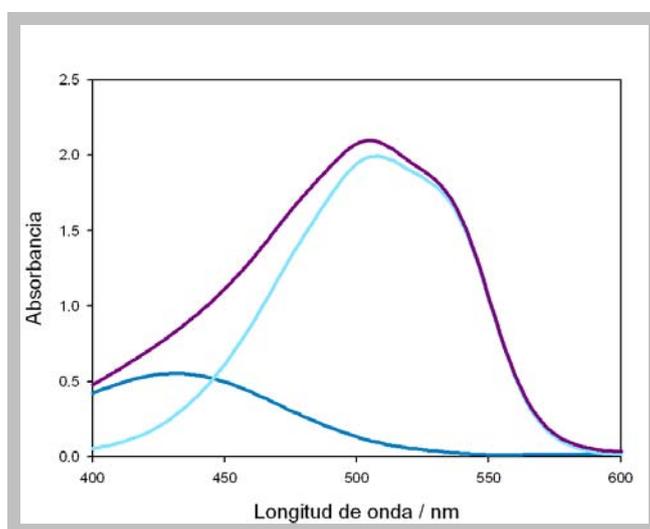


Figura 5.2. Espectro de absorción de una muestra de calibración de Fe(II) ($1,00 \text{ mg L}^{-1}$) en naranja de metilo ($16,00 \text{ mg L}^{-1}$) (celeste), una solución de púrpura de bromocresol ($12,00 \text{ mg L}^{-1}$) (azul) y una muestra de Fe(II) ($1,00 \text{ mg L}^{-1}$) en naranja de metilo ($16,00 \text{ mg L}^{-1}$) con adición de púrpura de bromocresol ($12,00 \text{ mg L}^{-1}$) como interferencia (violeta).

5.5.1. MATERIALES Y MÉTODOS

5.5.1.1. Equipamiento

Las medidas de absorción electrónica se realizaron en un espectrofotómetro Perkin Elmer Lambda 20 (Perkin Elmer, Waltham, Massachusetts, USA) empleando cubetas de cuarzo de 1,00 cm y ancho de ranura de 2 nm.

Todos los espectros se guardaron en formato ASCII y se transfirieron a una PC para su manipulación posterior.

5.5.1.2. Reactivos

Todas las experiencias se realizaron con drogas calidad para análisis y agua milliQ. Se preparó una solución stock de Fe(II) de 1000,0 mg L⁻¹ mediante la disolución de 3,5833 g de Fe(NH₄)(SO₄)₂·6H₂O (Cicarelli, San Lorenzo, Argentina) en agua, el agregado de 5,0 mL de H₂SO₄ y la dilución hasta la marca, en un matraz volumétrico de 500,00 mL. Por otra parte se prepararon una solución de naranja de metilo (Carlo Erba Reagenti, Milan, Italia) 1240,0 mg L⁻¹ y otra de púrpura de bromocresol (Carlo Erba Reagenti, Milan, Italia) 346,0 mg L⁻¹ disolviendo cantidades adecuadas de cada compuesto en agua. Además se prepararon soluciones de bromato de potasio (Cicarelli, San Lorenzo, Argentina) 0,1 mol L⁻¹ y ácido fosfórico (Cicarelli, San Lorenzo, Argentina) 3,0 mol L⁻¹.

5.5.1.3. Diseño del juego de entrenamiento

Se preparó un juego de entrenamiento compuesto por diecinueve soluciones estándares de Fe(II) con concentraciones comprendidas en el rango de 0,20 a 2,00 mg L⁻¹ cada 0,10 mg L⁻¹. Las muestras número 2, 5, 9, 12 y 18 se usaron para monitorizar el entrenamiento de la red MLP.

5.5.1.4. Diseño del juego de validación

Se prepararon catorce muestras de validación que requerían de la ventaja de segundo orden para ser resueltas debido a que estaban compuestas por Fe(II) en diferentes concentraciones, pero además contenían 6,00 o 12,00 mg L⁻¹ de púrpura de bromocresol. Estas muestras no se emplearon para entrenar ni monitorizar la red MLP.

5.5.1.5. Procedimiento experimental

El procedimiento cinético se adaptó del propuesto por Safavi y col (Safavi y col 2001). En matraces de 5,00 mL conteniendo cantidades adecuadas de solución stock de Fe(II) se agregaron 1,00 mL de solución de naranja de metilo 80,0 mg L⁻¹ y 1,00 mL de solución de ácido fosfórico 3,0 mol L⁻¹. Las muestras se llevaron a enrase con agua milliQ y se homogeneizaron. Posteriormente, se

transfirieron 2,00 mL de cada solución a una cubeta de cuarzo de 1,00 cm y pasados 10 s de la adición de 40,0 μL de solución de bromato de potasio 0,1 mol L^{-1} , se registraron veinte espectros consecutivos en el rango de longitudes de onda comprendido entre 400 y 600 nm, cada 2 nm, con una velocidad de barrido de 2880 nm min^{-1} . Se obtuvieron matrices de 101×20 datos por muestra.

5.5.2. RESULTADOS Y DISCUSIÓN

La Figura 5.3 muestra la superficie correspondiente a la evolución espectral-temporal de una solución conteniendo naranja de metilo (16,00 mg L^{-1}) durante la oxidación con bromato de potasio catalizada por 1,00 mg L^{-1} de Fe(II). Se puede observar como el colorante, presentando una banda de absorción de 400 a 600 nm, se degrada por oxidación siguiendo una cinética que depende de la concentración de Fe(II). Estos cambios permiten obtener datos de segundo orden espectrales-temporales para cada muestra experimental, que pueden ser procesados con algoritmos de segundo orden adquiriendo la ventaja de segundo orden. Sin embargo, al existir desviaciones de la relación lineal absorbancia-concentración, se necesita de un algoritmo no lineal para resolver este sistema de manera satisfactoria.

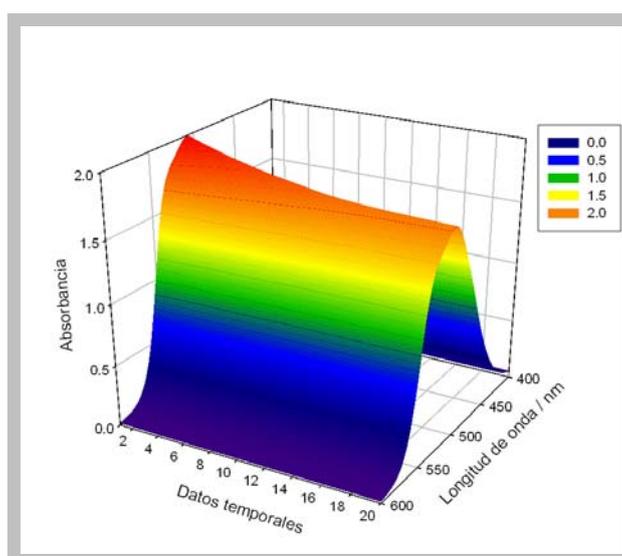


Figura 5.3. Gráfica tridimensional de absorbancia en función de longitud de onda y tiempo para la reacción entre naranja de metilo (16,00 mg L^{-1}) y bromato de potasio (2×10^{-3} mol L^{-1}), catalizada por 1,00 mg L^{-1} de Fe(II).

Con el objetivo de corroborar que estos datos no deberían ser tratados con algoritmos lineales, se aplicó U-PLS/RBL a las muestras de validación. La aplicación del procedimiento de RBL permitió concluir que con un solo componente se modelaban las potenciales interferencias. Este valor se estableció incrementando gradualmente N_{hoesp} hasta estabilizar los residuos finales s_u . La Tabla 5.1 muestra los resultados de la predicción. Como se puede observar, U-PLS/RBL no resultó apropiado para cuantificar el Fe(II) presente en las muestras, evidenciándose en la baja recuperación promedio alcanzada, con un error relativo de predicción de aproximadamente 30%.

Por el contrario, la aplicación de U-PCA/RBL-MLP arrojó mejores resultados. La Tabla 5.2 muestra los parámetros de entrenamiento del sistema. Como se puede observar, se necesita un gran número de ciclos para entrenarlo de manera satisfactoria, hecho que se debe a la marcada no linealidad presente. En la Tabla 5.1 se observan los resultados de la predicción del juego de validación (calculados como el promedio de los obtenidos de cinco MLPs diferentes entrenadas para este sistema) los cuales son más exactos que los arrojados por U-PLS/RBL. Este hecho se evidencia observando la Figura 5.4 en donde se grafican las regiones elípticas de confianza conjunta (EJCR) construidas a partir del ajuste por cuadrados mínimos ordinarios de las concentraciones nominales versus las predichas para cada método, y comparando el valor de la pendiente y la ordenada al origen con sus valores ideales de 1 y 0.

Tabla 5.1. Resultados obtenidos de la aplicación de U-PLS/RBL y U-PCA/RBL-MLP a las muestras de validación para el sistema Fe(II).

Muestra	Concentración nominal / mg L ⁻¹	U-PLS/RBL ^a		U-PCA/RBL-MLP ^b	
		Predicción / mg L ⁻¹	Recup. / %	Predicción / mg L ⁻¹	Recup. / %
1	0,40	0,42	105,0	0,31(4)	77,5
2	0,40	0,28	70,0	0,33(5)	82,5
3	0,40	0,23	57,5	0,34(3)	85,0
4	0,60	0,53	88,3	0,61(5)	101,7
5	0,60	0,51	85,0	0,61(5)	101,7
6	0,80	0,42	52,5	0,73(3)	91,3
7	0,80	0,41	51,3	0,74(5)	92,5
8	0,80	0,38	47,5	0,73(4)	91,3
9	1,00	0,97	97,5	0,97(5)	97,0
10	1,00	1,16	116,0	0,92(6)	92,0
11	1,30	1,28	98,5	1,25(2)	96,2
12	1,30	1,58	121,5	1,40(6)	107,7
13	1,90	1,91	100,5	1,90(4)	100,0
14	1,90	2,00	105,3	1,94(5)	102,1
Recup. promedio / % ^c			85(25)		94,2(8)
RMSEP y REP% ^d		0,35 mg L ⁻¹ ; 31,7%		0,13 mg L ⁻¹ ; 11,9%	

^a Número de variables latentes, 4 (establecidas por validación cruzada); número de componentes no esperados, 1.
^b Número de componentes no esperados, 1. Las predicciones son promedios de cinco ciclos diferentes de entrenamiento, con desviaciones estándares en la última cifra significativa entre paréntesis.
^c Desviación estándar entre paréntesis.
^d RMSEP, error cuadrático medio de predicción, REP, error relativo de predicción.

Tabla 5.2. Resultados del entrenamiento de la red MLP para el sistema Fe(II).

Parámetro	Valor
Arquitectura (neuronas de entrada-ocultas-salida) ^a	4-3-1
Número de épocas de entrenamiento ^b	5192-7310
Velocidad de aprendizaje	0,5
Momento	0,5
RMSET absoluto y relativo ^c	0,027 mg L ⁻¹ ; 2,5%
RMSEM absoluto y relativo ^c	0,028 mg L ⁻¹ ; 2,7%

^a Las neuronas de la capa de entrada se determinaron mediante PCA y corresponden al número de PCs que explican aproximadamente el 99% de la variancia del sistema. Las neuronas de la capa oculta se establecieron mediante prueba y error.

^b Rango de épocas empleadas en los cinco entrenamientos.

^c RMSE = error cuadrático medio, T = entrenamiento, M = monitorización. Los errores relativos porcentuales se calcularon con respecto a la concentración media de la calibración. Los valores mostrados son promedios obtenidos de cinco procedimientos diferentes de entrenamiento-monitorización.

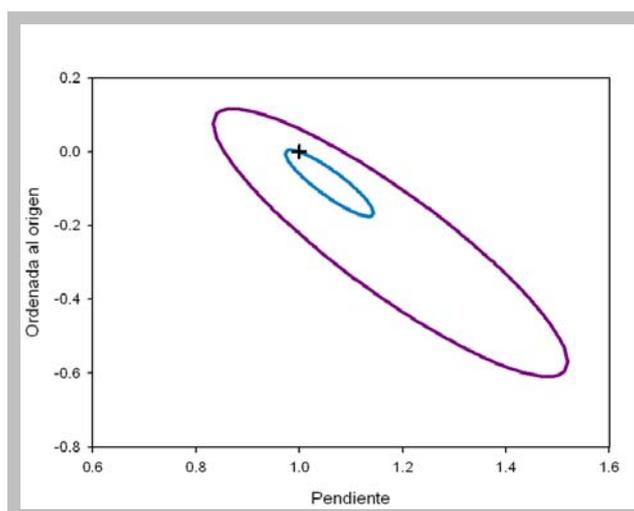


Figura 5.4. Región elíptica de confianza conjunta para la pendiente y para la ordenada al origen, usando el ajuste cuadrados mínimos ordinarios entre concentración nominal versus la predicha, para U-PCA/RBL-MLP (azul) y para U-PLS/RBL (violeta). La marca en forma de cruz indica el punto teórico (1,0).

Nuevamente se necesitó un solo componente para el procedimiento RBL, obtenido en base a un análisis similar de la variación de s_u con el incremento de N_{noesp} . U-PCA/RBL-MLP fue capaz de detectar, corregir y modelar satisfactoriamente el componente no esperado, lo cual se puede observar en las Figuras 5.5 A y B, tanto en la dimensión espectral como en la temporal.

Estas figuras muestran que los perfiles espectrales y cinéticos de la interferencia (púrpura de bromocresol), obtenidos por RBL, concuerdan de manera excelente con los experimentales. Por otra parte, se observa que de la aplicación de red MLP se obtiene una reducción considerable de la cifra de mérito REP, llegando a casi la mitad del valor obtenido con U-PLS/RBL.

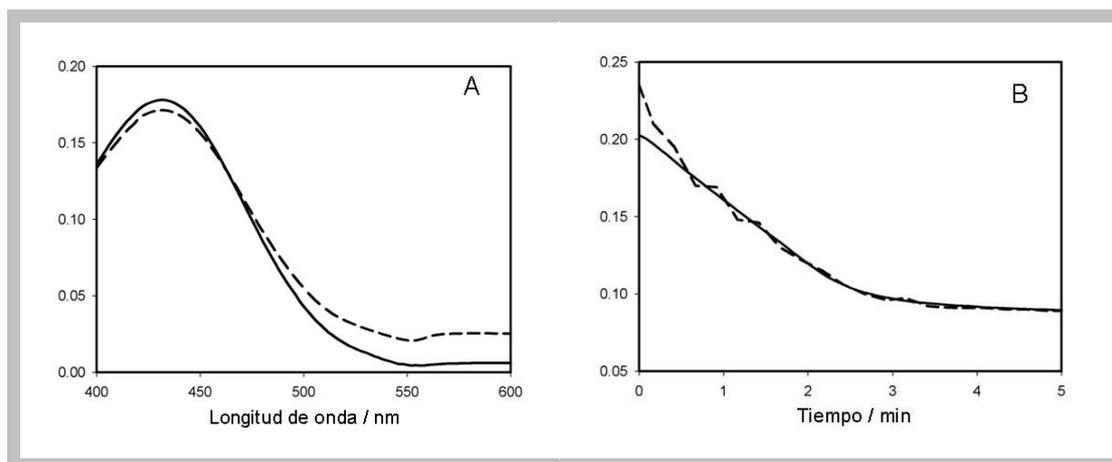


Figura 5.5. Perfiles espectrales (A) y temporales (B) para el componente no esperado en el sistema Fe(II). Los perfiles experimentales (líneas sólidas) y provistos por RBL (líneas de trazos) se normalizaron a la unidad. El procedimiento de RBL se aplicó a una muestra conteniendo naranja de metilo ($16,00 \text{ mg L}^{-1}$), bromato de potasio ($2 \times 10^{-3} \text{ mol L}^{-1}$), Fe(II) ($1,00 \text{ mg L}^{-1}$) y púrpura de bromocresol como componente no esperado ($12,00 \text{ mg L}^{-1}$).

Las muy buenas recuperaciones obtenidas muestran que la técnica podría ser extendida a la determinación de Fe(II) en mezclas complejas en presencia de otras interferencias, como fluidos biológicos o productos alimenticios.

5.6. SISTEMA LORATADINA-PSEUDOEFEDRINA SULFATO

La combinación de loratadina (LOR) y pseudoefedrina sulfato (PES) en productos medicinales es efectiva en el alivio de síntomas asociados con la rinitis alérgica estacional acompañada de congestión nasal (Brunton 2005). LOR es un antihistamínico no sedante de larga acción que pertenece al grupo de los antagonistas H-1, mientras que PES es un descongestivo nasal activo por vía oral (Brunton 2005).

En la literatura se informaron pocas metodologías analíticas para cuantificar LOR y PES. Ambos analitos se determinaron simultáneamente en preparaciones farmacéuticas por espectroscopia derivada (Mabrouk y col 2003, Onur y col 2000) y cromatografía líquida de alta resolución (CLAR) (Sane y col 2001, Vasudevan y col 2001). La cuantificación de LOR y PES en fluidos biológicos se llevó a cabo mediante la aplicación de cromatografía líquida acoplada a espectrometría de masas (Naidong 2003, Gergov y col 2003).

En el presente estudio se generaron datos de segundo orden no lineales empleando un sistema de análisis por inyección en flujo (FIA) con gradiente de pH y detección mediante arreglo de diodos. Como resultado, la determinación de ambos analitos se realizó en base a las diferentes características ácido-base y espectrales de cada droga evidenciadas en la Figura 5.6, la cual muestra los espectros de ambos y de la matriz en la que están inmersos en la preparación comercial, a pH 2,0 y 7,0. A causa de la relación no lineal existente entre las concentraciones y las absorbancias medidas en el rango estudiado para ambos analitos, debido a que uno de los dos (PES) está presente en una concentración mucho más alta en comparación con el otro (LOR), se trabajó bajo la hipótesis de que los métodos de calibración multivariada de segundo orden clásicos no podían ser aplicados de manera satisfactoria para resolver el presente problema analítico, y sí lo sería U-PCA/RBL-MLP.

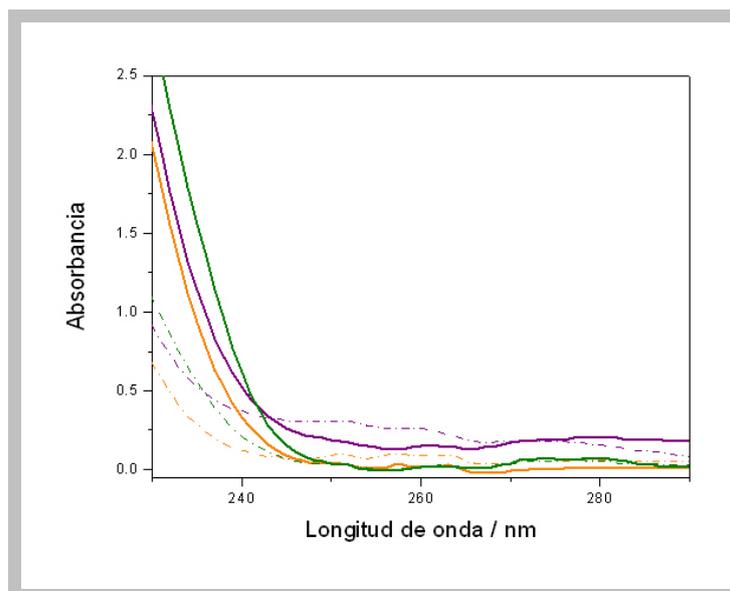


Figura 5.6. Espectro UV-Vis para LOR (línea violeta; $8,33 \text{ mg L}^{-1}$), PES (línea anaranjada; $100,00 \text{ mg L}^{-1}$) y una mezcla de los excipientes (línea verde) a valores de pH de 2,0 (líneas sólidas) y 7,0 (líneas de trazos).

5.6.1. MATERIALES Y MÉTODOS

5.6.1.1. Equipamiento

El sistema FIA de simple canal y el procedimiento cromatográfico se desarrollaron usando cinco módulos (degasificador, bomba, válvula de inyección, inyector automático y detector) de un instrumento Agilent 1100 Series (Agilent Technologies, Waldbronn, Alemania).

El sistema de inyección en flujo ilustrado en la Figura 5.7 consistió en un solvente transportador compuesto por ácido acético $0,2 \text{ mol L}^{-1}$ de pH=3,0 fluyendo a una velocidad de $0,2 \text{ mL min}^{-1}$, en el cual se realizaron inyecciones de $100 \mu\text{L}$ de muestra preparada en solución reguladora de acetato $0,2 \text{ mol L}^{-1}$ de pH=7,0. El gradiente de pH se generó en una cámara de mezcla de $200 \text{ cm} \times 0,12 \text{ mm}$ de diámetro interno y mostró ser muy reproducible. Para cada pico FIA se registraron espectros en el rango comprendido entre 200 y 300 nm cada 1 nm a intervalos regulares de tiempo de 1,6 s durante 176 s. Por lo tanto, se obtuvieron matrices de tamaño 111×101 , aunque se seleccionaron regiones antes de realizar la calibración multivariada.

Las mediciones cromatográficas se llevaron a cabo usando una columna ($4,6 \times 150 \text{ mm}$) Zorbax Eclipse XDB-C18 de tamaño de partícula $5 \mu\text{m}$. Los

cromatogramas se realizaron a temperatura ambiente, empleando como fase móvil una mezcla compuesta por 60% de (metanol-agua-ácido fosfórico-dihidrógeno fosfato de amonio 220:300:2:3 g, *v/v/v/w*) y 40% de acetonitrilo, a una velocidad de flujo de 1 mL min⁻¹ y detección ultravioleta a 247 nm (Mabrouk y col 2003).

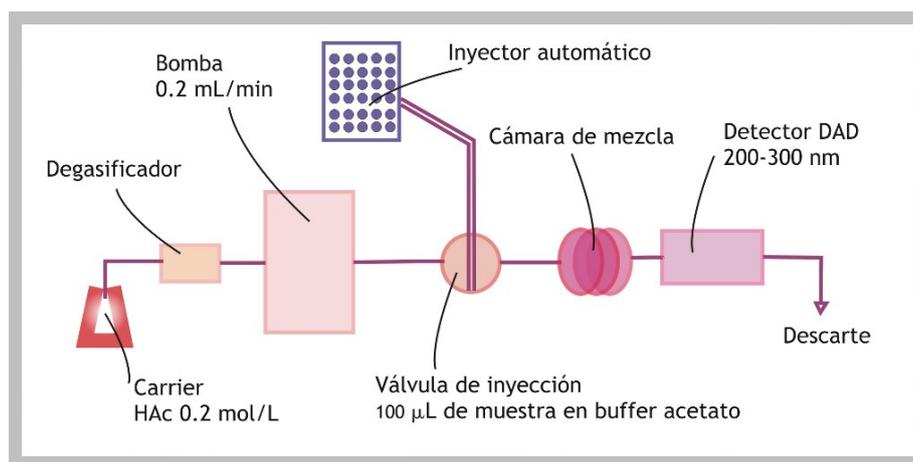


Figura 5.7. Sistema de inyección en flujo.

5.6.1.2. Reactivos

Todas las experiencias se realizaron con drogas calidad para análisis y agua milliQ. Los estándares de LOR y PES fueron donados por Química Montpellier, Buenos Aires, Argentina. Se prepararon soluciones stock de LOR 620,3 mg L⁻¹ y PES 6016,4 mg L⁻¹ mediante disolución de cada estándar en HCl 0,01 mol L⁻¹. Por otra parte se preparó una solución acuosa conteniendo los excipientes generalmente presentes en la muestra comercial: propilenglicol, sorbitol, ácido cítrico anhidro, benzoato de sodio y sacarosa, con concentraciones correspondientes a su valor informado.

Para crear el doble gradiente de pH se prepararon dos soluciones a partir de ácido acético glacial (Cicarelli, San Lorenzo, Argentina) e NaOH (Cicarelli, San Lorenzo, Argentina). La solución de ácido acético 0,2 mol L⁻¹ se usó como transportador y la solución reguladora de acetato 0,2 mol L⁻¹ se empleó para diluir las muestras.

5.6.1.3. Diseño del juego de entrenamiento

Para realizar la calibración se preparó un juego de veinticinco muestras correspondientes a un diseño factorial completo en cinco niveles con concentraciones comprendidas entre 6,66 y 10,00 mg L⁻¹ y 80,00 y 120,00 mg L⁻¹ para LOR y PES, respectivamente (ver Tabla 5.3).

5.6.1.4. Diseño del juego de monitorización

Con el objetivo de monitorizar el entrenamiento de la red MLP, se construyó un juego compuesto por nueve muestras correspondientes a un diseño central compuesto con concentraciones de LOR y PES diferentes a las empleadas para calibrar, las cuales se muestran en la Tabla 5.4.

Por otra parte se preparó un juego adicional de seis muestras que no se emplearon para entrenar ni monitorizar la red MLP sino para realizar predicciones que requieren de la ventaja de segundo orden. Estas mezclas estaban compuestas por cantidades apropiadas de LOR, PES y excipientes presentes en las muestras comerciales en sus concentraciones declaradas.

Todas las muestras se prepararon en matraces volumétricos de 2,00 mL, se llevaron a volumen con solución reguladora de acetato 0,2 mol L⁻¹ y se inyectaron en el sistema FIA dos veces.

5.6.1.5. Muestra real

La muestra analizada fue Decidex Plus Solución (Roemmers, Argentina). Este medicamento contiene 1,0 g L⁻¹ de LOR y 12,0 g L⁻¹ de PES. Antes de cada análisis se le realizó una dilución 1:120 con solución reguladora de acetato 0,2 mol L⁻¹ empleando material volumétrico.

5.6.1.6. Procedimiento cromatográfico para muestras reales

Las soluciones de trabajo de LOR y PES se prepararon disolviendo cantidades conocidas exactas de cada estándar en fase móvil y diluyendo en el mismo solvente para obtener concentraciones de 20 y 238 mg L⁻¹, respectivamente. Las muestras reales se diluyeron 1:50 en fase móvil. Todas

las soluciones se filtraron a través de membrana de 0,45 μm y se inyectaron 20 μL en el cromatógrafo.

Tabla 5.3. Composición del juego de calibración siguiendo un diseño experimental factorial completo en cinco niveles.

Muestra	Concentración nominal / mg L^{-1}	
	Loratadina	Pseudoefedrina sulfato
1	6,66	80,00
2	7,50	80,00
3	8,33	80,00
4	9,16	80,00
5	10,00	80,00
6	6,66	90,00
7	7,50	90,00
8	8,33	90,00
9	9,16	90,00
10	10,00	90,00
11	6,66	100,00
12	7,50	100,00
13	8,33	100,00
14	9,16	100,00
15	10,00	100,00
16	6,66	110,00
17	7,50	110,00
18	8,33	110,00
19	9,16	110,00
20	10,00	110,00
21	6,66	120,00
22	7,50	120,00
23	8,33	120,00
24	9,16	120,00
25	10,00	120,00

Tabla 5.4. Composición del juego de validación siguiendo un diseño experimental central compuesto.

Muestra	Concentración nominal / mg L ⁻¹	
	Loratadina	Pseudoefedrina sulfato
1	7,15	100,00
2	9,50	100,00
3	8,33	85,90
4	8,33	114,10
5	7,50	90,00
6	9,16	90,00
7	7,50	110,00
8	9,16	110,00
9	8,33	100,00

5.6.2. RESULTADOS Y DISCUSIÓN

5.6.2.1. Detección de no linealidades

Como consecuencia de la gran diferencia de concentración entre LOR y PES en la muestra real analizada, la dilución requerida para llevar a cabo la determinación de ambos analitos simultáneamente introduce no linealidades en el analito más concentrado (PES), pero también puede conducir a desviaciones del rango de concentración-absorbancia al analito menos concentrado (LOR).

Con el objetivo de detectar la presencia de no linealidad, se implementaron herramientas de detección cualitativa y cuantitativa a los espectros de las muestras de calibración seleccionados en el máximo del pico FIA, es decir, a datos de primer orden. En principio se construyó para cada analito el gráfico de diagnóstico universalmente más empleado, conocido como Gráfico de los residuos parciales aumentados de Mallows (APaRP), con el objetivo de evaluar una posible relación cuadrática entre alguno de los primeros factores y la concentración (Center 1998). Con este propósito se realizó una regresión entre un vector y que contenía las concentraciones individuales de analito y los primeros A factores de PLS-1 (LV) de la matriz de datos y el cuadrado de los primeros LV (Draper y Smith 1981):

$$y = b_0 + b_1LV_{1i} + \dots + b_{Ai}LV_{Ai} + b_{11}(LV_{1i})^2 + e_{APaRPi} \quad (5.13)$$

donde e_{APaRPi} es un vector que contiene los residuos del ajuste APaRP. Luego se obtuvo el gráfico de interés representando la suma $[e_{APaRP} + b_1LV_1 + b_{11}(LV_1)^2]$ en función de LV_1 (Drapper y Smith 1981). En el presente caso, los gráficos APaRP se construyeron incluyendo en el modelo los seis primeros LVs para ambos analitos. A partir de la observación de la Figura 5.8 (a-b) se puede evidenciar la presencia de un comportamiento no lineal leve para LOR, en contraste con un patrón de no linealidad cuadrática considerable para PES.

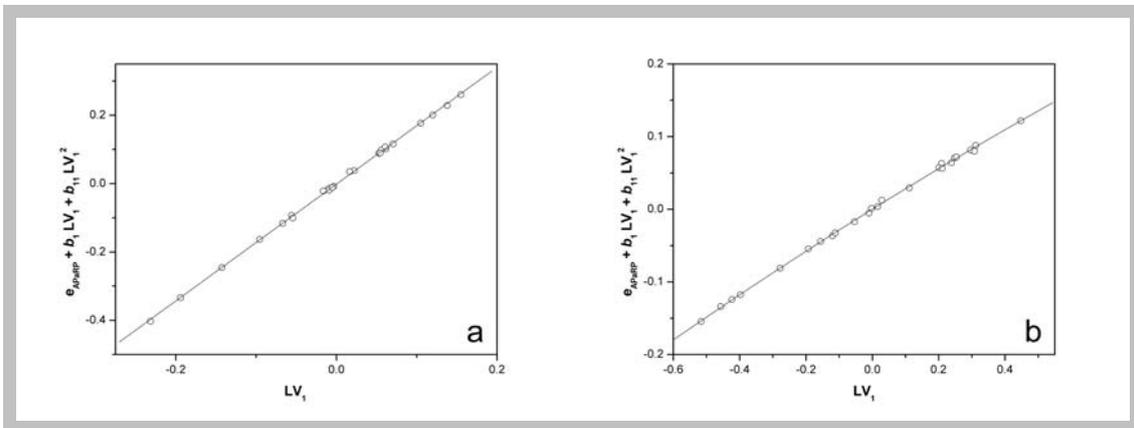


Figura 5.8. Gráfico de los residuos parciales aumentados de Mallows (APaRP) para LOR (a) y PES (b) cuando se incluyen en el modelo, para ambos, los seis primeros LVs.

Para evaluar cuantitativamente el grado de no linealidad se aplicaron las pruebas Runs y Durbin-Watson a los gráficos de Mallows. La primera examina el número de series de residuos consecutivos con el mismo signo, llamadas *runs*, indicando una tendencia cuando existen *runs* largas en base al cálculo del indicador z (Drapper y Smith 1981):

$$z = (u - \sigma + 0,5) / \sigma \quad (5.14)$$

donde u es el número de *runs* y:

$$u = 1 + 2n_+n_- / (n_+ + n_-) \quad (5.15)$$

$$\sigma^2 = 2n_+n_-(2n_+n_- - n_+ - n_-) / [(n_+ + n_-)^2(n_+ + n_- - 1)] \quad (5.16)$$

con n_+ y n_- como el número de residuos positivos y negativos. La no linealidad está presente cuando z excede el valor crítico de 1,96 (con un nivel de confianza del 95%). Por otra parte la prueba de Durbin-Watson considera la hipótesis nula (H_0) de que no hay correlación entre los residuos sucesivos versus la hipótesis alternativa de que existe correlación, por medio del cálculo de un estadístico d (Montgomery 1991):

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (5.17)$$

donde e_i es un dado residuo y e_{i-1} es el precedente (los residuos están ordenados de acuerdo al valor asociado en el vector y). La comparación de d con dos valores críticos, d_L (inferior) y d_U (superior), conduce a las siguientes conclusiones: (a) si $d < d_L$ se rechaza la hipótesis nula indicando correlación entre los residuos, (b) si $d_L < d < d_U$ la prueba no es concluyente y (c) si $d > d_U$ se considera que no hay correlación entre los residuos. Como se puede observar, los resultados cuantitativos presentados en la Tabla 5.5 confirman lo estimado visualmente mediante las gráficas APaRP, sugiriéndose la presencia de no linealidades leves para LOR y apoyando fuertemente el patrón no lineal seguido por PES.

En base al análisis anterior, se pudo establecer claramente que el presente problema analítico es no lineal y debe ser resuelto con una metodología de segundo orden capaz de tratar este tipo de datos. Por lo tanto se eligió U-PCA/RBL-MLP aunque, a los fines de comparar rendimientos, también se aplicaron U-PLS/RBL y MCR-ALS.

Tabla 5.5. Resultados de las pruebas Runs y Durbin-Watson aplicadas a los gráficos APaRP para evaluar el grado de no linealidad.

Componente	Prueba Runs		Prueba Durbin-Watson	
	Valor z^a	Conclusión	Valor d^a	Conclusión
LOR	2,36	No lineal	1,64	No concluyente
PES	7,33	No lineal	1,19	No lineal

^a Los valores críticos de z y d para las pruebas Runs y Durbin-Watson con $\alpha=0,05$ son, respectivamente, $z_{crit}=1,96$; $d_L=1,32$ y $d_U=1,92$.

5.6.2.2. Determinación de las arquitecturas de las redes MLPs

La optimización de las redes MLPs se realizó mediante prueba y error siguiendo dos estrategias diferentes para establecer el número inicial de neuronas de entrada. La metodología de adición por pasos, la cual consiste en comenzar con un número deliberadamente pequeño de variables de entrada (por ejemplo algunos pocos de los primeros PCs) e ir añadiendo nuevas variables de a una por vez hasta que el rendimiento de monitorización o predicción de la red no mejore más (Zupan y Gasteiger 1999), se aplicó para encontrar la red más apropiada para LOR. Por otra parte, la eliminación por pasos se implementó para establecer la mejor red para PES, empezando con un número deliberadamente grande de variables de entrada y removiendo gradualmente alguna de ellas hasta que el rendimiento de monitorización y/o predicción de la red deje de mejorar (Zupan y Gasteiger 1999). Además, para cada número de prueba de neuronas de entrada, se ensayaron neuronas ocultas en el rango de uno a nueve. Los parámetros de todas las redes que presentaban RMSET y RMSEM aceptables se guardaron con el objetivo de ser testeados posteriormente.

Con el propósito de evaluar la robustez de los resultados provistos por las redes MLPs, las cinco mejores redes para cada sistema se emplearon para predecir el juego de muestras de validación. La Tabla 5.6 muestra los parámetros surgidos del entrenamiento. Los porcentajes de recuperación obtenidos de la aplicación de todas las redes a la resolución de las muestras

de validación fueron analizados con las pruebas estadísticas de Bartlett y Tukey HSD (Diferencia Honestamente Significativa) (Montgomery 1991).

La prueba de Bartlett se empleó para verificar la homogeneidad de las variancias entre cada grupo de predicciones. Debido a que ambos valores de p -calculados, 0,54 y 0,82 para LOR y PES respectivamente, son superiores al valor crítico de 0,05; se puede concluir que no existe diferencia significativa entre las variancias cuando las muestras de validación se predicen con cualquiera de las cinco redes consideradas. Además, se estableció que no existe diferencia significativa entre las medias obtenidas para cada analito a través de la comparación múltiple por pares realizada con la prueba de Tukey, a un nivel de confianza del 95% (ver Figuras 5.9 A y B).

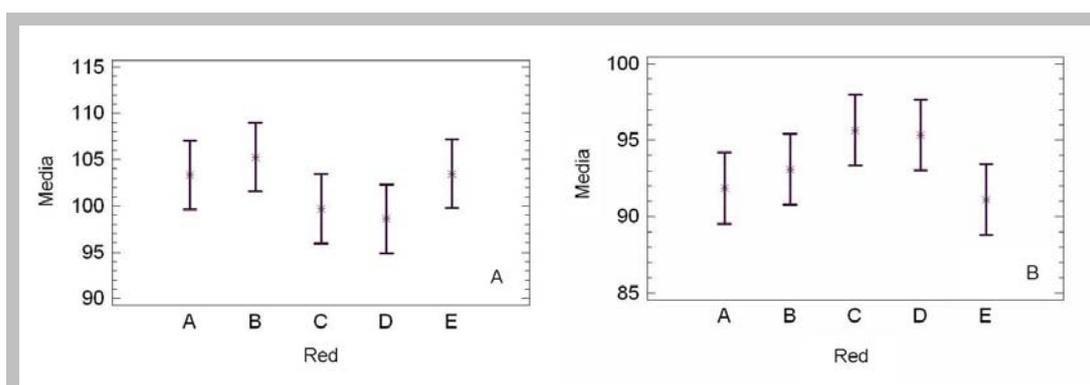


Figura 5.9. Medias e intervalos de Tukey HSD versus redes para A) LOR, con arquitecturas A=4-6-1, B=4-7-1, C=4-8-1, D=5-8-1 y E=5-9-1 y B) PES, con arquitecturas A=8-9-1, B=8-8-1, C=8-7-1, D=7-7-1 y E=7-5-1.

En base a la evidencia anterior, así como también a los RMSET y RMSEM similares entre cada grupo de redes seleccionadas, los cuales se muestran en la Tabla 5.6, se puede concluir que todas las redes testeadas podrían ser usadas para resolver adecuadamente el presente problema analítico. Sin embargo, se decidió trabajar con aquellas de arquitectura 4-8-1 y 8-7-1 para LOR y PES, respectivamente, debido a que presentaban los porcentajes de recuperación mayores al realizar la predicción de las muestras de validación (ver Tabla 5.6). Como se puede observar, otro hecho que avala las conclusiones obtenidas del análisis de no linealidad es la necesidad de una arquitectura más compleja para modelar adecuadamente PES.

Tabla 5.6. Parámetros de las redes MLPs

Parámetro ^a	Valor				
	Loratadina				
Arquitectura ^b	4-6-1	4-7-1	4-8-1	5-8-1	5-9-1
Ciclos de entrenamiento ^c	2953	1575	1285	1598	1182
RMSET absoluto / mg L ⁻¹ ; relativo / %	0,41; 4,9	0,44; 5,4	0,44; 5,3	0,46; 5,5	0,47; 5,6
RMSEM absoluto / mg L ⁻¹ ; relativo / %	0,27; 3,3	0,26; 3,1	0,29; 3,5	0,26; 3,2	0,41; 5,0
Recuperación media / % ^d	103,3(7)	105,3(8)	99,7(5)	98,6(6)	103,5(7)
	Pseudoefedrina sulfato				
Arquitectura ^b	8-9-1	8-8-1	8-7-1	7-7-1	7-5-1
Ciclos de entrenamiento ^c	681	1520	3000	1423	1402
RMSET absoluto / mg L ⁻¹ ; relativo / %	0,79; 0,79	0,76; 0,76	0,72; 0,72	0,90; 0,90	0,78; 0,78
RMSEM absoluto / mg L ⁻¹ ; relativo / %	2,0; 2,0	1,9; 1,9	1,8; 1,8	1,7; 1,7	1,7; 1,7
Recuperación media / % ^d	91,2(3)	95,3(4)	95,6(5)	93,1(4)	91,9(4)

^a RMSE = error cuadrático medio, T = entrenamiento, M = monitorización. Los errores relativos porcentuales se calculan con respecto a la concentración media de la calibración.

^b Las arquitecturas están dadas como neuronas de entrada, ocultas y de salida.

^c Velocidad de aprendizaje y momento igual a 0,5.

^d Promedio de veinte determinaciones, desviación estándar entre paréntesis.

5.6.2.3. Predicción en muestras sintéticas y reales mediante U-PCA/RBL-MLP

5.6.2.3.1. Muestras de validación

Las Tablas 5.7 y 5.8 muestran los resultados obtenidos para LOR y PES, respectivamente, cuando se analizaron las muestras de validación. Con respecto a U-PLS/RBL, se estimaron seis variables latentes de calibración a

través del procedimiento de validación cruzada (Haaland y Thomas 1988), indicando que deben considerarse variables latentes de PLS adicionales en un intento de compensar las no linealidades. Tres componentes no esperados fueron necesarios para disminuir los residuos de la predicción, hasta que se estabilizaron en un valor compatible con el ruido instrumental. Por otra parte, el análisis mediante SVD detectó cinco componentes para la aplicación adecuada de MCR-ALS. Como se puede observar, se obtienen los mejores resultados de predicción, además de los mejores parámetros estadísticos, cuando la resolución se realiza con U-PCA/RBL-MLP, considerando dos componente no esperados. Es interesante resaltar que los resultados provistos por los dos métodos lineales al sistema no lineal (PES) muestran una diferencia muy marcada con los obtenidos por U-PCA/RBL-MLP (aunque los de MCR-ALS son mejores), hecho que no debería sorprender considerando la no linealidad de la relación concentración-señal. Además, las recuperaciones obtenidas con los métodos lineales están fuera del rango aceptable del 95-105% sugerido por las agencias (AOAC 1993). Por otra parte, se obtienen predicciones aceptables para LOR al aplicar los tres algoritmos aunque se evidencia una mayor dispersión de los resultados con los métodos lineales.

Tabla 5.7. Concentraciones de las muestras de validación y resultados de la predicción empleando U-PLS/RBL, MCR-ALS y U-PCA/RBL-MLP para LOR.

Muestra	Nominal / mg L ⁻¹	Predicha / mg L ^{-1a}		
		U-PLS/RBL ^b	MCR-ALS ^c	U-PCA/RBL-MLP ^d
1	8,33	8,62	9,97	8,27
2	7,50	8,52	8,84	8,11
3	9,16	9,54	9,83	9,00
4	7,50	7,01	7,24	7,14
5	9,16	8,78	8,92	9,32
6	8,33	7,83	7,71	7,98
Recuperación media / % ^e		100,8(8)	105,2(12)	99,7(5)
RMSEP / mg L ⁻¹ , REP / % ^f		0,56; 6,7	0,95; 11,4	0,33; 4,0

^a Media de dos inyecciones realizadas en el sistema FIA para cada muestra.

^b Número de variables latentes = 6 (establecidas por validación cruzada), número de componentes no esperados= 3.

^c Número de componentes = 5 (establecido por SVD).

^d Número de componentes no esperados = 2.

^e Desviación estándar entre paréntesis.

^f RMSEP = error cuadrático medio de la predicción; REP = error relativo de la predicción.

Tabla 5.8. Concentraciones de las muestras de validación y resultados de la predicción empleando PLS/RBL, MCR-ALS y U-PCA/RBL-MLP para PES.

Muestra	Nominal / mg L ⁻¹	Predicha / mg L ^{-1a}		
		PLS/RBL ^b	MCR-ALS ^c	U-PCA/RBL-MLP ^d
1	114,10	102,51	122,42	104,13
2	90,00	83,67	97,12	82,80
3	90,00	84,48	94,80	84,96
4	110,00	93,80	124,68	113,44
5	110,00	95,66	110,30	110,15
6	100,00	87,81	102,86	92,93
Recuperación media / % ^e		89,5(3)	106,2(5)	96,5(5)
RMSEP / mg L ⁻¹ , REP / % ^f		11,7; 11,4	7,8; 7,6	6,3; 6,1

^a Media de dos inyecciones realizadas en el sistema FIA para cada muestra.

^b Número de variables latentes = 6 (establecidas por validación cruzada), número de componentes no esperados= 3.

^c Número de componentes = 5 (establecido por SVD).

^d Número de componentes no esperados = 2.

^e Desviación estándar entre paréntesis.

^f RMSEP = error cuadrático medio de la predicción; REP = error relativo de la predicción.

5.6.2.3.2. Muestra real

Finalmente, con el objetivo de corroborar la aplicabilidad del método investigado, se analizó una muestra real (cuatro réplicas) mediante U-PCA/RBL-MLP y los resultados se compararon con los obtenidos mediante CLAR (Mabrouk y col 2003). En el caso de LOR, la predicción media obtenida por U-PCA/RBL-MLP fue 0,95 g L⁻¹ (95,0 % de la cantidad rotulada). Esta concentración es estadísticamente comparable con la obtenida empleando el método CLAR: 0,94 g L⁻¹ (94,0 % de la cantidad rotulada) en base a una prueba *t* independiente. Esta prueba se realizó sobre los resultados de cuatro replicados procesados con cada método y arrojó un valor de probabilidad de

0,57; mucho mayor al límite de confianza elegido de 0,05. Con respecto a PES, la recuperación media obtenida por U-PCA/RBL-MLP fue 11,29 g L⁻¹ (94,1 % de la cantidad rotulada), la cual es también equivalente a la obtenida por CLAR: 11,57 g L⁻¹ (96,4 % de la cantidad rotulada) obteniéndose un valor p de 0,42. Finalmente, se emplearon U-PLS/RBL y MCR-ALS para cuantificar las muestras reales. Las diferencias entre los valores medios al determinar LOR y PES resultaron significativas con $p < 0,05$ para ambos algoritmos. Estos resultados están de acuerdo con los obtenidos del análisis de las muestras de validación.

En vista de los resultados obtenidos, la presente metodología puede ser recomendada para realizar análisis de control de calidad de rutina de LOR y PES en esta preparación farmacéutica.

5.6.2.4. Predicción de muestras sintéticas mediante U-PCA/RBL seguido de MLP, RBF y SVM

5.6.2.4.1. Consideraciones generales

La aplicación de la metodología U-PCA/RBL al sistema experimental bajo estudio comprende los siguientes pasos comunes: 1) establecer el número óptimo de *scores* de PCA para entrenar la red neuronal, 2) filtrar la contribución de los componentes no esperados en las muestras incógnitas y 3) entrenar los modelos de redes neuronales diferentes y predecir la concentración del analito de interés en cada caso.

Para estimar el valor de A , se comparan los residuos de la reconstrucción de la matriz de calibración en función del número creciente de *loadings* con la estimación del ruido instrumental del sistema. Generalmente, el residuo del análisis PCA decrece hasta que se hace indistinguible del nivel de ruido establecido mediante una prueba estadística adecuada como, por ejemplo, la prueba F . Esta incluye la comparación de la razón cuadrada (residuo PCA / ruido instrumental) con el valor crítico de F para un nivel de confianza del 95%, usando JK y J (o K) grados de libertad, respectivamente [el ruido instrumental se evalúa mediante réplicas de un espectro del blanco a las J (o K) longitudes de onda espectrales]. Se desprecia la pequeña disminución en los grados de libertad otorgada por la estimación de los A PCs.

Una vez que se estimó el valor de A , se ajusta el número de componentes no esperados mediante una gráfica de los residuos del proceso de RBL en función del valor creciente de N_{noesp} . Nuevamente, cuando los residuos se vuelven comparables al ruido instrumental N_{noesp} alcanza su valor óptimo. Este procedimiento se repite para cada nueva muestra analizada.

Finalmente, una vez estimados A y N_{noesp} , los *scores* de entrenamiento y de las muestras incógnitas se almacenan para el entrenamiento futuro de diferentes modelos ANN.

Como se explica en el apartado 5.6.2.2., el número de neuronas de la capa de entrada para las redes MLP se determinó mediante prueba y error siguiendo la metodología de adición por pasos y eliminación por pasos para LOR y PES, respectivamente (Zupan y Gasteiger 1999). Además, para cada número de prueba de neuronas de entrada, se ensayaron neuronas ocultas en el rango de uno a nueve. Los parámetros para las redes con arquitecturas 4-8-1 (LOR) y 8-7-1 (PES) se almacenaron para predecir las muestras de validación previamente sometidas al proceso de RBL.

Por otra parte, la complejidad de las redes RBF se controló según lo descrito por Orr (Orr 1999). Se seleccionó un subset de centros mediante selección hacia delante, de acuerdo al grado de reducción del error cuadrático, empleando el criterio de la validación cruzada generalizada, cuyo objetivo es minimizar el error cuadrático penalizado GCV:

$$GCV = \frac{l_{cal}}{(l_{cal} - \gamma)^2} SSE \quad (5.18)$$

donde l_{cal} es el número de muestras de calibración, SSE la suma de los errores cuadráticos y γ el número efectivo de parámetros del modelo, según lo definido por Orr (Orr 1999). Finalmente, los anchos gaussianos se ajustaron mediante prueba y error, empezando con un rango de valores de ancho y mejorando gradualmente el error cuadrático hasta que se encontró el ancho adecuado. Para construir de manera adecuada el modelo RBF se siguió la guía descrita por Orr (Orr 1999).

Con respecto a LS-SVM, la optimización se realizó mediante validación cruzada, empleando un procedimiento de búsqueda mediante una grilla con posibles valores de γ y σ^2 . Una vez realizado lo anterior, el número de funciones gaussianas y sus centros, los pesos lineales y los niveles de error se computan automáticamente. En general se adoptó el procedimiento recomendado por Pelckmans y col (Pelckmans y col 2003).

Se debe notar que tanto los parámetros del modelo RBF como del LS-SVM y la complejidad se pueden ajustar empleando los datos de calibración sin la necesidad de preparar un juego independiente de monitorización. Esto constituye una ventaja sobre el procedimiento MLP usual, aunque se puede dividir el juego original de calibración en dos juegos y emplear uno para entrenar y otro para monitorizar. Sin embargo, esta última metodología implica remover muestras del juego de calibración diseñado, lo cual no siempre es conveniente.

5.6.2.4.2. Predicción de las muestras de validación

Como se mencionó anteriormente se construyó una matriz de tamaño 111×101 por muestra (tanto para los estándares como para las muestras incógnitas), aunque se seleccionaron regiones de trabajo apropiadas en ambas dimensiones antes de aplicar calibración multivariada. La Figura 5.10 muestra la superficie correspondiente a la muestra de calibración número 2, la cual se graficó para 54 tiempos FIA/puntos de pH en la región comprendida entre 240 y 300 nm. Luego de seleccionar la mejor región de trabajo, la matriz se redujo a 54×61 datos por muestra. Como se puede observar en la Figura 5.10, existen cambios importantes en el espectro registrado a tiempo igual a 20 comparado con el obtenido a tiempo igual a 1. Estas variaciones pueden ser usadas de manera conveniente para adquirir la ventaja de segundo orden a través del procedimiento RBL.

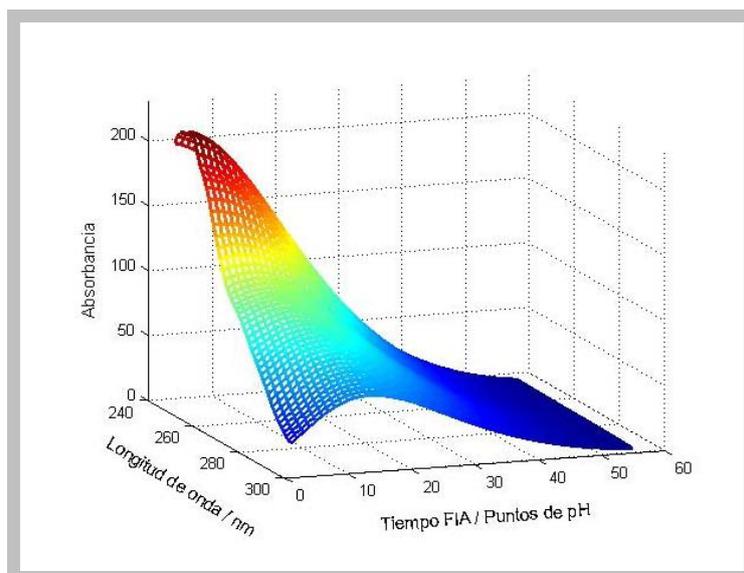


Figura 5.10. Superficie obtenida recogiendo los espectros UV a diferentes puntos de tiempo FIA para una de las muestras de calibración que contenía $7,50 \text{ mg L}^{-1}$ de LOR y $80,00 \text{ mg L}^{-1}$ de PES.

El número de componentes principales (A) que explican la variabilidad espectral del juego de calibración se evaluó a través de la comparación del residuo de U-PCA con el ruido instrumental de la manera discutida anteriormente. El nivel de ruido en este sistema se estimó a través de replicados del blanco como $0,1$ (unidades de absorbancia $\times 10^3$). Luego se calcularon los valores del parámetro F_{exp} (las razones cuadráticas entre los residuos de PCA y el ruido) para los primeros siete como: $527 - 15,2 - 9,0 - 5,8 - 4,8 - 1,4$ y $1,0$; mientras que $F_{\text{crit}(0,05;54 \times 61;54)} = 1,42$. Por lo tanto, el primer número de PCs para el cual F_{exp} es comparable a F_{crit} es 6. Es por esto que se seleccionaron los primeros seis PCs como entradas de las redes neuronales RBF y SVM.

Los resultados obtenidos de la aplicación de los tres modelos de redes neuronales testeados a las muestras de validación, se muestran en la Tabla 5.9. La primera muestra de validación tuvo que ser excluida por otorgar resultados anómalos al ser analizada mediante RBF y SVM. Todos requirieron de la ventaja de segundo orden para realizar una predicción exacta de las muestras. Los parámetros críticos son: el número de PCs necesarios para modelar los datos de calibración, el cual es grande en comparación con el número de analitos debido a no linealidades presentes en el sistema, y el

número de componentes no esperados que deben ser considerados para alcanzar un residuo comparable con el ruido instrumental para cada muestra de validación analizada. Las arquitecturas informadas en la Tabla 5.9 para todas las redes (MLP, RBF y SVM) implican que se necesitan modelos más complejos para el analito que muestra un comportamiento claramente no lineal (pseudoefedrina sulfato), indicado por el número de nodos ocultos alto.

El analito que presenta una relación casi lineal entre la señal y la concentración (loratadina), muestra mejores predicciones con MLP (REP de aproximadamente 4% versus REP de aproximadamente 7% para los otros dos modelos de redes). Para analizar el nivel de validez estadística de la diferencia de los valores de REP, se los comparó empleando una prueba de randomización (van der Voet 1994), que consiste en comparar dos modelos A y B, empleando la distribución de los errores de predicción como una medida de la exactitud. Con este objetivo, se calcula un estadístico $T_{calc} = \bar{d} = (1/n) \sum d_i$, en donde d_i es la diferencia de los errores cuadráticos de predicción para cada muestra, por ambos métodos analizados, es decir, $d_i = e_{Ai}^2 - e_{Bi}^2$ con $i=1, \dots, n$. A continuación, los d_i se someten a cambios de signo de manera aleatoria, un número determinado de veces, y se obtienen diferentes valores de $T = \bar{d}$, que se ordenan de mayor a menor. El nivel de significación se obtiene según $p = k/(m+1)$, donde m es el número de pruebas de randomización realizadas y k es la ubicación del T_{calc} dentro de los valores de randomización de T . Específicamente, los niveles de significación estimados para evaluar si los REPs para RBF y SVM eran mayores que el REP para MLP resultaron <0,01 y 0,03 respectivamente, ambos más bajos que 0,05 y por lo tanto significativos. Este hecho nos permite concluir que los resultados de la predicción obtenidos con las redes MLP son más exactos que los obtenidos mediante RBF y SVM.

Tabla 5.9. Predicciones y análisis estadístico para loratadina y pseudoefedrina sulfato.^a

Muestra	Loratadina ^b / mg L ⁻¹			
	Nominal	MLP	RBF	SVM
1	7,50	8,11	8,22	8,38
2	9,16	9,00	9,46	8,38
3	7,50	7,14	6,95	6,98
4	9,16	9,32	9,33	8,78
5	8,33	7,98	7,52	8,02
RMSEP / mg L ⁻¹		0,37	0,57	0,56
REP / %		4,4	6,7	6,7
Pseudoefedrina sulfato ^c / mg L ⁻¹				
1	90,00	82,80	94,52	101,18
2	90,00	84,96	87,16	87,86
3	110,00	113,44	110,47	112,01
4	110,00	110,15	103,54	99,37
5	100,00	92,93	101,03	101,94
RMSEP / mg L ⁻¹		5,3	3,8	6,5
REP / %		5,3	3,8	6,5

^a En todos los casos se usaron 2 componentes RBL no esperados. Las redes MLP contenían 4 nodos de entrada para LOR y 8 para PES. Las redes RBF y SVM contenían 6 nodos de entrada. Todas las redes contenían 1 nodo de salida.

^b MLP, 8 nodos ocultos; RBF, 2 nodos ocultos, ancho = 2; SVM, $\gamma = 294$, $\sigma^2 = 0,1$.

^c MLP, 7 nodos ocultos; RBF, 21 nodos ocultos, ancho = 1,35; SVM, $\gamma = 341$, $\sigma^2 = 244$.

El hecho de que las cifras de mérito para MPL son mejores se puede asociar a la menor complejidad de la relación señal/concentración para loratadina, indicando que es más probable encontrar una arquitectura MLP apropiada para modelar los datos. Por otra parte, en el caso de pseudoefedrina sulfato, cuya concentración muestra una relación no lineal con la señal, se obtienen resultados más exactos con RBF. Estos resultados sugieren una habilidad de predicción aumentada para este último modelo de red neuronal sobre los demás, en especial cuando los datos muestran un comportamiento no lineal más marcado.

5.7. CONCLUSIONES

La resolución satisfactoria de los diferentes problemas analíticos implica, entre otras cosas, el conocimiento de la naturaleza de la relación entre la señal y la concentración del analito. Su importancia radica en que permite seleccionar un algoritmo adecuado para el tratamiento de los datos.

U-PCA/RBL se puede usar de manera conveniente para filtrar la contribución de componentes no esperados en muestras incógnitas como paso previo a la predicción con diferentes redes neuronales. Esto genera una variedad de posibilidades en lo que concierne al procesamiento satisfactorio de datos de segundo orden no lineales, conduciendo a la obtención de la ventaja de segundo orden. El rendimiento de diferentes modelos de redes neuronales, juzgado a través de la exactitud de las predicciones en un juego de validación independiente, parece depender de la estructura particular de los datos del sistema bajo análisis. Es por esta razón que resulta importante tener disponible una variedad de procedimientos de redes neuronales para procesar datos de segundo orden no lineales, con el objetivo de seleccionar el modelo más apropiado para los datos bajo análisis.

El proceso de modelado se puede simplificar de manera considerable cuando se implementan las redes RBF y SVM, debido a que no requieren de la preparación de un juego de muestras para monitorizar el proceso de entrenamiento. Este aspecto representa una ventaja de estas nuevas metodologías sobre el concepto de la red neuronal MPL tradicional.

CAPÍTULO 6

Conclusiones

Con referencia a los objetivos propuestos para este trabajo de tesis, se concluye que:

- Se desarrollaron métodos de análisis que permitieron determinar especies químicas que formaban parte de muestras complejas, como fármacos y sus metabolitos en materiales biológicos, en preparaciones farmacéuticas y muestras ambientales, y aditivos en muestras alimenticias. Este desafío se pudo concretar satisfactoriamente gracias a la implementación de algoritmos quimiométricos adecuados, que permitieron extraer información relevante de cada muestra y cuantificar los analitos aún en presencia de interferencias no modeladas.
- El algoritmo U-PLS/RBL se aplicó a la resolución de un sistema biológico compuesto por tetraciclina en suero, en presencia de efecto matriz, y mostró ser una alternativa adecuada para extraer información útil de un sistema complejo, con baja sensibilidad con respecto al analito de interés, que requiere estricta adherencia a la ventaja de segundo orden. En términos generales, este tipo de análisis puede sugerirse para ser implementado en estudios de farmacocinética y biodisponibilidad de medicamentos, siendo ventajoso en el sentido de que se evitan pretratamientos de las muestras, muchas veces engorrosos, y permite una resolución exitosa aún en el caso en que se modifiquen las características espectrales del analito entre muestras y de que exista gran superposición espectral, baja sensibilidad, interacción entre componentes y presencia de componentes no esperados.
- De la evaluación del rendimiento analítico del algoritmo MCR-ALS frente a la resolución de sistemas compuestos por datos de segundo orden con distintos grados de solapamiento en la dimensión espectral o temporal, se observó que este algoritmo es capaz de proveer resultados satisfactorios, aún en presencia de solapamiento espectral severo y componentes no modelados. Además, a partir del análisis de los sistemas experimentales, se evidenció la importancia de seleccionar MCR-ALS como algoritmo de procesamiento, empleando el modo de aumento correcto dependiendo del grado de solapamiento espectral. Se pudo confirmar que este es un algoritmo muy

flexible que, de acuerdo a las características que presenten los datos, posibilita la aplicación de múltiples estrategias para lograr los objetivos y con la ventaja de proporcionar información con significado físico. Debido a los muy buenos resultados obtenidos, las metodologías desarrolladas podrían ser implementadas en el control de calidad de los aditivos tartrazina y azul brillante en productos alimenticios (gelatinas), y en la determinación de CBZ en estudios de farmacocinética y biodisponibilidad de medicamentos.

- De la comparación del algoritmo U-PLS/RBL frente a MCR-ALS en la resolución de un sistema compuesto por ocho tetraciclinas presentes como contaminantes en agua residual, cuyos datos presentaban gran complejidad y leve desviación a la trilinealidad, se pudo establecer que U-PLS/RBL, haciendo uso de su estructura de variable latente, puede manejar este tipo de datos generando resultados de predicción comparables a los obtenidos con MCR-ALS.

- Finalmente, se evaluó una nueva herramienta quimiométrica: U-PCA/RBL. A través del análisis de dos sistemas experimentales se demostró que este algoritmo se puede utilizar de manera conveniente para filtrar la contribución de componentes no esperados en muestras incógnitas, como paso previo a la predicción con diferentes redes neuronales. Esto genera una variedad de posibilidades en lo que concierne al procesamiento satisfactorio de datos de segundo orden no lineales, conduciendo a la obtención de la ventaja de segundo orden. Qué tipo de red neuronal (MLP, RBF, SVM) presentará mejor rendimiento dependerá de la estructura particular de los datos del sistema bajo análisis. En general, el proceso de modelado se puede simplificar de manera considerable cuando se implementan las redes RBF y SVM, debido a que no requieren de la preparación de un juego de muestras para monitorizar el proceso de entrenamiento. Este aspecto representa una ventaja de estas nuevas metodologías sobre el concepto de la red neuronal MPL tradicional. Los muy buenos resultados obtenidos al cuantificar Fe(II) mediante U-PLS/RBL-MLP sugieren que la técnica podría ser extendida a la determinación de este analito en mezclas complejas, en presencia de otras interferencias, como fluidos biológicos o productos alimenticios. Por otra parte, los

resultados alcanzados para el sistema loratadina-pseudoefedrina sulfato hacen recomendable el empleo de la metodología desarrollada en el control de calidad de rutina de preparaciones farmacéuticas.

- En el transcurso de la tesis se demostró que actualmente es factible proponer metodologías de análisis alternativas a las oficiales, que muchas veces requieren de instrumental complejo, grandes cantidades de solventes, largos tiempos de análisis de la muestra y costo económico elevado, y obtener resultados de rendimiento comparable. Se arribó a esta conclusión a través de la validación de las metodologías propuestas aplicadas a muestras artificiales y reales, lo cual implicó el cálculo de cifras de mérito y la aplicación de pruebas estadísticas apropiadas a cada análisis.

CAPÍTULO 7

Resumen

En la presente tesis se desarrollaron métodos de análisis que permitieron determinar especies químicas que formaban parte de muestras complejas, como fármacos y sus metabolitos en materiales biológicos, en preparaciones farmacéuticas y muestras ambientales, y aditivos en muestras alimenticias. Este desafío se pudo concretar satisfactoriamente gracias a la implementación de algoritmos quimiométricos adecuados, que permitieron extraer información relevante de cada muestra y cuantificar los analitos, aún en presencia de interferencias no modeladas. El eje principal de este trabajo se centró en la utilización de datos de segundo orden, procesados mediante algoritmos quimiométricos que aportaran la ventaja de segundo orden, con el objetivo de desarrollar metodologías que permitieran resolver de manera eficiente los diferentes problemas analíticos.

Los sistemas analizados involucran el uso de datos de segundo orden que, en algunos casos, presentan una relación señal-concentración lineal y en otros no lineal, con respecto al analito de interés. La clase de datos implicados condicionó el tipo de modelo implementado, el que fue seleccionado entre los algoritmos PARAFAC, U-PLS/RBL y MCR-ALS para resolver sistemas lineales, y U-PCA/RBL, seguido de la aplicación de diferentes redes neuronales artificiales, para resolver sistemas no lineales.

Concretamente, el algoritmo U-PLS/RBL se implementó para procesar datos experimentales de excitación-emisión de fluorescencia correspondientes a un sistema biomédico con baja sensibilidad con respecto al analito de interés, que requirió de la ventaja de segundo orden. Las muestras estaban compuestas por diferentes sueros humanos adicionados con el antibiótico tetraciclina. A raíz de que la interacción con el suero modifica de manera significativa las propiedades espectrales del analito, la calibración debió incluir a la matriz biológica. El rendimiento de U-PLS/RBL se comparó frente al clásico PARAFAC en muestras de validación preparadas en sueros diferentes a los que se usaron para calibrar y en presencia de drogas adicionadas, debido a que se planteó la hipótesis de que éstas podrían actuar como posibles interferencias. Se demostró que el nuevo algoritmo representa una interesante alternativa para procesar este tipo de datos y obtener resultados precisos y exactos.

Por otra parte, se evaluó el rendimiento del algoritmo MCR-ALS en la resolución de sistemas compuestos por datos de segundo orden con distintos grados de solapamiento en las dimensiones espectral o temporal. Para tal fin se estudiaron dos sistemas espectroscópicos-cinéticos y uno espectroscópico-electroforético: 1) la oxidación de naranja de metilo con bromato de potasio, catalizada por Fe(II) (el analito), en presencia de Zn(II) como interferencia, el cual también cataliza la reacción pero con diferente cinética, 2) la reacción de tartrazina (el analito) con bromato de potasio, catalizada por Fe(II) en presencia de azul brillante como interferencia, el cual reacciona produciendo un compuesto con un espectro de absorción muy similar al de la tartrazina y 3) un sistema en donde el analito de interés (carbamazepina) se encuentra solapado en la dimensión temporal con algunos de los componentes de la muestra. Los resultados mostraron un buen rendimiento analítico con respecto al analito de interés, aún en presencia de solapamiento espectral severo y componentes no modelados. Además, del análisis de estos sistemas experimentales, se evidenció la importancia de seleccionar MCR-ALS como algoritmo de procesamiento, empleando el modo de aumento correcto dependiendo del grado de solapamiento espectral.

A continuación, se presentó la primera aplicación de U-PLS/RBL a datos cromatográficos, con el objetivo de estudiar las potencialidades de su estructura de variable latente, especialmente en casos en donde, al menos en principio, puede proveer resultados comparables con el clásico MCR-ALS. Los datos experimentales correspondieron a corridas cromatográficas con detección mediante arreglo de diodos de muestras de agua residual que presentaban interferencias muy significativas, propias de las muestras medioambientales, debidas a la presencia de compuestos que coeluyen con el analito de interés. El rendimiento analítico se mejoró a través del pretratamiento de los datos, el cual incluyó la eliminación de la línea de base del cromatograma y el empleo de estandarización para compensar el factor de recuperación aplicado cuando la muestra se preconcentró. Se pudo establecer que U-PLS/RBL puede manejar este tipo de datos y generar resultados de predicción comparables a los obtenidos con MCR-ALS.

Para finalizar, se evaluó el algoritmo U-PCA/RBL en el análisis de datos experimentales de segundo orden no lineales. Este algoritmo es una nueva herramienta quimiométrica que se puede utilizar de manera conveniente para filtrar la contribución de componentes no esperados en muestras incógnitas, como paso previo a la predicción con diferentes redes neuronales. En principio se estudió la aplicabilidad de U-PCA/RBL-MLP a la resolución de dos sistemas experimentales en donde los analitos calibrados se desvían de la relación lineal concentración-respuesta. En el primer sistema, el Fe(II) actúa como catalizador de la oxidación del colorante naranja de metilo, siguiendo una cinética no lineal. Cabe aclarar que este sistema es similar al empleado cuando se evaluó el rendimiento del algoritmo MCR-ALS, pero debido a que las concentraciones en las que se encontraba el Fe(II) eran mayores, en esa ocasión el sistema resultó lineal. Este comportamiento se utilizó con el fin de determinar Fe(II) en muestras de naranja de metilo, en presencia del colorante púrpura de bromocresol como interferencia. El segundo sistema consistió en una preparación farmacéutica que contiene loratadina y pseudoefedrina sulfato. Este último analito se encuentra en una concentración más alta en comparación con el otro, lo cual introduce no linealidades al querer determinarlos de manera conjunta. Se generaron datos de segundo orden no lineales empleando un sistema de análisis por inyección en flujo con gradiente de pH y detección mediante arreglo de diodos. Debido a que las redes MLP construyen modelos de aproximación global a relaciones de entrada-salida no lineales, mientras que las redes RBF y SVM, empleando funciones gaussianas, tienden a construir aproximaciones locales, resultó de interés investigar si los nuevos modelos no lineales RBF y SVM podían aproximar mejor las no linealidades presentes en este último sistema. Por lo tanto, una vez filtradas las contribuciones de los componentes no esperados en las muestras mediante U-PCA/RBL, se modeló la relación *score*-concentración de los datos con las redes neuronales artificiales RBF y SVM. El rendimiento de los diferentes modelos de redes neuronales, juzgado a través de la exactitud de las predicciones en un juego de validación independiente, parece depender de la estructura particular de los datos del sistema bajo análisis.

En el transcurso de este trabajo se demostró que actualmente es factible proponer metodologías de análisis alternativas a las oficiales, que muchas veces requieren de instrumental complejo, grandes cantidades de solventes, largos tiempos de análisis de la muestra y costo económico elevado, y obtener resultados de rendimiento comparable. Se arribó a esta conclusión a través de la validación de las metodologías propuestas aplicadas a muestras artificiales y reales, lo cual implicó el cálculo de cifras de mérito y la aplicación de pruebas estadísticas apropiadas a cada análisis.

CAPÍTULO 8

Bibliografía

Absalan, G.; Safavi, A.; Maesum, S. (2001). Application of artificial neural networks as a technique for interference removal: kinetic-spectrophotometric determination of trace amounts of Se(IV) in the presence of Te(IV). *Talanta* 55: 1227 - 1233.

Andersen, C.M.; Bro, R. (2003). Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *J. Chemom.* 17: 200 - 215.

AOAC. (1993). Peer-verified method program, manual on policies and procedures. AOAC, Arlington, VA, USA.

Arancibia, J.A.; Escandar, G.M. (2003). Two different strategies for the fluorimetric determination of piroxicam in serum. *Talanta* 60: 1113 - 1121.

Arancibia, J.A.; Olivieri, A.C.; Bohoyo Gil, D.; Muñoz de la Peña, A.; Durán-Merás, I.; Espinosa Mansilla, A. (2006). Trilinear least-squares and unfolded-PLS coupled to residual trilinearization: New chemometric tools for the analysis of four-way instrumental data. *Chemom. Intell. Lab. Syst.* 80: 77 - 86.

Arancibia, J.A.; Olivieri A.C.; Escandar, G.M. (2002). First- and second-order multivariate calibration applied to biological samples: determination of anti-inflammatories in serum and urine. *Anal. Bioanal. Chem.* 374: 451 - 459.

Azubel, M.; Fernández, F.M.; Tudino, M.B.; Trocoli, O.E. (1999). Novel application and comparison of multivariate calibration for the simultaneous determination of Cu, Zn and Mn at trace levels using flow injection diode array spectrophotometry. *Anal. Chim. Acta* 398: 93 - 102.

Beltran, J.L.; Guiteras, J.; Ferrer, R. (1998). Three-Way Multivariate Calibration Procedures Applied To High-Performance Liquid Chromatography Coupled with Fast-Scanning Fluorescence Spectrometry Detection. Determination of Polycyclic Aromatic Hydrocarbons in Water Samples. *Anal. Chem.* 70: 1949 - 1955.

Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; Redón, M. (1995). Artificial Neural Networks for Multicomponent Kinetic Determinations. *Anal. Chem.* 67: 4477 - 4483

Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; Redon, M. (1995). Partial least-squares regression for multicomponent kinetic determinations in linear and non-linear systems. *Anal. Chim. Acta* 303: 309 - 320.

Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; Redon, M.; Villegas, N. (1996). Artificial neural networks and partial least squares regression for pseudo-first-order with respect to the reagent multicomponent kinetic-spectrophotometric determinations *Analyst* 121: 395 - 400.

Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; Riba, J. (1994). Kinetic spectrophotometric method for analyzing mixtures of metal ions by stopped-flow injection analysis using partial least-squares regression. *Anal. Chem.* 66: 2905 - 2911.

Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; Riba, J.; Rovira, E. (1993). Kinetic spectrophotometric determination of Ga(III)-Al(III) mixtures by stopped-flow injection analysis using principal component regression. *Talanta* 40: 261 - 267.

Booksh, K. S.; Kowalski, B. R. (1994). Theory of Analytical Chemistry. *Anal. Chem.* 66: 782A - 791A.

Borin, A.; Ferrão, M.F.; Mello, C.; Maretto, D.A.; Poppi, R.J. (2006). Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. *Anal. Chim. Acta* 579: 25 - 32.

Bosco, M.V.; Garrido, M.; Larrechi, M.S. (2006). Determination of phenol in the presence of its principal degradation products in water during a TiO₂-

photocatalytic degradation process by three-dimensional excitation-emission matrix fluorescence and parallel factor analysis. *Anal. Chim. Acta* 559: 240 - 247.

Boudart, M. (1968). *Kinetics of chemical processes*. Prentice-Hall, Englewood, USA. 124 - 127.

Bouhsain, Z.; Garrigues, S.; de la Guardia, M. (1996). Simultaneous stopped-flow determination of paracetamol, acetylsalicylic acid and caffeine in pharmaceutical formulations by Fourier transform infrared spectrometry with partial least-squares data treatment. *Analyst* 121: 1935 - 1938.

Braga, J.W.B.; Bottoli, C.B.G.; Jardim, I.C.S.F.; Goicoechea, H.C.; Olivieri, A.C.; Poppi, R.J. (2007). Determination of pesticides and metabolites in wine by high performance liquid chromatography and second-order calibration methods. *J. Chromatogr. A* 1148: 200 - 210.

Brereton, R.G. (1990). *Chemometrics: Applications of Mathematics and Statistics to Laboratory Systems*. New York. Ellis Horwood. Capítulo 1.

Bro, R. (1998). *Multi-way analysis in the food industry. Models, algorithms and applications*. Doctoral Thesis. University of Amsterdam. The Netherlands.

Bro, R. (1996). Multiway calibration. *Multilinear PLS. J. Chemom.* 10: 47 - 61.

Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemom. Intell. Lab. Syst.* 38: 149 - 171.

Bro, R.; Kiers, H.A.L. (2003). A new efficient method for determining the number of components in PARAFAC models. *J. Chemom.* 17: 274 - 286.

Bro, R.; Sidiropoulos, N.D.; Smilde, A.K. (2002). Maximum likelihood fitting using simple least squares algorithms. *J. Chemom.* 16: 387 - 400.

Brunton, L.; Lazo J.; Parker K. (2005). Goodman and Gilman's The Pharmacological Basis of Therapeutics. McGraw-Hill Professional, New York.

Cao, Y.Z.; Mo, C.Y.; Long, J.G.; Chen, H.; Wu, H.L.; Yu, R.Q. (2002). Spectrofluorometric resolution of closely overlapping drug mixtures using chemometrics methods. *Anal. Sci.* 18: 333 - 336.

Carlin, M.; Kavli; T.; Lillekjendlie, B. (1994) A comparison of four methods for non-linear data modelling. *Chemom. Intell. Lab. Syst.* 23: 163 - 177.

Carrol, J.D.; Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika* 35: 283 - 319.

Center, V.; de Noord, O.E.; Massart, D.L. (1998). Detection of nonlinearity in multivariate calibration. *Anal. Chim. Acta* 376: 153 - 168.

Chauchard, F.; Cogdill, R.; Roussel, S.; Roger, J.M.; Bellon-Maurel, V. (2004). Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes. *Chemom. Intell. Lab. Syst.* 71: 141 - 150.

Chen, Z.P., Wu, H.L.; Jiang, J.H.; Li, Y.; Yu, R.Q. (2000). A novel trilinear decomposition algorithm for second-order linear calibration. *Chemom. Intell. Lab. Syst.* 52: 75 - 86.

Christensen, J.; Miquel Becker, E.; Frederiksen, C.S. (2005). Fluorescence spectroscopy and PARAFAC in the analysis of yogurt. *Chemom. Intell. Lab. Syst.* 75: 201 - 208.

Choi, K.J.; Kim, S.G.; Kim, C.W.; Kim, S.H. (2007). Determination of antibiotic compounds in water by on-line SPE-LC/MSD. *Chemosphere* 66: 977 - 984.

Choi, B.K.; Hercules, D.M.; Gusev, A.I. (2001). LC-MS/MS signal suppression effects in the analysis of pesticides in complex environmental matrices. *Fresenius J. Anal. Chem.* 369: 370 - 377.

Cinquina, A.L.; Longo, F.; Anastasi, G.; Giannetti, L.; Cozzani, R. (2003). Validation of a high-performance liquid chromatography method for the determination of oxytetracycline, tetracycline, chlortetracycline and doxycycline in bovine milk and muscle. *J. Chromatogr. A* 987: 227 - 233.

Cladera, A.; Gómez, E.; Estela, J.M.; Cerdá, J.; Cerdá, V. (1993a). Computer method for the simultaneous kinetic determination of compounds in mixtures based on the use of diode-array spectrophotometry. *Anal. Chim. Acta* 272: 339 - 344.

Cladera, A.; Gómez, E.; Estela, J.M.; Cerdá, V. (1993b). Resolution of simultaneous kinetic spectrophotometric processes by factor analysis. *Anal. Chem.* 65: 707 - 715.

Coello, J.; Maspoch, S.; Villegas, N. (2000). Simultaneous kinetic-spectrophotometric determination of levodopa and benserazide by bi- and three-way partial least squares calibration. *Talanta* 53: 627 - 637.

Cogdill, R.P.; Schimleck, L.R.; Jones, P.D.; Peter, G.F.; Daniels, R.F.; Clark, A. (2004). Estimation of the physical wood properties of *Pinus taeda* L. radial strips using least squares support vector machines. *J. Near Infrared Spectrosc.* 12: 263 - 270.

Comas, E.; Gimeno, R.A.; Ferré, J.; Marcé, R.M.N.; Borrull, F.; Rius, FX. (2004). Quantification from highly drifted and overlapped chromatographic peaks using second-order calibration methods. *J. Chromatogr. A* 1035: 195 - 202.

Cuadros Rodríguez, L.; García Campaña, A.M.; Jimenez Linares, C.; Román Ceba, M. (1993). Estimation of performance characteristics of an Analytical Method using the data set of the calibration experiment. *Anal. Lett.* 26: 1243 - 1258.

Currie, L.A. (1999). Nomenclature in evaluation of analytical methods including detection and quantification capabilities: (IUPAC Recommendations 1995). *Anal. Chim. Acta.* 391: 105 - 126.

Damiani, P.C.; Nepote, J.A.; Bearzotti, M.; Olivieri A.C. (2004). A test field for the second-order advantage in bilinear least-squares and parallel factor analyses: fluorescence determination of ciprofloxacin in human urine. *Anal. Chem.* 76: 2798 -2806.

Danzer, K.; Currie, L.A (1998). Guidelines for calibration in analytical chemistry. Part I. Fundamentals and single component calibration (IUPAC Recommendations 1998). *Pure Appl. Chem.* 70: 993 - 1014.

Daszykowski, M.; Walczak, B. (2006). Use and abuse of chemometrics in chromatography. *Trends Anal. Chem.* 25: 1081 - 1096.

De Juan, A.; Casassas, E.; Tauler, R. (2000). Soft modeling of analytical data. Meyers, R.A. *Encyclopedia of Analytical Chemistry*. Wiley, Chichester, UK. 9800 - 9837.

De Juan, A.; Tauler, R. (2003). Chemometrics applied to unravel multicomponent processes and mixtures: Revisiting latest trends in multivariate resolution. *Anal. Chim. Acta* 500: 195 - 210.

de la Guardia, M.; Khalaf, K.D.; Hasan, B.A.; Morales-Rubio, A.; Arias, J.J.; García-Fraga, J.M.; Jiménez, A.I.; Jiménez, F. (1996). Simultaneous kinetic spectrophotometric determination of five phenolic compounds by reaction with p-aminophenol, using partial least squares data treatment. *Analyst* 121:

1321 - 1326.

Derks, E.P.P.A.; Sanchez Pastor, M.S.; Buydens, L.M.C. (1995). Robustness analysis of radial base function and multi-layered feed-forward neural network models. *Chemom. Intell. Lab. Syst.* 28: 49 - 60.

Despagne, F.; Massart, D.L. (1998). Neural networks in multivariate calibration. *Analyst* 123: 157R - 178R.

Drapper, N.R.; Smith, H. (1981). *Applied Regression Analysis*, 2nd Edition. Wiley, New York, USA.

Durán-Merás, I.; Espinosa Mansilla, A.; Salinas López, F. (1995). Simultaneous kinetic spectrophotometric determination of 2-furfuraldehyde and 5-hydroxymethyl-2-furfuraldehyde by application of a modified Winkler's method and partial least squares calibration. *Analyst* 120: 2567 - 2571.

Eilers, P.H.C. (2004). Parametric Time Warping. *Anal. Chem.* 76: 404 - 411.

Eilers, P.H.C.; Currie, I.D.; Durbán, M. (2006). Fast and compact smoothing on large multidimensional grids. *Comp. Stat. Data Anal.* 50: 61 - 76.

Escandar, G.M.; Faber, N.M.; Goicoechea, H.C.; Muñoz de la Peña, A.; Olivieri, A.C.; Poppi, R.J. (2007). Second- and third-order multivariate calibration: data, algorithms and applications. *Trends Anal. Chem.* 26: 752 - 765.

Escandar, G.M.; González Gómez, D.; Espinosa Mansilla, A.; Muñoz de la Peña, A.; Goicoechea, H.C. (2004). Determination of carbamazepine in serum and pharmaceutical preparations using immobilization on a nylon support and fluorescence detection. *Anal. Chim. Acta.* 506: 161 - 170.

Espinosa Mansilla, A.; Muñoz de la Peña, A.; Goicoechea, H.C.; Olivieri, A.C. (2004). Two Multivariate Strategies Applied to Three-Way Kinetic Spectrophotometric Data for the Determination of Mixtures of the Pesticides Carbaryl and Chlorpyrifos. *Appl. Spectrosc.* 58: 83 - 90.

Esteves da Silva, J.C.G.; Litao, J.M.M.; Costa, F.S.; Ribeiro, J.L.A. (2002). Detection of verapamil drug by fluorescence and trilinear decomposition techniques. *Anal. Chim. Acta.* 453: 105 - 115.

Esteves da Silva, J.C.G.; Oliveira, C.J.S. (1999). Parafac decomposition of three-way kinetic-spectrophotometric spectral matrices corresponding to mixtures of heavy metal ions. *Talanta.* 49: 889 - 897.

Evgeniou, T.; Pontil, M.; Poggio, T. (2000). Regularization Networks and Support Vector Machines. *Adv. Comput. Math.* 13: 1 - 50.

Faber, N(K). M.; Bro, R; Hopke, P.K. (2003). Recent developments in CANDECOMP/PARAFAC algorithms: a critical review. *Chemom. Intell. Lab. Syst.* 65: 119 - 137.

Feudale, R.N.; Woody, N.A.; Tan, H.; Myles, A.J.; Brown, S.D.; Ferré, J. (2002) Transfer of multivariate calibration models: a review. *Chemom. Intell. Lab. Syst.* 64: 181 - 192.

Fidêncio, P.H.; Poppi, R.J.; Carlos de Andrade, J. (2002) Determination of organic matter in soils using radial basis function networks and near infrared spectroscopy. *Anal. Chim. Acta.* 453: 125 - 134.

Fischbacher, C.; Jagemann, K.U.; Danzer, K.; Muller, U.A.; Papenkorcht, L.; Schuler, J. (1997). Enhancing calibration models for noninvasive near-infrared spectroscopic blood glucose determinations. *Fresenius J. Anal. Chem.* 359: 78 - 82.

Fraga, C.G.; Corley, C.A. (2005). The chemometric resolution and quantification of overlapped peaks from comprehensive two-dimensional liquid chromatography. *J. Chromatogr. A* 1096: 40 - 49.

Fraga, C.G.; Prazen, B.J.; Synovec, R.E. (2001). Objective Data Alignment and Chemometric Analysis of Comprehensive Two-Dimensional Separations with Run-to-Run Peak Shifting on Both Dimensions. *Anal. Chem.* 73: 5833 - 5840.

Frank, I.E. (1995). Modern nonlinear regression methods. *Chemom. Intell. Lab. Syst.* 27: 1 - 9.

Friedman, J.H. (1984). Variable Span Smoother. Technical Report No. 5. Department of Statistics. Stanford University.

García, I.; Ortiz, M.C.; Sarabia, L.; Aldama, J.M. (2007). Validation of an analytical method to determine sulfamides in kidney by HPLC-DAD and PARAFAC2 with first-order derivative chromatograms. *Anal. Chim. Acta* 587: 222 - 234.

García, J.M.; Jiménez, A.I.; Arias, J.J.; Khalaf, K.D.; Morales-Rubio, A.; de la Guardia, M. (1995). Application of the partial least-squares calibration method to the simultaneous kinetic determination of propoxur, carbaryl, ethiofencarb and formetanate. *Analyst* 120: 313 - 317.

García-Reiriz, A.; Damiani, P.C.; Olivieri, A.C. (2007). Analysis of amoxicillin in human urine by photo-activated generation of fluorescence excitation-emission matrices and artificial neural networks combined with residual bilinearization. *Anal. Chim. Acta* 588: 192 - 199.

García-Reiriz, A.; Damiani, P.C.; Olivieri, A.C. (2007). Different strategies for the direct determination of amoxicillin in human urine by second-order

multivariate analysis of kinetic-spectrophotometric data. *Talanta* 71: 806 - 815.

Gergov, M.; Ojanpera, I.; Vuori, E. (2003). Simultaneous screening for 238 drugs in blood by liquid chromatography-ionspray tandem mass spectrometry with multiple-reaction monitoring. *J Chromatogr. B* 795: 41 - 53.

Gimenez, D.; Sarabia, L.; Ortiz, M.C. (2005). The maintenance of a PARAFAC calibration and the second-order property: application to the determination of ciprofloxacin in presence of enrofloxacin by excitation-emission fluorescence. *Anal. Chim. Acta* 544: 327 - 336.

Goicoechea, H.C.; Collado, M.S.; Satuf, M.L.; Olivieri, A.C. (2002). Complementary use of partial least-squares and artificial neural networks for the non-linear spectrophotometric analysis of pharmaceutical samples. *Anal. Bioanal. Chem.* 374: 460 - 465.

Goicoechea, H. C.; Olivieri, A. C. (1999). Enhanced synchronous spectrofluorimetric determination of tetracycline in blood serum by chemometric analysis. Comparison of partial least-squares and hybrid linear analysis calibrations. *Anal. Chem.* 71: 4361 - 4368.

Goicoechea, H.C.; Olivieri, A.C. (2005). New Robust Bilinear Least Squares Method for the Analysis of Spectral-pH Matrix Data. *Appl. Spectrosc.* 59: 926 - 933.

Guterres, M.V.; Volpe, P.L.; Ferreira, M.C. (2004). Multiway Calibration for Creatinine Determination in Human Serum Using the Jaffé Reaction. *Appl. Spectrosc.* 58: 54 - 60.

Haaland, D.M.; Thomas, E.V. (1988). Partial Least-Squares Methods for Spectral Analyses. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information. *Anal. Chem.* 60: 1193 - 1202.

Hamscher, G.; Sezesny, S.; Hoper, H.; Nau, H. (2002). Determination of Persistent Tetracycline Residues in Soil Fertilized with Liquid Manure by High-Performance Liquid Chromatography with Electrospray Ionization Tandem Mass Spectrometry *Anal. Chem.* 74: 1509 - 1518.

Hanrahan, G.; Montes, R.; Gomez, F.A. (2008). Chemometric experimental design based optimization techniques in capillary electrophoresis: a critical review of modern applications *Anal. Bional. Chem.* 390: 169 - 179.

Harshman, R.A. (1972). Determination and proof of minimum uniqueness conditions for PARAFAC. *UCLA Work. Pap. Phon.* 22: 111 - 117.

Harshman, R.A. (1970). Foundations of the PARAFAC procedure: model and conditions for an 'explanatory' multi-mode factor analysis. *UCLA Work. Pap. Phon.* 6: 1 - 84.

Havel, J.; Jiménez, F.; Bautista, R.D.; Arias León, J.J. (1993). Evaluation of multicomponent kinetic analysis data by a partial least squares calibration method. *Analyst* 118: 1355 - 1360.

Haykin, S. (1999). *Neural networks. A comprehensive foundation.* 2nd Edition. Prentice-Hall, Upper Saddle River, NJ.

Hergert, L.A.; Escandar, G.M. (2003) Spectrofluorimetric study of the α -cyclodextrin-ibuprofen complex and determination of ibuprofen in pharmaceutical preparations and serum. *Talanta* 60: 235 - 246.

Hirsch, R.; Ternes, T.; Haberer, K.; Kratz, K. (1999). Occurrence of antibiotics in the aquatic environment. *Sci. Total Environ.* 225: 109 - 118.

Hu, L.Q, Wu, H.L.; Ding, Y.J.; Fang, D.M.; Xia, A.L.; Yu, R.Q. (2006). Alternating asymmetric trilinear decomposition for three-way data arrays analysis. *Chemom. Itell. Lab. Syst.* 82: 145 - 153.

ICH, Harmonised Tripartite Guideline, Q2A and Q2B, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceutical for Human Use, 1994.

István, K.; Rajkó, R.; Keresztury, G. (2006). Towards the solution of the eluent elimination problem in high-performance liquid chromatography-infrared spectroscopy measurements by chemometric methods. *J. Chromatogr. A* 1104: 154 - 163.

Jaumot, J.; Gargallo, R.; De Juan, A.; Tauler, R. (2005). A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemom. Intell. Lab. Syst.* 76: 101 - 110.

Jiao, L.; Gao, S.; Zhang, F.; Li, H. (2008). Quantification of components in overlapping peaks from capillary electrophoresis by using continuous wavelet transform method. *Talanta* 75: 1061 - 1067.

Kaczmarek, K.; Walczak, B.; de Jong, S.; Vandeginste, B.G.M. (2005). Baseline reduction in two dimensional gel electrophoresis images. *Acta Chromatogr.* 15: 82 - 95.

Karthikeyan, K.G.; Meyer, M.T. (2006). Occurrence of antibiotics in wastewater treatment facilities in Wisconsin, USA. *Sci. Total Environ.* 361: 196 - 207.

Kiers, H.A.L. (1997). Weighted least squares fitting using ordinary least squares algorithms. *Psychometrika* 62: 251 - 266.

Kim, S.C.; Carlson, K. (2007). Quantification of human and veterinary antibiotics in water and sediment using SPE/LC/MS/MS. *Anal. Bioanal. Chem.* 387: 1301 - 1315.

Kloepfer, A.; Quintana, J.B.; Reemtsma, T. (2005). Operational options to reduce matrix effects in liquid chromatography-electrospray ionisation-mass spectrometry analysis of aqueous environmental samples. *J. Chromatogr. A* 1067: 153 - 160.

Kolpin, D. W.; Furlong, E.T.; Meyer, M.T.; Thurman, E.M.; Zaugg, S.D.; Barber, L.B.; Buxton, H.T. (2002). Pharmaceuticals, Hormones, and Other Organic Wastewater Contaminants in U.S. Streams, 1999 - 2000: A National Reconnaissance. *Environ. Sci. Technol.* 36: 1202 -1211.

Kraly, J.; Abul Fazal, Md.; Schoenherr, R.M.; Bonn, R.; Harwood, M.M.; Turner, E.; Jones, M.; Dovichi, N.J. (2006). Bioanalytical Applications of Capillary Electrophoresis *Anal. Chem.* 78: 4097 - 4110.

Li, H.; Hou, J.; Wang, K.; Zhang, F. (2006). Resolution of multicomponent overlapped peaks: A comparison of several curve resolution methods. *Talanta* 70: 336 - 343.

Li, H.; Zhang, F.; Havel, J. (2003). Quantification of analytes in overlapping peaks from capillary electrophoresis using multivariate curve resolution-alternating least squares methods. *Electrophoresis* 24: 3107 - 3115.

Linder, M.; Sundberg, R. (2002). Precision of prediction in second-order calibration, with focus on bilinear regression methods. *J. Chemom.* 16: 12 - 27.

Linder, M.; Sundberg, R. (1998). Second-order calibration: bilinear least squares regression and a simple alternative. *Chemom. Intell. Lab. Syst.* 421: 59 - 178.

Lindsey, M.L.; Meyer, M.T.; Thurman, E.M. (2001). Analysis of Trace Levels of Sulfonamide and Tetracycline Antimicrobials in Groundwater and Surface

Water Using Solid-Phase Extraction and Liquid Chromatography/Mass Spectrometry. *Anal. Chem.* 72: 4640 - 4646.

Lopes, J.A.; Menezes, J.C. (2004). Multivariate monitoring of fermentation processes with non-linear modelling methods. *Anal. Chim. Acta* 515: 101 - 108.

Magni, D.M.; Olivieri, A.C.; Bonivardi, A.L. (2005). Artificial neural networks study of the catalytic reduction of resazurin: stopped-flow injection kinetic-spectrophotometric determination of Cu(II) and Ni(II). *Anal. Chim. Acta* 528: 275 - 284.

Mabrouk, M.M.; El-Fataty, H.M.; Hammad, S.; Aziz, A.; Wahbi, A.A.M. (2003). Simultaneous determination of pseudoephedrine sulfate, dexbrompheniramine maleate and loratadine in pharmaceutical preparations using derivative spectrophotometry and ratio spectra derivative spectrophotometry. *J. Pharm. Biomed. Anal.* 33: 597 - 604.

Mahedero, M.C.; Mora Díaz, N.; Muñoz de la Peña, A.; Espinosa Mansilla, A.; González Gómez, D.; Bohoyo Gil, D. (2005). Strategies for solving matrix effects in the analysis of sulfathiazole in honey samples using three-way photochemically induced fluorescence data. *Talanta* 65: 806 - 813.

Marsili, N.R.; Lista, A.; Fernandez Band, B.S.; Goicoechea, H.C.; Olivieri, A.C. (2005). Evaluation of complex spectral-pH three-way arrays by modified bilinear least-squares: determination of four different dyes in interfering systems. *Analyst* 130: 1291 - 1298.

MATLAB 6.0 (R12), The MathWorks Inc., Natick, Massachusetts, USA, 2000.

MATLAB 7.0 (R14), The MathWorks Inc., Natick, Massachusetts, USA, 2005.

Martens, H.; Naes, T. (1989). *Multivariate Calibration*. Wiley, Chichester, UK. 116 - 163.

Mas, S.; Fonrodona, G.; Tauler, R.; Barbosa, J. (2007). Determination of phenolic acids in strawberry samples by means of fast liquid chromatography and multivariate curve resolution methods. *Talanta* 71: 1455 - 1463.

Massart, D.L.; Vandeginste, B.G.M.; Buydens, L.MC.; De Jong, S.; Lewi, P.J.; Smeyers-Verbeke, J. (1997). *Handbook of Chemometrics and Qualimetrics*. Amsterdam. Elsevier. Capítulo 1.

Miao, X.S.; Bishay, F.; Chen, M.; Metcalfe, C.D. (2004). Occurrence of Antimicrobials in the Final Effluents of Wastewater Treatment Plants in Canada. *Environ. Sci. Technol.* 38: 3533 - 3541.

Moberg, L.; Robertsson, G.; Karlberg, B. (2001). Spectrofluorimetric determination of chlorophylls and pheopigments using parallel factor analysis. *Talanta* 54: 161 - 170.

Montgomery, D.C. (1991). *Design and analysis of experiments*. Wiley, New York, USA.

Mottola, H. (1988). *Kinetic Aspects of Analytical Chemistry*, Wiley, New York, USA.

Muñoz, G.; De Juan A. (2007). pH- and time-dependent hemoglobin transitions: A case study for process modelling. *Anal. Chim. Acta* 595: 198 - 208.

Muñoz de la Peña, A.; Espinosa-Mansilla, A.; Acedo Valenzuela, M.I.; Goicoechea, H.C.; Olivieri, A.C. (2002). Comparative study of net analyte signal-based methods and partial least squares for the simultaneous

determination of amoxicillin and clavulanic acid by stopped-flow kinetic analysis. *Anal. Chim. Acta* 463: 75 - 88.

Muñoz de la Peña, A.; Espinosa Mansilla, A.; González Gómez, D.; Olivieri, A.C.; Goicoechea, H.C. (2003). Interference-Free Analysis Using Three-Way Fluorescence Data and the Parallel Factor Model. Determination of Fluoroquinolone Antibiotics in Human Serum. *Anal. Chem.* 75: 2640 - 2646.

Nahorniak, M.L.; Cooper, G.A.; -Ch. Kim, Y.; K Booksh, S. (2005). Three- and four-way parallel factor (PARAFAC) analysis of photochemically induced excitation-emission kinetic fluorescence spectra. *Analyst* 130: 85 - 93.

Naidong, W. (2003). Bioanalytical liquid chromatography tandem mass spectrometry methods on underivatized silica columns with aqueous/organic mobile phases. *J Chromatogr. B* 796: 209 - 224.

Nazi, A.; Ghasemi, J.; Yazdanipur, A. (2005). PARAFAC Decomposition of Three-Way Kinetic-Spectrophotometric Spectral Matrices Based on Phosphomolybdenum Blue Complex Chemistry for Nitrite Determination in Water and Meat Samples. *Anal. Lett.* 38: 2377 - 2392.

Ni, Y.; Huang, C.; Kokot, S. (2004). Application of multivariate calibration and artificial neural networks to simultaneous kinetic-spectrophotometric determination of carbamate pesticides. *Chemom. Intell. Lab. Syst.* 71: 177 - 193.

Öhman, J.; Geladi, P.; Wold, S. (1990a). Residual bilinearization. Part 1: Theory and algorithms. *J. Chemom.* 4: 79 - 90.

Öhman, J.; Geladi, P.; Wold, S. (1990b). Residual bilinearization. Part 2: Application to HPLC - diode array data and comparison with rank annihilation factor analysis. *J. Chemom.* 4: 135 - 146.

Olivieri, A.C. (2005b). A combined artificial neural network/residual bilinearization approach for obtaining the second-order advantage from three-way non-linear data. *J. Chemom.* 19: 615 - 624.

Olivieri, A.C. (2008). Analytical Advantages of Multivariate Data Processing. One, Two, Three, Infinity?. *Anal. Chem.* 80: 5713 - 5720.

Olivieri, A.C. (2005a). On a versatile second-order multivariate calibration method based on partial least-squares and residual bilinearization: Second-order advantage and precision properties. *J. Chemom.* 19: 253 - 265.

Olivieri A.C. (2004). Sample-specific standard prediction errors in three-way parallel factor analysis (PARAFAC) exploiting the second-order advantage. *J. Chemom.* 18: 363 - 371.

Olivieri, A.C.; Arancibia, J.A.; Muñoz de la Peña, A.; Durán-Merás, I.; Espinosa Mansilla, A. (2004). Second-Order Advantage Achieved with Four-Way Fluorescence Excitation-Emission-Kinetic Data Processed by Parallel Factor Analysis and Trilinear Least-Squares. Determination of Methotrexate and Leucovorin in Human Urine. *Anal. Chem.* 76: 5657 - 5666.

Olivieri, A.C.; Goicoechea, H.C.; Iñón, F.A. (2004). MVC1: An integrated Matlab toolbox for first-order multivariate calibration. *Chemom. Intell. Lab. Syst.* 73: 189 - 197.

Onur, F.; Yucesoy, C.; Dermis, S.; Kartal, M.; Kokdil, G. (2000). Simultaneous determination of pseudoephedrine sulfate, dexbrompheniramine maleate and loratadine in pharmaceutical preparations using derivative spectrophotometry and ratio spectra derivative spectrophotometry. *Talanta* 51: 269 - 279.

Orr, M.J.L. (1999). Matlab functions for radial basis function networks. Technical report. Institute for Adaptive and Neural Computation, Division of Informatics, Edinburgh University.

Orr, M.J.L. (1995). Regularisation in the selection of radial basis function centres. *Neural Comp.* 7: 606 - 623.

Pelckmans, K.; Suykens, J.A.K.; Van Gestel, T.; De Brabanter, J.; Lukas, L.; Hamers, B.; De Moor, B.; Vandewalle, J. (2003). ESAT-SCD-SISTA Technical Report 02-145. LS-SVMLab Toolbox User's Guide. Department of Electrical Engineering, Leuven-Heverlee, Belgium.

Peré Trepát, E.; Hildebrandt, A.; Barceló, D.; Lacorte, S.; Tauler, R. (2004). Fast chromatography of complex biocide mixtures using diode array detection and multivariate curve resolution. *Chemom. Intell. Lab. Syst.* 74: 293 - 303.

Peré Trepát, E.; Lacorte, S.; Tauler, R. (2005). Solving liquid chromatography mass spectrometry coelution problems in the analysis of environmental samples by multivariate curve resolution. *J. Chromatogr. A* 1096: 111 - 122.

Pereira Alves, S.; Mares Brum, D.; Castello Branco de Andrade, E' dira; Duarte Pereira Netto, Annibal. (2008). Determination of synthetic dyes in selected foodstuffs by high performance liquid chromatography with UV-DAD detection. *Food Chem.* 107: 489 - 496.

Perez-Bendito, D.; Silva, M. (1988). *Kinetic Methods in Analytical Chemistry*. Horwood, Chichester.

Pettersson, A.K.; Karlberg, B. (1997). Simultaneous determination of orthophosphate and arsenate based on multi-way spectroscopic-kinetic data evaluation. *Anal. Chim. Acta* 354: 241 - 248.

Poiger, H.; Schlatter, C. (1976). Fluorimetric determination of tetracyclines in biological materials. *Analyst* 101: 808 - 814.

Prazen, B.J.; Synovec, R.E.; Kowalski, R.B. (1998). Standardization of Second-Order Chromatographic/Spectroscopic Data for Optimum Chemical Analysis. *Anal. Chem.* 70: 218 - 225.

Ramis Ramos, G.; García Álvarez Coque, M.C. (2001). *Quimiometría. Síntesis*, Madrid, España. Capítulo 1.

Reverete, S.; Borrull, F.; Pocurull, E.; Marce, R.M. (2003). Determination of antibiotic compounds in water by solid-phase extraction-high-performance liquid chromatography-(electrospray) mass spectrometry. *J. Chromatogr. A* 1010: 225 - 232.

Rodríguez-Cuesta, M.J.; Boqué, R.; Rius, F.X.; Picón Zamora, D.; Martínez Galera, M.; Garrido Frenich, A. (2003). Determination of carbendazim, fuberidazole and thiabendazole by three-dimensional excitation-emission matrix fluorescence and parallel factor analysis. *Anal. Chim. Acta* 491: 47 - 56.

Ruckebusch, C. ; De Juan, A. ; Duponchel, L. ; Huvenne, J.P. (2006). Matrix augmentation for breaking rank-deficiency: A case study. *Chemom. Intell. Lab. Syst.* 80: 209 - 214.

Ruyck, H.; De Ridder, H. (2007). Determination of tetracycline antibiotics in cow's milk by liquid chromatography/tandem mass spectrometry. *Rapid Comm. Mass Spectrom.* 21: 1511 - 1520.

Safavi, A.; Absalan, G.; Maesum, S. (2001). Simultaneous determination of V(IV) and Fe(II) as catalyst using "neural networks" through a single catalytic kinetic run., *Anal. Chim. Acta* 432: 229 - 233.

Sanchez, E.; Kowalski, B.R. (1986). Generalized rank annihilation factor analysis. *Anal Chem.* 58: 496 - 499.

Sanchez, E.; Kowalski, B.R. (1990). Tensorial resolution: A direct trilinear decomposition. *J. Chemom.* 4 : 29 - 45.

Sane, R.T.; Francis, M.; Khedkar, S.; Pawar, S.; Moghe, A. (2001). Simultaneous HPTLC determination of pseudoephedrine sulphate and loratadine from their combined dosage form. *Indian-Drugs* 38: 436 - 438.

Santiago Valverde, R.; Gil García, M.D.; Martínez Galera, M.; Goicoechea, H.C. (2006). Determination of tetracyclines in surface water by partial least squares using multivariate calibration transfer to correct the effect of solid phase preconcentration in photochemically induced fluorescence signals. *Anal. Chim. Acta* 562: 85 - 93.

Saurina, J; Leal, C.; Compañó, R.; Granados, M.; Tauler, R.; Prat, M.D. (2000) Determination of triphenyltin in sea-water by excitation-emission matrix fluorescence and multivariate curve resolution. *Anal. Chim. Acta* 409: 237 - 245.

Saurina, J.; Tauler, R. (2000). Strategies for solving matrix effects in the analysis of triphenyltin in sea-water samples by three-way multivariate curve resolution. *Analyst* 125: 2038 - 2043.

Sena, M.M.; Trevisan, M.G; Poppi, R.G. (2005). PARAFAC: a chemometric tool for multi-dimensional data treatment. Applications in direct determination of drugs in human plasma by spectrofluorimetry. *Quim. Nova* 28: 910 - 920.

Sentellas, S.; Saurina, J. (2003). Chemometrics in capillary electrophoresis. Part B: Methods for data analysis. *J. Sep. Sci.* 26: 1395 - 1402.

Sentellas, S.; Saurina, J.; Hernández-Cassou, S.; Galceran, M.T.; Puignou, L. (2001). Resolution and quantification in poorly separated peaks from capillary zone electrophoresis using three-way data analysis methods. *Anal. Chim. Acta* 431: 49 - 58.

Sidiropoulos, N.D.; Bro, R. (2000). On the uniqueness of multilinear decomposition of N-ways arrays. *J. Chemom.* 14: 229 - 239.

Sinha, A.E.; Fraga, C.G.; Prazen, B.J.; Synovec, R.E. (2004). Trilinear chemometric analysis of two-dimensional comprehensive gas chromatography-time-of-flight mass spectrometry data. *J. Chromatogr. A* 1027: 269 - 277.

Sinha, A.E.; Hope, J.L.; Prazen, B.J.; Fraga, C.G.; Nilsson, E.J.; Synovec, R.E. (2004). Multivariate selectivity as a metric for evaluating comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry subjected to chemometric peak deconvolution. *J. Chromatogr. A* 1056: 145 - 154.

Sinha, A.; Prazen, B.J.; Synovec, R.E. (2004b). Trends in chemometric analysis of comprehensive two-dimensional separations. *Anal. Bioanal. Chem.* 378: 1948 - 1951.

Smilde, A.; Bro, R.; Geladi, P. (2004). *Multi-way Analysis: Applications in the Chemical Sciences*. Wiley, New York, USA.

Smith, C.M.; Reynard, A.M. (1993). *Farmacología. Médica Panamericana*, Buenos Aires, Argentina.

STATGRAPHICS Plus. (1994-2000). Statistical Graphics Corp., Herndon, Virginia, USA.

Stueber, M.; Reemtsma, T. (2004). Evaluation of three calibration methods to compensate matrix effects in environmental analysis with LC-ESI-MS. *Anal. Bioanal. Chem.* 378: 910 - 916.

Suykens, J.A.K.; Vanderwalle, J. (1999). Least Squares Support Vector Machine Classifiers. *Neural Process. Lett.* 9: 293 - 300.

Suykens, J.A.K.; van Gestel, T.; de Brabanter, J.; de Moor, B.; Vandewalle, J. (2002). *Least-Squares Support Vector Machines*. World Scientific, Singapore.

Thissen, U.; Pepers, M.; Üstün, B.; Melssen, W.J.; Buydens, L.M.C. (2004). Comparing support vector machines to PLS for spectral regression applications. *Chemom. Intell. Lab. Syst.* 73: 169 - 179.

Thissen, U.; Üstün, B.; Melssen, W.J.; Buydens, L.M.C. (2004). Multivariate Calibration with Least-Squares Support Vector Machines. *Anal. Chem.* 76: 3099 - 3105.

Trevisan, M.G.; Poppi, R.J. (2003). Determination of doxorubicin in human plasma by excitation-emission matrix fluorescence and multi-way analysis. *Anal. Chim. Acta* 493: 69 - 81.

Vaananen, T.; Koskela, H.; Hiltunen, Y.; Ala-Korpela, M. (2002). Application of Quantitative Artificial Neural Network Analysis to 2D NMR Spectra of Hydrocarbon Mixtures. *J. Chem. Inf. Comput. Sci.* 42: 1343 - 1346.

van der Linden, W.E. (1989). Definition and classification of interferences in analytical procedures. *Pure Appl. Chem.* 61: 91 - 95.

van der Voet, H. (1994). Comparing the predictive accuracy of models using a simple randomization test. *Chemom. Intell. Lab. Syst.* 25: 313 - 323.

van Mispelaar, V.G. (2005). *Chromametrics*. Doctoral Thesis. University of Amsterdam, The Netherlands.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer, New York, USA.

Vasudevan, M.; Ravisankar, S.; Sathiyarayanan, A.; Chandan, R.S. (2001). Simultaneous estimation of loratadine and pseudoephedrine by RP-HPLC method. *Indian-Drugs* 38: 276 - 278.

Wahba, G. (1990). *Spline Models for Observational Data*,. Vol. 59 of CBMSNSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia, PA, USA.

Wegman, E.J.; Wright, I.W. (1983). Splines in Statistics. *J. Am. Stat. Assoc.* 78: 351 - 365.

Wentzell, P.D.; Kazayannis, M.I.; Crouch, S.R. (1989). Simultaneous kinetic determinations with the kalman filter. *Anal. Chim. Acta* 224: 263 - 274.

Wiberg, K.; Jacobsson, S.P. (2004). Parallel factor analysis of HPLC-DAD data for binary mixtures of lidocaine and prilocaine with different levels of chromatographic separation. *Anal. Chim. Acta* 514: 203 - 209.

Windig, W.; Guilment, J. (1991). Interactive self-modeling mixture analysis. *Anal. Chem.* 63: 1425 - 1432.

Wise, B.M.; Gallagher, N.B.; Bro, R.; Shaver, J.M.; Windig, W.; Koch, R.S. (2005). *PLS_Toolbox* Version 3.5 for use with MATLAB. Eigenvector Research, Inc. 830 Wapato Lake Road, Manson, WA 98831, USA.

Wold, S. (1974). Spline Functions in Data Analysis. *Technometrics* 16: 1 - 11.

Wold, S.; Geladi, P.; Esbensen, K.; Öhman, J. (1987). Multiway principal components and PLS analysis. *J. Chemom.* 1: 41 - 56.

Wu, H.L.; Shibukawa, M.; Oguma, K. (1998). An alternating trilinear decomposition algorithm with application to calibration of HPLC-DAD for

simultaneous determination of overlapped chlorinated aromatic hydrocarbons
J. Chemom. 12: 1 - 26.

Wu, X.; Cai, R.; Zhang, L.; Xiao, M. (2001). Two-rate method for simultaneous kinetic determination based on detecting the final product in successive reactions. Anal. Chim. Acta 448: 257 - 263.

Xia, A.L.; Wu, H.L.; Fang, D.M.; Ding, Y.J.; Hu, L.Q.; Yu, R.Q. (2005). Alternating penalty trilinear decomposition algorithm for second-order calibration with application to interference-free analysis of excitation-emission matrix fluorescence data. J. Chemom. 19: 65 - 76.

Xia, A.L.; Wu, H.L.; Li, S.F.; Zhu, S.H.; Hu, L.Q.; Yu, R.Q. (2007). Alternating penalty quadrilinear decomposition algorithm for an analysis of four-way data arrays. J. Chemom. 21: 133 -144.

Xie, Y.L.; Baeza-Baeza, J.J.; Ramis-Ramos, G. A. (1996). Comparative study of several chemometric methods applied to the treatment of two-way kinetic-spectral data for mixture resolution. Anal. Chim. Acta 321: 75 - 95.

Xiong, R.; Velasco, A.; Silva, M.; Pérez Bendito, D. (1991). Performance of the Kalman filter algorithm in differential reaction-rate methods. Anal. Chim. Acta 251: 313 - 319.

Yang, S.; Cha, J.; Carlson, K. (2005). Simultaneous extraction and analysis of 11 tetracycline and sulfonamide antibiotics in influent and effluent domestic wastewater by solid-phase extraction and liquid chromatography-electrospray ionization tandem mass spectrometry. J. Chromatogr. A 1097: 40 - 53.

Yang, Y.D.; Veltkamp, D.J.; Kowalski B.R. (1991). Multivariate Instrument Standardization. Anal. Chem. 63: 2750 -2756.

Ye, Z.; Weinberg, H.S.; Meyer, M.T. (2007). Trace Analysis of Trimethoprim and Sulfonamide, Macrolide, Quinolone, and Tetracycline Antibiotics in Chlorinated Drinking Water Using Liquid Chromatography Electrospray Tandem Mass Spectrometry. *Anal. Chem.* 79: 1135 - 1144.

Zhang, F.; Li, H. (2006). Resolution of overlapping capillary electrophoresis peaks by using chemometric analysis: Improved quantification by using internal standard. *Chemom. Intell. Lab. Syst.* 82: 184 - 192.

Zhang, Y.; Li, H.; Hou, A.; Havel, J. (2006). Artificial neural networks based on principal component analysis input selection for quantification in overlapped capillary electrophoresis peaks. *Chemom. Intell. Lab. Syst.* 82: 165 - 175.

Zhu, J.; Snow, D.D.; Cassada, D.A.; Monson, S.J.; Spalding, R.F. (2001). Analysis of oxytetracycline, tetracycline, and chlortetracycline in water using solid-phase extraction and liquid chromatography-tandem mass spectrometry. *J. Chromatogr. A* 928: 177 -186.

Zupan, J.; Gasteiger, J. (1999). *Neural Networks in Chemistry and Drug Design*. Wiley, New York, USA.

Zupan, J.; Novic, M.; Ruisánchez, I. (1997). Kohonen and counterpropagation artificial neural networks in analytical chemistry. *Chemom. Intell. Lab. Syst.* 38: 1 - 23.