Revista Electrónic@ de Medioambiente. UCM.
IUCA
ISSN 1886-3329

ESTUDIO MULTIVARIANTE DE LA CALIDAD DEL AGUA: APLICACIÓN AL RÍO JÚCAR EN EL PERÍODO 1990-2013

Mª Isabel LÓPEZ RODRÍGUEZ
Departamento de Economía Aplicada
Universidad de Valencia
maria.I.Lopez@uv.es

Daniel G. PALACÍ LÓPEZ[1]
Universidad Politécnica de Valencia
dapalpe@etsii.upv.es

Recibido: 6 de marzo del 2014
Enviado a evaluar: 13 de marzo del 2014
Aceptado: 2 de junio del 2014

RESUMEN

Se realiza un estudio de la calidad del agua del río Júcar en el período 1990-2013, haciendo uso de técnicas de análisis multivariante. Este estudio es tanto longitudinal como transversal, es decir, la calidad del agua se evalúa tanto a lo largo del río, observando así las variaciones en sus características durante su trayectoria, desde el nacimiento del Júcar hasta su desembocadura, como a través del tiempo, detectando de esta forma el cambio en sus propiedades y calidad durante el período considerado. Se propone además el uso de las herramientas de análisis multivariante como forma de detectar anomalías en las propiedades del agua no visibles mediante técnicas univariantes, así como para poder discernir cuándo el agua cumple los criterios de calidad exigidos.

Palabras clave: calidad, agua, control de calidad, análisis multivariante, proyección en estructuras latentes, análisis discriminante, Confederación Hidrográfica del Júcar

WATER QUALITY MULTIVARIATE ANALYSIS: APLICATION TO JÚCAR RIVER DURING THE 1990-2013 PERIOD

SUMMARY

A study of the Júcar River's water quality in the period 1990-2013 is performed using multivariate analysis tools. This study is both of longitudinal and transversal nature, ie, water quality is assessed along the river, thus observing the variations in its characteristics

[1] Personal Contratado FPI.

during its trajectory from Júcar's birth to its mouth, and over time, thus detecting changes in its properties and quality over the aforementioned period of time. The use of multivariate analysis tools is also proposed as a way to detect abnormalities in water properties not visible when using univariate techniques, as well as to discriminate when the water being analyzed meets the quality standards.

Keywords: quality, water quality control, multivariate analysis, projection on latent structures, discriminant analysis, Júcar River Basin

ÉTUDE MULTIVARIÉE DE LA QUALITÉ DE L'EAU: APPLICATION DE LA RIVIÈRE JÚCAR DURANT LA PÉRIODE 1990-2013

RÉSUMÉ

L'étude a pour objet l'étude de la qualité de l'eau de la rivière Júcar durant la période 1990-2013, elle fait appel aux techniques d'analyse multivariée. Cette étude prend en compte aussi bien les aspects longitudinaux comme transversaux, c'est-à-dire que la qualité de l'eau est évaluée aussi bien le long de la rivière : en observant les variations de ses caractéristiques pendant son parcours depuis sa source jusqu'à son embouchure, comme au cours du temps : en détectant de cette manière le changement de ses propriétés et de sa qualité pendant la période considérée. L'étude met en œuvre l'application d'outils d'analyse multivariée comme méthode de détection d'anomalies des propriétés de l'eau invisibles par l'utilisation d'autres techniques univariées et comme système pour savoir quand l'eau satisfait tous les critères de qualité requis.

Mots clés: qualité, eau, control de qualité, analyse multivariée, projection de structures latentes, analyse discriminante, Confédération Hydrographique du Júcar.

1. INTRODUCCIÓN

El estudio de la calidad del agua, y en especial de aquellas aguas destinadas al uso o consumo humano, resulta imprescindible para garantizar su buen estado y la seguridad de todas aquellas personas que vayan a hacer uso de ésta, así como las condiciones óptimas de la misma para mantener la biodiversidad de las especies que habitan su entorno.

Una clara muestra de este interés se refleja en las múltiples investigaciones que tienen como objetivo principal analizar la calidad del agua (Beamonte et al., 2004, 2007; Karavoltsos et al., 2007; Shrestha y Kazama, 2006); definir y/o utilizar indicadores que permitan llevar a cabo dicho análisis (Beamonte et al., 2004, 2010, 2012), estudiar la vertiente económica que conlleva la calidad de la misma (Hernández et al, 2009; Saz et al 2009; Sevilla y Torregrosa, 2009-2011, Sevilla et al, 2010), e incluso ha dado lugar a líneas de investigación en las que se deja patente la importancia de su estudio como una pieza clave que permita el crecimiento de los distintos destinos turísticos (Vera, 2006; Sotelo, 2013).

El interés indicado queda de manifiesto en la continua monitorización de la calidad del agua llevada a cabo por organismos oficiales, como la Confederación Hidrográfica del Júcar (CHJ, 2005, 2008), que ha proporcionado los datos empleados en el presente trabajo.

En esa línea, en el trabajo que se presenta se pretende estudiar y conocer las posibles relaciones existentes entre los distintos "parámetros de calidad de las aguas" establecidos según la Normativa del Plan Hidrológico de Cuenca del Júcar, así como la relación entre dichos parámetros y la clasificación que, según el *Anexo 4* de la misma normativa, y tal como se muestra a continuación en la Tabla 1, se hace del agua de acuerdo con su aptitud para ser usada con fines agrícolas.

Tabla 1. Concentración límite/parámetro con objeto de clasificar el agua[2]

PARÁMETROS	Unidades	Buena	Admisible	Mediocre	Mala
SALINIDAD					
Permeabilidad* (Ci-Sj)	(i+j)	2-3	4	5-6	>=7-8
Cloruros	mg/l Cl	50	200	500	>=1100
TOXICIDAD					
Boro	mg/l B	0,7	1,0	3,0	>3,0
VARIOS					
pH	-	6-9	6-9	6-9	<6-9<
Sólidos en suspensión	mg/l	20	60	120	>120
DBO_5	mg/l	20	40	60	>60

Fuente: Extracto del Anexo 4 de la Normativa de la Cuenca Hidrográfica del Júcar.

Concretando, se persigue, como un primer objetivo, ofrecer la posibilidad de detectar en el futuro posibles anomalías en las propiedades del agua que no serían detectables si se hiciese uso, únicamente, de lo establecido en la normativa vigente. Para ello se va a hacer uso de técnicas de análisis multivariante a la información proporcionada relativa a la monitorización de la calidad del agua y, más concretamente, la correspondiente al río Júcar y sus afluentes.

Por otro lado, debe tenerse en cuenta que la monitorización mencionada se lleva a cabo realizando medidas de diferentes parámetros mediante diversas técnicas, cada una de las cuales lleva asociado un coste y por tanto lo ideal, en el caso general, sería que cada uno de los parámetros monitorizados proporcionen una información diferente y lo menos correlacionada posible con el resto. Además, la relación existente entre diferentes parámetros de calidad permitirá comprobar que los valores recogidos durante el control de calidad del agua son lógicos, es decir, que si por ejemplo la medida de conductividad da valores elevados, las concentraciones que compuestos que aumenten la conductividad también deberán ser elevadas.

Por tanto, es de esperar que no todos los parámetros medidos estén fuertemente correlacionados entre sí, pero que sí exista un cierto grado de correlación entre algunos de ellos. Un segundo objetivo, por tanto, será llevar a cabo esta comprobación.

Como un tercer objetivo se plantea el estudio de la evolución de la calidad del agua, y de los parámetros asociados con esta, tanto espacial (a lo largo del río Júcar) como temporal. Tanto este objetivo como los dos anteriores pueden alcanzarse mediante la realización de un Análisis de Componentes Principales (PCA).

[2] (*) Consideración conjunta de Conductividad y S.A.R., expresada como suma de los subíndices (i+j) de las respectivas calidades Ci y Sj

Finalmente se trata de validar el uso de otra herramienta de análisis multivariante, el Análisis Discriminante mediante Mínimos Cuadrados Parciales (PLS-DA) para poder predecir cuándo la calidad de una masa de agua será buena y cuándo "no buena" (es decir, admisible, mediocre o mala) para uso agrícola, de acuerdo con los criterios del Anexo 4 de la Normativa del Plan Hidrológico de Cuenca del Júcar, recogidos en la Tabla 1.

Así, la estructura del trabajo es la siguiente: en el epígrafe 2 se exponen los métodos de análisis empleados. En el epígrafe 3 se detalla todo lo relativo al tratamiento de los datos, es decir, la forma en que se han preparado los datos de partida para posibilitar el uso de las herramientas multivariantes citadas. En el epígrafe 4 se exponen los resultados obtenidos y finalmente, en el epígrafe 5, se recogen las principales conclusiones del trabajo.

2. METODOLOGÍA

Para alcanzar los objetivos propuestos se hace uso de dos herramientas de análisis multivariante, antes mencionadas, el PCA y el PLS-DA, de manera que con la finalidad de facilitar el seguimiento del trabajo se realiza una breve exposición ambos métodos.

El PCA tiene por objetivo principal "reducir la información" contenida en una matriz de datos, partiendo de un número inicial N_0 de variables, entre las cuales existen correlaciones, y obteniendo un conjunto de N ($N<N_0$) variables incorrelacionadas (componentes principales o factores), siendo cada una de ellas una combinación lineal de las variables originales.

Esta "reducción de la información" se refleja en dos aspectos fundamentales, ya que permite:

- Observar las posibles relaciones existentes entre las diferentes observaciones o individuos de la matriz de datos.
- Observar las posibles relaciones entre las variables contenidas en la matriz de datos.

Dada una matriz de datos X, que contiene las variables "explicativas", es decir los parámetros de control de calidad en este estudio, y una matriz de datos "Y", que contiene las variables "respuesta", es decir la clasificación del agua según su aptitud para uso agrícola, el PCA es una herramienta que se puede utilizar para explicar la variabilidad en X ó en Y, maximizando la varianza en X ó en Y, según cuál de las dos matrices de datos se esté analizando. Por tanto, no es la herramienta de análisis multivariante más adecuada para la elaboración de un modelo que relacione las variables X e Y.

Por su parte, el PLS requiere la distinción de dos matrices de datos X e Y. En este caso, se pretende encontrar las relaciones entre conjuntos de datos multivariantes X e Y, para lo cual se extraen diferentes componentes principales para ambos grupos de datos, de modo que se maximice la correlación entre X e Y. Usando el PLS, las componentes principales de X son ortogonales entre sí, pero no necesariamente es así para las componentes principales de Y. En el caso del PLS-DA, las variables Y son de tipo cualitativo, y el modelo obtenido permitirá estimar la probabilidad de que un individuo X_i pertenezca a cada uno de los posibles grupos definidos por las distintas variables Y_j.

Tanto al hacer uso del PCA como del PLS deben dividirse las observaciones en dos grupos distintos: un grupo de entrenamiento, y un grupo de validación. El grupo de entrenamiento suele estar constituido por entre un 60% y un 70% de todas las

observaciones disponibles, y sirve para obtener el modelo que se validará con las restantes, mediante comparación entre los valores predichos por el modelo para los individuos del grupo de validación (no incluidos durante la construcción del modelo), con las observaciones reales para dichos individuos.

Además de los grupos de entrenamiento y de validación, se hace uso también de dos parámetros para estimar la capacidad explicativa y predictiva de los modelos obtenidos mediante PCA y PLS. Estos parámetros son R^2 y Q^2. El primero de ellos se obtiene del mismo modo que al construir modelos mediante métodos de regresión clásicos, es decir, a partir de las diferencias entre los valores de la variable respuesta predichos por el modelo para los individuos observados, y los valores reales para los mismos, construyéndose el modelo a partir de los mismos individuos. Para el cálculo de Q^2 se emplea en el presente trabajo el método *leave-one-out*, de modo que si se dispone de *n* individuos, se construyen *n* modelos con (*n-1*) observaciones, excluyendo cada vez uno de los individuos, para el cual se predice el valor esperado utilizando el modelo para cuya construcción no ha sido utilizada dicho individuo. El valor de Q^2 se estima de forma similar a R^2, pero comparando los valores reales observados para los *n* individuos con las *n* predicciones obtenidas por el método *leave-one-out*.

3. TRATAMIENTO Y PREPARACIÓN DE LOS DATOS

El primer paso, en la aplicación de las técnicas multivariantes citadas en el epígrafe anterior, es la preparación de los datos de partida. Dichos datos son los valores medidos de los diferentes parámetros de calidad de las aguas de los que, en algún momento, se ha hecho un seguimiento entre enero de 1990 y enero de 2013, a lo largo de las diferentes estaciones de control del río Júcar y de sus afluentes. Se dispone del día, mes y año en que se realizó cada medida para cada parámetro en cada estación, en caso de haberse realizado. En total, se parte de la medida de 412 parámetros diferentes, medidos con diferentes frecuencias temporales y en distintas estaciones, muchas de las cuales han sufrido además al menos un cambio en su código identificativo a lo largo del período para el cual se dispone de las mencionadas medidas.

Inicialmente, cada fila de la matriz que forma la base de datos de que se parte presenta la información tal como se ve en la Tabla2.a. Por tanto, en cada fila de esta matriz se dispone sólo de la medida de un único parámetro, mientras que muchas de las columnas proporcionan información innecesaria para el estudio planteado. Puesto que la información organizada de esta manera no puede ser utilizada para los análisis deseados, el primer paso es transformar la matriz inicial de los datos del formato anterior a uno en cuyas filas la información aparezca del modo en que se esquematiza en la Tabla 2.b.

Con la tabla de datos reordenada, se deben escoger las variables que se estudiarán, así como los individuos para los cuales desea hacerse el estudio. Debido al gran tamaño de la base de datos de la que se dispone, se propone estudiar las características de las aguas a lo largo del río Júcar, excluyendo sus afluentes. Esto se hace con objeto de cubrir uno de los objetivos previamente presentados: estudiar la posible relación entre las características del agua del río y la distancia recorrida por el agua desde el nacimiento del río hasta su desembocadura.

Tabla 2. Estructura de las filas de la base de datos (a) y estructura deseada para el estudio (b)

Nombre y código estación nacional y europeo (3 columnas)	Nombre y código tramo (2 columnas)	Coordenadas estación (2 columnas)	Código masa de agua	Provincia y código y nombre municipio y cuenca (5 columnas)	…	Año, mes, día y hora de medición	Nombre, medida y unidades de medida parámetro medido (3 columnas)

(a)

Código estación	Fecha medición	Medida parámetro 1	Medida parámetro 2	…	Medida último parámetro	Calidad masa de agua según norma

(b)

Fuente: Elaboración Propia a partir de los datos proporcionados por la Confederación Hidrográfica del Júcar

Una vez llevada a cabo la transformación mencionada, se dispone de una tabla con una gran cantidad de datos faltantes. Esto se debe a que el seguimiento de la mayoría de los parámetros se realiza con frecuencias muy dispares, haciendo una medición cada mes, cada tres meses, cada seis meses o incluso una sola vez cada año o cada dos años, o sólo en determinados casos puntuales, sin seguir una cierta periodicidad. Además, en cada mes suelen tomarse medidas de varios parámetros, pero en distintos días. Por ello, se agrupan los datos por meses en lugar de por días. Esto supone una simplificación aceptable que no afectará a la precisión de las conclusiones extraídas en el estudio, y que sin embargo facilita enormemente el análisis del problema planteado.

Por otra parte, deberán excluirse todos aquellos parámetros cuyo valor no cambia en ningún momento ni para distintas estaciones, puesto que serán incapaces de explicar la variabilidad de los resultados obtenidos. Con ello se pasará de trabajar con 412 parámetros, a 186. Hay que tener en cuenta que esta reducción es considerable debido a que se ha restringido el estudio a las estaciones que forman parte del recorrido principal de río Júcar sin sus afluentes.

A continuación se escogen todas aquellas observaciones para las cuales se dispone de la clasificación del agua en buena, admisible, mediocre o mala. Posteriormente se eliminan de nuevo aquellos parámetros que no varían para el conjunto de observaciones, y aquellos para los que hay más de un 70% de datos faltantes, pasando a trabajar con 48 variables, entre las que se incluyen dos nuevas variables creadas a partir de la fecha y del código de la estación de cada observación:

- La primera, la variable “fecha relativa”, hace referencia al número de meses transcurridos desde el primero en que se dispone de medidas para las observaciones y variables considerados.

- La segunda, la variable “orden”, referente a la posición de la correspondiente estación a lo largo del recorrido principal del río Júcar (la primera estación es la más cercana al nacimiento del río, y la última a su desembocadura).

La matriz de datos resultante posee aún un gran número de huecos, por haberse incluido en ella variables con hasta un 70% de datos faltantes. Una posible forma de solucionar este inconveniente es imputar para los datos faltantes el valor medio de la correspondiente variable. Sin embargo, una alternativa que induce a menor error es la estimación de los valores faltantes mediante la construcción de modelos multivariantes. En este caso, el valor de R^2 de los modelos construidos para las imputaciones es superior al 65% para la mayoría de casos, no siendo inferior al 55% en ninguno de ellos.

Hay que tener en cuenta que aunque estas capacidades de predicción pueden parecer bajas a priori, el error producido al llevar cabo las estimaciones es mucho menor al introducido al sustituir los valores faltantes por las medias de las correspondientes variables. En efecto, se comprobó que las conclusiones del análisis difieren en gran medida según se use un método de imputación o el otro.

Las variables finalmente incluidas en los análisis realizados en este estudio se presentan en la Tabla 3.

Tabla 3. Parámetros de control de calidad de las aguas

Código	*Nombre del parámetro*	*Código*	*Nombre del parámetro*
A026	Alcalinidad	A193	Fosfatos
A037	Amoniaco no ionizado	A197	Fosforo total
A039	Amonio total	A222	Hidrocarburos visibles
A047	Aspecto	A242	Magnesio
A071	Bicarbonatos	A252	Mercurio
A072	Boro	A267	Nitratos
A080	Cadmio	A269	Nitritos
A081	Calcio	A270	Nitrogeno Kjeldahl
A082	Carbonatos	A285	Oxigeno disuelto "in situ"
A083	Carbono orgánico total	A328	PH
A091	Caudal instantáneo	A337	Plomo
A093	Cianuros	A338	Potasio
A095	Cloro total	A349	Saturación de oxigeno disuelto
A108	Cloruros	A357	Sodio
A112	Cobre	A358	Sólidos en suspensión
A113	Coliformes fecales	A359	Sulfatos
A114	Coliformes totales 37ºC	A377	Temperatura agua "in situ"
A117	Conductividad eléctrica a 20ºC	A378	Temperatura del aire "in situ"
A121	Cromo	A379	Tensoactivos aniónicos
A124	Cromo VI	A412	Zinc
A133	DBO_5	B345	Ratio absorción de sodio
A134	DQO	S327	Permeabilidad
A158	Dureza total	c_total	Calidad del agua para uso agrícola

Fuente: Elaboración Propia

A estas variables, parámetros medidos para el control de la calidad del agua, se añaden la "fecha relativa" y el "orden" de la estación de cuyas medidas se disponen, tal como se mencionó anteriormente.

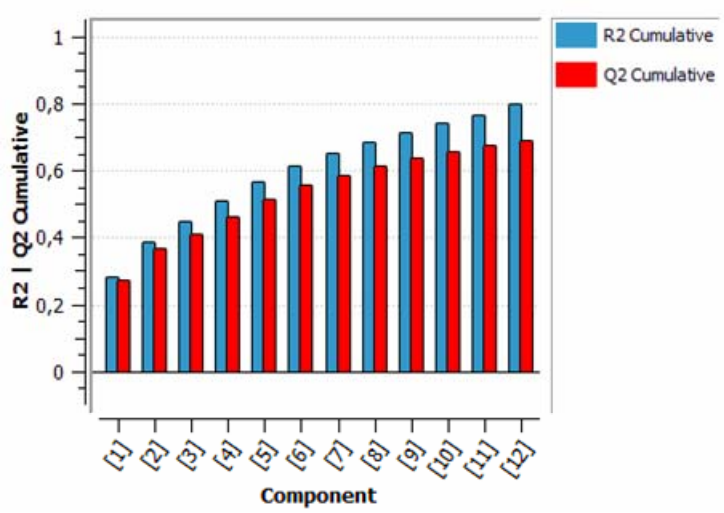
Los parámetros cuyo código comienza por la letra "A" son relativas al agua, los que lo hacen con la "B" a la biota y los que empiezan por"S" a los sedimentos.

4. RESULTADOS

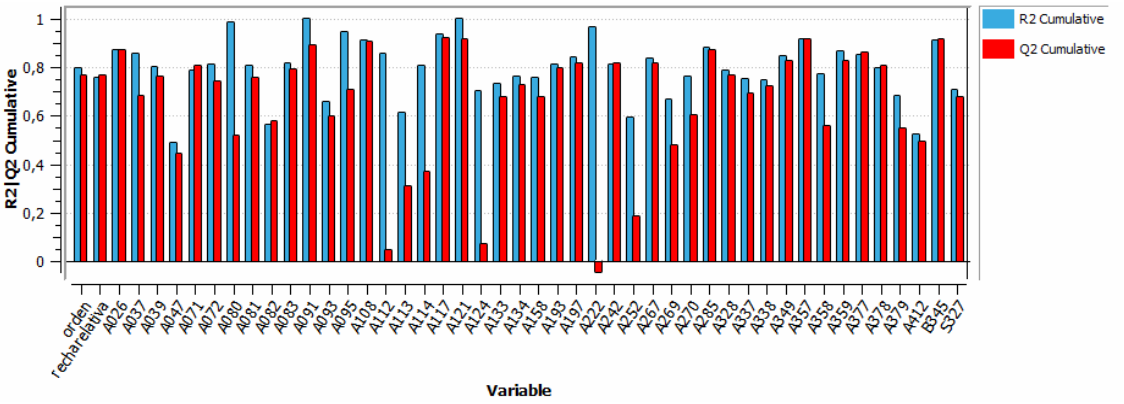
4.1. ANÁLISIS DE LA ESTRUCTURA DE RELACIONES ENTRE LOS PARÁMETROS DE CALIDAD

Tal y como se indicó en la introducción, en primer lugar se lleva a cabo el análisis de todo el conjunto de datos seleccionado, con el fin de poder encontrar relaciones entre los distintos parámetros de calidad de las aguas, así como para detectar posibles datos anómalos y outliers. Para ello se hace uso de un Análisis de Componentes Principales (PCA).

Figura 1. Variabilidad explicada (a) total, según el número de componentes y (b) de cada variable, para el modelo con 12 componentes



(a)



(b)

Fuente: Elaboración Propia

Un primer paso consiste en la constatación de que la utilización de dicha técnica es aplicable a los datos. Con este fin se obtiene el valor del KMO, que resulta ser de 0.787, así como el resultado del test de esfericidad de Bartlett, que proporciona un p-valor muy inferior al 1%. Por tanto, el PCA es aplicable, rechazándose la hipótesis de incorrelación entre las variables (parámetros de calidad).

Tras esta comprobación se construye un modelo capaz de explicar el 80% de la variabilidad total de los datos, que consta de 12 componentes principales. En la Figura 1.a puede verse el tanto por uno acumulado de la variabilidad que puede ser explicada (R^2) y predicha (Q^2) en función del número de componentes consideradas para la construcción del modelo. Por otra parte, la Figura 1.b muestra qué proporción de variabilidad de cada uno de los parámetros puede ser explicada y predicha por el modelo, respectivamente.

Se ve, de acuerdo con la Figura 1.b, cómo el modelo permite explicar como mínimo el 50% de la variabilidad de cada uno de los parámetros.

Conviene, en cualquier caso, llevar a cabo la interpretación de las distintas componentes, observándose en la Tabla 4 qué parámetros están relacionados positivamente y cuáles negativamente de manera significativa o muy significativa (sombreado) con cada una de las componentes principales.

Tabla 4. Tabla de relación entre parámetros y componentes (12)[3]

Variable/Parámetro	*C1*	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*	*C9*	*C10*	*C11*	*C12*
Orden	(+)		(+)				(+)			(-)		
Fecha relativa		(+)			(-)	(+)					(+)	
Alcalinidad	(+)											
Amoniaco no ionizado		(-)					(+)			(-)	(-)	
Amonio total	(+)	(-)										
Aspecto					(+)							
Bicarbonatos	(+)				(-)						(+)	
Boro		(+)			(-)	(+)	(+)				(+)	
Cadmio												(-)
Calcio					(+)	(+)	(-)			(+)		
Carbonatos			(-)		(+)		(+)				(-)	
Carbono orgánico total		(-)		(+)			(+)			(+)	(-)	
Caudal instantáneo			(+)					(+)				
Cianuros		(-)	(+)								(+)	
Cloro total				(+)								
Cloruros	(+)	(+)										
Cobre						(+)						
Coliformes fecales		(-)								(-)		
Coliformes totales 37ºC		(-)								(-)	(-)	
Conductividad eléctrica 20ºC	(+)	(+)										
Cromo			(+)						(+)			
Cromo VI												(+)
DBO_5		(-)								(+)	(+)	
DQO	(+)	(-)				(-)				(+)	(+)	

[3] *Debe notarse que, a pesar de que el peso del parámetro "Hidrocarburos visibles" es bajo para todas las componentes, su variabilidad queda explicada en más del 80% por la 9ª componente

Variable/Parámetro	***C1***	***C2***	***C3***	***C4***	***C5***	***C6***	***C7***	***C8***	***C9***	***C10***	***C11***	***C12***
Dureza total	(+)				(+)							
Fosfatos	(+)						(+)					
Fosforo total	(+)						(+)			(+)		
Hidrocarburos visibles*												
Magnesio	(+)	(+)			(+)							
Mercurio						(-)						(+)
Nitratos	(+)											
Nitritos	(+)	(-)									(+)	
Nitrógeno Kjeldahl		(-)				(+)	(+)			(+)	(+)	
Oxigeno disuelto "in situ"	(-)	(+)			(+)		(+)					
PH	(-)						(+)					
Plomo		(-)										
Potasio	(+)	(+)										
Saturación de oxígeno disuelto	(-)	(+)					(+)					
Sodio	(+)	(+)										
Sólidos en suspensión										(+)	(-)	
Sulfatos					(+)	(+)	(-)			(+)		
Temperatura agua "in situ"			(+)									
Temperatura del aire "in situ"			(+)									
Tensoactivos aniónicos	(+)											
Zinc						(+)				(+)	(-)	
Ratio absorción de sodio	(+)	(+)										
Permeabilidad		(+)	(+)		(+)					(-)		

Fuente: Elaboración Propia

De cuya observación se deduce que:

-La 1ª componente está relacionada con la conductividad y la presencia de sales en el agua, y su evolución a lo largo del río. Concretamente, señala un aumento en la conductividad y concentración de sales desde el inicio hasta la desembocadura del Júcar.

- La 2ª componente tiene relación con los procesos biológicos que tienen lugar dentro del agua, y su disminución a lo largo del tiempo. Es decir, indica que en promedio, entre 1990 y 2013, se han reducido los procesos biológicos que afectan negativamente a la calidad del agua. Aunque es posible que en algunos tramos se haya producido un aumento o no haya habido descenso, por término medio ha habido una disminución.

-La 3ª componente es la que indica un aumento de la temperatura a lo largo del río, y de la permeabilidad de los sedimentos en el lecho del mismo. Al aumentar la temperatura, aumenta la solubilidad de las sales en el agua, algo que se observa en la primera componente. Además, una mayor permeabilidad del lecho del río evita la filtración de las sales a través del mismo.

-La 4ª componente se refiere al carbono orgánico y al cloro totales, indicando que un mayor valor del primero suele ir acompañado de valores más altos del segundo y viceversa.

-La 5ª componente hace referencia a la cantidad de sustancias responsables del pH del agua disueltas en ésta, a su aspecto, y a la evolución de ambos con los años. En términos generales indica un empeoramiento con los años del aspecto de las aguas del Júcar.

-La 6ª componente contiene principalmente información relativa a la cantidad de mercurio

presente en el agua y su disminución con el tiempo. También indica un aumento de la contaminación por desechos orgánicos con el tiempo.

-La 7ª componente se corresponde con la cantidad de oxígeno disuelto, relacionado con las sustancias presentes que influyen en el pH, y su evolución a lo largo del río. De nuevo se tiene un aumento de la contaminación por desechos orgánicos, esta vez a lo largo del río.

-La 8ª componente explica únicamente el caudal en cada punto.

-La 9ª componente corresponde al cromo presente en el agua.

-La 10ª componente corresponde principalmente a los sólidos en suspensión, que vendrán influenciados por la presencia de otras sustancias que influyen en su capacidad de dilución, y que varía a lo largo del recorrido del Júcar, con mayor presencia en zonas cercanas al nacimiento.

-La 11ª componente está relacionada con el nitrógeno presente en el agua y el aumento de su presencia a lo largo de los años.

-La 12ª componente contiene principalmente información sobre el cadmio presente. También señala que masas de agua con grandes cantidades de cadmio suelen presentar pequeñas cantidades de mercurio.

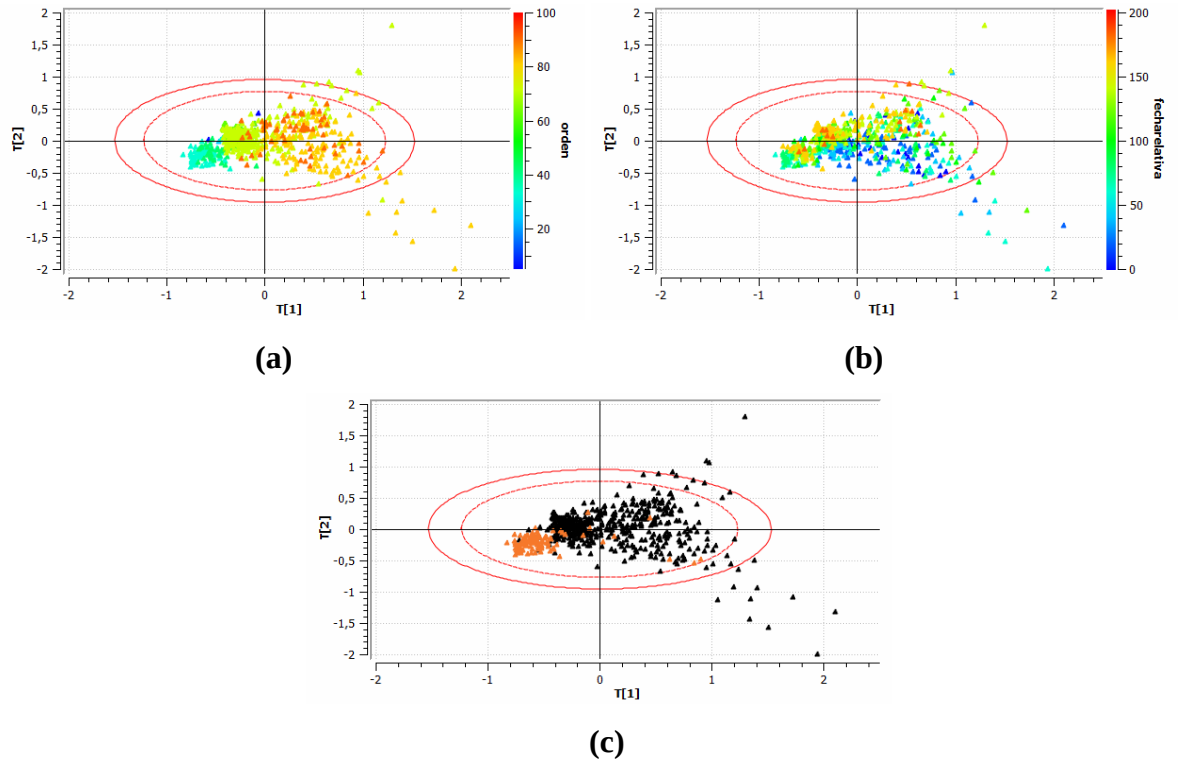
Visualmente, es posible observar de forma intuitiva parte de la información que dan algunas de estas componentes. Por ejemplo, si se considera el gráfico de los Scores de las observaciones frente a las componentes C1 y C2, añadiendo a la representación un esquema de colores en función de la posición (orden) de la estación, de la "fecha relativa" de la medida y de si la medida corresponde a agua de calidad buena o no, se obtienen los gráficos mostrados en la Figura 2.

En la Figura 2.a se muestra que todas las observaciones correspondientes a estaciones situadas cerca del nacimiento del río (valores de "orden" más bajos) se encuentran agrupadas, mientras que el resto están separadas de éstas y más dispersas.

En la Figura 2.b se detecta que las observaciones en fechas más cercanas a 1990, es decir, las más tempranas de las que se dispone información, se encuentran más dispersas, mientras que las más cercanas a 2013, esto es, las más recientes, aunque también presentan cierta dispersión, se encuentran más agrupadas.

Pero además, si se comparan la Figura 2.a y la Figura 2.c se ve claramente que aquellas observaciones donde el agua se clasificó como buena coinciden casi perfectamente con las observaciones de las estaciones más cercanas al nacimiento del Júcar. Esto podría explicarse por el hecho de que cuanto más cerca del nacimiento del río, el agua ha sufrido en menor medida la acción del hombre.

Figura 2. Visualización gráfica de los datos.[4]



Fuente: Elaboración Propia

A modo de resumen, el PCA realizado deja ver claramente la influencia de las variables espacial y temporal sobre la calidad del agua, además de las relaciones existentes, si las había, entre el resto de parámetros medidos. Se ha observado que, en término medio, parece haber una mejora en la calidad del agua con el transcurso de los años, pero se ha detectado también la influencia del hombre sobre esta misma calidad a lo largo del Júcar, pues la calidad del agua sólo se ha clasificado como "buena" en los puntos más cercanos al nacimiento del río.

4.2. RELACIÓN ENTRE LOS PARÁMETROS DE CALIDAD Y LA CLASIFICACIÓN DEL AGUA

En este subepígrafe del trabajo se estudia la influencia de todos los parámetros anteriores sobre la calidad del agua, a excepción de la variable "fecha relativa", haciendo uso de la regresión por cuadrados mínimos parciales en forma de análisis discriminante (PLS-DA).

Concretamente el análisis se centra en la elaboración de un modelo capaz de predecir, a partir de los parámetros de cuya información se dispone, si el agua puede considerarse o no de buena calidad para uso agrícola. De las observaciones disponibles, 106 corresponden con casos en los que la calidad del agua es "buena" y 479 en las que la calidad es "no buena".

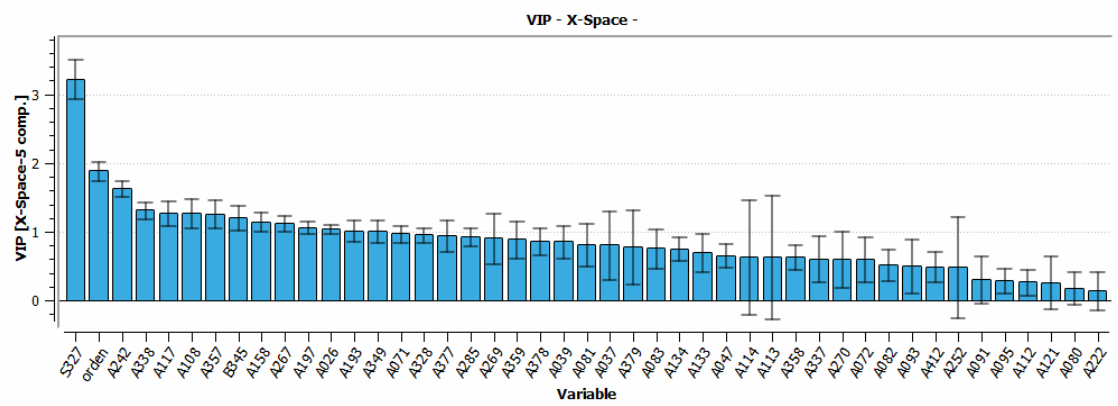
[4] Visualización gráfica de los datos según (a) posición relativa en el río, (b) fecha relativa y (c) calidad del agua para uso agrícola (buena, en naranja, o "no buena", en negro)

La exclusión de la variable "fecha relativa" se debe a que sus valores, para la matriz de datos de que se dispone, varían dentro de los posibles para el rango temporal entre enero de 1990 y enero de 2013. Por tanto, para cualquier observación posterior a enero de 2013 "fecha relativa" tomaría un valor superior al que toma en cualquiera de las observaciones que se tienen para este estudio, invalidando cualquier resultado obtenido para futuras observaciones si se incluye esta variable en el modelo.

Para el set de "entrenamiento" se toman al azar 81 observaciones para las cuales la calidad del agua sea "buena" (se le asigna el valor 1) y 383 observaciones para las que el agua haya sido "no buena" (con valor asignado 0) y el resto de observaciones se utilizan para el set de validación. Este es un proceso iterativo, de modo que el modelo obtenido finalmente presente buena capacidad predictiva, lo cual se comprueba haciendo uso del test de validación. De esta forma se obtiene finalmente un modelo que se considera adecuado, con 5 componentes, una capacidad explicativa (R^2) del 80% y una capacidad predictiva (Q^2) de casi el 75% para el set de entrenamiento.

Con el fin de detectar cuáles son las variables que más importancia tienen en el modelo, se obtiene el gráfico de importancias que se muestra en la Figura 3. El gráfico de importancias muestra el peso medio de cada parámetro en las componentes principales del modelo, ponderado por la variabilidad explicada por cada una de las componentes. Por tanto, ofrece una estimación de cuán importante es cada una de las variables originales para el modelo global construido.

Figura 3. Gráfico de importancias de las variables X en el modelo PLS-DA.



Fuente: Elaboración Propia

De cuya observación de deduce que las únicas variables que, según este modelo, se puede decir que no tienen una clara influencia a la hora de determinar si la calidad del agua es buena o "no buena" para uso agrícola son coliformes fecales (A113) y totales a 37ºC (A114), mercurio (A252), cromo (A121), cadmio (A080), hidrocarburos visibles (A222) y caudal instantáneo (A091). Por el contrario, el resto de parámetros influyen, en mayor o menor medida, en esta clasificación.

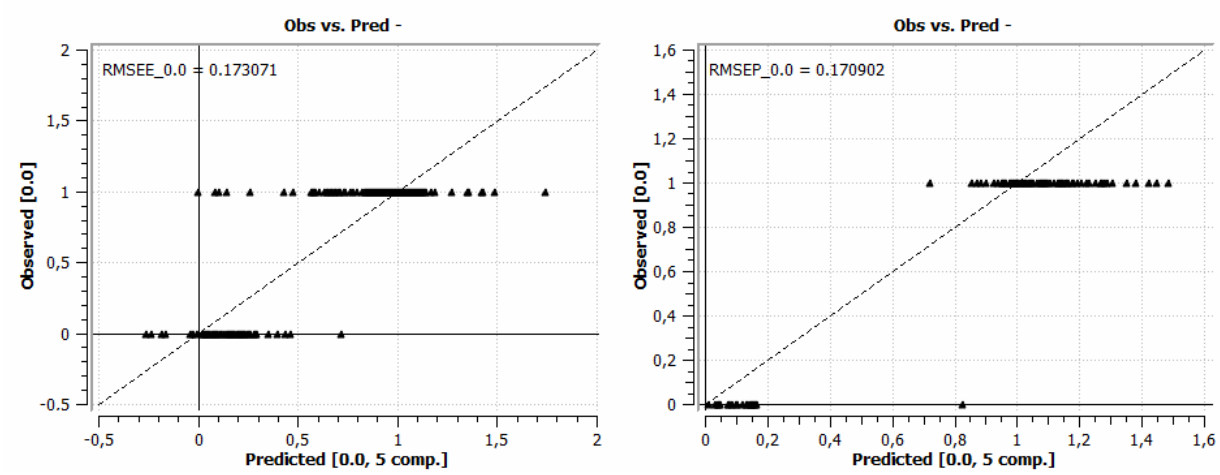
Debe mencionarse también que los parámetros que la normativa vigente utiliza para clasificar el agua según su calidad para uso agrícola son la salinidad (mediante la permeabilidad, que depende a su vez de la conductividad, el sodio, el magnesio y el calcio, y los cloruros), la toxicidad (boro) y otros parámetros como son el pH, los sólidos en

suspensión y la DBO_5.

Si además, y atendiendo a la información aportada por la Figura 3, se observa cuál es el conjunto de las variables más importantes para la construcción de este modelo, se puede comprobar que todos los parámetros considerados por la normativa para clasificar el agua se encuentran dentro de este conjunto. Igualmente se pueden localizar como ¿MÁS MENOS? importantes aquellas variables que, tal como se vio en el PCA realizado en el epígrafe anterior, estaban relacionadas con dichos parámetros.

En la Figura 4 se muestra una representación de los valores observados (1 para el agua clasificada como buena, y 0 para el agua clasificada como "no buena") frente a los predichos (probabilidad de que la masa de agua sea clasificada como buena).

Figura 4. Observaciones frente a predicciones del modelo PLS-DA.[5]



(a) **(b)**

Fuente: Elaboración Propia

Puesto que el modelo presentado proporciona la probabilidad de que la masa de agua analizada posea calidad buena para riego, debe escogerse a partir de qué valor de probabilidad predicho se aceptará que el agua sea apta para riego. Así, escogiendo una probabilidad del 50% como punto de corte:

- En el set de entrenamiento, señalaría como aguas de calidad "no buena" 7 casos en que es buena, y como buena en 1 ocasión en que es "no buena".

- En el set de predicción, señalaría como agua de calidad buena en 1 caso en que es "no buena".

De modo que se comete en total un 1,7% de errores, equivocándose en un 1,67% de los casos en que debería haber identificado el agua como buena, y en un 1,88% de los casos en que debería haber identificado el agua como "no buena".

Así, el PLS-DA se presenta como una herramienta capaz de ofrecer una buena

[5] Observaciones frente a predicciones del modelo PLS-DA para el set de entrenamiento (a) y el de validación (b)

capacidad de discriminación entre masas de agua buenas y "no buenas" para el consumo agrícola, que además detecta como parámetros más importantes para dicha discriminación a aquellos que en la práctica se utilizan realmente para llevar a cabo la clasificación. Adicionalmente, al llevarse a cabo el análisis multivariante de los datos permite tener en cuenta la estructura de relaciones entre los parámetros de calidad de las aguas, lo que facilita la detección de determinadas observaciones anómalas que no serían detectadas como tales al hacer uso de técnicas de análisis univariantes.

5. CONCLUSIONES

El estudio de la estructura de relaciones entre los parámetros de control de calidad de las aguas mediante análisis de componentes principales ha permitido comprobar que el 80% de la información aportada inicialmente por las 48 variables consideradas puede ser explicada por 12 de estas componentes, cuya interpretación es diferente para cada una y coherente con la naturaleza del estudio. Así cabe mencionar que se detecta un aumento en la conductividad y concentración de sales desde el inicio hasta la desembocadura del Júcar y una disminución, por término medio, de los procesos biológicos dentro del agua a lo largo del tiempo. También se concluye que, en términos generales, se ha producido un empeoramiento con los años del aspecto de las aguas del Júcar y un aumento de la contaminación por desechos orgánicos a lo largo del río. Cabe subrayar, también, que las observaciones donde el agua se clasificó como buena coinciden con las de las estaciones más cercanas al nacimiento del Júcar.

En cuanto al estudio de la evolución de la calidad del agua se puede concluir que el PCA realizado deja ver, además de las relaciones existentes entre los parámetros de calidad, la influencia de las variables espacial y temporal sobre la calidad del agua. Así, se observa que la calidad del agua del Júcar sufre un empeoramiento desde su nacimiento hasta su desembocadura, y se detecta una mejora en el control de la calidad del agua entre 1990 y 2013.

Cabe destacar, por otra parte, que el uso de la técnica multivariante citada ha permitido, además de lograr una reducción de las dimensiones a un 25% de las iniciales, aprovechar la existencia de una estructura de relaciones definidas para poder detectar observaciones anómalas que no podrían verse como tales mediante métodos univariantes.

En cuanto al análisis discriminante PLS-DA llevado a cabo, se concluye su buena capacidad para discriminar entre masas de aguas buenas y "no buenas", al aportar un modelo capaz de predecir, a partir de los parámetros de calidad, si el agua puede clasificarse como buena o no, con un alto grado de acierto (98,3% de acierto global). Además las componentes en este modelo están fuertemente relacionadas con los parámetros usados por la Normativa para la clasificación de las aguas atendiendo a su calidad para uso agrícola, lo que corrobora la bondad del estudio realizado.

6. BIBLIOGRAFÍA

BEAMONTE, E.; CASINO, A.; VERES, E. J. (2004). *La calidad del agua en ciertas estaciones de control del canal Júcar-Turia (período 1994-2001)*, en Revista Española de estudios Agrosociales y Pesqueros, número 201; pp. 105-126.

BEAMONTE, E.; BERMÚDEZ, J.; CASINO, A.; VERES, E. (2004). *Un indicador global para la calidad del agua. Aplicación a las aguas supeficiales de la Comunidad Valenciana,* en Revista Estadística Española, Volumen 46, número 156; pp. 189-204.

BEAMONTE, E.; BERMÚDEZ, J.; CASINO, A.; VERES, E. (2007). *A statistical study of the quality of suarce wáter intended for human consumption near Valencia (Spain)*, en Journal of Environmental Management, número 83; pp. 307-314.

BEAMONTE, E.; CASINO, A.; VERES, E. (2010). *Water quality indicators: Comparison of a probabilistic index and general quality index. The case of the Confederación Hidrográfica del Júcar (Spain),* en Ecological Indicators, número 10; pp.1049 -1054.

BEAMONTE, E.; CASINO, A.; VERES, E. J. (2012). *Análisis de la calidad general del agua superficial en la cuenca hidrográfica del Júcar: periodo 2000-2009*, en M+A. Revista Electrónic@ de Medioambiente, número 12; pp. 18-32.

CHJ (2005). *Informe para la Comisión Europea sobre los artículos 5 y 6 de la Directiva Marco del Agua. Demarcación Hidrográfica del Júcar.* Confederación Hidrográfica del Júcar, Valencia.

CHJ (2008). *Plan de recuperación del río Júcar.* Confederación Hidrográfica del Júcar, Valencia.

HERNÁNDEZ, F.; MOLINOS, M.; SALA, R. (2009). *Valoración Económica de los Beneficios Ambientales del Proceso de Depuración de Aguas Residuales* en Rect@, Volumen 17.

KARAVOLTSOS, S.; SAKELLARI,A.; MIHOPOULOS, N.; DASSENAKIS, M.; SCOULLOS, M.J. (2008). *Evaluation of the quality of drinking water in regions of Greece,* en Desalination, número 224: pp. 317-329

SAZ, S.; HERNÁNDEZ, F. ; SALA, R. (2009). *Estimación del valor económico de la calidad del agua de un río mediante una doble aproximación: una aplicación de los principios económicos de laDirectiva Marco del Agua,* en Economía Agraria y Recursos Naturales", Volumen 9, número 1; pp. 37-63

SEVILLA, M.; TORREGROSA, T. (2009-2011). *Alternativas hídricas y agricultura.* Documento de trabajo del Instituto Interuniversitario de Economía Internacional (IEI). Alicante.

SEVILLA, M.; TORREGROSA, T.; MORENO, L. (2010). *Un panorama sobre la economía del agua,* en Estudios de Economía Aplicada, Volumen 28, número 2; pp. 265-304

SHRESTHA, S.; KAZAMA, F. (2007). *Assesment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan*, en Enviroment Modelling & Software, número 22; pp. 464-475

SOTELO, M. (2013). *Territorio y medio ambiente en la Comunidad de Madrid. Las infraestructuras históricas, nuevos paisajes culturales del agua,* en M+A. Revista Electrónic@ de Medioambiente, Volumen 14, número 1; pp. 87-115.

VERA, J.F. (2006). *Agua y modelo de desarrollo turístico: la necesidad de nuevos criterios para la gestión de los recursos,* en Boletín de la Asociación de Geógrafos Españoles (A.E.E.) número 42; pp. 155-178