



Universidad Nacional de Rosario
Facultad de Ciencias Bioquímicas y Farmacéuticas

Instituto de Química Rosario
Área Química Analítica General

**Herramientas Quimiométricas para analizar
conjuntamente datos de orden cero, primero y
segundo aplicadas a bases de datos de
monitoreo ambiental**

Tesis para optar al título de Licenciado en Biotecnología
Andrés Martínez Bilesio

Director
Dr. Alejandro García Reiriz

2018

Abreviaturas y símbolos

a_{in}	Concentración relativa del componente n en la i^{ma} muestra
$\mathbf{a}_n$	Vector columna resultante de la descomposición de PARAFAC correspondiente a las concentraciones relativas del n^{mo} componente en todas las I muestras (dimensiones $I \times 1$)
$\mathbf{A}$	Matriz de <i>scores</i> de PARAFAC relacionados con las concentraciones relativas de los componentes en cada muestra
ALS	Cuadrados mínimos alternantes
b_{jn}	Intensidad relativa del componente n en el j^{mo} modo de emisión
$\mathbf{b}_n$	Vector columna resultante de la descomposición de PARAFAC correspondiente al perfil de emisión del n^{mo} componente (dimensiones $J \times 1$)
$\mathbf{B}$	Matriz de <i>loadings</i> de PARAFAC del modo de emisión
c_{kn}	Intensidad relativa del componente n en el k^{mo} modo de excitación
$\mathbf{c}_n$	Vector columna resultante de la descomposición de PARAFAC correspondiente al perfil de excitación del n^{mo} componente (dimensiones $K \times 1$)
$\mathbf{C}$	Matriz de <i>loadings</i> de PARAFAC del modo de excitación
CONCORDIA	Test de consistencia del núcleo de PARAFAC
$\mathbf{D}$	Matriz de datos aumentada en columnas de MCR-ALS
DTLD	Descomposición trilineal directa para inicialización de PARAFAC
$\mathbf{E}$	Matriz de error no modelado
I	Canal Ibarlucea
I	Cantidad de muestras involucradas en la construcción del arreglo tridimensional de datos de PARAFAC

<i>J</i>	Número de datos en el modo de emisión en el arreglo tridimensional de PARAFAC
<i>K</i>	Número de datos en el modo de excitación en el arreglo tridimensional de PARAFAC
<i>L</i>	Arroyo Ludueña
<i>loading</i>	Vector de cargas
<i>M</i>	Matriz de datos global generada a partir de la fusión de los datos
max	Máximo
MCR	Resolución de curvas multivariantes
MEE	Matriz de emisión-excitación de fluorescencia
min	Mínimo
MinMax	Pre-procesamiento de normalización por MinMax
MOD	Materia orgánica disuelta
MOP	Materia orgánica particulada
<i>N</i>	Número de componentes de PARAFAC
<i>n</i>	Un componente particular de PARAFAC
NIR	Infrarrojo cercano
<i>N-way</i>	Datos multivía
ONU	Organización de las Naciones Unidas
PARAFAC	Análisis de factores paralelos
PCA	Análisis de componentes principales
RN9	Ruta Nacional Número 9
RN33	Ruta Nacional Número 33
RN34	Ruta Nacional Número 34
RNA012	Ruta Nacional A012

RP34S	Ruta Provincial Número 34S
S	Canal Salvat
S	Matriz de <i>scores</i> de MCR-ALS
<i>score</i>	Vector de puntuaciones
SVD	Descomposición en valores singulares
UA	Unidades arbitrarias
UAF	Unidades arbitrarias de fluorescencia
UV-Vis	Ultravioleta-visible
X_i	i^{ma} matriz de dimensiones $J \times K$ del arreglo tridimensional de datos de PARAFAC
X	Arreglo tridimensional de datos de PARAFAC
Y^T	Matriz de <i>loadings</i> de MCR-ALS

Difusión de resultados

Como fruto del siguiente trabajo de tesis de grado, se lograron enviar los resultados obtenidos en forma de posters al 'IX Congreso Argentino de Química Analítica (IXCAQA)' Rio Cuarto, Córdoba, Argentina, 07-10 de Noviembre de 2017:

- Estudio Quimiométrico de los sedimentos suspendidos de la cuenca del arroyo Ludueña utilizando espectroscopia de infrarrojo cercano (NIR). Marianela Batistelli, Andres Martinez Bilesio y Alejandro García Reiriz. IQUIR-CONICET, Dpto. Qca. Analítica, FBioyF, UNR.
- Herramientas quimiométricas para analizar conjuntamente datos de orden cero, primero y segundo aplicadas a bases de datos de monitoreo ambiental. Andrés Martínez Bilesio, Marianela Batistelli y Alejandro García Reiriz. IQUIR-CONICET, Dpto. Qca. Analítica, FBioyF, UNR.

También se encuentra en proceso de evaluación un manuscrito enviado a la revista *Microchemical Journal* en Febrero de 2018:

- Development of a fast and inexpensive method for detecting the main sediment sources in a river basin. Marianela Batistelli, Andrés R. Martínez Bilesio y Alejandro G. García-Reiriz. IQUIR-CONICET, Dpto. Qca. Analítica, FBioyF, UNR.

Resumen

Los protocolos de monitoreo ambiental normalmente generan grandes bases de datos, haciendo necesario su procesamiento mediante técnicas estadísticas multivariadas para obtener información relevante y una correcta interpretación. Debido al gran avance del instrumental analítico, no solo es posible realizar mediciones puntuales sino también de órdenes superiores, las cuales se proponen fusionar en este trabajo.

Las diferentes técnicas analíticas experimentales reflejan distintas propiedades de un sistema, que a la vez pueden estar interrelacionadas. Es por ello que se desarrolló una metodología para estudiar de forma conjunta los datos medidos de distintos órdenes sobre un mismo sistema.

Como sistema de estudio se eligió la cuenca del arroyo Ludueña, que se encuentra ubicada en el sur de la provincia de Santa Fe. Se tomaron muestras de forma periódica en diferentes sitios de la cuenca. Sobre las muestras obtenidas se realizaron mediciones puntuales de conductividad y de pH (datos de orden cero), que se fusionaron con los espectros de infrarrojo cercano (NIR) de los sedimentos suspendidos de las muestras (datos de primer orden) y las matrices de excitación-emisión (MEEs) de fluorescencia de la materia orgánica disuelta (datos de segundo orden).

Por medio de distintos algoritmos quimiométricos fue posible extraer la información de los datos de primer y segundo orden acerca de las concentraciones relativas de los componentes de cada técnica. Estas reflejan la diferente composición de las muestras medidas y no pueden ser calibradas a un valor absoluto debido a que los fluoróforos y los sedimentos son de composición indefinida y, por lo tanto, no existen patrones de los mismos. A continuación, con estos valores se construyó una nueva base de datos que también incluyó las variables de pH y conductividad. Para un correcto procesamiento de la misma se realizó un escalado y luego se analizaron los datos con un modelo trilineal. Este último análisis permitió diferenciar las variables afectadas por las

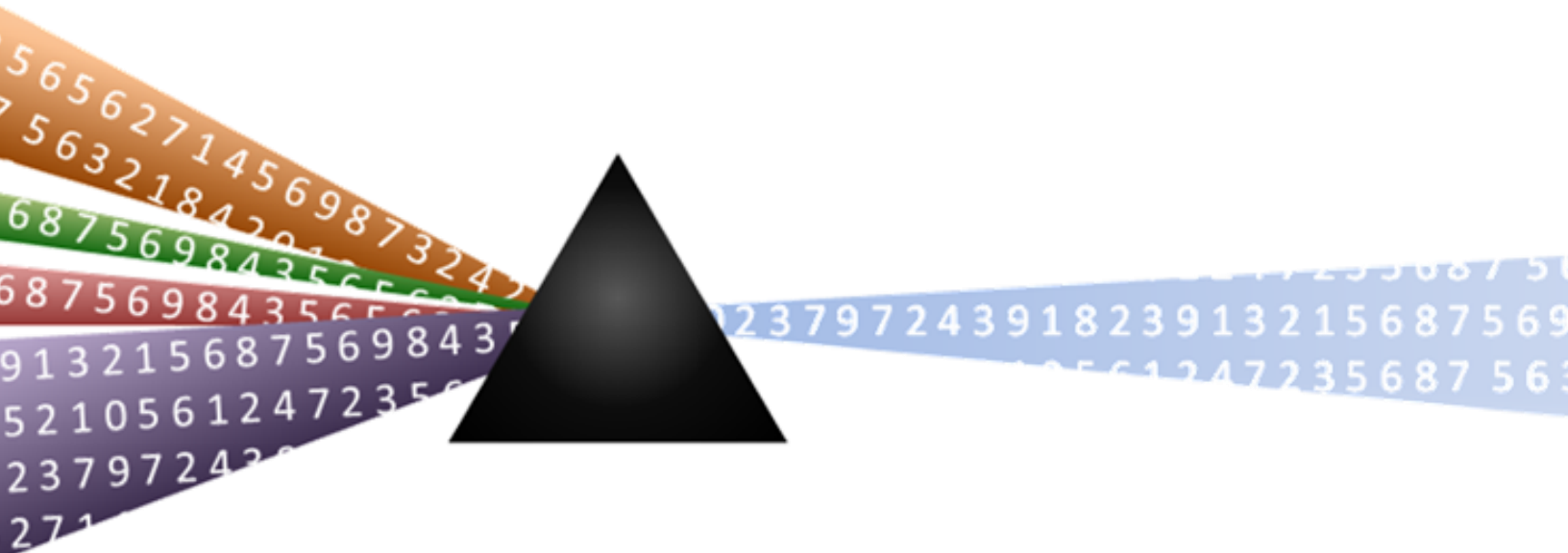
actividades antrópicas de las de origen natural, agrupando aquellas con un comportamiento similar, es decir, un origen en común. Adicionalmente, fue posible representar su distribución sobre los sitios de muestreo y observar su evolución a través de las distintas campañas.

Índice

Abreviaturas y símbolos.....	I
Difusión de resultados	IV
Resumen	V
1. Introducción	1
1.1 Antecedentes	2
1.2 Planteo de la hipótesis de trabajo.....	5
2. Objetivos	12
3. Materiales y métodos	14
3.1 Materiales y métodos experimentales.....	15
3.1.1 Presentación del sistema en estudio.....	15
3.1.2 Sitios de muestreo y realización de campañas.....	15
3.1.3 Toma de muestra.....	18
3.1.4 Procesamiento de las muestras.....	18
3.1.5 Instrumental	20
3.1.6 Mediciones de pH y conductividad	21
3.1.7 Mediciones de absorbancia UV-Visible (UV-Vis)	21
3.1.8 Adquisición de las MEEs de fluorescencia.....	21
3.2 Materiales y métodos quimiométricos	22
3.2.1 Resolución de curvas multivariantes por cuadrados mínimos alternantes (MCR-ALS)	22
3.2.2 Resolución de curvas multivariantes por cuadrados mínimos alternantes (MCR-ALS) para modelos trilineales	24
3.2.3 Análisis de factores paralelos (PARAFAC).....	28
3.2.4 Pre-procesamiento de las MEEs	28
3.2.5 Software.....	29

4. Resultados y Discusión	30
4.1 Resultados obtenidos a partir de los datos de NIR	31
4.2 Análisis de los datos de fluorescencia	34
4.3 Fusión y escalado de los datos	41
4.4 Análisis de la matriz de datos global	46
5. Conclusiones	55
6. Referencias bibliográficas	57
 Anexo	 65
A.1 Procesamiento de las muestras	65
A.2 Instrumental.....	66

1. Introducción



1.1. Antecedentes

Durante mucho tiempo, las políticas ambientales no fueron consideradas entre los tópicos prioritarios discutidos globalmente en las cumbres de organismos internacionales. Fue así como, a lo largo del siglo XX, la sociedad fue testigo de una gran cantidad de desastres ecológicos, tales como la contaminación de la Bahía de Minamata (Japón), el gran smog londinense, los escapes de sustancias químicas de Seveso (Italia) o de Bophal (India), y el accidente de Chernóbil (actual Ucrania), entre otros.

La temática ambiental fue ganando importancia hacia fines de los años 60 e inicio de los años 70, impulsada sobre todo por constataciones científicas, la maduración de movimientos de la sociedad civil, y publicaciones de gran impacto. Sin embargo, hasta el inicio de los años 70, los esfuerzos internacionales en materia ambiental habían sido dirigidos a problemas específicos, con poca atención a la relación entre ambiente y desarrollo económico-social.

No fue entonces hasta junio de 1972 cuando se llevó a cabo la primer gran conferencia convocada por la Organización de las Naciones Unidas (ONU) sobre cuestiones ambientales internacionales. Dicha 'Conferencia de Naciones Unidas sobre el Ambiente Humano', también conocida como 'Cumbre de la Tierra de Estocolmo', fue la primera en tratar el tema ambiental de manera amplia y global, marcando un punto de inflexión en el desarrollo de su política internacional [1]. Desde entonces, la preocupación pública respecto al cuidado del ambiente ha ido en aumento. La organización periódica de nuevas cumbres y conferencias ha permitido introducir el concepto de desarrollo sostenible y evidenciar su estrecha relación con la preservación de los recursos, propiciando de este modo el desarrollo de una nueva conciencia ambiental.

A escala global, las problemáticas ambientales han generado hasta la fecha un gran impacto en la salud y el bienestar de las comunidades, así como también en la biodiversidad y en la calidad y disponibilidad de los recursos naturales. Por lo tanto, cada vez son más los estudios y trabajos orientados a la preservación y control del ambiente.

En este contexto mundial, surge la necesidad de desarrollar metodologías y técnicas orientadas a la preservación y el control de los recursos naturales; se generan entonces los denominados protocolos de monitoreo ambiental.

Un monitoreo ambiental es realizado a efectos de determinar el estado de conservación de los recursos naturales. Con este propósito se debe cuantificar la presencia de uno o varios contaminantes o variables en el ambiente, a fin de ser utilizados como indicadores. Es por lo tanto una herramienta de medición y evaluación en el tiempo del estado de un determinado ecosistema en estudio, y permite de este modo:

- Identificar fortalezas y debilidades.
- Identificar claves para la toma de decisiones estratégicas.
- Cuantificar los objetivos.
- Comparar las metas propuestas con los resultados.

El diseño de un plan de monitoreo es complejo, ya que involucra evaluaciones previas y un conocimiento profundo del sistema. Implica llevar a cabo diversas tareas, tales como: selección de los sitios, tipo y frecuencia de muestreo, selección de variables críticas para ser evaluadas con métodos de medición adecuados, creación de un protocolo de trabajo, procesamiento y análisis de los datos obtenidos.

Los monitoreos ambientales convencionales son tediosos e insumen mucho tiempo y dinero en la recolección y procesamiento de las muestras, así como también en las distintas determinaciones realizadas. Sin embargo, seleccionando variables específicas que reflejen el estado y la evolución de un sistema, es posible llevar a cabo un monitoreo de manera simple y eficiente.

En el caso particular de los ecosistemas acuáticos, el crecimiento sostenido de la población ha dado lugar a un aumento preocupante de la contaminación. Es entonces de suma importancia poder desarrollar protocolos sencillos para monitorear, controlar y proteger dichos ambientes. Por lo general, el monitoreo completo de un sistema hidrológico genera una gran cantidad de datos, debido fundamentalmente a que implica

mediciones de diferentes variables o analitos de interés en distintos puntos de muestreo a lo largo del tiempo. De esta manera, la extracción de información relevante a partir de la base de datos generada se vuelve muy compleja, y es necesario su procesamiento mediante técnicas estadísticas multivariadas. Los métodos quimiométricos han surgido entonces como un enfoque práctico para procesar toda esta información e interpretarla correctamente [2,3].

En la actualidad, la innovación y el avance tecnológico han posibilitado adquirir datos de mayor complejidad. Gracias al avance de la Química Analítica, el instrumental moderno permite no solo realizar mediciones puntuales (como pH, conductividad, etc.), sino también de órdenes superiores (tales como la adquisición de espectros o matrices de datos). De este modo, es posible obtener datos de diferentes estructuras dimensionales o modos. En la Figura 1 se muestran las distintas estructuras de los datos que pueden ser generadas al aplicar sobre una muestra una técnica analítica en particular. Hablamos de datos de orden cero cuando al medir una propiedad o parámetro se generan datos puntuales; datos de primer orden cuando la medición genera datos vectoriales; datos de segundo orden cuando la medición genera datos matriciales; y, finalmente, datos de tercer orden cuando la medición origina un arreglo tridimensional de datos.

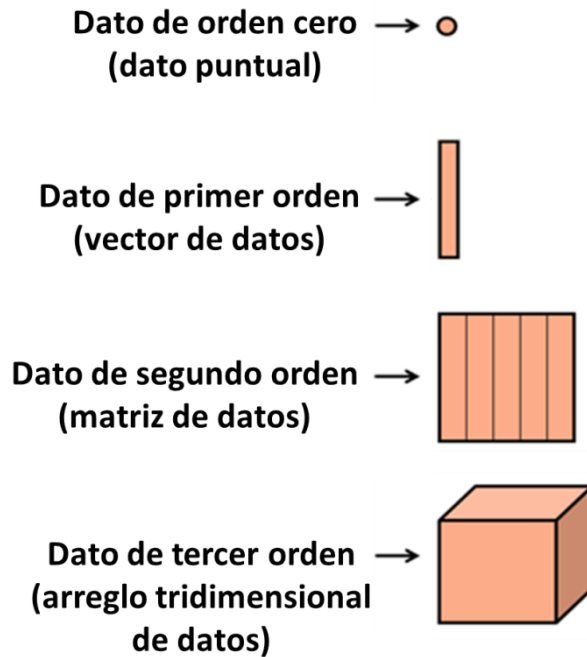


Figura 1. Representación de la estructura de los datos de una muestra.

1.2. Planteo de la hipótesis de trabajo

En el presente trabajo de tesis de grado se propone desarrollar una novedosa metodología que permita fusionar y explotar las ventajas que ofrecen los datos de distintos órdenes. Se plantea como objetivo la construcción de una base de datos ambiental, pero que a diferencia de los monitoreos convencionales que realizan mediciones puntuales, en este caso involucra conjuntamente datos de orden cero, primero y segundo. De este modo, se espera obtener una visión global del estado del sistema ambiental, en una forma más rápida y sencilla para el laboratorista. Existe un trabajo previo de nuestro grupo de investigación donde se fusionan únicamente datos de orden cero con datos de segundo orden [4]. En esta oportunidad, se pretende extender este abordaje también a datos de primer orden. En la Figura 2 se muestra un esquema que resume el flujo de datos de la metodología propuesta.

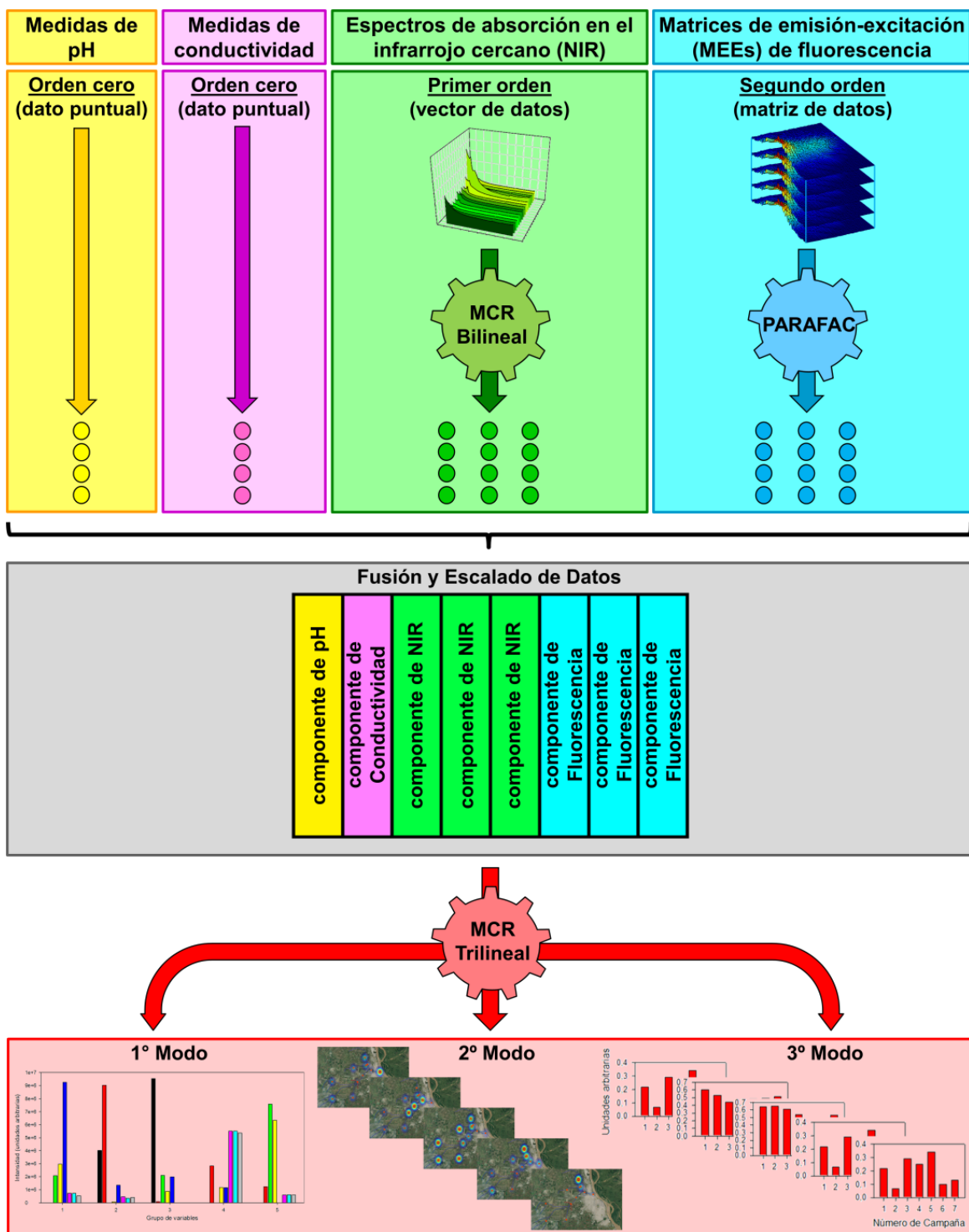


Figura 2. Esquema del flujo de datos de la metodología propuesta. El sector amarillo corresponde a las mediciones de pH (datos de orden cero). El sector fucsia corresponde a las determinaciones de conductividad (datos de orden cero). El sector verde corresponde a los espectros de absorción NIR (datos de primer orden). El sector celeste corresponde a las MEEs de fluorescencia (datos de segundo orden). El sector gris corresponde a la fusión y el escalado de los datos. Finalmente, el sector rojo corresponde al análisis de la matriz de datos global generada en la etapa previa.

Para corroborar la metodología de trabajo aquí propuesta, se seleccionaron distintas técnicas que permiten generar datos de diferentes órdenes. Como datos de orden cero se escogieron las mediciones de pH y de conductividad. Si bien estas no son específicas, aportan una noción básica respecto a la calidad del agua de un determinado sitio y permiten a su vez compararla con la de otros sitios. Estas técnicas poseen la ventaja de ser sumamente simples en cuanto al manejo experimental y relativamente accesibles para cualquier laboratorio. Estas determinaciones se encuentran representadas dentro del sector amarillo y del sector fucsia de la Figura 2, respectivamente.

Las técnicas de infrarrojo cercano (NIR) y de fluorescencia permiten adquirir datos multivariados. Por medio de la espectroscopia NIR es factible obtener datos vectoriales, es decir, datos de primer orden conformados por el espectro que refleja la absorbancia a las distintas longitudes de onda. Las mediciones de fluorescencia, por su parte, permiten acceder a datos de segundo orden generados a partir de la medición de la intensidad de fluorescencia a distintas longitudes de onda de emisión y de excitación. Ambas técnicas analíticas resultan de gran interés desde el punto de vista ambiental, debido a que por medio de la espectroscopia NIR es factible estudiar la composición y concentración del material particulado suspendido que da origen a los sedimentos de una cuenca [5], y por medio de la espectroscopia de fluorescencia es posible monitorear la materia orgánica disuelta (MOD) [6].

Entre los parámetros a determinar, los sedimentos son el elemento principal por el cual se determina la dinámica hidrológica de una cuenca. Estos influyen en la red trófica de alimentos como así también en la química del cuerpo de agua que los contiene. Son participantes activos del almacenamiento y transporte de nutrientes y contaminantes orgánicos y metálicos. Por ello, a través de su estudio, es posible determinar las causas que afectan las características fisicoquímicas del lecho del río o arroyo [7-9]. Dado que están directamente relacionados con la calidad del agua, al estudiar la variabilidad y distribución de los mismos es posible inferir las propiedades del cuerpo de agua que los contiene.

La espectroscopia NIR se ha transformado en una técnica muy utilizada para una amplia cantidad de determinaciones, debido a que presenta una serie de ventajas con respecto a las técnicas convencionales tales como: rapidez de análisis, no ser destructiva, requerir de poca o ninguna preparación de la muestra a medir, no utilizar reactivos y ser confiable. Por estas razones es que progresivamente está siendo implementada en las industrias químicas, petroquímicas, alimenticias y agrícolas [10]. También ha adquirido un papel importante en estudios de monitoreo ambiental a través del estudio de los sedimentos y su relación con el agua [5,11-13].

Recientemente, nuestro grupo ha propuesto un método alternativo rápido, eficaz y libre de residuos para determinar las principales fuentes de los sedimentos de la cuenca del arroyo Ludueña, permitiendo estudiar su distribución geográfica y su evolución en el tiempo. Dicha estrategia está basada en espectroscopía NIR asistida por los algoritmos análisis de componentes principales (PCA) y resolución de curvas multivariantes por cuadrados mínimos alternantes (MCR-ALS) para el análisis de los datos. Ese trabajo se realizó conjuntamente con el que aquí se presenta y ha dado origen a la tesis de grado de la estudiante Marianela Batistelli. Los aspectos más relevantes del mismo se encuentran citados en el *Anexo*. Tanto la adquisición como el procesamiento de los espectros NIR de primer orden se encuentran representados en el sector verde de la Figura 2.

Como se mencionó anteriormente, además de las determinaciones efectuadas sobre el material particulado, se ha propuesto realizar mediciones de matrices de emisión-excitación (MEEs) de fluorescencia, las cuales aportan información relevante acerca de la MOD. Teniendo en cuenta que la MOD presenta distintas propiedades fluorescentes que dependen de su origen, las MEEs brindan información valiosa acerca de las actividades antropogénicas en una cuenca hidrográfica [6].

Al igual que los sedimentos, la MOD está compuesta por una gran variedad de sustancias orgánicas. Estas sustancias pueden ser clasificadas tanto por la localización de su origen como por la naturaleza de su origen. En relación al lugar donde se generan, pueden clasificarse como: 1) autóctonas, generadas dentro del cuerpo de agua; o 2)

alóctonas, incorporadas al sistema desde el exterior. Por otro lado, con respecto a la fuente de su origen, pueden ser de origen natural o antropogénico (derivadas de las actividades humanas). El primer grupo de estos compuestos incluye principalmente a las sustancias derivadas de los ácidos húmicos. El segundo grupo, por su parte, representa la MOD con alto contenido proteico, es decir, derivada de péptidos y proteínas disueltas, que provienen de la descarga de aguas residuales industriales y aguas servidas no tratadas [14-16].

Por medio de la espectroscopía de fluorescencia es posible caracterizar la MOD de muestras de diferentes orígenes, evaluar la calidad del agua, monitorear el nivel de contaminación de áreas comprometidas y distinguir regiones con un gran impacto antrópico de otras no tan afectadas [17-19]. Las aguas naturales contienen una mezcla de fluoróforos que hacen compleja su identificación mediante espectros de fluorescencia unidimensionales [6]. Por lo tanto, como ya se mencionó anteriormente, una excelente alternativa analítica es adquirir las MEEs de fluorescencia, las cuales permiten obtener información mucho más rica con respecto a la presencia y la clase de fluoróforos disueltos. Las MEEs comenzaron a estudiarse en la década de 1990, distinguiendo compuestos húmicos y no húmicos en aguas naturales [6,20,21].

Con el fin de extraer información de índole química de los datos bidimensionales registrados en las MEEs, se pueden aplicar varias metodologías. Una de las más populares es construir un arreglo de señales tridimensionales o cubo de datos a partir de los datos adquiridos de MEEs, seguido por análisis de factores paralelos (PARAFAC) [22]. Este algoritmo permite identificar sustancias húmicas y compuestos similares a proteínas en muestras de agua [23], caracterizar la MOD presente en lagos y suelos [24], identificar contaminantes antropogénicos y trazas de elementos metálicos en aguas [25], estudiar la descarga de efluentes en ríos y aguas marinas asociado con rangos de salinidad y nutrientes [26], detectar ácidos fúlvicos y proteínas ricas en triptófano en descargas de aguas residuales [27] y clasificar muestras de agua basándose únicamente en el contenido de ácidos húmicos [28]. Otra opción disponible para el análisis de las MEEs es el algoritmo

MCR-ALS previamente mencionado [29]. Por lo tanto, ambos métodos pueden aplicarse con la misma finalidad [4]. MCR-ALS puede funcionar de un modo similar a PARAFAC si se aplican las restricciones adecuadas que se profundizarán en las siguientes secciones. Al igual que para las técnicas anteriores, el procesamiento de los datos de segundo orden se encuentra representado en el sector celeste de la Figura 2.

Los algoritmos quimiométricos antes mencionados, aplicados a datos de primer o segundo orden, brindan información cualitativa y cuantitativa de las posibles fuentes de la señal medida. De esta manera, con la información cualitativa (almacenada en los *loadings*) sería posible identificarlas, mientras que con la información cuantitativa (contenida en los llamados *scores*) se podría construir una nueva base de datos ambiental que permita estudiar el comportamiento de todas las mediciones en forma conjunta (sector gris de la Figura 2).

Para analizar el comportamiento de los fluoróforos, los sedimentos, el pH y la conductividad conjuntamente en esta nueva base de datos a través de las campañas (el tiempo) y en los distintos sitios de muestreo (el espacio), es posible aplicar nuevamente MCR-ALS pero con un enfoque parecido al empleado en los estudios convencionales de monitoreo ambiental [4,30]. Este algoritmo permitiría agrupar las variables con similar comportamiento, proporcionando así su distribución espacial y temporal conjunta. De esta manera, sería posible hacer una interpretación global sobre su origen e interrelación, tal como se observa en el sector rojo de la Figura 2. Este hecho representaría una ventaja sumamente importante, ya que tanto la MOD como los sedimentos son afectados por el pH y la conductividad del medio que los contiene, y adicionalmente, la MOD puede generar material sedimentario por actividad biológica o viceversa.

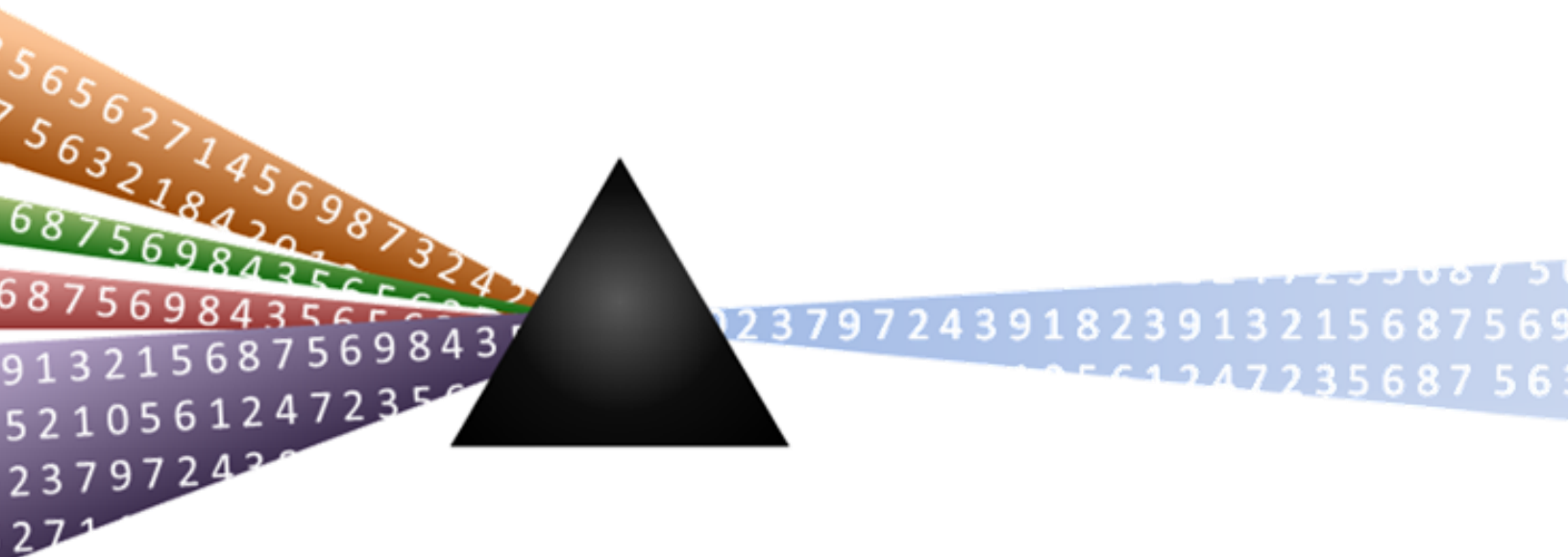
Por lo tanto, MCR-ALS se ha convertido en una poderosa herramienta quimiométrica con aplicaciones crecientes en el análisis de bases de datos de monitoreo ambiental [31]. Recientemente ha sido validado para la identificación de patrones de contaminación ambiental en aguas superficiales [32] y se han propuestos abordajes similares para la resolución e interpretación de fuentes de contaminación en varias

cuencas hidrográficas alrededor del planeta [33]. Como se mencionó anteriormente, un trabajo previo de nuestro grupo ha permitido demostrar la utilidad de este algoritmo para el análisis conjunto de datos de orden cero y de segundo orden [4].

Al aplicar MCR-ALS a bases de datos ambientales, la suposición básica es que cada valor de una variable medida en una muestra en particular se debe a la suma de las contribuciones de fuentes independientes e individuales de distinto origen. Cada una de estas fuentes se caracteriza por un perfil de composición química particular y se distribuye entre las muestras de un modo diferente. Como resultado de aplicar este método quimiométrico, sería posible identificar el origen de las fuentes de contaminación (ya sean puntuales o difusas), como así también caracterizar sus perfiles de distribución entre muestras (geográfico y temporal).

De esta manera, los algoritmos quimiométricos pueden representar una herramienta muy útil y poderosa para el manejo, análisis e interpretación de sistemas con una gran cantidad de información. Sin embargo, al tratarse de datos de distintos órdenes, llevar a cabo su análisis conjunto reviste una alta complejidad, no solo por la fusión en sí misma, sino también por ser datos de un monitoreo ambiental que son intrínsecamente ruidosos. Así, su tratamiento y estudio requiere distintos abordajes como resultado de las diversas metodologías analíticas empleadas. Por lo tanto, se plantea un gran desafío a la hora de desarrollar una estrategia que permita fusionar y analizar conjuntamente todos los datos medidos de distintos órdenes de un mismo sistema. En el marco del actual proyecto de tesis de grado se propone analizar la factibilidad de esta metodología utilizando como sistema modelo la cuenca del arroyo Ludueña de la provincia de Santa Fe, Argentina.

2. Objetivos



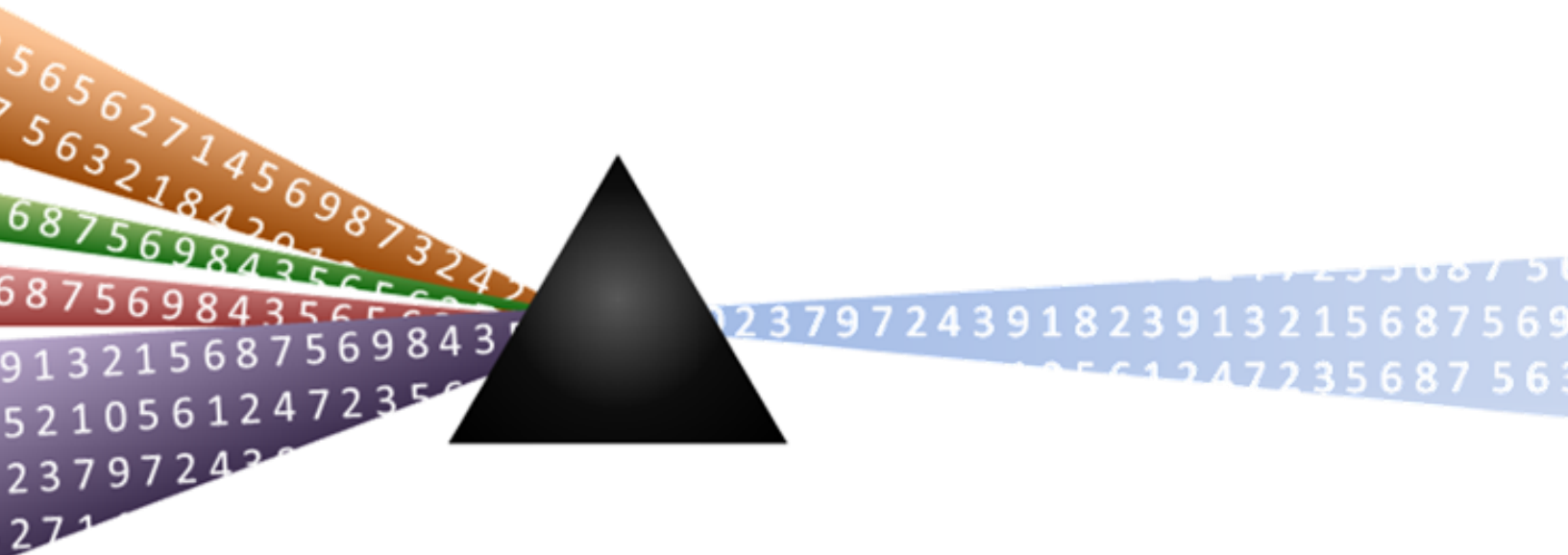
Generales:

- Optimizar métodos de fusión de datos y análisis conjunto.
- Desarrollar un protocolo de monitoreo que permita reflejar el impacto antropogénico sobre una cuenca de una forma simple, rápida y económica.
- Desarrollar un método que permita evidenciar e interpretar las variaciones a lo largo del tiempo y el espacio en las distintas campañas de un monitoreo ambiental.

Específicos:

- Proponer una nueva estrategia para fusionar la información extraída por métodos quimiométricos a partir de los datos de orden cero, primero y segundo, a fin de establecer sus posibles relaciones y su distribución en el espacio y el tiempo, y así poder detectar potenciales fuentes de contaminación.
- Investigar las zonas más vulnerables y afectadas por las actividades antrópicas en la cuenca del arroyo Ludueña.
- Estimar la distribución geográfica y el comportamiento temporal de las variables medidas, a fin de contribuir a la evaluación de la calidad ambiental del agua superficial de la región bajo estudio.

3. Materiales y métodos



3.1. Materiales y métodos experimentales

3.1.1. Presentación del sistema en estudio

Como sistema modelo para corroborar la metodología de trabajo aquí propuesta se escogió la cuenca del arroyo Ludueña. Esta se encuentra ubicada en el sur de la provincia de Santa Fe, donde atraviesa parte de la ciudad de Rosario y de otras poblaciones. Su red de cursos permanentes y transitorios tiene una longitud de aproximadamente 120 Km. El área de la cuenca es de 717 Km², y se desarrolla casi enteramente al sur de la Ruta Nacional N°9 (RN9). Su cauce está altamente modificado debido a que presenta intervenciones antrópicas tales como terraplenes de rutas y ferrocarriles, alcantarillas y puentes, y posee una represa de retención de crecidas ubicada aproximadamente 1 Km aguas arriba de la intersección del arroyo Ludueña con la RN9. Parte del recorrido urbano del arroyo se encuentra entubado y desemboca en el río Paraná. En las inmediaciones del cruce de dicho arroyo con la Avenida de Circunvalación de Rosario, recibe el aporte del canal Ibarlucea y del canal Salvat (véase la Figura 3).

Debido al constante crecimiento de actividades humanas en sus inmediaciones y al desarrollo de nuevos barrios, gran parte de los suelos de la cuenca han sido impermeabilizados. Por esta razón, su caudal se incrementa drásticamente en periodos de precipitaciones. Además, en sus márgenes presenta densos asentamientos irregulares que por medio de canales clandestinos aportan aguas residuales no tratadas a la cuenca [34,35].

3.1.2. Sitios de muestreo y realización de campañas

Una etapa crítica en el diseño de un monitoreo ambiental es la selección de la ubicación y cantidad de sitios de muestreo. Por este motivo, en la etapa inicial del trabajo de tesis de grado desarrollado se estudió la geografía de la cuenca, a fin de diseñar un protocolo de muestreo que permitiera reflejar su comportamiento global. Se seleccionaron puntos antes y después de las distintas ramificaciones de los canales. Se

escogieron sitios que fueran accesibles, es decir, que presenten caminos rurales y/o puentes que permitieran realizar la toma de muestra.

Se definieron así once sitios de muestreo, de los cuales siete están ubicados a lo largo del arroyo Ludueña propiamente dicho, dos sobre el canal Salvat y los dos restantes sobre el canal Ibarlucea. La ubicación y distribución de todos ellos se muestra en detalle en la Figura 3.

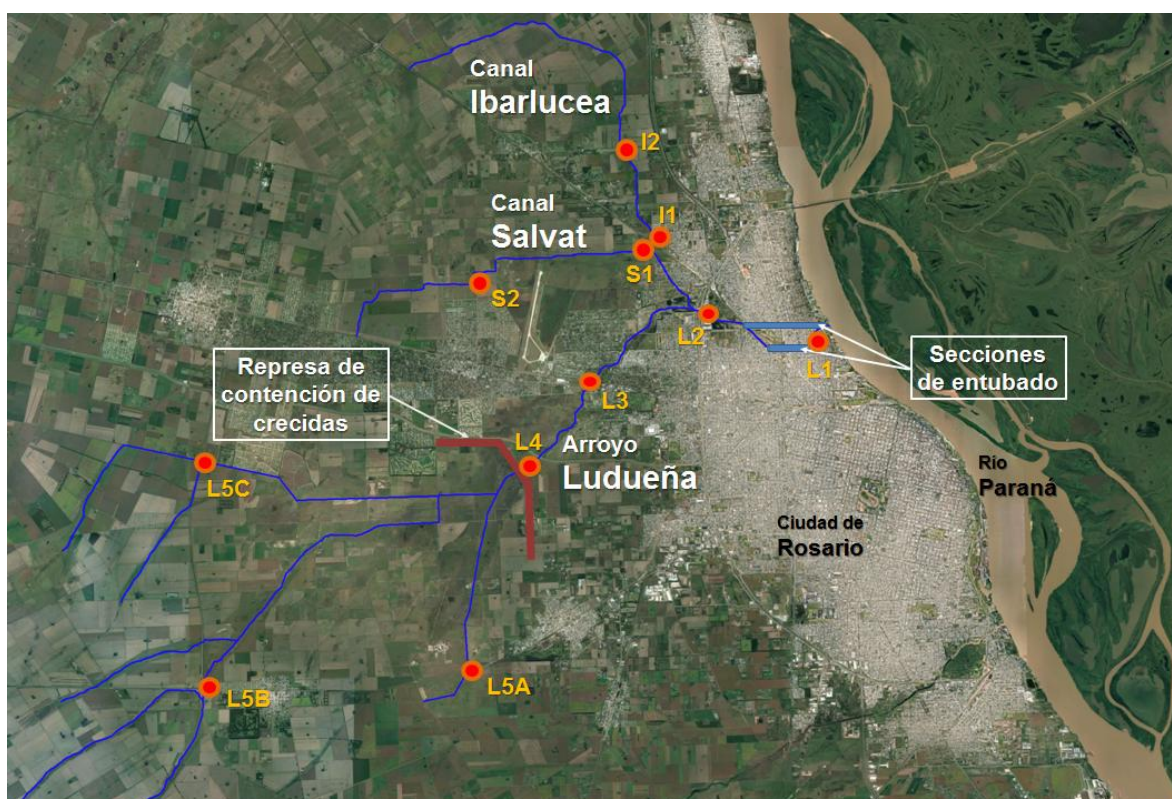


Figura 3. Sistema en estudio: Cuenca del arroyo Ludueña. Se muestra la localización del arroyo Ludueña junto a los canales principales, el canal Ibarlucea y el canal Salvat. Se indican además los distintos sitios de muestreo junto al arroyo y a los canales principales. Se representan también las secciones de entubado y la represa de contención de crecidas.

A lo largo del trabajo, se realizaron campañas de monitoreo periódicas con un intervalo de 45 días aproximadamente. Debido a que el caudal de la cuenca es muy variable con las precipitaciones, y a su vez estas 'limpian' por arrastre el contenido de los sedimentos y contaminantes de los diferentes canales, la selección de la fecha exacta de campaña fue variable. Se tuvo en cuenta que hayan transcurrido un mínimo de cinco días

desde la última precipitación. Se realizaron un total de siete campañas de muestreo desde principios de noviembre de 2016 hasta fines de septiembre de 2017, con fecha en los días 7/11/16, 12/12/16, 16/02/17, 24/04/17, 6/06/17, 24/07/17 y 19/09/17. En un mismo día se recolectaron todas las muestras de una campaña, tomando por duplicado cada una de ellas.

El muestreo se llevó a cabo comenzando en la desembocadura del arroyo Ludueña sobre el río Paraná, donde se efectuó la primera toma de muestra (L1) desde el puente situado en la calle Nansen de la ciudad de Rosario. Luego se prosiguió corriente arriba del arroyo, donde se tomó la segunda muestra (L2) posterior a la afluencia de los canales Ibarlucea y Salvat, en el sitio ubicado sobre la Colectora Este de la Avenida 25 de Mayo (Av. de Circunvalación). Sobre el canal Ibarlucea se recolectó la tercer muestra (I1), en el cruce de este con la Ruta Nacional N° 34 (RN34). Por medio de un camino rural se accedió al canal Salvat, donde se realizó la siguiente toma (S1). Luego se colectó la segunda muestra del canal Ibarlucea (I2) en el cruce del mismo con la Ruta Provincial N° 34 S (RP34S), entre las localidades de Ibarlucea y Granadero Baigorria. La tercera muestra del arroyo Ludueña (L3) se recolectó en su intersección con la Ruta Nacional N° 9 (RN9). La segunda toma de muestra del canal Salvat (S2) se realizó en la ciudad de Funes, en el cruce de calle Tucumán con la RP34S. Posteriormente, se procedió a la cuarta toma de muestra del arroyo Ludueña (L4) en la represa de contención. Luego, sobre la Ruta Nacional A012 (RNA012), se colectó la quinta muestra del Arroyo Ludueña (L5C) en la localidad de Roldán, próximo al barrio Tierra de Sueños III. Las muestras restantes (L5A y L5B) se tomaron en extensiones del arroyo, una de ellas en la intersección de la RNA012 con la Ruta Nacional N° 33 (RN33) en la localidad de Zavalla, y la otra sobre esta misma ruta en las proximidades de la localidad de Pérez.

En la Tabla 1 se muestran los datos de latitud y longitud de la localización de los diferentes sitios de muestreo.

Sitio de muestreo	Latitud	Longitud
L1	32°54'32.7" Sur	60°40'53.1" Oeste
L2	32°54'0.67" Sur	60°43'21.3" Oeste
L3	32°55'27.1" Sur	60°45'43.9" Oeste
L4	32°57'02.5" Sur	60°47'19.7" Oeste
L5A	33°00'42.2" Sur	60°48'27.0" Oeste
L5B	33°01'05.9" Sur	60°54'15.0" Oeste
L5C	32°56'54.0" Sur	60°54'19.9" Oeste
S1	32°53'03.6" Sur	60°44'51.0" Oeste
S2	32°53'37.3" Sur	60°48'18.6" Oeste
I1	32°52'50.6" Sur	60°44'27.8" Oeste
I2	32°51'07.9" Sur	60°45'07.1" Oeste

Tabla 1. Datos de latitud y longitud de los sitios de muestreo.

3.1.3. Toma de muestra

Al llegar al sitio de muestreo, se seleccionó la zona del cauce donde el agua circulaba sin interrupciones. En esta zona se realizó la toma de muestra, arrojando un recipiente sostenido con la ayuda de una cuerda en contra de la corriente del agua. Inicialmente, con el agua recolectada se efectuó un enjuague de todo el material utilizado. Luego se procedió a tomar la muestra y, con la ayuda de un embudo, se trasvasó a una botella de vidrio color caramelo de un litro. Todas las muestras se tomaron por duplicado.

Una vez finalizado el muestreo en su totalidad, las botellas se almacenaron a 4°C para su conservación y posterior análisis.

3.1.4. Procesamiento de las muestras

Con el objetivo de realizar las diferentes determinaciones, el día posterior a las distintas campañas de muestreo se trataron las muestras como se describe a continuación. Se colocaron en un baño termostático a 30 °C hasta que alcanzaran temperatura ambiente. Posteriormente, se las agitó manualmente y se les aplicó ultrasonido durante 5 minutos, para asegurar de este modo su restitución a las condiciones originales.

A continuación, se fraccionaron en tubos de ensayo porciones de aproximadamente 10 mL de volumen de cada muestra a fin de efectuar las mediciones de pH y de conductividad; el material utilizado se enjuagó previamente con las muestras a medir.

Luego, el volumen restante de cada muestra (aproximadamente 900 mL) se filtró mediante un tamiz o malla metálica de 4 mm de tamaño de apertura. En este paso se eliminaron los restos de crustáceos, algas y residuos.

De esta forma se obtuvo lo que llamamos a continuación como 'filtrado primario'. Este se fraccionó en dos partes iguales, una destinada a la adquisición de los espectros de absorción NIR, y otra para la obtención de las MEEs de fluorescencia.

El tratamiento de las muestras para la obtención de sus respectivos espectros NIR se detalla en el *Anexo*.

Para las mediciones por medio de la técnica de fluorescencia, se tomaron 200 mL del filtrado primario de cada muestra para ser filtradas nuevamente. Este procedimiento se llevó a cabo a fin de remover la totalidad del material particulado de las muestras y así disminuir las posibles interferencias que pudieran causar las partículas sólidas sobre las medidas de fluorescencia. De no efectuarse este paso, las mismas podrían absorber la luz utilizada para excitar las moléculas de la MOD como así también la luz emitida, y adicionalmente causar efecto Tyndal debido al fenómeno de dispersión. La remoción de este material fue llevada a cabo en dos etapas. Inicialmente, se realizó una filtración con vacío empleando membranas de celulosa (papel de filtro) con la especificación 'Delta Nº 365 - papel blando para filtración rápida'. Luego, el filtrado resultante fue sometido nuevamente a un proceso de filtración con vacío, utilizando filtros de membrana de nylon 'Agilent Technologies - 0,45 μm ' de 47 mm de diámetro.

Finalmente, el filtrado definitivo se trasvasó a botellas de vidrio color caramelo de 250 mL, y se almacenó de nuevo en heladera a 4°C hasta realizar la adquisición de datos.

3.1.5. Instrumental

■ Conductímetro

Las medidas de conductividad se efectuaron con un conductímetro Horiba portátil modelo D-24 (Horiba Ltd. - Kioto, Japón), conectado a una celda de conductividad modelo 9382-10D del mismo proveedor.

■ Equipo para mediciones de pH (Electrodo y pHmetro)

Las medidas de pH se realizaron empleando un pHmetro modelo Orion 420 (Thermo Fisher Scientific - Waltham, USA), equipado con un electrodo combinado de membrana de vidrio sensible a protones y electrodo de referencia Ag/AgCl (Boeco - Hamburgo, Alemania).

■ Espectrofotómetro

Las medidas de absorbancia se realizaron en un Espectrofotómetro Beckman DU 640 (Beckman Coulter Inc. - Fullerton, USA), utilizando una celda de cuarzo de 10 mm de paso óptico.

■ Espectrómetro de reflectancia de infrarrojo cercano (NIR)

El equipamiento y la metodología empleada en las mediciones de los espectros NIR se detallan en el *Anexo*.

■ Espectrofluorómetro

Las mediciones de las matrices de fluorescencia se realizaron en un Espectrofotómetro de Fluorescencia Cary Eclipse de Varian (Varian Inc. - Palo Alto, USA), equipado con dos monocromadores Czerny-Turner y una lámpara de flash de xenón. El equipo estaba conectado al microprocesador de una PC por una interfaz serie IEE 488 (GPIB). Para las mediciones se utilizó una celda de cuarzo de 10 mm de paso óptico con las cuatro caras pulidas.

3.1.6. Mediciones de pH y conductividad

Para realizar las mediciones puntuales (de orden cero) se utilizaron las fracciones de 10 mL de cada muestra separadas previo a su filtrado. Dichas mediciones se efectuaron sobre las muestras a temperatura ambiente previamente homogeneizadas. Por lo tanto, los parámetros de pH y conductividad se midieron directamente en las muestras originales (sin filtrar), empleando el instrumental antes mencionado oportunamente calibrado.

3.1.7. Mediciones de absorbancia UV-Visible (UV-Vis)

A fin de seleccionar una dilución apropiada que permita evitar el efecto de filtro interno en las medidas de fluorescencia, se adquirieron los espectros UV-Vis a una velocidad de 1200 nm/min de las muestras completamente filtradas y se registraron los valores de absorbancia entre 220 y 500 nm. Para realizar la correcta dilución de cada muestra, se tuvo en cuenta que la absorbancia a 370 nm posterior a las diluciones sea menor que 0,01 UA. Se seleccionó esta longitud de onda como parámetro de control debido a que es la longitud de onda más baja a la cual emiten fluorescencia la mayoría de los fluoróforos de la MOD. Las diluciones se efectuaron empleando material volumétrico calibrado y agua destilada en cantidad suficiente para llevar cada muestra a su volumen final correspondiente. Finalmente, las muestras ya filtradas y diluidas se almacenaron nuevamente a 4°C hasta la adquisición de las MEEs de Fluorescencia.

3.1.8. Adquisición de las MEEs de fluorescencia

Las matrices de fluorescencia se registraron en los siguientes rangos: excitación, 220-481 nm cada 3 nm; emisión, 280-600 nm cada 5 nm. Por lo tanto, cada matriz tenía un tamaño de $88 \times 65 = 5720$ datos puntuales. La velocidad de escaneo utilizada fue de 12000 nm/min. El voltaje del detector se fijó a 700 V y el slit, tanto del monocromador de excitación como del monocromador de emisión, fue de 5 nm. Estas condiciones se aplicaron para todas las muestras procesadas.

3.2. Materiales y métodos quimiométricos

3.2.1. Resolución de curvas multivariantes por cuadrados mínimos alternantes (MCR-ALS)

Las técnicas multivariadas son métodos de análisis de gran utilidad para el estudio de sistemas en los cuales la información se encuentra no resuelta. Es decir, existen señales solapadas o superpuestas debido a que la señal medida es la sumatoria total de diferentes fuentes que la originan, impidiendo que por medio de la observación directa de los datos medidos sea posible obtener conclusiones confiables. Entre todos los métodos disponibles, MCR-ALS se destaca por ser uno de los algoritmos más flexibles, ya que permite aplicar una variedad de restricciones brindando al algoritmo una gran versatilidad para adaptarse a distintos tipos de problemas. Por lo tanto, es una herramienta quimiométrica muy potente con aplicación creciente en distintas áreas de la química y la biología [36-40].

Este método trabaja con el conjunto de los datos experimentales dispuestos en una única matriz de datos **D** comúnmente llamada matriz aumentada, debido a que se extiende la longitud de las columnas al ir colocando los datos de cada muestra uno debajo del otro. La construcción de esta matriz **D** puede realizarse con datos de muestras de primer orden, es decir, un vector de datos por cada muestra, uniéndolos de forma que queden uno debajo del otro.

Para resolver la matriz **D** y encontrar las fuentes que dan origen a la información que contiene, los llamados 'componentes', se realiza una descomposición bilineal de los datos. Esta descomposición puede escribirse en forma matricial de acuerdo con la siguiente expresión:

$$D = SY^T + E \quad (1)$$

La ecuación (1) describe la descomposición de la matriz **D** en dos matrices, la matriz de cargas o matriz de *loadings* **Y^T** y la matriz de puntuaciones o matriz de *scores* **S**.

La matriz de *loadings* Y^T identifica las principales fuentes de varianza de los datos, es decir, describe el perfil de los distintos componentes en el primer modo. La matriz de *scores* S proporciona la intensidad o concentración relativa de los *loadings* en cada muestra incluida en el armado de la matriz aumentada, es decir, brinda la información del segundo modo. Así, S es la matriz de *scores* relacionada con los objetos o muestras, mientras que Y^T es la matriz de *loadings* relacionada con las variables o sensores. Cada vector de S está asociado con un vector de Y^T a través de un producto que representa un componente. En la Figura 4 se muestra una representación gráfica del modelo bilineal.

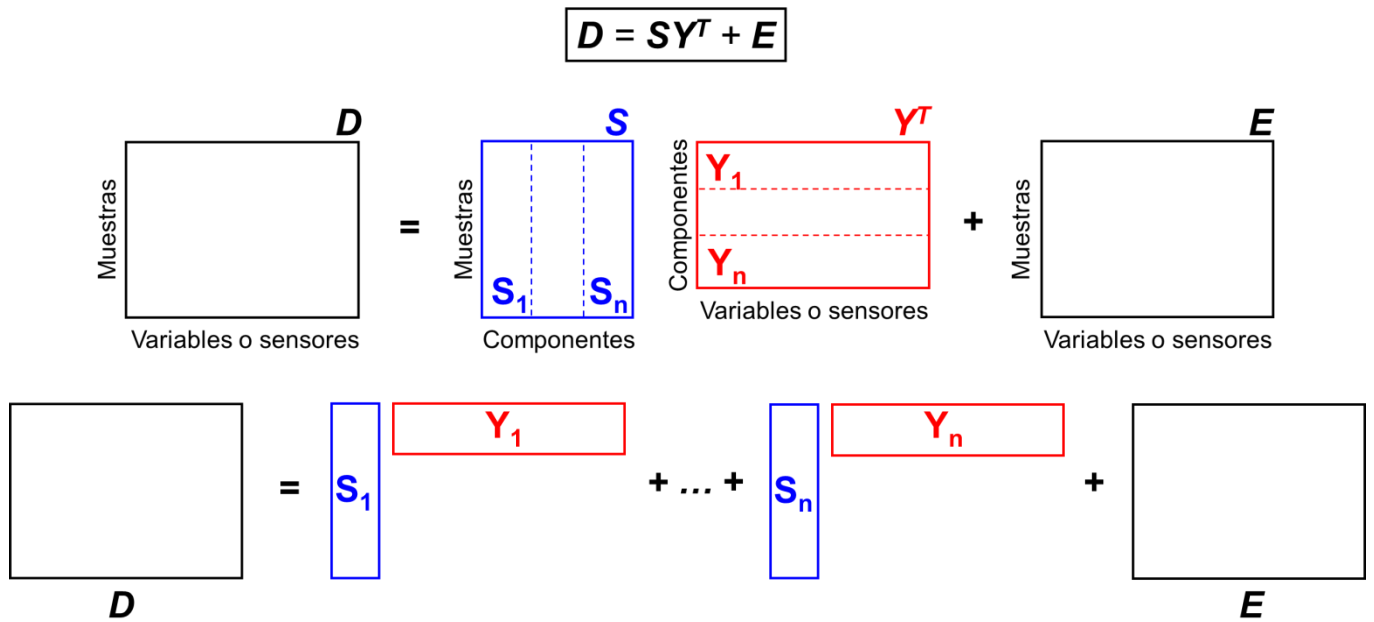


Figura 4. Representación gráfica del modelo bilineal de MCR-ALS. En la parte superior se muestra la descomposición en forma matricial, mientras que en la parte inferior se la muestra en forma vectorial.

El modelo bilineal descrito por la ecuación (1) asume que las fuentes principales de la varianza de los datos experimentales pueden ser explicadas por un pequeño número de ‘componentes’ que están definidos por dos matrices reducidas (*scores* y *loadings*). Asimismo, asume que los datos medidos para una muestra en particular son la suma de un reducido número de contribuciones de estos ‘componentes’.

Dado que la solución de la ecuación (1) no es única, su descomposición matricial debe ser llevada a cabo bajo ciertas restricciones. Dicha descomposición es similar a la de PCA, la cual se realiza bajo restricciones de ortogonalidad, normalización de los *loadings* y máxima varianza explicada para los sucesivos componentes extraídos. Bajo estas restricciones, PCA proporciona soluciones únicas. Sin embargo, estas soluciones son una combinación lineal abstracta de las verdaderas fuentes de varianza experimental y, aunque son muy útiles para la exploración y resumen de datos, en muchos casos pueden ser demasiado complejas para su interpretación ambiental [41].

Para la resolución de la ecuación (1) MCR utiliza el método iterativo de cuadrados mínimos alternantes (ALS). Durante esta etapa pueden aplicarse diferentes restricciones. Una de ellas es la restricción de no-negatividad para los perfiles en ambos modos (*loadings* y *scores* aumentados), otorgándole de esta forma significado fisicoquímico a los resultados obtenidos, ya que los *scores* están directamente relacionados con la intensidad o concentración de los 'componentes' en cada muestra y los *loadings* describen el primer modo de los datos.

A diferencia de la PCA, la descomposición bilineal de la matriz realizada por MCR-ALS utiliza restricciones naturales más suaves y como resultado la interpretación de los perfiles de *loadings* y *scores* es más sencilla y adaptable a los distintos sistemas a estudiar [6,32].

3.2.2. Resolución de curvas multivariantes por cuadrados mínimos alternantes (MCR-ALS) para modelos trilineales

Otra restricción importante que puede ser implementada durante la optimización iterativa de ALS es la condición de trilinealidad. De esta forma se obliga a que la matriz aumentada de *scores* $\mathbf{S}$ pueda ser descompuesta linealmente en los dos modos restantes.

Esta restricción puede ser aplicada cuando la matriz aumentada $\mathbf{D}$ se ha construido a partir de datos de segundo orden que presentan una relación de linealidad en un modo

adicional a los presentados en la sección anterior. Es decir, presentan linealidad en tres modos. A lo largo de este trabajo de tesis de grado se presentarán dos sistemas donde es factible exigir la trilinealidad de la resolución de MCR-ALS. El primero son los datos de MEEs de fluorescencia medidos en todas las muestras. Estos datos son naturalmente trilineales, ya que la señal depende del perfil de emisión, el perfil de excitación y la concentración (tres modos) de los distintos fluoróforos en cada muestra. El segundo sistema donde se exigirá trilinealidad es en el análisis global de los datos. Aquí la matriz aumentada está conformada por la base de datos ambiental, donde el primer modo son las distintas variables medidas en cada muestra, el segundo es la dimensión temporal reflejada por las distintas campañas de muestreo, y el tercer modo son los sitios de muestreo. En las siguientes secciones se explicará con más detalle cada uno de estos sistemas a resolver.

Para aplicar el modelo de MCR-ALS con la restricción de trilinealidad se realiza el mismo procedimiento comentado anteriormente para la obtención de los *loadings* del primer modo y la matriz de *scores* aumentados. Sin embargo, para asegurar la trilinealidad, cada columna de la matriz S se pliega apropiadamente en cada paso iterativo de la optimización con ALS para dar una matriz de dimensiones correspondientes al segundo y tercer modo. En cada iteración se realiza la descomposición en valores singulares (SVD) de esta matriz de *scores* plegada, generando los *loadings* en el segundo y tercer modo para un determinado componente. Estos dos nuevos *loadings* describen la variación de los componentes en los dos modos restantes. Luego se realiza el producto de Kronecker de estos dos nuevos vectores de *loadings*, para así obtener el nuevo vector de *scores* aumentados que sustituye a la columna correspondiente de la matriz de *scores* S (original).

A lo largo de los últimos años, esta metodología ha sido empleada con éxito en distintos sistemas ambientales [30].

3.2.3. Análisis de factores paralelos (PARAFAC)

Luego de adquirir los datos de MEEs de fluorescencia de segundo orden para un conjunto de I muestras, cada una de las matrices $\mathbf{X}_i$ de dimensiones $J \times K$ (siendo J el número de datos en el modo de emisión y K el número de datos en el modo de excitación) se agrupan en un arreglo tridimensional de datos denominado matriz $\mathbf{X}$, cuyas dimensiones son entonces $I \times J \times K$. En el caso que $\mathbf{X}$ pueda ajustarse correctamente al modelo trilineal de PARAFAC, puede ser descrita en términos de tres vectores (designados como $\mathbf{a}_n$, $\mathbf{b}_n$ y $\mathbf{c}_n$) para cada componente, reuniendo así las concentraciones relativas o *scores* ($I \times 1$) para el componente n , y los perfiles en ambos modos ($J \times 1$) y ($K \times 1$), respectivamente. En el modelo de PARAFAC, la expresión que define un elemento genérico X_{ijk} del arreglo $\mathbf{X}$ es [42]:

$$X_{ijk} = \sum_{n=1}^N a_{in} b_{jn} c_{kn} + E_{ijk} \quad (2)$$

donde N es el número total de componentes químicos (o fluoróforos) que producen la señal o respuesta, a_{in} es la concentración relativa del componente n en la i^{ma} muestra, y b_{jn} y c_{kn} son las intensidades en los modos j y k . Así, a_{in} , b_{jn} y c_{kn} son los elementos de los vectores columna $\mathbf{a}_n$, $\mathbf{b}_n$ y $\mathbf{c}_n$, los cuales corresponden, respectivamente, a las concentraciones relativas ($I \times 1$), los perfiles de emisión ($J \times 1$) y los perfiles de excitación ($K \times 1$) para cada uno de los N componentes. Los valores de E_{ijk} son los elementos del arreglo $\mathbf{E}$, que es un término de error residual de las mismas dimensiones que $\mathbf{X}$. Los vectores columna $\mathbf{a}_n$, $\mathbf{b}_n$ y $\mathbf{c}_n$ son almacenados dentro de la correspondiente matriz de puntuaciones o *scores* $\mathbf{A}$ (conteniendo las concentraciones relativas de los componentes), y dentro de las matrices de los vectores de carga o *loadings* $\mathbf{B}$ y $\mathbf{C}$ (con las columnas generalmente normalizadas a la unidad). En la Figura 5 se representa la construcción del arreglo tridimensional de los datos a partir de las MEEs de fluorescencia y la descomposición trilineal de PARAFAC.

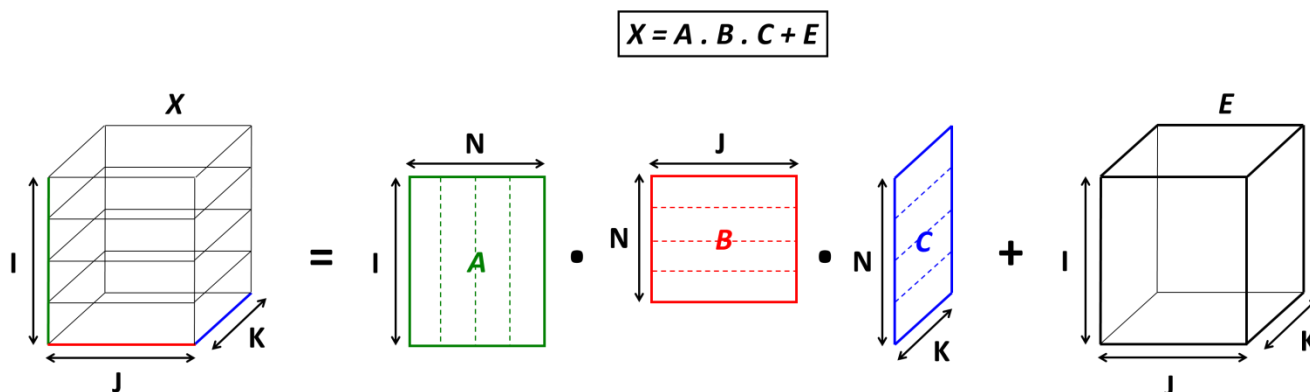


Figura 5. Representación gráfica de la descomposición del modelo trilineal de PARAFAC. **X**: arreglo tridimensional de los datos construido a partir de las MEEs de fluorescencia; **A**: matriz de *scores* relacionados con las concentraciones relativas de los componentes en cada muestra; **B**: matriz de *loadings* del modo de emisión; **C**: matriz de *loadings* del modo de excitación; y **E**: matriz de error no modelado.

El modelo descrito en la ecuación (2) define una descomposición de **X** que suministra los perfiles o *loadings* en ambos modos (**B** y **C**) y las concentraciones relativas o *scores* (**A**) de los componentes individuales de toda la mezcla, sean o no entidades químicas conocidas. La descomposición se logra generalmente a través de un esquema de minimización mediante ALS [22,43].

Las pasos críticos que deben tenerse en cuenta al aplicar el modelo de PARAFAC a datos multivariados son: 1) seleccionar la forma de inicialización del algoritmo, 2) establecer el número de factores (componentes) que generan la señal o respuesta, 3) seleccionar las restricciones del ajuste por mínimos cuadrados para obtener perfiles interpretables, y 4) identificar los componentes a partir de la información proporcionada por el modelo.

La inicialización de PARAFAC para el estudio de arreglos tridimensionales se puede hacer utilizando: 1) *loadings* proporcionados por descomposición trilineal directa (DTLD) [44], 2) perfiles espectrales para componentes puros previamente conocidos, o 3) *loadings* que brindan el mejor ajuste luego de pequeñas pruebas de PARAFAC que involucran DTLD y varios conjuntos aleatorios de *loadings*. Todas estas opciones están

disponibles en el paquete PARAFAC de Bro [45], de las cuales en este trabajo se utilizó la tercera.

Existen varias restricciones disponibles para aplicar durante el ajuste de PARAFAC mediante ALS que pueden servir para recuperar perfiles de componentes físicamente reconocibles. La restricción de no-negatividad en los tres modos cumple esta función, permitiendo que el ajuste converja al mínimo con el significado físico de los varios mínimos que pueden existir para los sistemas linealmente dependientes.

El número óptimo de factores puede estimarse por varios métodos, entre ellos el test de CORCONDIA (test de consistencia del núcleo), una herramienta diagnóstica que considera un parámetro interno de PARAFAC conocido como consistencia del núcleo [44, 46]. Este análisis involucra la observación de los cambios del parámetro de la consistencia del núcleo a medida que se aumenta el número de prueba de componentes. Se considera que un modelo es apropiado si al agregar más componentes no mejora el ajuste considerablemente. El número de componentes es tomado como el número más grande para el cual el parámetro es mayor que aproximadamente 50.

La identificación de los constituyentes químicos bajo análisis se consigue comparando los perfiles estimados con los de componentes puros conocidos, siempre que estén disponibles en forma pura o a partir de la literatura. Esto es necesario ya que los componentes obtenidos por la descomposición de $\mathbf{X}$ se ordenan de acuerdo con su contribución a la varianza espectral global.

3.2.4. Pre-procesamiento de las MEEs

Las señales correspondientes al Rayleigh y a los armónicos de segundo orden no son bilineales, es decir, no pueden ser descriptas como la combinación de perfiles espectrales simples de excitación y emisión. Por lo tanto, deben ser removidas antes del procesamiento de los datos. Con el fin de remover estas contribuciones indeseables, se aplicó una rutina escrita en MATLAB para reemplazar en las MEEs registradas de cada

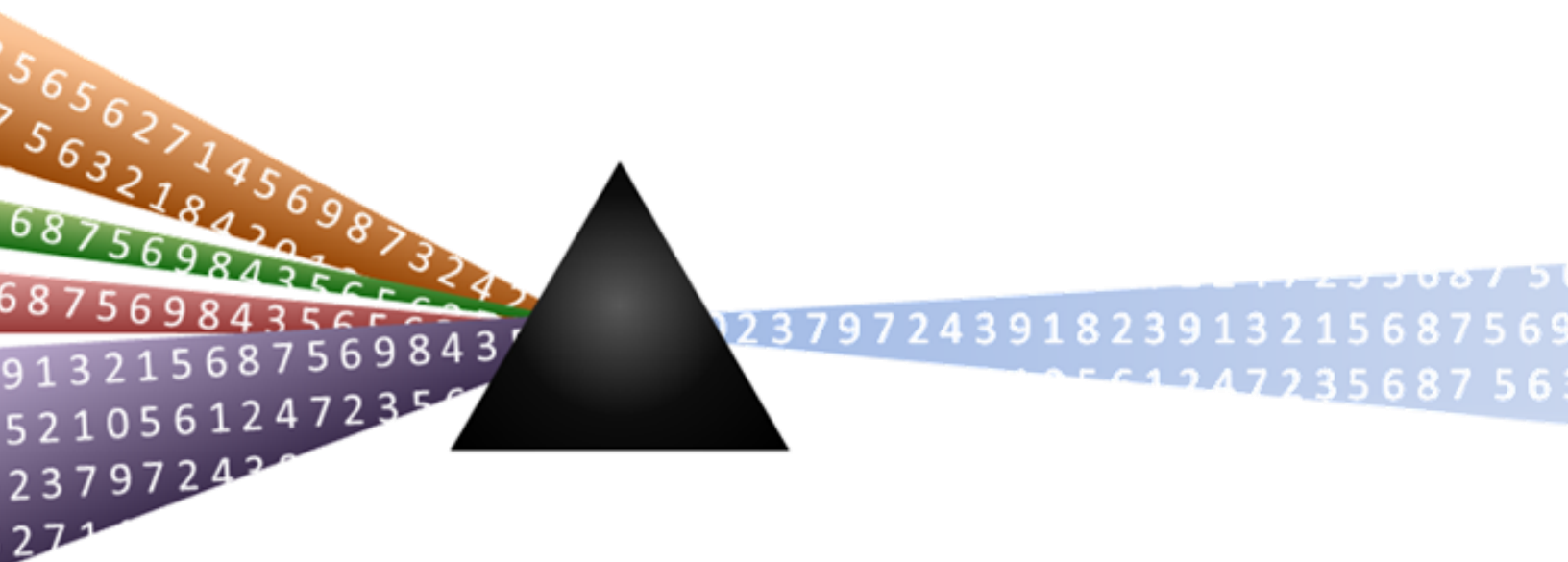
muestra los datos del área afectada por dicha dispersión con una estimación polinómica [47].

Este método de interpolación no-lineal involucra los siguientes pasos: 1) la localización de los rangos de longitudes de onda de excitación a los que aparecen las señales de dispersión, ya sean para el Rayleigh o para los armónicos de segundo orden, 2) la eliminación de las señales de fluorescencia a esas longitudes de onda, y 3) su reemplazo mediante valores de interpolación polinómica, utilizando los puntos adyacentes a la región eliminada como referencia para la estimación de las constantes polinómicas. Una vez que cada MEE fue corregida, se procedió a la construcción del correspondiente arreglo de datos necesario para MCR-ALS trilineal o para PARAFAC.

3.2.5. Software

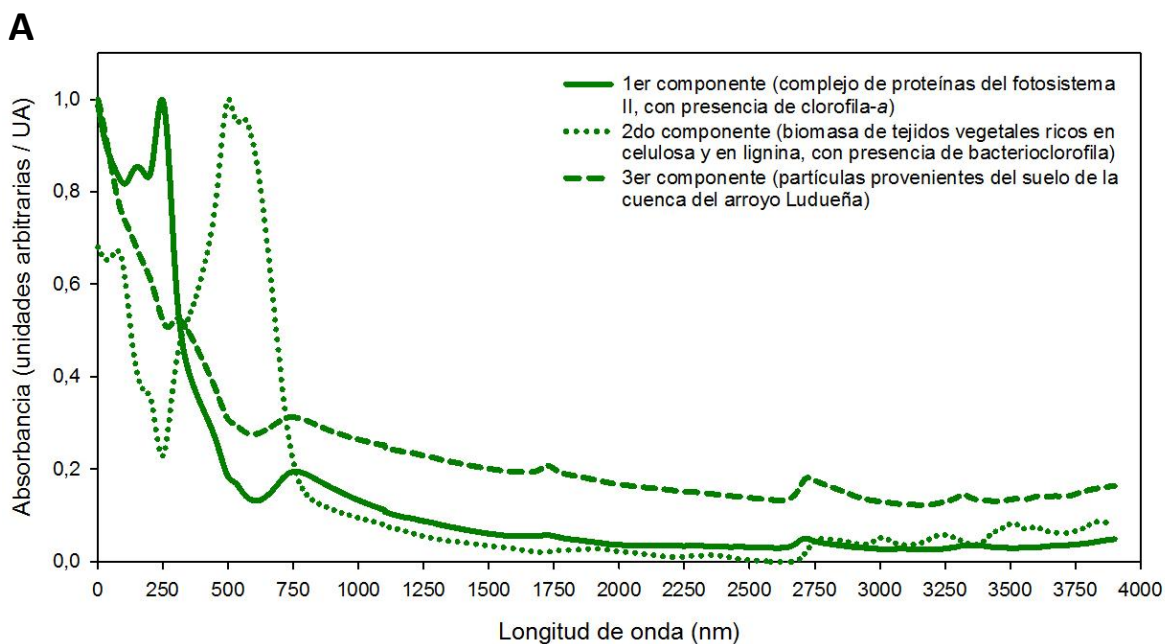
Todas las rutinas empleadas para realizar los cálculos en esta tesis de grado fueron escritas en MATLAB [48]. PARAFAC se aplicó utilizando las rutinas disponibles en internet gracias a Bro [45]. Para la corrección de la dispersión Rayleigh y la dispersión Raman se adaptó la rutina provista por Zepp [47]. Asimismo, con el fin de aplicar MCR-ALS, se emplearon los códigos de interfaz gráfica disponible en internet gracias a Tauler [49,50]. Las representaciones espaciales geo-referenciadas se realizaron con una rutina desarrollada por nuestro grupo de trabajo, utilizando como base imágenes aéreas obtenidas a partir de Google Maps [51].

4. Resultados y Discusión



4.1. Resultados obtenidos a partir de los datos de NIR

Esta etapa del trabajo, correspondiente al sector verde la Figura 2, pertenece a la tesis de grado citada en el *Anexo*, donde se tuvo como objetivo localizar e identificar las fuentes de los sedimentos del sistema en estudio. Al aplicar el modelo bilineal de MCR-ALS al conjunto de espectros de NIR se lograron identificar tres fuentes principales de sedimentos, obteniendo el espectro que caracteriza a cada una de ellas y el aporte a cada muestra recolectada. Los resultados más importantes de este análisis están reflejados en la Figura 6, donde se muestran los perfiles (*loadings*) y el aporte a cada muestra (*scores*) de las tres fuentes principales de sedimentos. Para más detalles, remitirse al trabajo de tesis de grado previamente mencionado.



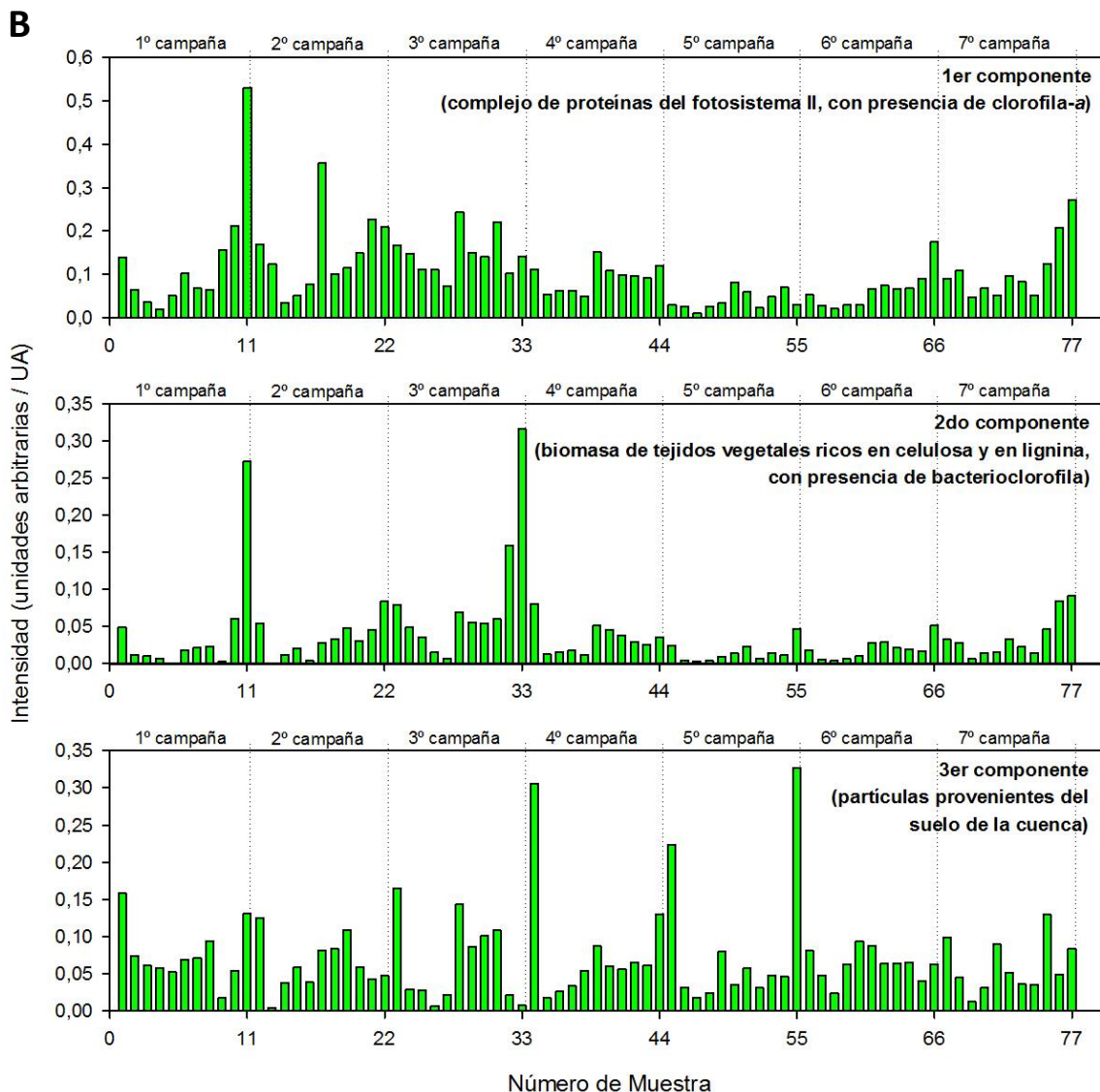


Figura 6. **A)** Perfiles (*loadings*) de los tres patrones principales de sedimentos descritos. **B)** Aportes de estos patrones (*scores*) a cada una de las muestras obtenidas.

MCR-ALS se aplicó con la restricción de no-negatividad, simplificando de este modo la interpretación de los resultados obtenidos. La varianza explicada fue del 99,7% para un modelo de tres componentes. El primer componente (98,0% de la varianza total explicada) presenta una señal espectral (*loading*) en la misma región que absorbe el principal complejo de proteínas del fotosistema II [52]. Esta señal es consistente con la presencia de organismos que contienen clorofila-*a*, la cual es clave para realizar la fotosíntesis oxigénica y presenta un pico de absorción característico a 680 nm. Por lo tanto, este componente representa el material sedimentario aportado por organismos

que desarrollan fotosíntesis oxigénica y proliferan en periodos de altas temperaturas, en áreas con altas cargas de nitratos y fosfatos.

El segundo componente (1,0% de la varianza total explicada) presenta dos regiones relevantes en su espectro (*loading*). La región a elevadas longitudes de onda está relacionada al espectro característico de biomasa de tejidos vegetales ricos en celulosa y lignina [53,54]. Por otra parte, la región entre 750 y 900 nm es consistente con el espectro de la bacterioclorofila [52]. Esta última aseveración se basa en que el principal complejo encargado de captar la luz en bacterias verdes, bacterias púrpuras, heliobacterias y clorobacterias, absorbe en la región entre 770 y 870 nm, teniendo lugar en este caso la fotosíntesis anoxigénica [55]. El material correspondiente a biomasa proviene de los tejidos vegetales que se desarrollan debido a la gran carga de nutrientes del agua (eutrofización de la cuenca).

Finalmente, el espectro (*loading*) del tercer componente (0,7% de la varianza total explicada) se corresponde con el espectro NIR de las partículas del suelo de la cuenca, incorporadas al sistema por el viento o por resuspensión debida al flujo de corriente. Esta conclusión fue corroborada comparando este *loading* con espectros de varias muestras de suelo de la zona, obteniendo siempre el mismo resultado.

En definitiva, como resultado de este análisis se obtuvieron las matrices de *loadings* y *scores*. A partir de los *loadings* se pudo caracterizar el material particulado (Figura 4-A), mientras que los *scores* fueron almacenados dentro de la matriz **S** que contiene la información cuantitativa (Figura 4-B). Esta matriz será utilizada posteriormente para realizar la fusión y el análisis de datos junto a la información proveniente de las otras técnicas analíticas propuestas.

4.2. Análisis de los datos de fluorescencia

Para abordar el objetivo de analizar globalmente los datos medidos, luego de realizar el análisis de los datos de espectroscopía NIR mencionado en la sección anterior, se debió identificar y estimar la proporción de los fluoróforos presentes en cada muestra recolectada a partir de los datos de las MEEs de fluorescencia. Esta etapa del trabajo está representada en el sector celeste de la Figura 2.

Normalmente, la descomposición de las señales de fluorescencia a partir de muestras de patrones conocidos es una tarea compleja. En el trabajo de tesis de grado aquí desarrollado se plantea un desafío aún mayor para los algoritmos quimiométricos al proponer resolver las MEEs de fluorescencia, las cuales contienen una variedad de compuestos con espectros superpuestos de identidad totalmente desconocida, siendo por ende una tarea todavía más difícil. En este sentido, el registro de las MEEs de fluorescencia proporciona una gran cantidad de información que, sin embargo, debe procesarse adecuadamente mediante algoritmos apropiados a fin de alcanzar una deconvolución exitosa de la contribución individual de los diversos componentes de una muestra. Para cumplir con este propósito, dichos datos pueden ser procesados convenientemente con PARAFAC [22] o MCR-ALS (con restricción de trilinealidad), debido a que ambos algoritmos resuelven las señales de las muestras en una primera etapa sin la necesidad de utilizar patrones. Sin embargo, sólo se informan los resultados obtenidos con PARAFAC, debido a que ambos algoritmos, bajo estas restricciones, trabajan en forma similar y generan resultados semejantes. Más allá de esto, se seleccionó PARAFAC por ser el algoritmo tradicionalmente utilizado para analizar datos de fluorescencia.

Previo a este análisis, se corrigió la señal del Rayleigh de las matrices de fluorescencia individuales utilizando el método de descrito por Zepp [47], dado que la misma no responde a la descomposición trilineal exigida por los algoritmos a aplicar. Este método identifica las zonas de las matrices de datos afectadas por las señales de Rayleigh y Raman de primer y segundo orden, y las reemplaza por una estimación polinómica ajustada utilizando los datos restantes. Una vez corregidas estas señales, las MEEs de

todas las campañas se ordenaron una sobre la otra en un solo cubo de datos, formando así un arreglo tridimensional de tamaño $83 \times 65 \times 154$ (excitación $\times$ emisión $\times$ sitio-campaña-duplicados). Posteriormente, se analizó con PARAFAC este cubo de datos.

Los pasos más importantes para obtener un análisis de PARAFAC exitoso son: la selección del modo de inicialización, la elección del número de componentes y aplicar restricciones adecuadas durante la fase de ajuste de mínimos cuadrados. Se seleccionó como modo de inicialización un método que realiza una serie de pequeñas ejecuciones a partir de valores DTLD o de números elegidos al azar. También se aplicó la restricción de no-negatividad durante la fase del ajuste de mínimos cuadrados de PARAFAC. Ambas opciones están disponibles en el paquete PARAFAC disponible en la web y son fácilmente aplicables [45].

Para estimar el número de componentes de PARAFAC, se calcularon los valores de la desviación estándar residual en función de un número creciente de componentes. Como se muestra en la Figura 7, el ajuste residual decrece significativamente al pasar del primer al tercer componente, y luego continúa disminuyendo levemente (3,6252; 2,8321; 2,4121; 2,3105; 2,2416; 2,1770 y 2,1506 unidades arbitrarias de fluorescencia o UAF son los respectivos valores de ajuste residual del modelo de PARAFAC de uno a siete componentes). También se aplicó el test de CORCONDIA, obteniéndose los valores 100,0000; 77,9808; 63,5066; 42,5800; 3,1866; 3,3854 y 3,3853 unidades arbitrarias de fluorescencia al pasar de uno a siete componentes.

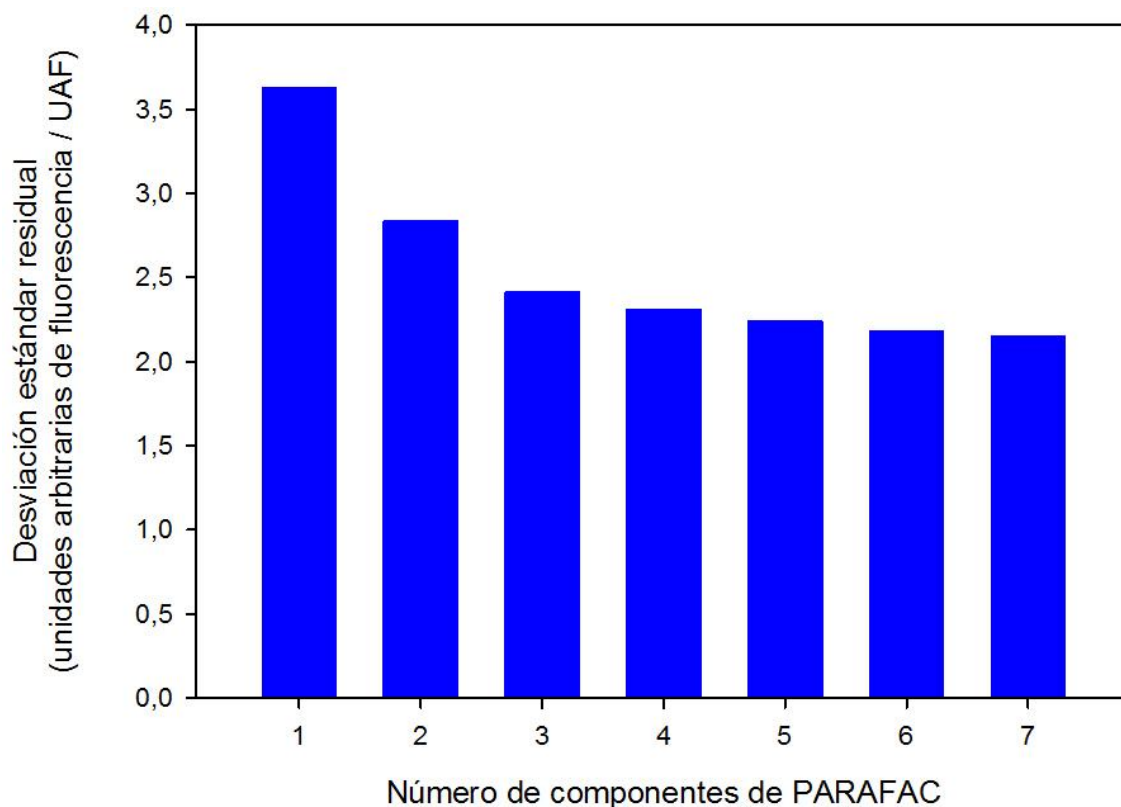
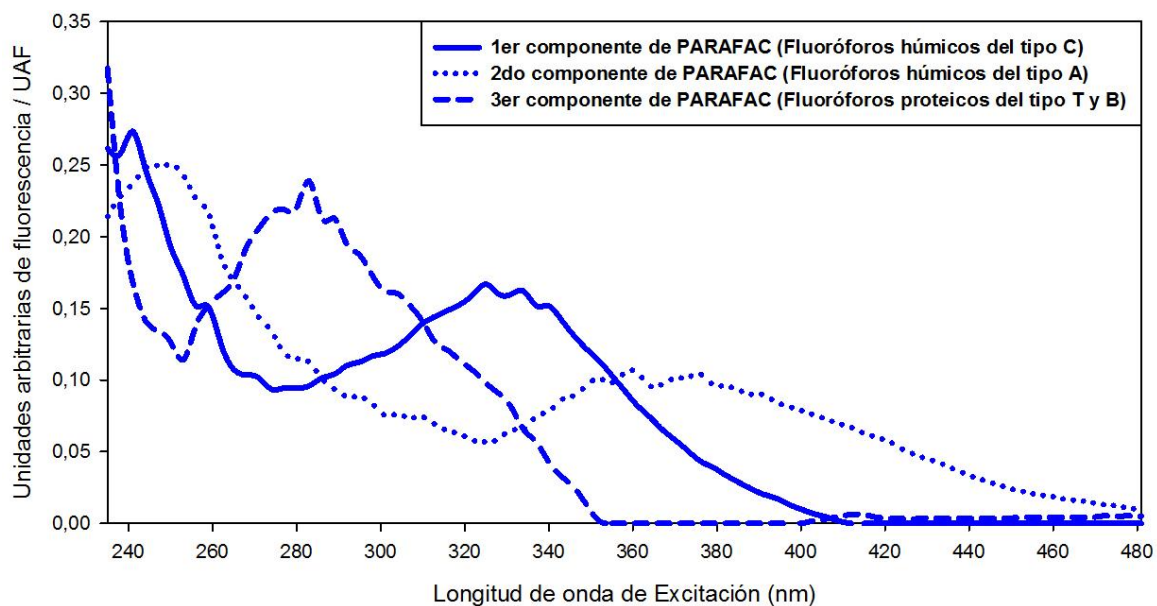
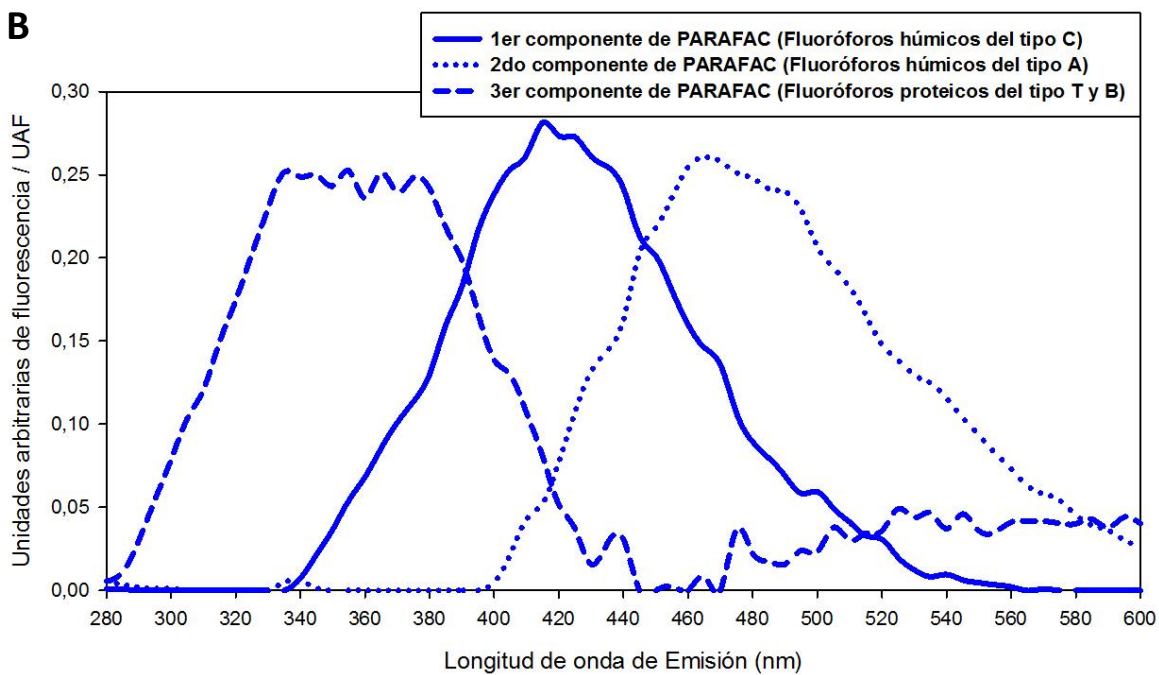


Figura 7. Valores de ajuste de desviación estándar residual, expresados en UAF, en función del número creciente de componentes de PARAFAC.

Por lo tanto, se seleccionó un modelo de PARAFAC de tres componentes. A partir de este análisis, se pudo obtener la proporción de cada componente en cada muestra, así como también los perfiles de excitación y emisión que los caracterizan (Figura 8). Posteriormente, se realizó su identificación a través de datos bibliográficos [56]. Por lo tanto, los diferentes componentes de PARAFAC representan a uno o varios fluoróforos con comportamiento similar en las muestras.

A**B**

C

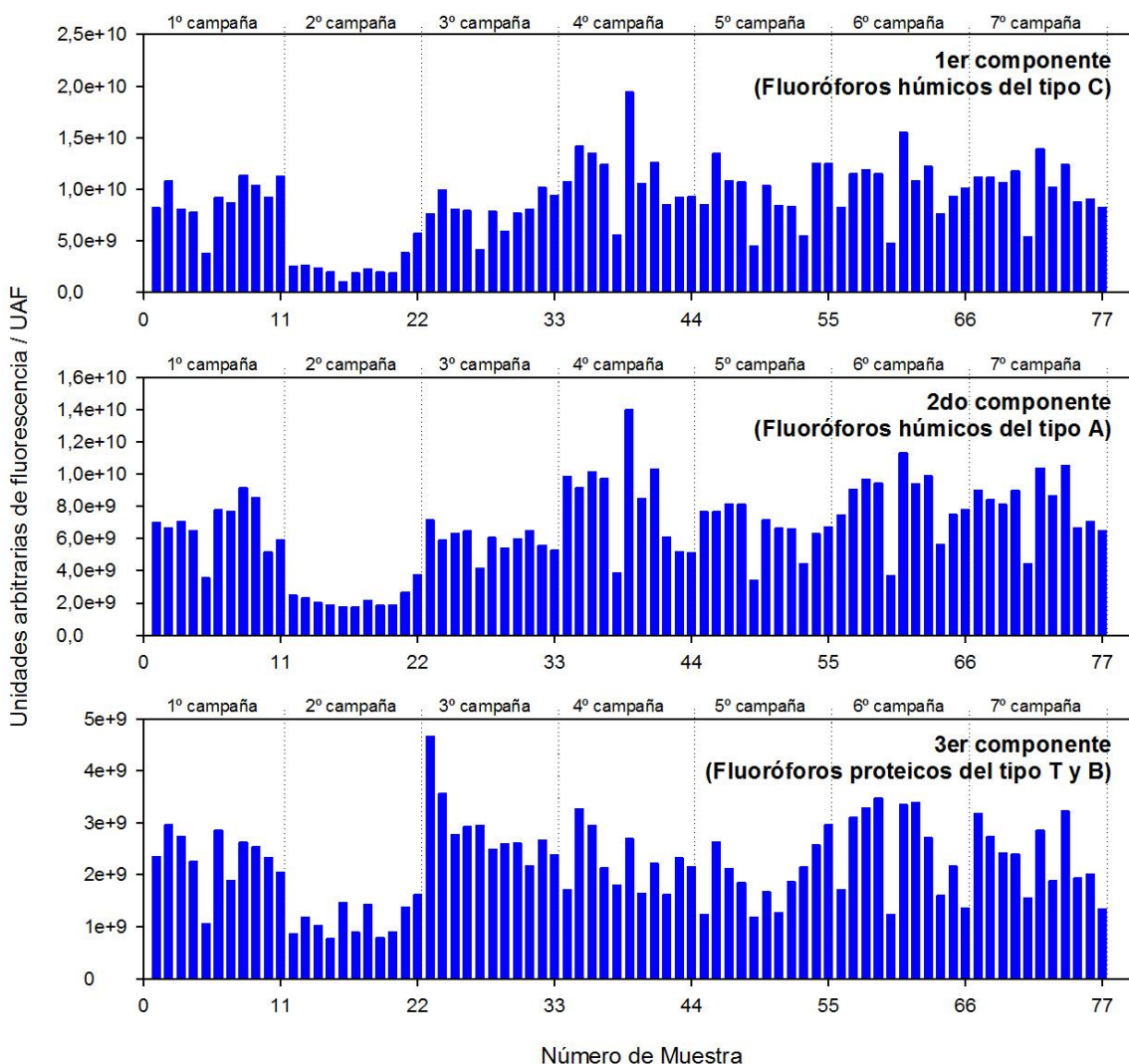


Figura 8. Resultados del análisis de PARAFAC de los datos de MEEs de fluorescencia. **A)** Espectros de excitación de fluorescencia de los tres componentes de PARAFAC. **B)** Espectros de emisión de fluorescencia de los tres componentes de PARAFAC. **C)** Distribución de los *scores* de los tres componentes de PARAFAC en cada muestra recolectada.

Los ácidos húmicos son uno de los principales fluoróforos que normalmente se encuentran en los sistemas acuáticos. Estos no son un único compuesto, sino son una mezcla compleja de compuestos estructuralmente relacionados. Por medio de datos bibliográficos es posible su caracterización al encontrarse tabuladas las regiones espectrales donde son excitados y donde emiten fluorescencia [56]. La información específica sobre los fluoróforos del tipo o clase húmica muestra que tienen un máximo de

emisión de fluorescencia concentrado en el rango de 400-500 nm, pero los perfiles espectrales de excitación se distribuyen entre dos regiones diferenciadas: 237-260 nm para los húmicos del tipo A, que son los de origen autóctono, y 300-370 nm para los húmicos del tipo C, que son alóctonos. En cuanto a los fluoróforos con un alto contenido proteico, se pueden distinguir dos clases: los fluoróforos del tipo B (ricos en tirosina), con excitación en el rango de 225-237 nm y a 275 nm y emisión entre 309-321 nm, y los fluoróforos del tipo T (ricos en triptófano), también con excitación en el rango de 225-237 nm y a 275 nm pero con emisión entre 340-381 nm [56].

Tal como se observa en la Figura 8, los perfiles para el primer componente de PARAFAC corresponden a los espectros de sustancias húmicas del tipo C (con excitación a 300-370 nm y emisión a 400-500 nm). Estas son características de ambientes poco contaminados debido a que son sustancias húmicas de origen natural. Están asociadas con la presencia de material orgánico de origen alóctono, es decir, que no se generan dentro del cuerpo de agua sino que son incorporadas por factores naturales. Normalmente, provienen del suelo cercano a los bordes de la cuenca y se transportan al seno de la misma por las lluvias y los vientos.

Por otro lado, los perfiles para el segundo componente de PARAFAC concuerdan con las características espectrales de las sustancias húmicas del tipo A (con excitación a 230-260 nm y emisión en la región de 400-500 nm, véase la Figura 8). Su presencia está asociada con la materia orgánica autóctona, es decir, generada dentro del curso de agua por la actividad biológica. Son características de ambientes que no se ven afectados, y constituyen la mayor proporción de sustancias húmicas de la MOD.

Finalmente, el tercer componente de PARAFAC puede asociarse con los fluoróforos del tipo T y B conjuntamente. Estos son característicos de ambientes con un gran impacto antropogénico, ya que presentan una elevada fracción proteica correspondiente al aminoácido triptófano (fluoróforos del tipo T, excitación a 225-237 / 275 nm y emisión a 340-381 nm, véase la Figura 8) y al aminoácido tirosina (fluoróforos del tipo B, excitación a 225-237 / 275 nm y emisión a 309-321 nm, véase la Figura 8). Se trata entonces de

sustancias proteicas, no-húmicas, de origen antropogénico, cuya presencia en los cursos de agua está ligada a la materia orgánica proveniente de efluentes industriales y/o aguas residuales sin tratar.

Una vez identificados los distintos componentes de PARAFAC, la matriz de *scores* **A** de tamaño 154x3 se multiplicó por el inverso del factor de dilución que se había aplicado previamente a cada muestra para evitar el fenómeno de filtro interno. De esta forma, los datos de fluorescencia quedan condensados dentro de la matriz de *scores* **A** corregida. Como dicha matriz contiene la información del aporte de cada fluoróforo a cada muestra en las condiciones iniciales, es utilizada para su posterior análisis en forma conjunta con la información proveniente de las otras técnicas propuestas en este trabajo.

De este manera, la matriz de *scores* **A** corregida por el factor de dilución se combinó en una matriz de datos única junto con los valores de pH y conductividad y los *scores* de MCR-ALS bilineal de los datos de NIR, a fin de analizarlos globalmente por MCR-ALS con trilinealidad [50], como si cada componente de PARAFAC o MCR-ALS fuera una variable medida directamente. Sin embargo, para analizar conjuntamente todos los datos y evitar que algunos tengan más importancia que otros debido a sus diferentes magnitudes, se aplicó previamente un escalado apropiado de los datos, tal como se detalla en la siguiente sección.

4.3. Fusión y escalado de los datos

Para proseguir con el análisis global es necesario fusionar los datos obtenidos por cada una de las diferentes técnicas en una única base de datos. Con este objetivo se creó una nueva matriz de datos, donde las filas son los distintos sitios de muestreo en todas las campañas por duplicado y las columnas son todas las variables medidas. Los datos de conductividad y pH, al ser de orden cero, se pudieron colocar directamente uno al lado de otro en forma de columnas. Sin embargo, como se explicó previamente, los datos medidos son de distintos órdenes, de manera que resulta imposible construir esta nueva matriz en forma directa. Por tal motivo, se utilizan los *scores* de los análisis de los datos de NIR y fluorescencia, los cuales resumen en datos de orden cero la información medida por estas técnicas.

Una vez que los datos de las MEEs de fluorescencia y de los espectros de NIR fueron respectivamente analizados por PARAFAC y MCR-ALS bilineal, se obtuvieron los *scores* que contienen las proporciones de cada fluoróforo (o grupo de fluoróforos) y de cada fuente de sedimentos en todas las muestras. Esta información es almacenada en la matriz de *scores A* de PARAFAC, por un lado, y en la matriz de *scores S* de MCR-ALS, por el otro. Las filas en estas matrices se corresponden con los sitios de muestreo de cada campaña, mientras que las columnas se relacionan con cada componente. La matriz de *scores A* corregida de PARAFAC tiene un tamaño de 154x3 que brinda la información de las 3 principales fuentes de fluoróforos a lo largo de toda la cuenca. Esta matriz se multiplicó por el factor de la dilución aplicada a las muestras previamente a su medición, para de esta forma obtener la proporción real de los fluoróforos en cada muestra. Por otra parte, la matriz de *scores S* también tiene un tamaño de 154x3, dado que se identificaron tres fuentes principales de sedimentos. Por lo tanto, con las matrices de *scores A* corregida y *S* es posible construir la nueva base de datos global. Se agregaron dos columnas adicionales, una correspondiente a las mediciones de pH y otra correspondiente a las medidas de conductividad, las cuales se muestran en la Figura 9.

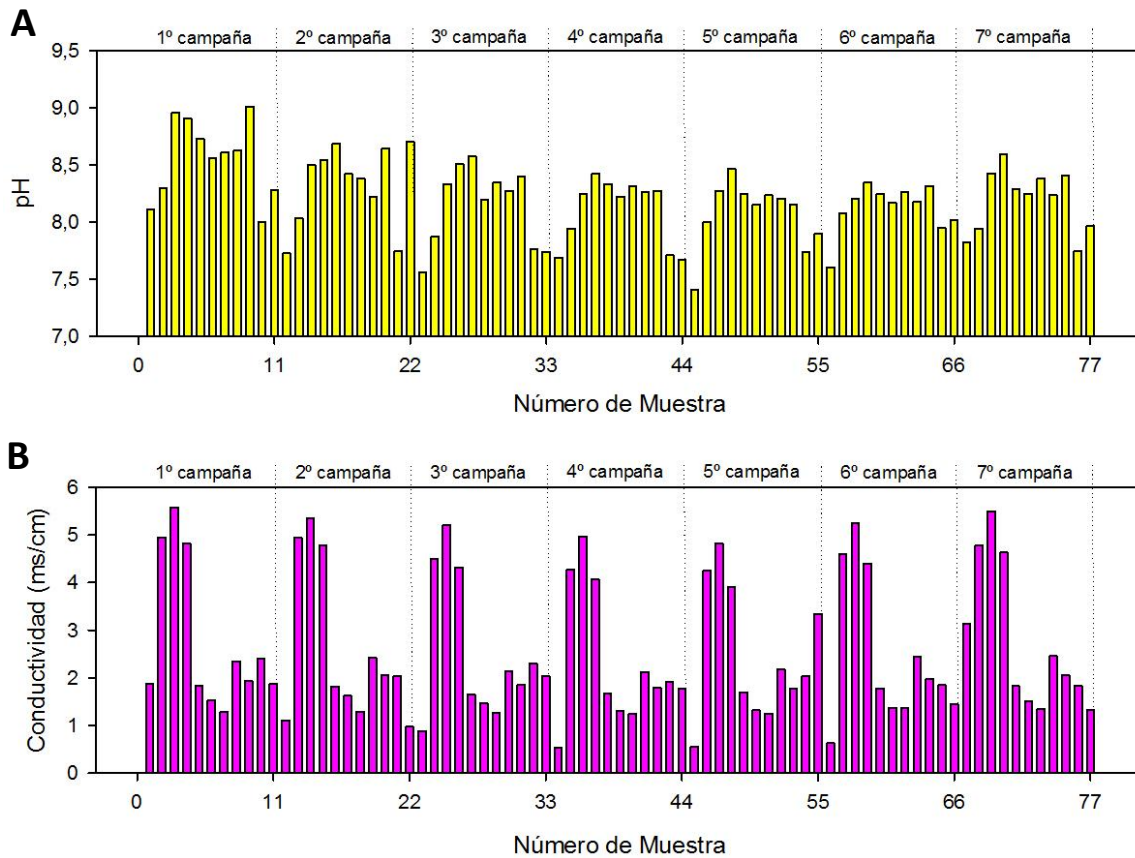


Figura 9. Mediciones puntuales de A) pH y B) Conductividad.

De este modo, se construyó una nueva matriz de datos global llamada **M** de dimensiones 154x8, tal como se observa en la Figura 10. Esta etapa del trabajo está representada en el sector gris de la Figura 2.

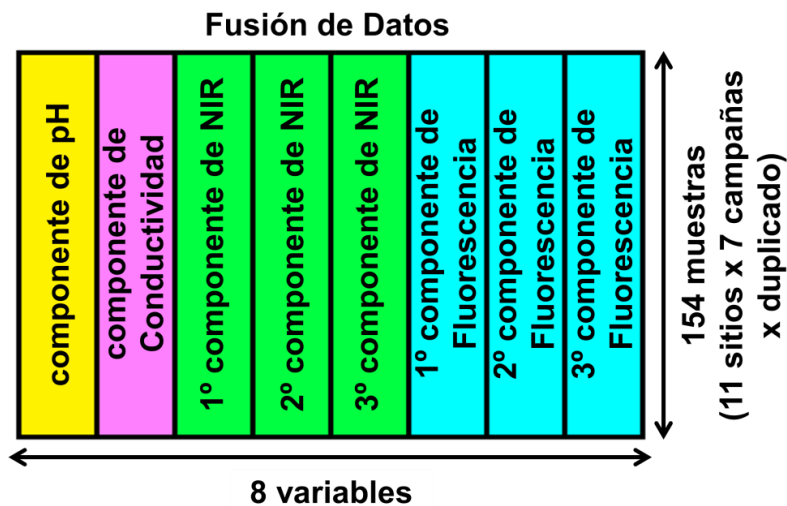


Figura 10. Fusión de los datos y construcción de la nueva matriz global **M** de dimensiones 154x8.

Un punto crítico en el análisis conjunto de las distintas variables es el escalado de las mismas. Una vez que todos los datos de los fluoróforos y los sedimentos se fusionaron con los de pH y conductividad en un mismo arreglo, se procedió a corregir la escala. Este 'escalado' se sustenta principalmente en dos hechos:

- La distribución de la mayoría de las variables ambientales no es estadísticamente normal, sino logarítmica. Esto puede ser producto de fenómenos climáticos y/o antropogénicos que tienen lugar en el sistema en estudio y generan tal dispersión, y adicionalmente, se debe tener en cuenta la heterogeneidad entre los distintos puntos de la cuenca, siendo este un factor sumamente relevante.
- La gran diferencia de escala entre las distintas variables que conforman la matriz global **M**. Así, durante el análisis podría darse mayor importancia a aquellas variables de magnitudes mayores, llevando de este modo a consideraciones erróneas.

A los fines de efectuar un escalado óptimo, se probaron distintos tratamientos matemáticos sobre la matriz de datos global **M**. De entre ellos, la transformación de aplicar raíz cuadrada seguida de MinMax resultó ser el pre-tratamiento más adecuado para esta clase de datos, tal como puede observarse en la Figura 11-A. La expresión algebraica específica de esta transformación es:

$$x_{transf} = \frac{\sqrt{x_i} - \min(\sqrt{x})}{\max(\sqrt{x}) - \min(\sqrt{x})}$$

donde x es un vector correspondiente a los valores de cada una de las ocho variables, $\max(\sqrt{x})$ y $\min(\sqrt{x})$ son el máximo y el mínimo de la transformación de raíz cuadrada del vector x , respectivamente, mientras que x_i y x_{transf} son los elementos crudos y transformados. Este procedimiento se llevó a cabo para cada columna de la matriz **M**.

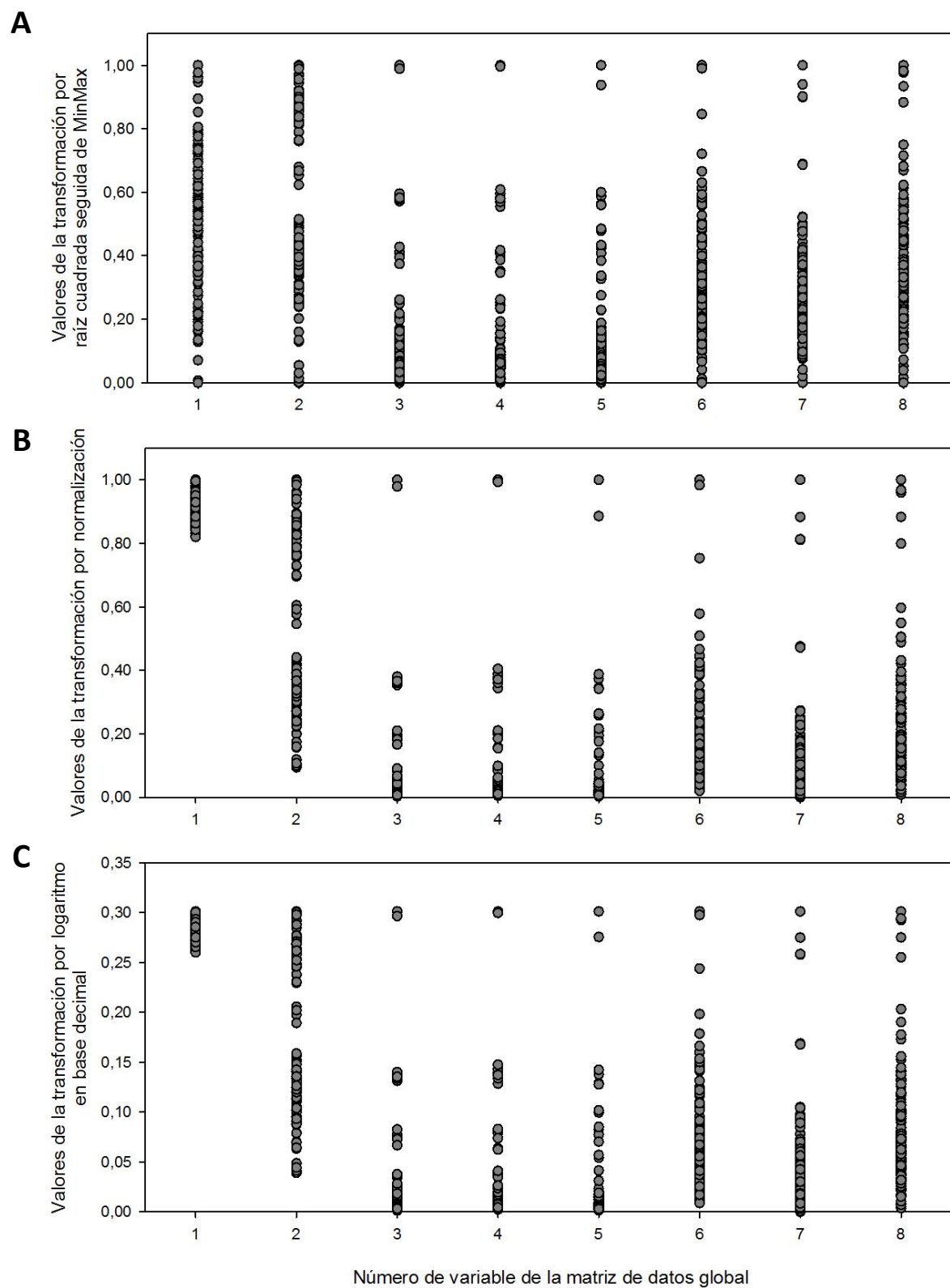


Figura 11. Distintos tratamientos de escalado sobre la matriz de datos global. **A)** Raíz cuadrada seguida de MinMax. **B)** Normalización. **C)** Logaritmo en base decimal.

Además de este tratamiento, también se evaluaron otros como la normalización y la función logarítmica en base decimal. Sin embargo, como se muestra en las Figuras 11-B y 11-C, estas transformaciones no lograron corregir con éxito la distribución logarítmica de los datos y no resultaron tan apropiadas como la mencionada previamente.

La matriz obtenida después del pre-procesamiento de raíz cuadrada seguida de MinMax es la denominada matriz aumentada ***D***, debido a que contiene la información de una campaña debajo de la otra. Finalmente, esta matriz global ***D*** se analizó por MCR-ALS trilineal, obteniéndose los resultados que se comentan en la siguiente sección.

4.4. Análisis de la matriz de datos global

Una vez obtenida la matriz de datos global D y realizado el escalado apropiado, se procedió a efectuar su análisis mediante MCR-ALS. A diferencia del enfoque desarrollado previamente con este mismo algoritmo, y dado que la matriz aumentada está compuesta por tres modos (variables×campañas×sitios de muestreo), en esta oportunidad no solo se aplicó la condición de no-negatividad sino también la trilinealidad. La obtención de los *loadings* del primer modo se realiza exactamente igual que al aplicar el modelo bilineal. Sin embargo, en este caso los *loadings* representan el aporte de las distintas variables a un determinado ‘componente’ de MCR-ALS y se interpretan como potenciales fuentes de contaminación. El modelo trilineal se implementa como una restricción durante la optimización iterativa con ALS en el método de MCR-ALS. De esta forma se exige que no solo los *loadings* de los componentes obtenidos sean los mismos en todas las muestras, como ocurre en los modelos bilineales, sino que adicionalmente su distribución entre los sitios de muestreo sea equivalente en todas las campañas, permitiendo únicamente que varíe la intensidad total entre ellas. De esta manera, una característica distintiva de la restricción de trilinealidad es su menor flexibilidad, lo que conlleva a un porcentaje menor de varianza explicada. Sin embargo, presenta la ventaja de resolver el análisis separando los distintos componentes en los tres modos (grupos de variables, distribución espacial o sitios de muestreo, y tiempo o campañas), haciendo mucho más simple la interpretación en términos de su distribución geográfica y temporal.

Teniendo en cuenta estas características, y con el objetivo de estudiar la distribución espacial y temporal conjunta de todas las variables medidas, se propuso aplicar el modelo de MCR-ALS trilineal con no-negatividad sobre la matriz de datos global. Esta etapa del trabajo está representada en el sector rojo de la Figura 2. Con el fin de diferenciar las dos secciones del trabajo donde se aplica MCR-ALS, y dada la interpretación que tienen los *loadings* de los componentes en cada caso, durante este análisis global de los datos se prefiere hablar de ‘grupos de variables’ en lugar de ‘componentes’. Durante la optimización de MCR-ALS, las variables que tienen un origen común o presentan un

comportamiento similar son colocadas dentro del mismo grupo y, de este modo, es posible obtener información valiosa acerca de la composición de potenciales fuentes de contaminación (sean puntuales o difusas), como así también de la localización y evolución temporal de cada grupo de variables.

Como se comentó anteriormente en el apartado 3.2.1 de *Materiales y métodos* referido a MCR-ALS, una etapa clave del modelado consiste en la selección del número de componentes que justifican la varianza de los datos. Se seleccionaron para el análisis cinco grupos de variables a través de la descomposición en valores singulares (SVD) de la matriz de datos global escalada, dado que el agregado de grupos adicionales no afectaba significativamente la magnitud de los autovalores de la SVD y, adicionalmente, se obtiene una óptima interpretación de los *loadings* (11,80; 4,29; 3,25; 2,12; 1,80; 1,01; 0,21 y 0,10 son los valores en unidades arbitrarias desde uno a ocho grupos, respectivamente, los cuales se muestran en la Figura 12).

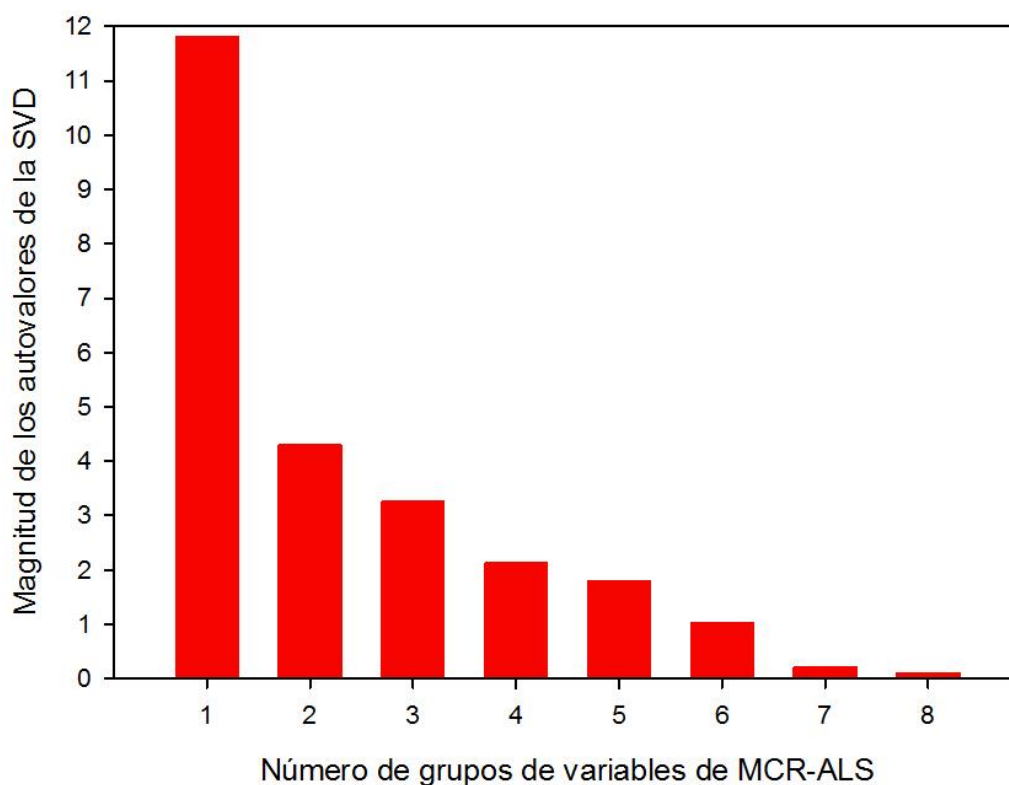


Figura 12. Magnitud de los autovalores de la SVD en función del número de grupos de variables, resultantes del análisis por MCR-ALS aplicando las restricciones de no-negatividad y trilinealidad.

A partir de los cinco grupos de variables seleccionados, se obtuvo un 95,9% total de varianza explicada, siendo entonces posible interpretar las agrupaciones reflejadas en la Figura 13:

- Dentro del primer grupo (77,8% de varianza explicada) se distingue una elevada contribución correspondiente a las partículas provenientes del suelo de la cuenca, acompañada de aportes menores de las otras fuentes de sedimentos (el complejo de proteínas del fotosistema II, con presencia de clorofila-*a*, y la biomasa de tejidos vegetales ricos en celulosa y lignina, con presencia de bacterioclorofila).
- Por otra parte, el segundo grupo de variables (9,4% de varianza explicada) se destaca principalmente por los aportes de elevada conductividad y valores de pH más alcalinos, junto a una pequeña proporción de sedimentos de las partículas provenientes del suelo de la cuenca.
- En tanto, el tercer grupo de variables (5,2% de varianza explicada) se caracteriza por la contribución de los valores de pH alcalinos y, en menor medida, de las tres fuentes de sedimentos identificadas (el complejo de proteínas del fotosistema II, con presencia de clorofila-*a*, la biomasa de tejidos vegetales ricos en celulosa y lignina, con presencia de bacterioclorofila, y por último las partículas provenientes del suelo de la cuenca del arroyo Ludueña).
- Adicionalmente, en el cuarto grupo de variables (1,3% de varianza explicada) se aprecia un gran aporte de todos los tipos de fluoróforos (los húmicos del tipo C, los húmicos del tipo A y los fluoróforos del tipo T y B), junto a una contribución de elevada conductividad y en menor medida de sedimentos (en particular, de la biomasa de tejidos vegetales ricos en celulosa y lignina, con presencia de bacterioclorofila, y de las partículas provenientes del suelo de la cuenca del arroyo Ludueña).
- Finalmente, el quinto grupo de variables (2,3% de varianza explicada) se distingue por la contribución de fuentes de sedimentos (principalmente del complejo de proteínas del fotosistema II, con presencia de clorofila-*a*, y de la biomasa de tejidos vegetales ricos en celulosa y lignina, con presencia de bacterioclorofila), junto a un

pequeño aporte de elevada conductividad y de fluoróforos (húmicos del tipo C, húmicos del tipo A y fluoróforos del tipo T y B).

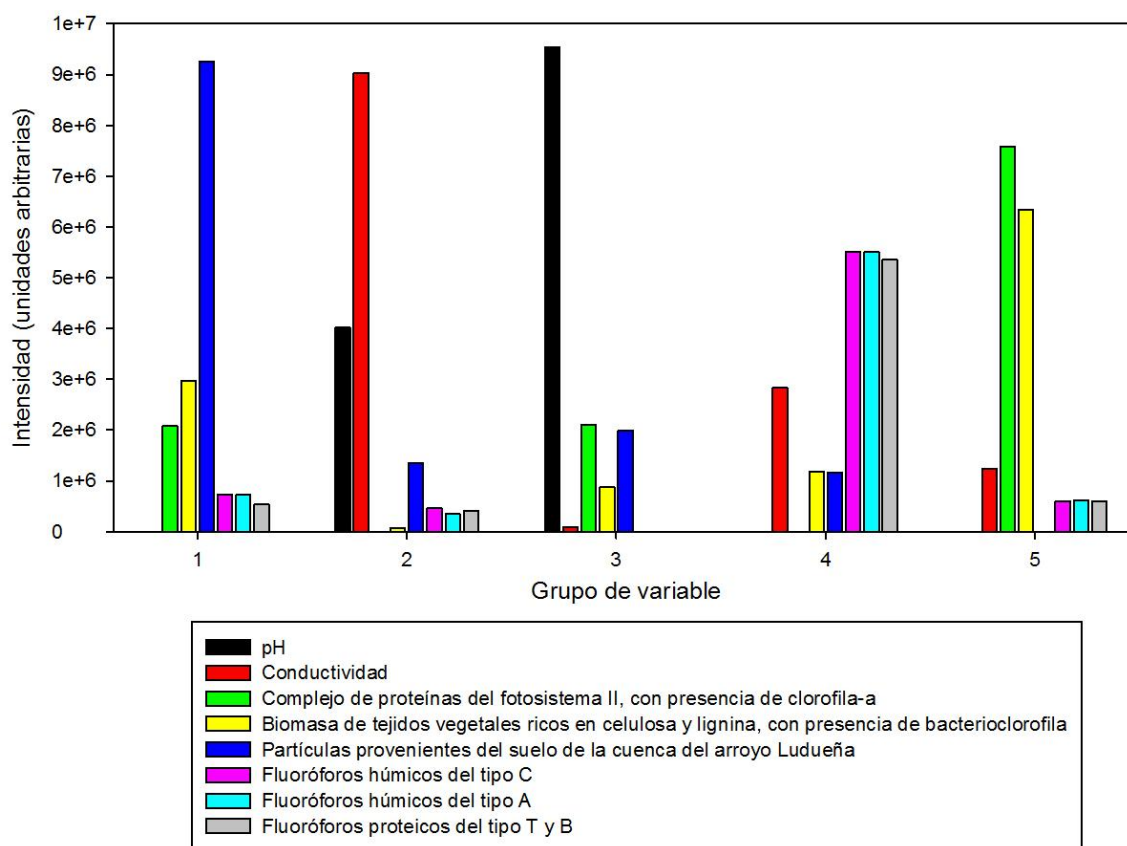


Figura 13. Agrupación de variables resultante del análisis por MCR-ALS aplicando las restricciones de no-negatividad y trilinealidad.

Una vez identificados los principales grupos de variables se estudió su correspondiente distribución espacial y temporal, tal como puede observarse en las Figuras 14 a 18.

El primer grupo de variables (77,8%) se localiza en toda la cuenca, pero presenta una mayor intensidad fundamentalmente en las regiones cercanas a la desembocadura del arroyo Ludueña sobre el río Paraná (Figura 14-A). El agua en este sitio en particular es una mezcla proveniente de ambos cursos, y en el caso del río Paraná se caracteriza por tener altas concentraciones de partículas del suelo en forma coloidal. En cuanto a la evolución temporal de este grupo de variables (Figura 14-B), se puede concluir que está

presente durante todo el período de muestreo pero con mayor intensidad en la cuarta y quinta campaña (24/04/17 y 06/06/17 respectivamente). Estas campañas corresponden a los meses de mayor viento y menor cantidad de lluvia registrados por el Servicio Meteorológico Nacional Argentino [57], de manera que el material particulado del suelo puede ser incorporado fácilmente en la cuenca y de forma natural.

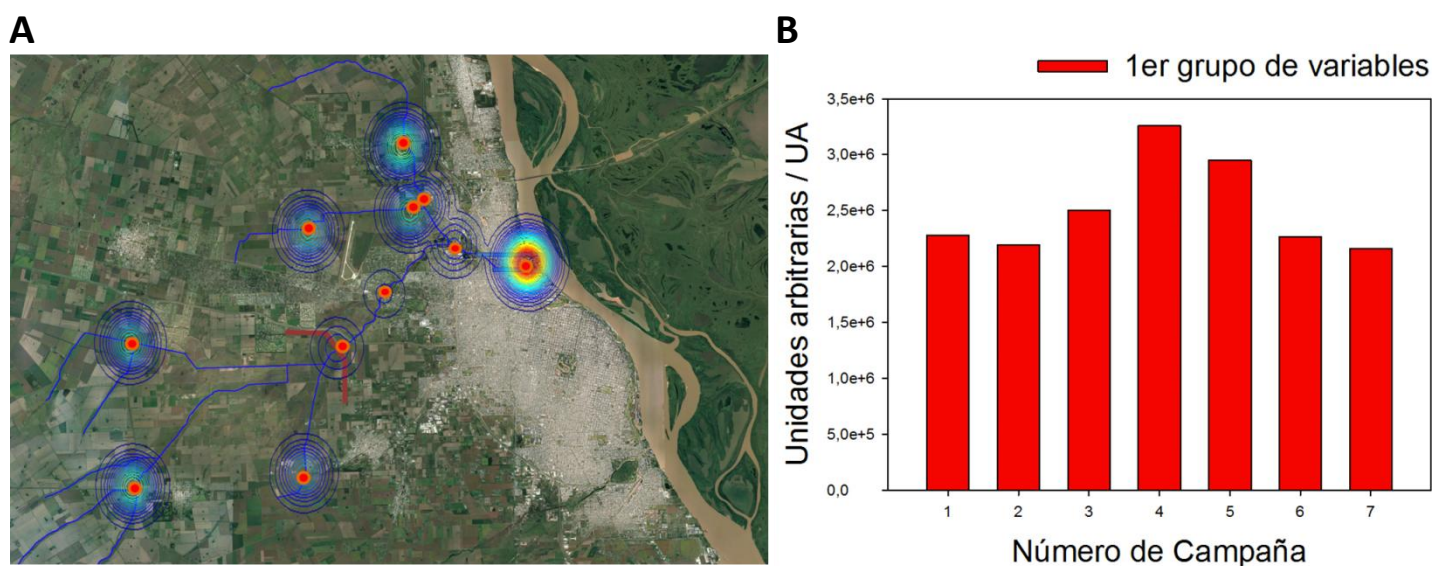
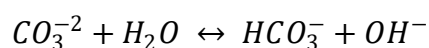


Figura 14. Primer grupo de variables resultantes del análisis por MCR-ALS aplicando las restricciones de no-negatividad y trilinealidad. **A)** Distribución geográfica en los distintos sitios de muestreo. **B)** Distribución temporal a lo largo de todas las campañas.

Por su parte, el segundo grupo de variables (9,4%) se ubica principalmente en las regiones donde la cuenca presenta las características de humedal (origen difuso, Figura 15-A). Esto permite la acumulación y depósito de sales en el suelo de esta zona, justificando con ello el aporte de la elevada conductividad observada en este grupo a lo largo de todas las campañas. A su vez, este grupo de variables también presenta un aporte de pH elevado. Este incremento es coherente con el hecho de que gran parte de las sales de los suelos se encuentran normalmente en forma de carbonatos. Estos participan del siguiente equilibrio ácido-base, aportando alcalinidad al medio [58]:



Adicionalmente, este grupo de variables presenta un aporte de carga sedimentaria proveniente de las partículas del suelo de la zona. Por este motivo, se puede interpretar a este grupo de variables como de origen geológico natural, siendo el material sedimentario el que aporta los iones responsables del aumento de la conductividad y el carbonato que alcaliniza el medio. En lo que respecta a la evolución temporal de este grupo de variables (Figura 15-B), se puede concluir que está presente de manera prácticamente constante a lo largo de todo el período de muestreo, mostrando solo leves variaciones entre una campaña y otra.

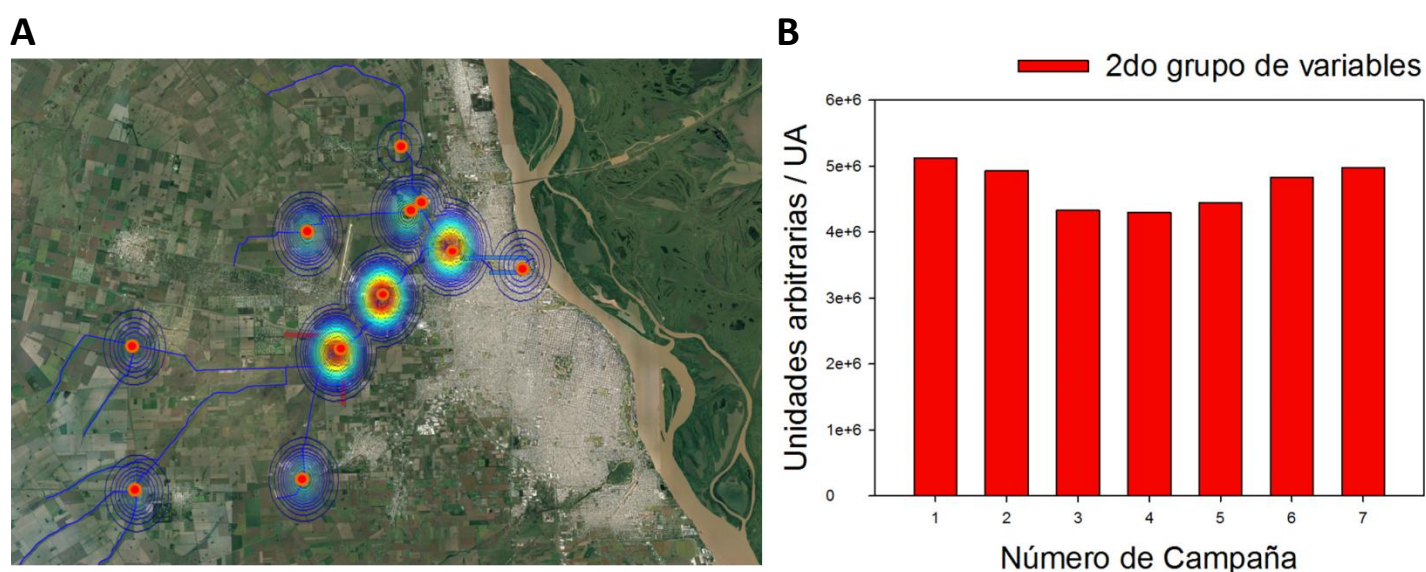
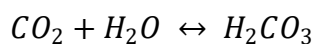


Figura 15. Segundo grupo de variables resultantes del análisis por MCR-ALS aplicando las restricciones de no-negatividad y trilinealidad. **A)** Distribución geográfica en los distintos sitios de muestreo. **B)** Distribución temporal a lo largo de todas las campañas.

El tercer grupo de variables (5,2%), si bien se encuentra distribuido por toda la cuenca (origen difuso), se localiza preferentemente en las áreas más altas que tienen menor flujo (Figura 16-A). Estas zonas permiten una mayor proliferación de algas y otros organismos fotosintéticos, aportando entonces materia orgánica particulada (MOP) de origen autóctono que contribuye a la acumulación de los sedimentos. De esta manera, se explicaría el comportamiento cíclico o estacional que tienen estos sedimentos naturales aportados por organismos que desarrollan fotosíntesis oxigénica y que proliferan en periodos de altas temperaturas (Figura 16-B), en áreas con altas cargas de nitratos y

fosfatos (eutrofización natural de la cuenca). Asimismo, la característica principal de este grupo es la contribución de valores elevados de pH. Esto se debe a la elevada tasa de fotosíntesis que exhiben los organismos en estas áreas. La alcalinización del medio viene dada como resultado de la asimilación de dióxido de carbono (CO_2) a través de la actividad fotosintética. Este proceso biológico fija el dióxido de carbono haciendo que disminuya su concentración en el cuerpo de agua. De esta forma, se da la alcalinización del medio a medida que por el principio de Le Châtelier se desplazan hacia la izquierda los siguientes equilibrios [58]:



Así, una disminución en la concentración del dióxido de carbono disuelto justifica la observación respecto a los valores elevados de pH.

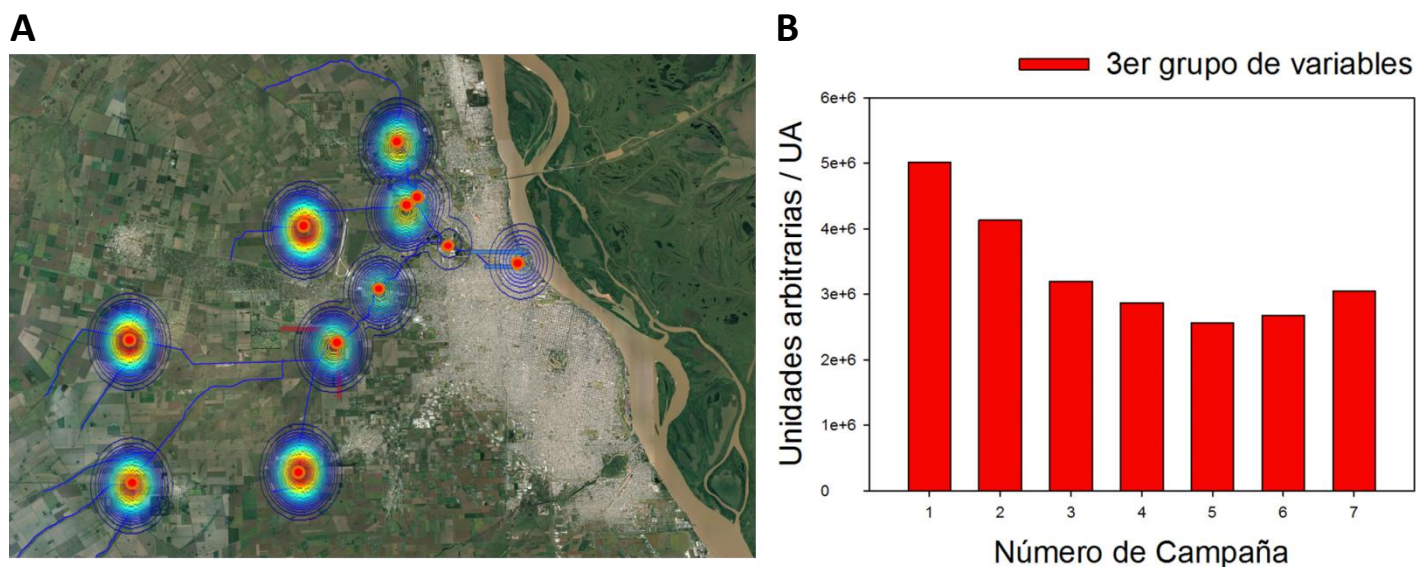


Figura 16. Tercer grupo de variables resultantes del análisis por MCR-ALS aplicando las restricciones de no-negatividad y trilinealidad. **A)** Distribución geográfica en los distintos sitios de muestreo. **B)** Distribución temporal a lo largo de todas las campañas.

El cuarto grupo de variables (1,3%) se localiza específicamente a lo largo del canal Ibarlucea (fuente puntual, véase la Figura 17-A), y tiene elevadas concentraciones sobre todo en la tercera y quinta campaña (16/02/17 y 6/06/17 respectivamente, véase la Figura

17-B). Este comportamiento es típico de las actividades antropogénicas, dado que no presentan una tendencia definida. Se trata entonces de sustancias de origen alóctono que se incorporan al sistema producto de la descarga de aguas residuales no tratadas, ya sea a través de conexiones clandestinas o del vaciado de camiones atmosféricos. Esto también permite explicar el aporte de elevada conductividad.

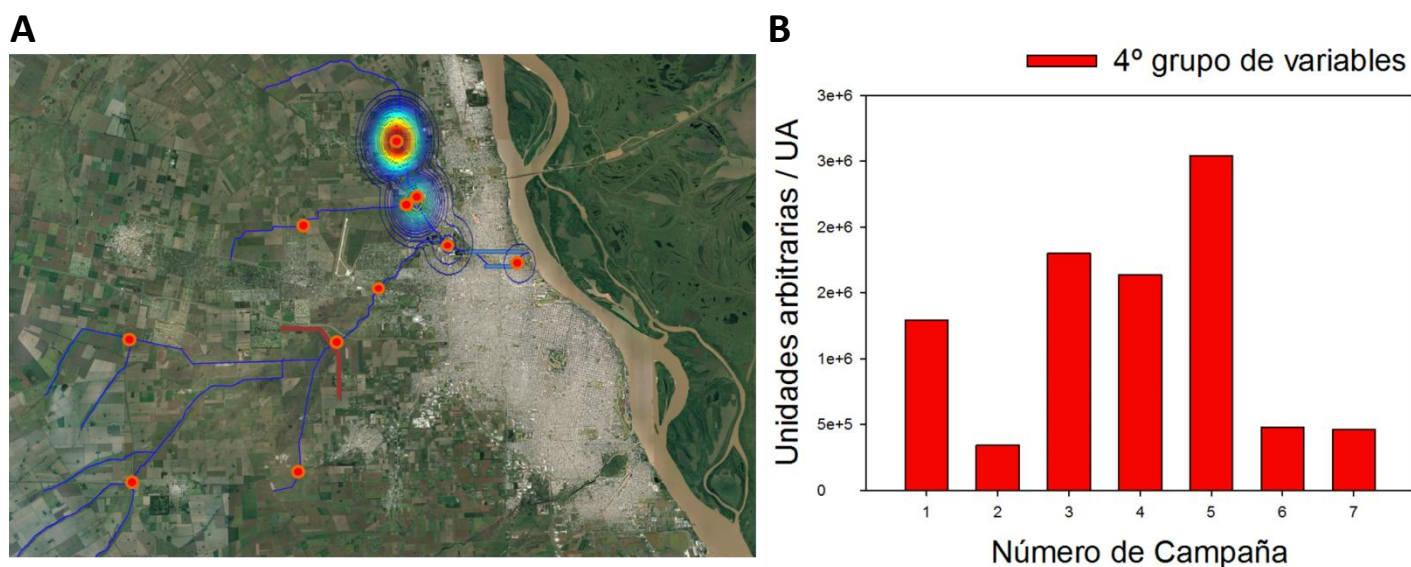


Figura 17. Cuarto grupo de variables resultantes del análisis por MCR-ALS aplicando las restricciones de no-negatividad y trilinealidad. **A)** Distribución geográfica en los distintos sitios de muestreo. **B)** Distribución temporal a lo largo de todas las campañas.

Por último, el quinto grupo de variables (2,3%) se localiza principalmente en áreas cercanas al canal Ibarlucea y a la localidad de Zavalla (Figura 18-A). En estas zonas existen varias lagunas de tratamiento de fluidos (una perteneciente al relleno sanitario de Ricardone y otras dos, próximas a la ciudad de Zavalla, de actividades agropecuarias). Este hecho permite explicar el gran aporte correspondiente a los sedimentos ricos en bacterioclorofila, tratándose entonces de partículas de origen alóctono que se incorporan al sistema como resultado de las actividades antropogénicas. En cuanto a la distribución temporal de este grupo (Figura 18-B), el comportamiento es estacional, observándose mayor carga de MOP en los meses de temperaturas más elevadas. Esta observación es compatible con la naturaleza de los sedimentos, debido a que son originados por organismos que desarrollan fotosíntesis oxigénica y anoxigénica.

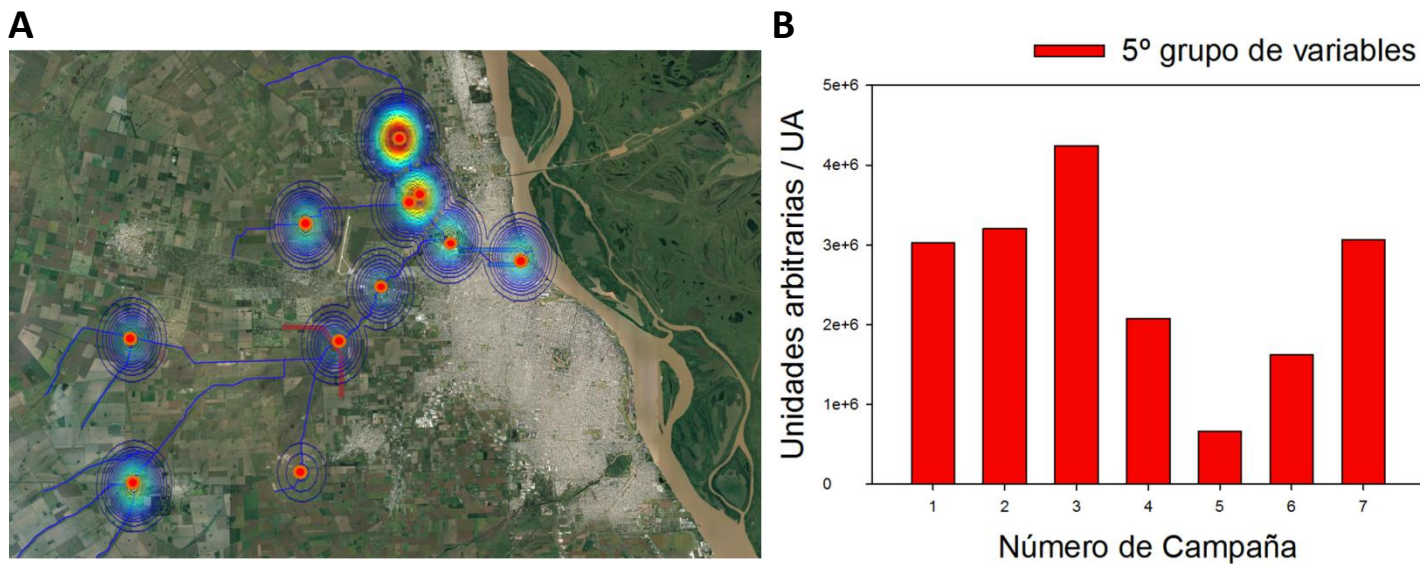
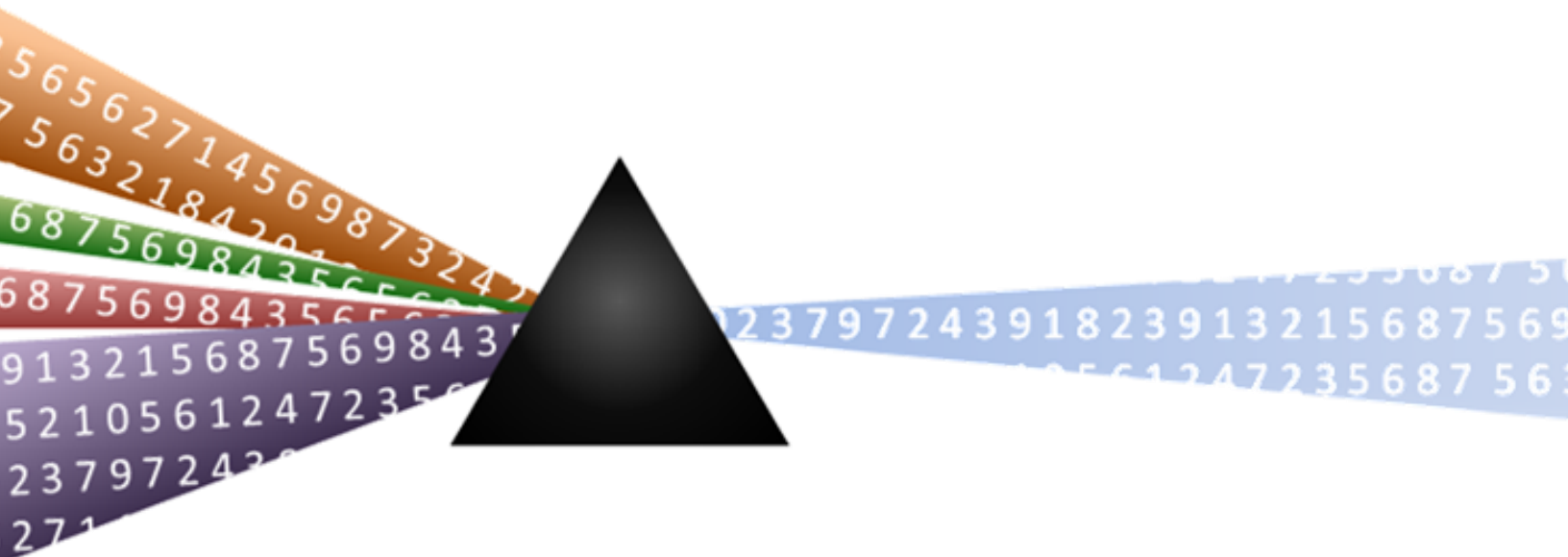


Figura 18. Quinto grupo de variables resultantes del análisis por MCR-ALS aplicando las restricciones de no-negatividad y trilinealidad. **A)** Distribución geográfica en los distintos sitios de muestreo. **B)** Distribución temporal a lo largo de todas las campañas.

5. Conclusiones



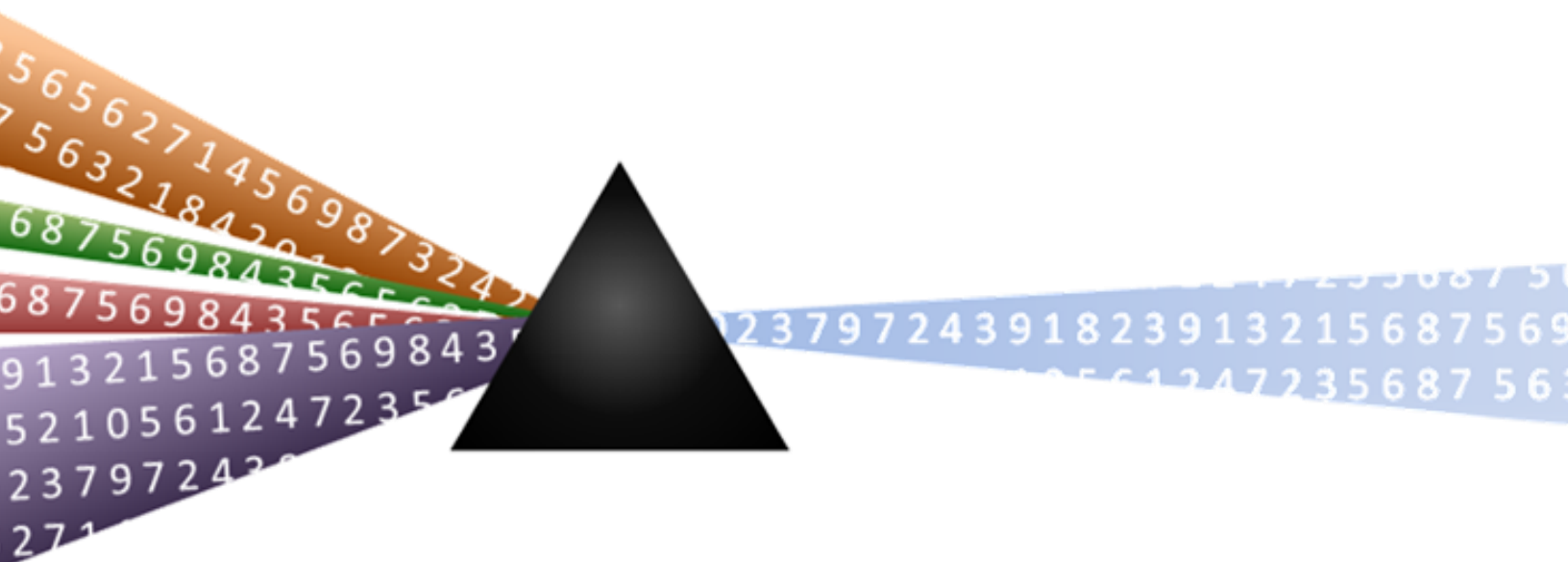
El presente trabajo de tesis de grado ha permitido realizar importantes contribuciones en el área de la quimiometría y de la química analítica ambiental. Dentro del campo de la quimiometría, se ha logrado desarrollar y optimizar exitosamente una novedosa metodología de trabajo que permite fusionar correctamente datos de distintos órdenes y analizarlos en forma conjunta. Por otra parte, en el área ambiental, fue posible desarrollar un nuevo protocolo de trabajo para realizar monitoreos ambientales a través del seguimiento de la MOD y el material particulado.

Específicamente, se corroboró la eficiencia y versatilidad del algoritmo MCR-ALS, dado que aplicando diferentes restricciones es capaz de analizar correctamente datos de distintos órdenes. Por lo tanto se comprobó que MCR-ALS, utilizado en forma única o en combinación con PARAFAC, constituye una poderosa herramienta para la fusión de datos. De esta manera, es posible condensar la información de los datos de primer y segundo orden y luego fusionarlos con los datos de orden cero para realizar el análisis en forma conjunta.

Como conclusiones particulares del sistema en estudio, se pudo observar que la zona más vulnerable se localiza en el canal Ibarlucea, debido a que en sus sitios de monitoreo se encuentra un mayor aporte de los grupos de variables con impacto antropogénico. También se pudo constatar el comportamiento estacional del material particulado de origen biológico y la presencia de otros componentes de origen geológico. Adicionalmente, fue posible evidenciar las relaciones que existen entre las variables medidas en forma directa con la MOD y el material particulado.

Finalmente, es relevante destacar que esta metodología de análisis y procesamiento de datos posee un gran potencial, ya que es factible su extrapolación a datos de órdenes superiores (*N-way*), quedando así planteada para aplicarse en futuros trabajos.

6. Referencias bibliográficas



-
- [1] Organización de las Naciones Unidas (1972) Informe de la Conferencia de las Naciones Unidas sobre el Medio Humano. Número de venta: S.73.II.A.14 y corrección, Capítulo 1. Estocolmo, Suecia.
- [2] Felipe-Sotelo, M.; Gustems, L.; Hernández, I.; Terrado, M. & Tauler, R. (2006) Investigation of geographical and temporal distribution of tropospheric ozone in Catalonia (North-East Spain) during the period 2000–2004 using multivariate data analysis methods. *Atmospheric Environment*, 40, 7421-7436.
- [3] Hotelling, H. (1936) Relations between two sets of variates. *Biometrika*, 28, 321-377.
- [4] García-Reiriz, A. G. (2014) Parallel factor analysis and multivariate curve resolution as data fusion tools to supervise a stream. *Chemom Intell Lab Syst*, 137, 120-127.
- [5] Poulenard, J.; Perrette, Y.; Fanget, B.; Quetin, P.; Trevisan, D. & Dorioz, J. M. (2009) Infrared spectroscopic tracing of sediment sources in a small rural watershed (French Alps). *Sci of the Total Environ*, 407, 2808-2819.
- [6] Coble, P. G. (1996) Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Mar Chem*, 51, 325-346.
- [7] Navarro, J. M. & Thompson, R. J. (1995) Seasonal fluctuations in the size spectra, biochemical composition and nutritive value of the seston available to a suspension-feeding bivalve in a subarctic environment. *Marine Ecology Progress Series*, 125, 95-106.
- [8] Haag, I.; Kern, U. & Westrich, B. (2011) Erosion investigation and sediment quality measurements for a comprehensive risk assessment of contaminated aquatic sediments. *Sci of the Total Environ*, 266, 249-257.
- [9] Zimmerman, J. K. H.; Vondracek, B. & Westra J. V. (2003) Agricultural land use effects on sediment *loading* and fish assemblages in two Minnesota (USA) watersheds. *Environ Manage*, 32, 93-105.

-
- [10] Shepherd, K. D. & Walsh, M. G. (2007) Infrared spectroscopy—Enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries. *Journal of Near Infrared Spectroscopy*, 15, 1-19.
- [11] Nilsson, M. B.; Dåbakk, E.; Korsman, T. & Renberg, I. (1996) Quantifying relationships between near-infrared reflectance spectra of lake sediments and water chemistry. *Environ Sci Technol*, 30, 2586-2590.
- [12] Cozzolino, D. (2016) Near infrared spectroscopy as a tool to monitor contaminants in soil, sediments and water—State of the art, advantages and pitfalls. *Trends in Environmental Analytical Chemistry*, 9, 1-7.
- [13] Korsman, T.; Nilsson, M.; Öhman, J. & Renberg, I. (1992) Near-Infrared reflectance spectroscopy of sediments: a potential method to infer the past pH of lakes. *Environ Sci Technol*, 26, 2122-2126.
- [14] Elkins, K. M. & Nelson, D. J. (2001) Fluorescence and FT-IR spectroscopic studies of Suwannee river fulvic acid complexation with aluminum, terbium and calcium. *J Inorg Biochem*, 87, 81-96.
- [15] Patel-Sorrentino, N.; Mounier, S.; Lucas, Y. & Benaim, J. Y. (2004) Effects of UV-visible irradiation on natural organic matter from the Amazon basin. *Sci of the Total Environ*, 321, 231-239.
- [16] Stedmon, C. A.; Markager, S. & Bro, R. (2003) Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Mar Chem*, 82, 239-254.
- [17] Baker, A. (2002) Spectrophotometric discrimination of river dissolved organic matter. *Hydrol Process*, 16, 3203-3213.

-
- [18] Cammack, W. K. L.; Kalff, J.; Prairie, Y. T. & Smith, E. M. (2004) Fluorescent dissolved organic matter in lakes: relationship with heterotrophic metabolism. *Limnol and Oceanogr*, 49, 2034-2045.
- [19] Clark, C. D.; Jimenez-Morais, J.; Jones, G.; Zanardi-Lamardo, E.; Moore, C. A. & Zika, R. G. (2002) A time resolved fluorescence study of dissolved organic matter in a riverine to marine transition zone. *Mar Chem*, 78, 121-135.
- [20] Coble, P. G.; Schultz C. A. & Mopper, K. (1993) Fluorescence contouring analysis of DOC intercalibration experiment samples: a comparison of techniques. *Mar Chem*, 41, 173-178.
- [21] De Souza-Sierra, M. M.; Donard, O. F. X.; Lamotte, M.; Bellin, C. & Ewald M. (1994) Fluorescence spectroscopy of coastal and marine waters. *Mar Chem*, 47, 127-144.
- [22] Bro, R. (1997) PARAFAC. Tutorial and applications. *Chemom Intell Lab Syst*, 38, 149-171.
- [23] Kowalczyk, P.; Cooper, W. J.; Durako, M. J.; Kahn, A. E.; Gonsior, M. & Young, H. (2010) Characterization of dissolved organic matter fluorescence in the South Atlantic Bight with use of PARAFAC model: relationships between fluorescence and its components, absorption coefficients and organic carbon concentrations. *Mar Chem*, 118, 22-36.
- [24] Fellman, J. B.; Miller, M. P.; Cory, R. M.; D'Amore, D. V. & White, D. (2009) Characterizing dissolved organic matter using PARAFAC modeling of fluorescence spectroscopy: a comparison of two models. *Environ Sci Technol*, 43, 6228-6234.
- [25] Henderson, R. K.; Baker, A.; Murphy, K. R.; Hambly, A.; Stuetz, R. M. & Khan, S. J. (2009) Fluorescence as a potential monitoring tool for recycled water systems: a review. *Water Res*, 43, 863-881.

-
- [26] Gao, L.; Fan, D.; Li, D. & Cai, J. (2010) Fluorescence characteristics of chromophoric dissolved organic matter in shallow water along the Zhejiang coasts, southeast China. *Mar Environ Res*, 69, 187-197.
- [27] Mostofa, K. M. G.; Wu, F.; Liu, C. Q.; Fang, W. L.; Yuan, J.; Ying, W. L.; Wen, L. & Yi, M. (2010) Characterization of Nanming River (southwestern China) sewerage-impacted pollution using an excitation-emission matrix and PARAFAC. *Limnology*, 11, 217-231.
- [28] Hall, G. J. & Kenny, J. E. (2007) Estuarine water classification using EEM spectroscopy and PARAFAC-SIMCA. *Anal Chim Acta*, 581, 118-124.
- [29] Esteves da Silva, J. C. G.; Tavares, M. J. C. G. & Tauler, R. (2006) Multivariate curve resolution of multidimensional excitation-emission quenching matrices of a Laurentian soil fulvic acid. *Chemosphere*, 64, 1939–1948.
- [30] García-Reiriz, A. G.; Olivieri, A. C.; Teixidó, E.; Ginebreda, A. & Tauler, R. (2014) Chemometric modeling of organic contaminant sources in surface waters of a Mediterranean river basin. *Environmental Science: Processes & Impacts*, 16, 124-134.
- [31] Tauler, R.; Barceló, D. & Thurman, E. M. (2000) Multivariate correlation between concentrations of selected herbicides and derivatives in outflows from selected U.S. midwestern reservoirs. *Environ Sci Technol*, 34, 3307-3314.
- [32] Terrado, M.; Barceló, D. & Tauler, R. (2009) Quality assessment of the multivariate curve resolution alternating least squares (MCR-ALS) method for the investigation of environmental pollution patterns in surface water. *Environ Sci Technol*, 43, 5321-5326.

-
- [33] Farnham, I. M.; Singh, A. K.; Stetzenbach, K. J. & Johannesson, K. H. (2006) Treatment of nondetects in multivariate analysis of groundwater geochemistry data. *Chemom Intell Lab Syst*, 60, 265-281.
- [34] Secretaría de Ambiente y Espacio Público, Municipalidad de Rosario (2016) Plan Ambiental Rosario. Sudamérica Impresos, Rosario, Argentina.
- [35] Basile, P. A.; Riccardi, G. A.; Stenta, H. & García, M. (2005) Evaluación de procesos de erosión/sedimentación en el sistema hidrológico del arroyo Ludueña. Segundo Simposio Regional sobre Hidráulica de Ríos.
- [36] Tauler, R. (1995) Multivariate curve resolution applied to second order data. *Chemom Intell Lab Syst*, 30, 133-146.
- [37] Tauler, R.; Smilde, A. & Kowalski, B. (1995) Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J Chemometr*, 9, 31-58.
- [38] Calvo, N. L.; Maggio, R. M. & Kaufman, T. S. (2018) Chemometrics-assisted solid-state characterization of pharmaceutically relevant materials. Polymorphic substances. *Journal of Pharmaceutical and Biomedical Analysis*, 147, 518-537.
- [39] Mortera, P.; Zuljan, F. A.; Magni, C.; Bortolato, S. A. & Alarcón, S. H. (2018) Multivariate analysis of organic acids in fermented food from reversed-phase high-performance liquid chromatography data. *Talanta*, 178, 15-23.
- [40] Gorrochategui, E.; Jaumot, J.; Lacorte, S. & Tauler, R. (2016) Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow. *TrAC Trends in Analytical Chemistry*, 82, 425-442.
- [41] Jolliffe, T. (2002) *Principal Component Analysis*, Second Edition. Springer Series in Statistics, Springer, New York.

-
- [42] Leurgans, S. & Ross, R. T. (1992) Multilinear models: applications in spectroscopy. *Statist Sci*, 7, 289-319.
- [43] Paatero, P. (1997) Monitoring the kinetics of the ion-exchange resin catalysed esterification of acetic acid with ethanol using near infrared spectroscopy with partial least squares (PLS) model. *Chemom Intell Lab Syst*, 38, 223-242.
- [44] Bro, R. (1998) Multi-way analysis in the food industry. University of Amsterdam, Netherlands (Doctoral Thesis).
- [45] <http://www.models.kvl.dk/algorithms>
- [46] Bro, R. & Kiers, H. A. L. (2003) A new efficient method for determining the number of components in PARAFAC models. *J Chemometr*, 17, 274-286.
- [47] Zepp, R. G.; Sheldon, W. M. & Moran, M. A. (2004) Dissolved organic fluorophores in southeastern US coastal waters: correction method for eliminating Rayleigh and Raman scattering peaks in excitation-emission matrices. *Mar Chem*, 89, 15-36.
- [48] MATLAB 7.0, The Math Works Inc., Natick, Massachusetts, USA, 2007.
- [49] <http://www.mcrals.info/>
- [50] 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.
- [51] © 2017 Google LLC. Todos los derechos reservados.  Google Maps es una marca comercial de Google LLC.
- [52] Wiggli, M.; Smallcombe, A. & Bachofena, R. (1999) Reflectance spectroscopy and laser confocal microscopy as tools in an ecophysiological study of microbial mats in an alpine bog pond. *J of Microbiological Methods*, 34, 173-182.

-
- [53] Li, X.; Sun, C.; Zhou, B. & He, Y. (2015) Determination of hemicellulose, cellulose and lignin in moso bamboo by near infrared spectroscopy. *Scientific Reports*, 5, 17210.
- [54] Galvez-Sola, L.; Moral, R.; Perez-Murcia, M. D.; Perez-Espinosa, A.; Bustamante, M. A.; Martinez-Sabater, E. & Paredes, C. (2010) The potential of near infrared reflectance spectroscopy (NIRS) for the estimation of agroindustrial compost quality. *Sci of the Total Environ*, 408, 1414-1421.
- [55] Bryant, D. A. & Frigaard, N. U. (2006) Prokaryotic photosynthesis and phototrophy illuminated. *Trends in Microbiology*, 14, 488-496.
- [56] Hudson, N.; Baker, A. & Reynolds, D. (2007) Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters-a review. *River Res Applic*, 23, 631-649.
- [57] <https://www.smn.gob.ar/>
- [58] Marín Galvín, R. (2003) *Fisicoquímica y microbiología de los medios acuáticos. Tratamiento y control de calidad de aguas*. Ediciones Díaz de Santos, Madrid, España.

Anexo

Los siguientes fragmentos fueron extraídos de la Tesina de grado de la estudiante Marianela Batistelli, titulada “Estudio de las fuentes de los sedimentos de la cuenca del arroyo Ludueña utilizando espectroscopia de infrarrojo cercano (NIR) y análisis quimiométrico”:

A.1 Procesamiento de las muestras

Como primer paso para la deposición de los sedimentos en los filtros se retiraron las muestras de agua de la heladera, se colocaron en un baño de agua para que tomen temperatura ambiente. Posteriormente se las agitó manualmente y se les aplicó ultra sonido durante 5 minutos, para asegurar que se restituyan las condiciones originales. Nuevamente se agitaron manualmente y se filtró la totalidad de su volumen (aproximadamente 900 mL) a presión normal por una malla metálica, cuyo tamaño de apertura es de 4 mm. En este paso se retiraron los restos de crustáceos, algas y residuos. De ese filtrado se tomaron 200 mL con una probeta y se filtraron a presión reducida, a través de filtros de fibra de vidrio marca Munktell, con tamaño de poro de 1.2 y de 47 mm de diámetro.

Los filtros se colocaron en placas de vidrio rotuladas y se incubaron en estufa a 50°C para su secado. Luego, se guardaron en desecadores con silica-gel toda la noche. Al día siguiente, antes de ser medidos, se colocan nuevamente en estufa a 50°C durante 30 minutos y luego en desecador para asegurar su deshidratación. Una vez que alcanzaron temperatura ambiente, se procede a la obtención de sus respectivos espectros NIR.

A.2 Instrumental

Las mediciones de los espectros NIR se realizaron en un Espectrómetro de reflectancia de infrarrojo cercano DS2500 de la empresa Foss, equipado con un monocromador de barrido continuo del espectro visible e infrarrojo cercano utilizando una red de difracción holográfica móvil de tipo pre-dispersivo, una lámpara de tungsteno y ocho detectores: cuatro de silicio (400-1100) nm y los restantes de sulfuro de plomo (1100-2500) nm. Los espectros se registraron en una copa de cuarzo de 5,8 cm de diámetro provista por el proveedor del equipo. Las lecturas se realizaron en un rango de medición establecido por el mismo de 400 a 2500 nm con un intervalo de 0,5 nm. El equipo está conectado a una computadora externa mediante la cual se obtienen los datos espectrales.

Para cada muestra el equipo realiza automáticamente siete mediciones en diferentes regiones del filtro, informando un único espectro promedio.

Para obtener el espectro de los filtros blanco y de los filtros con las muestras de sedimentos suspendidos se colocó cada filtro en dos posiciones distintas modificando mediante rotación manual la disposición del filtro sobre la copa para de este modo asegurar que las lecturas sean en diferentes zonas del mismo. Es decir, los datos utilizados posteriormente son un promedio de los catorce espectros obtenidos.

Los datos espectrales obtenidos se encuentran en formato '.nir' específico del fabricante del equipo se convierten a formato '.txt' por medio de una rutina escrita en MATLAB desarrollada por nuestro grupo de trabajo.