



GOBIERNO  
DE ESPAÑA

MINISTERIO  
DE MEDIO AMBIENTE  
Y MEDIO RURAL Y MARINO

CONFEDERACIÓN  
HIDROGRÁFICA  
DEL EBRO



# ESTUDIO ESTADÍSTICO DE DATOS FÍSICO-QUÍMICOS Y BIOLÓGICOS POR TIPOS DE MASAS DE AGUA DE LA CUENCA DEL EBRO

Julio-2010

Concha DURÁN (Confederación Hidrográfica del Ebro)

Pilar GARGALLO (Universidad de Zaragoza)

Cristina PINTOR (Confederación Hidrográfica del Ebro)

Manuel SALVADOR (Universidad de Zaragoza)



## INDICE

RESUMEN .....	5
1. OBJETIVO DEL ESTUDIO .....	7
2. PLANTEAMIENTO DEL PROBLEMA.....	8
3. ANÁLISIS ESTADÍSTICO DE LOS PARÁMETROS BIOLÓGICOS .....	11
3.1. ANÁLISIS ESTADÍSTICO DEL INDICADOR IBMWP .....	11
3.2 ANÁLISIS ESTADÍSTICO DEL INDICADOR IPS .....	17
3.3 CONSTRUCCIÓN DEL INDICADOR DE CALIDAD CONJUNTO .....	23
4. METODOLOGIA ESTADISTICA.....	25
5. DETERMINACIÓN DE LOS UMBRALES DE CALIDAD.....	28
5.1. UMBRALES DE CALIDAD PARA LOS INDICADORES GENERALES .....	28
5.2. UMBRALES DE CALIDAD PARA LOS INDICADORES DE MATERIA ORGÁNICA .....	50
5.3. UMBRALES DE CALIDAD PARA LOS INDICADORES DE NUTRIENTES.....	57
6. CONCLUSIONES.....	92



## **RESUMEN**

A partir de los datos biológicos y físico- químicos tomados entre los años 2002- 2008 se determinan estadísticamente los límites de algunos parámetros físico-químicos correspondientes al cambio de estado ecológico Bueno/ Moderado. Para ello se utilizan como referencia los límites establecidos según un indicador biológico conjunto formado por los índices IBMWP e IPS, estos índices son usados habitualmente en ríos para establecer el estado ecológico según los indicadores biológicos. El principal problema de los índices biológicos es el elevado coste que supone la toma de muestras y la determinación de los taxones hasta los niveles adecuados para cada índice, frente a la relativa sencillez de la toma de muestra de los parámetros físico químicos, por ello sería de gran utilidad determinar los umbrales físico-químicos correspondientes al cambio de estado Bueno/Moderado según los indicadores biológicos. Como novedad en este trabajo se estudia la conveniencia de tener en cuenta las tipologías de las masas de agua establecidas según la IPH en la determinación de dichos umbrales.



## 1. OBJETIVO DEL ESTUDIO

En este trabajo se realiza una determinación estadística de los umbrales de algunos parámetros físico- químicos indicadores de la calidad ecológica de los ríos de la cuenca del Ebro. De esta forma se busca optimizar los procedimientos sistemáticos de determinación del estado ecológico de las masas de agua, basados habitualmente en el estudio de sus parámetros biológicos complementados con la información proporcionada por parámetros físico-químicos característicos (pH, conductividad, nutrientes, etc.) que suelen ser más fáciles de determinar.

La construcción de dichos umbrales se basa en el estudio de la relación de los parámetros físico-químicos del agua con sus parámetros biológicos a partir de la cual, se determinan los valores de los parámetros físico-químicos asociados a un estado determinado a partir de parámetros biológicos. A diferencia de estudios anteriores, sin embargo, el análisis se realiza teniendo en cuenta la tipología de la masa de agua la cual, al hacer más homogénea la información proporcionada por las muestras obtenidas, puede potencialmente aumentar la precisión de la estimación de los umbrales obtenidos mejorando, por tanto, los procedimientos sistemáticos de determinación del estado ecológico

El estudio de la relación existente entre los parámetros físico-químicos y los biológicos se realiza mediante la utilización de modelos de regresión logística estimados a partir de muestras de agua obtenidas en diferentes puntos de muestreo de la cuenca a lo largo del tiempo. En dichos modelos se toma como variable dependiente un indicador de la calidad del agua construido a partir de los parámetros biológicos IBMWP e IPS y que toma como umbrales los utilizados por la Confederación Hidrográfica del Ebro (CHE en adelante) teniendo en cuenta la tipología del río en el que se ha tomado la muestra. La determinación del modelo se realiza aplicando procedimientos de selección de variables con el fin de evitar la existencia de problemas de multicolinealidad que distorsionen la estimación de los parámetros del mismo. A partir de modelos estimados, y para un nivel de riesgo dado, se determinan, para cada tipología, los umbrales de calidad de los parámetros físico-químicos. Así mismo se evalúa el comportamiento del umbral empíricamente, comparándolo con los umbrales utilizados actualmente por la CHE así como con procedimientos que utilizarían un único umbral para todas las tipologías.

El informe se organiza de la siguiente manera:

- Sección 2: se describen las variables y los datos utilizados en el estudio
- Sección 3 se realiza un análisis estadístico exploratorio de los parámetros biológicos

- Sección 4 se expone la metodología estadística utilizada para la determinación de los umbrales
- Sección 5 se muestran los resultados obtenidos al aplicar dicha metodología.
- Sección 6 concluye exponiendo los resultados más relevantes así como las líneas de investigaciones futuras que el equipo pretende abordar en trabajos posteriores.

## 2. PLANTEAMIENTO DEL PROBLEMA

Los datos corresponden a la medición, en distintos puntos de la cuenca del Ebro, de diversos parámetros biológicos y físico-químicos en muestras tomadas en los ríos de dicha cuenca. Las observaciones se realizaron en el periodo de tiempo comprendido entre los años 2002 y 2008, con una frecuencia quincenal-mensual, en el caso de las variables físico-químicas, y anual en el caso de los indicadores de calidad biológicos, debido al mayor coste y grado de dificultad de la medición de éstos últimos. Los datos finalmente analizados corresponden a los valores medianos anuales de ambos tipos de parámetros siendo el número total de observaciones igual a 2.759. De dichas observaciones algunas debieron ser eliminadas por las siguientes razones:

- no se disponía de información sobre la tipología de la masa de agua (221 observaciones correspondientes a canales y embalses en su mayor parte),
- no se disponía de información de ninguno de los dos indicadores biológicos (741)
- todas las muestras estaban clasificadas en un estado superior al bueno, este es el caso de la tipología 111 (ríos de montaña silíceo, 41 observaciones), este hecho hacía inviable estadísticamente el determinar un umbral de calidad (se necesitaría disponer de muestras de agua de baja calidad).

De esta forma la muestra quedó reducida a **1766 observaciones**. El **elevado porcentaje de datos ausentes** (29,20%) hace que los resultados obtenidos deban tomarse con las debidas precauciones y ponen de manifiesto la necesidad de mejorar el proceso de obtención de datos dado que, la mayor parte de los datos ausentes se concentran en los años 2004 (152 observaciones) y 2008 (134 observaciones), años en los que hay un menor número de muestras.

**Tabla 2.1: Frecuencia de observaciones por año**

<b>Año</b>	<b>Frecuencia</b>	<b>Porcentaje</b>
2002	166	9,4
2003	130	7,4
2004	219	12,4
2005	300	17,0
2006	321	18,2
2007	331	18,7
2008	299	16,9
Total	1766	100,0

En la Tabla 2.1 se muestra el número de observaciones por año, apreciándose que durante los años previos a 2005 dicho número fue significativamente menor.

Los indicadores biológicos considerados en el estudio fueron dos de los utilizados por la CHE para la determinación del estado ecológico en ríos y aparecen listados en la Tabla 2.2. En dicha tabla se muestran, además, los umbrales utilizados por la CHE, por tipología, para determinar si una masa de agua tipo río cumple los requisitos mínimos para ser declarada en buen estado ecológico, según los parámetros biológicos, de acuerdo con lo indicado en la Directiva Marco del Agua.

**Tabla 2.2: Indicadores biológicos y umbrales de calidad biológicos entre los estados Bueno y Moderado utilizados en el estudio**

<b>Tipología</b>	<b>IBMWP</b> (Macroinvertebrados)	<b>IPS</b> (Diatomeas)
109: Ríos mineralizados de baja montaña mediterránea	Estado inferior a Bueno: IBMWP $\leq$ 94	Estado inferior a Bueno: IPS $\leq$ 12.6
112: Ríos de montaña mediterránea calcárea	Estado inferior a Bueno: IBMWP $\leq$ 101	Estado inferior a Bueno: IPS $\leq$ 11.9
115:Ejes mediterráneos-continentales poco mineralizados	Estado inferior a Bueno: IBMWP $\leq$ 101	Estado inferior a Bueno: IPS $\leq$ 11.9
116:Ejes mediterráneo-continentales mineralizados	Estado inferior a Bueno: IBMWP $\leq$ 101	Estado inferior a Bueno: IPS $\leq$ 11.9
117:Grandes ejes en ambiente mediterráneo	Estado inferior a Bueno: IBMWP $\leq$ 101	Estado inferior a Bueno: IPS $\leq$ 11.9
126:Ríos de montaña húmeda calcárea	Estado inferior a Bueno: IBMWP $\leq$ 95	Estado inferior a Bueno: IPS $\leq$ 12.2
127:Ríos de alta montaña	Estado inferior a Bueno: IBMWP $\leq$ 103	Estado inferior a Bueno: IPS $\leq$ 13.1

Los indicadores físico-químicos considerados en el estudio fueron nueve y aparecen listados en la Tabla 2.3. Más concretamente se han utilizado tres indicadores generales: pH (PH), conductividad (COND20) y oxígeno disuelto en concentración (O2), uno de materia orgánica: demanda química de oxígeno (DQO) y cinco indicadores de nutrientes: amonio (NH4), nitritos (NO2), nitratos (NO3), fosfatos (PO4) y fósforo total (P\_TOT).

**Tabla 2.3: Indicadores físico-químicos utilizados en el estudio**

<b>Tipo de indicador</b>	<b>Nombre</b>	<b>Significado</b>
Generales	PH	pH
	COND20	Conductividad a 20°C (μS/cm)
	O2	Oxígeno disuelto (en concentración) (mg/L O2)
Materia orgánica	DQO	Demanda Química de Oxígeno (al dicromato) (mg/L O2)
Nutrientes	NH4	Amonio Total (mg/L NH4)
	NO2	Nitritos (mg/L NO2)
	NO3	Nitratos (mg/L NO3)
	PO4	Fosfatos (mg/L PO4)
	P_TOT	Fósforo Total (mg/L P)

Nuestro objetivo se concreta en establecer umbrales para cada uno de estos indicadores que permitan predecir el estado ecológico de una masa con la mayor fiabilidad posible a partir de estos parámetros, analizando si tener en cuenta la tipología del río de la que dicha muestra fue extraída mejora significativamente los resultados obtenidos frente a no tener en cuenta dicha información. Para tal fin se ha utilizado la metodología estadística que se describe en la sección 4. Previamente en la sección siguiente realizamos un estudio estadístico-descriptivo de los datos utilizados en el estudio.

En el presente estudio los términos “mala calidad” y “buena calidad” del agua se utilizan para referirse a “estado inferior a bueno” o “superior a bueno” de la masa de agua respectivamente, según el indicador al que se esté aludiendo en cada momento.

### 3. ANÁLISIS ESTADÍSTICO DE LOS PARÁMETROS BIOLÓGICOS

En esta sección se lleva a cabo un análisis estadístico del comportamiento de los indicadores biológicos considerados en el estudio, a saber IBMWP e IPS, según la tipología del río. Su finalidad es analizar si dicho comportamiento es lo suficientemente heterogéneo como para justificar un procedimiento de determinación de umbrales de calidad que tenga en cuenta la tipología del río.

El estudio se lleva a cabo tanto desde un punto de vista gráfico como numérico utilizando el paquete estadístico SPSS 15.0 como herramienta de cálculo.

#### 3.1. ANÁLISIS ESTADÍSTICO DEL INDICADOR IBMWP

Se observa, en primer lugar, la existencia de un alto porcentaje de datos ausentes (36,5%, ver Tabla 3.1.1) en la muestra analizada consecuencia de que dicho indicador sólo se ha medido a partir del año 2004 debido a problemas presupuestarios. Por tipologías llama la atención el escaso tamaño muestral de la tipología 116 (16 observaciones válidas de 23 posibles) que es, junto con las tipologías 117 (55,4% de datos ausentes) y 115 (46%) las que mayores porcentajes de datos ausentes tienen superando, en todos los casos el 40% de las observaciones. Sin embargo, se decidió mantener dichas tipologías en el estudio al disponer de muestras de ambos tipos (contaminadas y no contaminadas). No obstante, en el caso de los ríos de tipología 116 y dado el escaso tamaño muestral disponible, los resultados que hacen referencia a esta tipología deben tomarse con precaución.

**Tabla 3.1.1: Datos ausentes en el indicador IBMWP por tipologías**

Tipología	Válidos		Perdidos		Total	
	Casos	Porcentaje	Casos	Porcentaje	Casos	Porcentaje
109	198	64,3%	110	35,7%	308	100,0%
112	313	68,2%	146	31,8%	459	100,0%
115	147	54,0%	125	46,0%	272	100,0%
116	16	69,6%	7	30,4%	23	100,0%
117	54	44,6%	67	55,4%	121	100,0%
126	267	65,1%	143	34,9%	410	100,0%
127	127	73,4%	46	26,6%	173	100,0%
Total	1122	63,5%	644	36,5%	1766	100,0%

En la Tabla 3.1.2 se muestran los valores de algunos estadísticos descriptivos de este indicador por tipologías. Así mismo, en las Figuras 3.1.1 a 3.1.3 se muestran, por tipologías, el diagrama de puntos, el diagrama de cajas y el diagrama de barras de error del 95% de confianza para la media del indicador IBMWP, respectivamente. Finalmente, en las Tablas 3.1.3 a 3.1.5

se presentan los resultados obtenidos al aplicar los contrastes de Kruskal-Wallis (Tabla 3.1.3), mediana (Tabla 3.1.4) y Mann-Whitney (Tabla 3.1.5) dos a dos utilizando, en este último caso, un nivel de significación del 5% y el procedimiento de Bonferroni para tener en cuenta la simultaneidad de los contrastes.

**Tabla 3.1.2: Análisis estadístico-descriptivo del indicador IBMWP por tipologías**

Tipología	Casos	Media	Mediana	Desv. típ.	Mínimo	Máximo	Asimetría	Curtosis
109	198	87,34	80,00	43,85	6,00	222,50	0,56	-0,12
112	313	124,47	118,00	51,27	8,00	288,50	0,55	0,02
115	147	94,26	91,00	41,57	12,00	210,50	0,21	-0,49
116	16	49,53	55,00	18,11	21,00	77,00	-0,15	-1,33
117	54	79,57	75,00	30,63	25,00	159,00	0,31	-0,42
126	267	148,94	149,00	54,49	10,00	317,00	0,08	0,12
127	127	145,15	143,00	34,78	10,00	241,00	-0,04	1,54
<b>Total</b>	1122	118,89	116,00	53,70	6,00	317,00	0,38	-0,09

Todos los contrastes rechazan la hipótesis de homogeneidad de comportamiento del indicador IBMWP por tipologías, observándose (ver Figura 3.1.3 y Tabla 3.1.5) que los niveles más bajos del indicador se dan en la tipología 116 con niveles medios por debajo de 50, seguida por las tipologías 115, 117 y 109 todas ellas con niveles medios de IBMWP por debajo de 100. Los ríos de tipología 112 ocupan niveles intermedios correspondiendo los más altos a las tipologías 126 y 127 cuyos niveles medios son superiores a 140.

**Tabla 3.1.3: Contraste de Kruskal-Wallis de comparación del indicador IBMWP por tipologías**

Casos	Tipología						
	198	313	147	16	54	267	127
<b>Rango Promedio</b>	366,23	591,80	413,78	121,63	307,21	739,81	750,91
<b>Chi-cuadrado</b>	292,19						
<b>gl</b>	6						
<b>Pvalor asintót.</b>	0,0000						

**Tabla 3.1.4: Contraste de la mediana de comparación del indicador IBMWP por tipologías**

	Tipología						
	109	112	115	116	117	126	127
> Mediana	43	158	41	0	6	202	109
<= Mediana	155	155	106	16	48	65	18

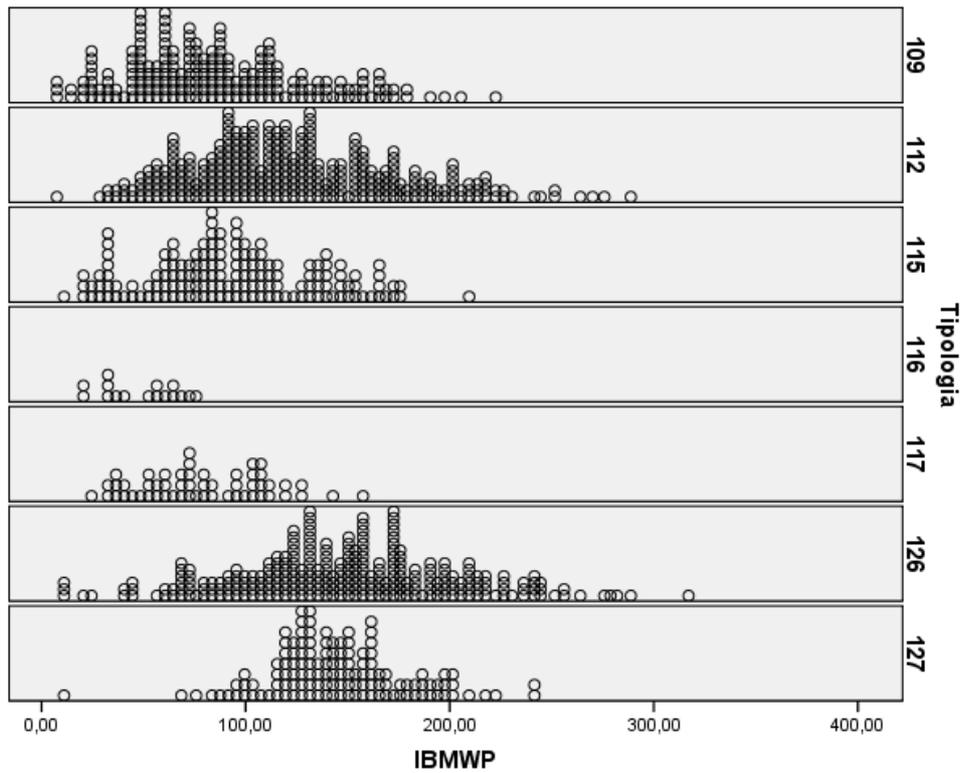
N 1122  
 Mediana 116,00  
 Chi-cuadrado 276,28  
 gl 6  
 Pvalor 0,000  
 asintót.

**Tabla 3.1.5: Resultados de los contrastes de Mann-Whitney aplicados dos a dos al indicador IBMWP por tipologías**

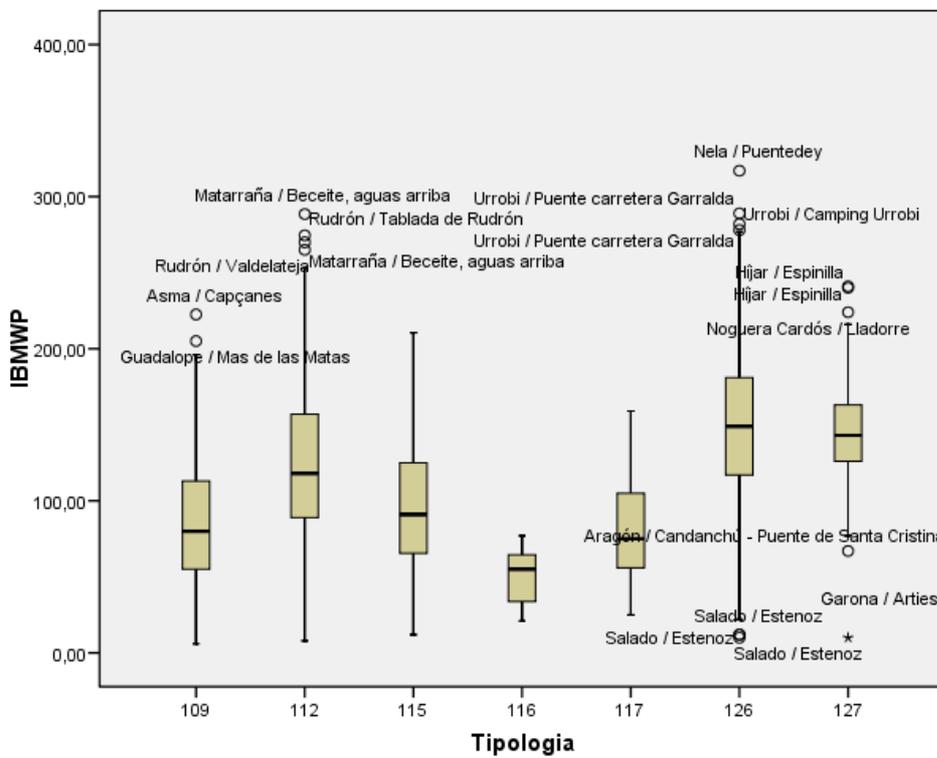
Tipología	112	115	116	117	126	127
109	< <sup>+</sup>		>		<	<
112		> <sup>++</sup>	>	>	<	<
115			>		<	<
116				<	<	<
117					<	<
126						

<sup>+</sup> El nivel mediano del indicador IBMWP en los ríos de la tipología 109 es significativamente inferior al de los ríos de la tipología 112

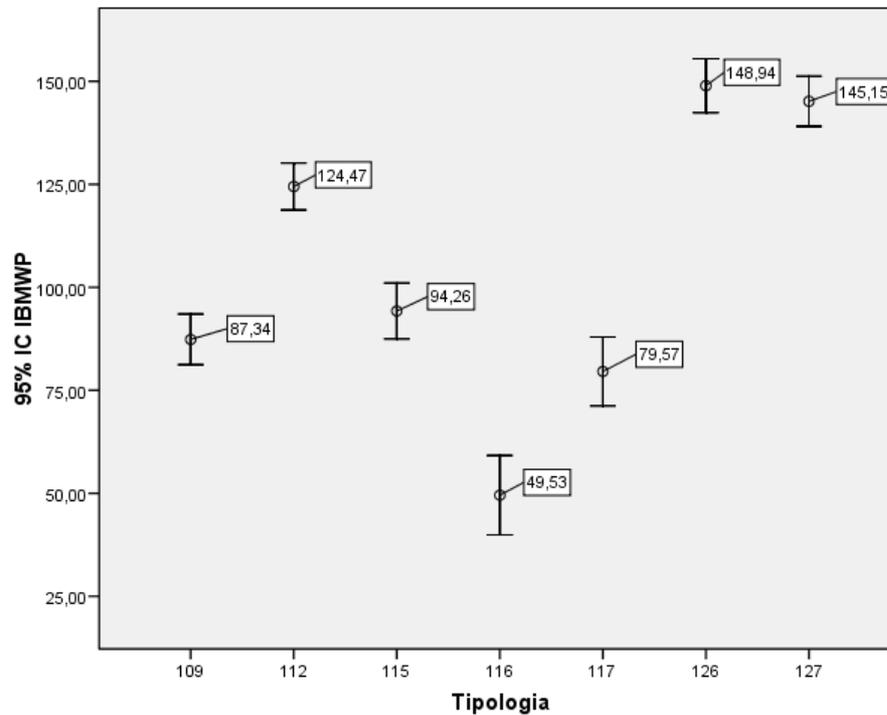
<sup>++</sup> El nivel mediano del indicador IBMWP en los ríos de la tipología 112 es significativamente superior al de los ríos de la tipología 115



**Figura 3.1.1: Diagrama de puntos del indicador IBMWP por tipologías**

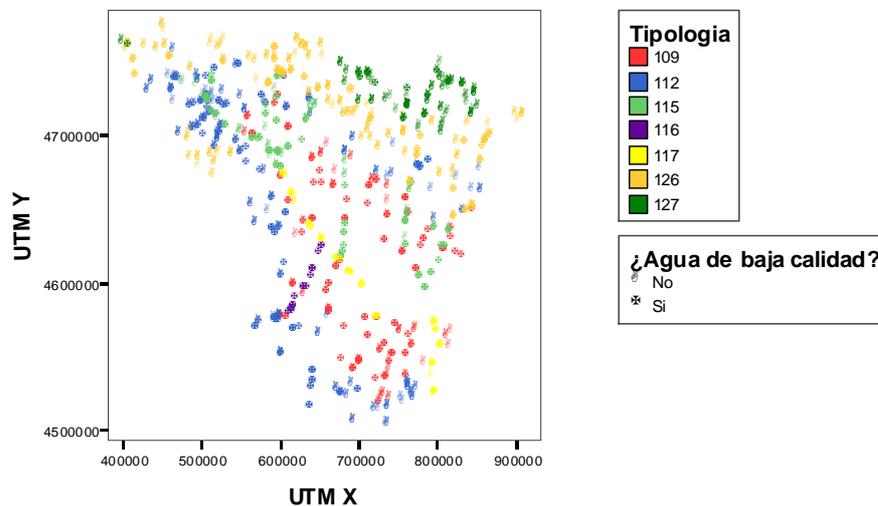


**Figura 3.1.2: Diagrama de cajas del indicador IBMWP**



**Figura 3.1.3: Barras de error del 95% del indicador IBMWP por tipologías (encuadrados los valores medios del indicador en cada tipología)**

Finalmente en la Figura 3.1.4 y en la Tabla 3.1.6 se analiza la distribución de los niveles de calidad del agua tanto a nivel geográfico (Figura 3.1.4) como por tipologías (Tabla 3.1.6).



**Figura 3.1.4: Distribución geográfica de los niveles de calidad biológica de las aguas a partir del indicador IBMWP tal y como se detalla en la Tabla 2.2**

Se observa (ver Figura 3.1.4) que la mayor parte de los puntos de muestreo con calidad de agua alta se concentra en la zonas de montaña (norte y noroeste) de la cuenca que corresponden a las tipologías 126 y 127. Por el contrario, los puntos de muestreo situados a lo largo del cauce del río Ebro, pertenecientes a las tipologías 109, 115, 116 y 117, tienden a mostrar peores niveles de calidad. Este hecho queda puesto de manifiesto en la Tabla 3.1.6 en la que se muestran los porcentajes de muestras de agua contaminada y sin contaminar por tipología, así como los residuos tipificados de un contraste de la  $\chi^2$  cuadrado de independencia. Con respecto al indicador IBMWP, el nivel de calidad de las aguas es, en general bueno, al haber un 61,1% de puntos de muestreo con agua de buena calidad. Sin embargo dichas distribución no es uniforme siendo las tipologías anteriores las que muestran niveles de calidad muy inferiores con porcentajes de muestras de agua contaminada superiores al 60% en todos los casos destacando, en particular, los ríos de la tipología 116, en los que el 100% de las observaciones son de baja calidad.

**Tabla 3.1.6: Nivel de calidad del agua por tipologías según el indicador IBMWP**

			¿Agua de baja calidad?		Total
			No	Si	No
<b>Tipología</b>	109	Recuento	75	123	198
		% de Tipología	37,9%	62,1%	100,0%
		Residuos tipificados	-4,2	5,3	
	112	Recuento	198	115	313
		% de Tipología	63,3%	36,7%	100,0%
		Residuos tipificados	,5	-,6	
	115	Recuento	58	89	147
		% de Tipología	39,5%	60,5%	100,0%
		Residuos tipificados	-3,4	4,2	
	116	Recuento	0	16	16
		% de Tipología	,0%	100,0%	100,0%
		Residuos tipificados	-3,1	3,9	
	117	Recuento	16	38	54
		% de Tipología	29,6%	70,4%	100,0%
		Residuos tipificados	-3,0	3,7	
	126	Recuento	225	42	267
		% de Tipología	84,3%	15,7%	100,0%
		Residuos tipificados	4,8	-6,1	
	127	Recuento	114	13	127
		% de Tipología	89,8%	10,2%	100,0%
		Residuos tipificados	4,1	-5,2	
Total	Recuento	686	436	1122	
	% de Tipología	61,1%	38,9%	100,0%	

### 3.2 ANÁLISIS ESTADÍSTICO DEL INDICADOR IPS

Se observa, en primer lugar, la existencia de un porcentaje moderado de datos ausentes (27,1%) (ver Tabla 3.2.1) que es consecuencia de la inexistencia de observaciones respecto a este indicador en el año 2004 debido a problemas presupuestarios. Por tipologías vuelve a llamar la atención, en primer lugar, el escaso tamaño muestral de la tipología 116 (15 observaciones de 23 posibles) que es, además, la que más datos ausentes tiene (34,8%) seguida de la tipología 109 con un 33,8% . Se decidió mantener, sin embargo, a dicha tipología en el estudio al disponer de muestras de ambos tipos (contaminadas y no contaminadas) aunque, dado el escaso tamaño muestral disponible, los resultados que hacen referencia a esta tipología deben tomarse con precaución.

**Tabla 3.2.1: Datos ausentes en el indicador IPS por tipologías**

Tipología	Válidos		Perdidos		Total	
	Casos	Porcentaje	Casos	Porcentaje	Casos	Porcentaje
109	204	66,2%	104	33,8%	308	100,0%
112	322	70,2%	137	29,8%	459	100,0%
115	205	75,4%	67	24,6%	272	100,0%
116	15	65,2%	8	34,8%	23	100,0%
117	93	76,9%	28	23,1%	121	100,0%
126	308	75,1%	102	24,9%	410	100,0%
127	140	80,9%	33	19,1%	173	100,0%
Total	1287	72,9%	479	27,1%	1766	100,0%

En la Tabla 3.2.2 se muestran los valores de algunos estadísticos descriptivos de este indicador por tipologías. Así mismo, en las Figuras 3.2.1 a 3.2.3 se muestran, respectivamente, el diagrama de puntos, el diagrama de cajas y el diagrama de barras de error para la media al 95% del indicador IPS por tipologías. Finalmente, en las Tablas 3.2.3 a 3.2.5 se presentan los resultados obtenidos al aplicar los contrastes de Kruskal-Wallis (Tabla 3.2.3), mediana (Tabla 3.2.4) y Mann-Whitney (Tabla 3.2.5) dos a dos utilizando, en este último caso, un nivel de significación del 5% y el procedimiento de Bonferroni para tener en cuenta la simultaneidad de los contrastes.

**Tabla 3.2.2: Análisis estadístico-descriptivo del indicador IPS por tipologías**

Tipología	Casos	Media	Mediana	Desv. típ.	Mínimo	Máximo	Asimetría	Curtosis
109	204	13,85	15,25	4,12	1,20	19,60	-1,12	0,53
112	322	15,26	16,00	3,27	1,30	19,80	-1,51	2,67
115	205	12,85	12,90	3,56	2,20	19,80	-0,34	-0,01
116	15	12,89	13,30	1,63	9,10	14,70	-1,17	1,07
117	93	9,63	10,10	2,94	1,90	16,40	-0,23	-0,24
126	308	17,01	17,60	2,62	4,20	20,00	-1,68	4,11
127	140	18,41	18,90	1,59	11,50	20,00	-2,21	5,59
Total	1287	14,98	15,90	3,91	1,20	20,00	-1,00	0,52

Todos los contrastes rechazan la hipótesis de homogeneidad de comportamiento del indicador IPS por tipologías, observándose (ver Figura 3.2.3 y Tabla 3.2.5) que los niveles más bajos del indicador se dan en la tipología 117 con niveles medios por debajo de 10, seguidos de los ríos de las tipologías 115, 116 y 109, todos ellos con niveles medios por debajo de 14. Los ríos de tipología 112 ocupan niveles intermedios correspondiendo los más altos a las tipologías 126 y 127 cuyos niveles medios son superiores a 17.

**Tabla 3.2.3: Contraste de Kruskal-Wallis de comparación del indicador IPS por tipologías**

Casos	Tipología						
	204	322	205	15	93	308	140
<b>Rango Promedio</b>	527,66	643,50	413,42	342,57	181,96	849,94	1038,45
Chi-cuadrado	504,78						
Gl	6						
Pvalor asintót.	0,0000						

**Tabla 3.2.4: Contraste de la mediana de comparación del indicador IPS por tipologías**

	Tipología						
	109	112	115	116	117	126	127
<b>&gt; Mediana</b>	77	162	42	0	1	223	130
<b>&lt;= Mediana</b>	127	160	163	15	92	85	10

N 1287  
 Mediana 15,90  
 Chi-cuadrado 352,26  
 gl 6  
 Pvalor asintót. ,000

**Tabla 3.2.5: Resultados de los contrastes de Mann-Whitney aplicados dos a dos al indicador IPS por tipologías**

Tipología	112	115	116	117	126	127
109	< <sup>+</sup>	> <sup>++</sup>		>	<	<
112		>	>	>	<	<
115				>	<	<
116				>	<	<
117					<	<
126						<

<sup>+</sup> El nivel mediano del indicador IPS en los ríos de la tipología 109 es significativamente inferior al de los ríos de la tipología 112

<sup>++</sup> El nivel mediano del indicador IPS en los ríos de la tipología 109 es significativamente superior al de los ríos de la tipología 115

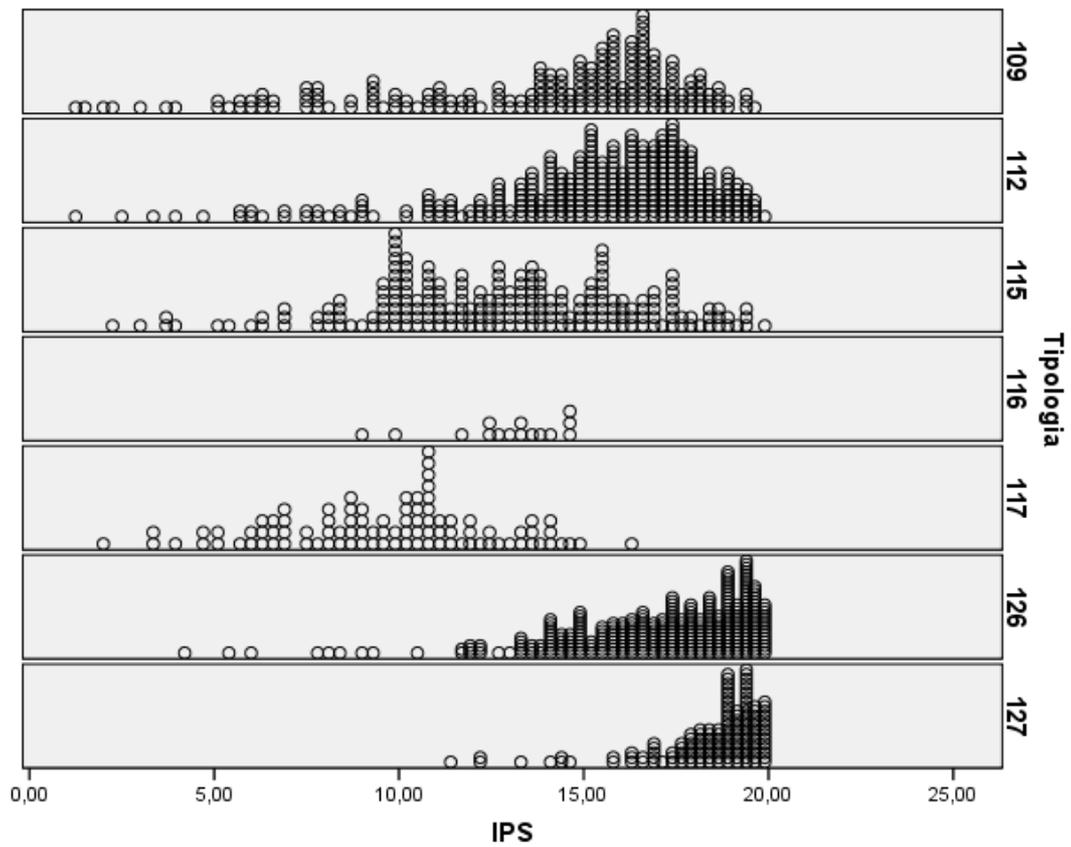


Figura 3.2.1: Diagrama de puntos del indicador IPS por tipologías

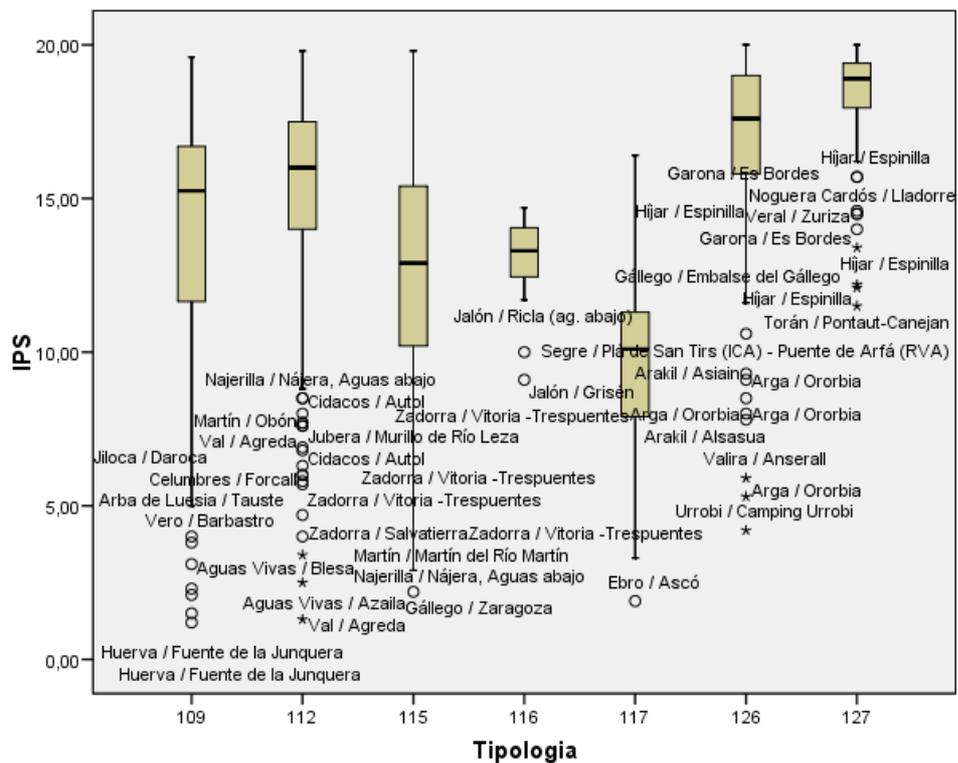
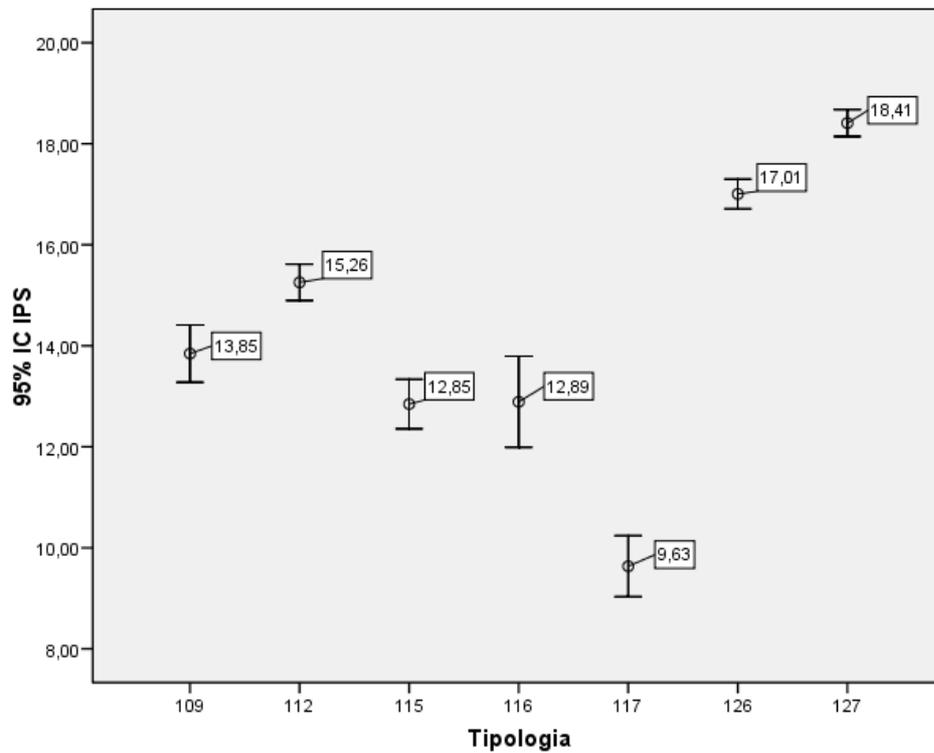
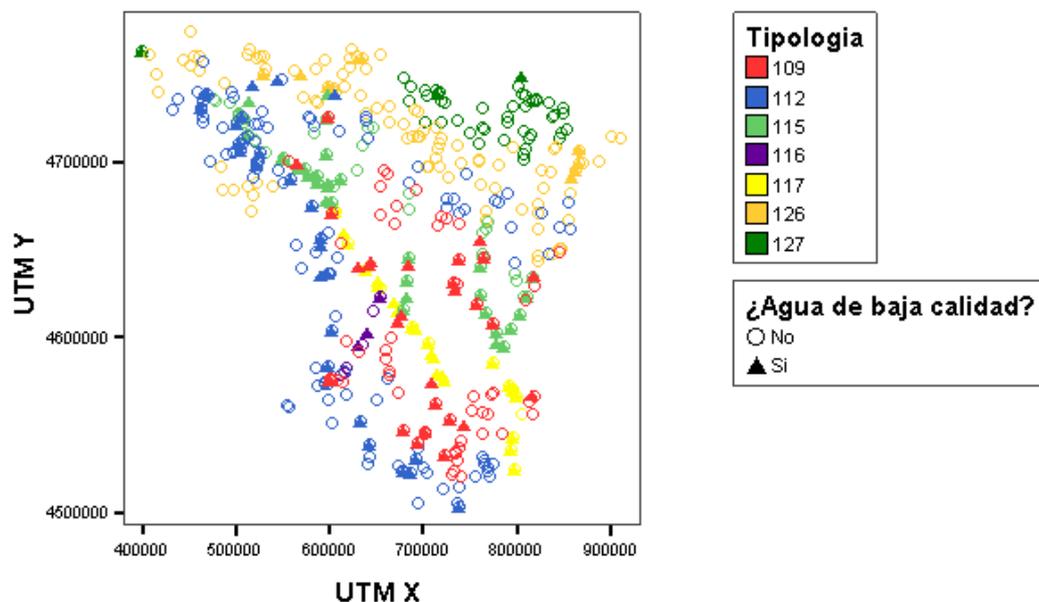


Figura 3.2.2: Diagrama de cajas del indicador IPS



**Figura 3.2.3: Barras de error del 95% del indicador IPS por tipologías (enmarcados los valores medios del indicador en cada tipología)**

Finalmente en la Figura 3.2.4 y en la Tabla 3.2.6 se analiza la distribución de los niveles de calidad del agua tanto a nivel geográfico (Figura 3.2.4) como por tipologías (Tabla 3.2.6).



**Figura 3.2.4: Distribución geográfica de los niveles de calidad de las aguas a partir del indicador IPS tal y como se detalla en la Tabla 2.2**

Se observa (ver Figura 3.2.4) que la mayor parte de los puntos de muestreo con calidad de agua alta se concentra en la zonas de montaña (norte y noroeste) de la cuenca que corresponden a las tipologías 126 y 127. Por el contrario, los puntos de muestreo situados a lo largo del cauce del río Ebro y muy particularmente los pertenecientes a las 115 y, sobre todo, la 117, tienden a mostrar peores niveles de calidad. Este hecho queda puesto de manifiesto en la Tabla 3.2.6 en la que se muestran los porcentajes de muestras de agua contaminada y sin contaminar por tipología, así como los residuos tipificados de un contraste de la  $\chi^2$  cuadrado de independencia. Con respecto al indicador IPS, el nivel de calidad de las aguas es, en general bueno, al haber un 78,4% de puntos de muestreo con agua de buena calidad. Sin embargo, dicha distribución no es uniforme destacando, en este sentido, los ríos de la tipología 117 para los cuales el 78.5% de las muestras de agua obtenida fueron de baja calidad.

**Tabla 3.2.6: Nivel de calidad del agua por tipologías según el indicador IPS**

			¿Agua de baja calidad?		Total
			No	Si	No
Tipología	109	Recuento	145	59	204
		% de Tipología	71,1%	28,9%	100,0%
		Residuos tipificados	-1,2	2,2	
	112	Recuento	281	41	322
		% de Tipología	87,3%	12,7%	100,0%
		Residuos tipificados	1,8	-3,4	
	115	Recuento	122	83	205
		% de Tipología	59,5%	40,5%	100,0%
		Residuos tipificados	-3,1	5,8	
	116	Recuento	12	3	15
		% de Tipología	80,0%	20,0%	100,0%
		Residuos tipificados	,1	-,1	
	117	Recuento	20	73	93
		% de Tipología	21,5%	78,5%	100,0%
		Residuos tipificados	-6,2	11,8	
	126	Recuento	292	16	308
		% de Tipología	94,8%	5,2%	100,0%
		Residuos tipificados	3,3	-6,2	
	127	Recuento	137	3	140
		% de Tipología	97,9%	2,1%	100,0%
		Residuos tipificados	2,6	-5,0	
Total	Recuento	1009	278	1287	
	% de Tipología	78,4%	21,6%	100,0%	

### 3.3 CONSTRUCCIÓN DEL INDICADOR DE CALIDAD CONJUNTO

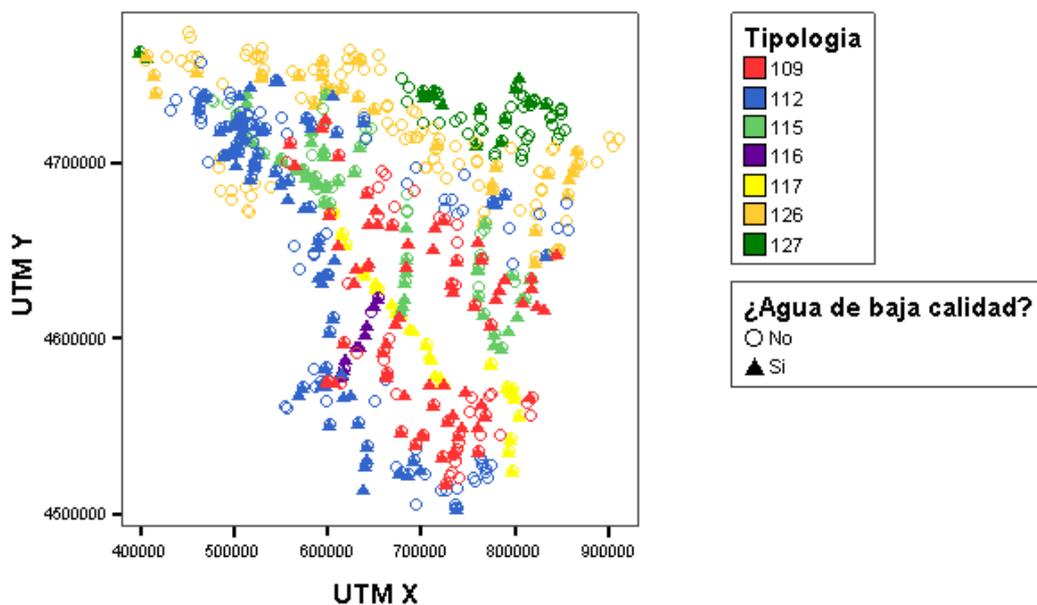
Dada la gran cantidad de datos ausentes existentes en cada uno de los indicadores biológicos por separado y con el fin de aprovechar al máximo la información disponible, se decidió construir un indicador de calidad conjunto que reflejara, con la mayor fiabilidad posible, la información contenida en los 2 indicadores biológicos anteriores. Dicho indicador se construyó de la siguiente manera: *diremos que un punto de muestreo cumple con los niveles de calidad mínimo exigibles desde un punto de vista biológico si verifica los requisitos listados en la columna umbrales de calidad de la Tabla 2.2 para todos los indicadores biológicos del estudio observados en dicho punto*. En otro caso diremos que no cumple dichos requisitos, a no ser que no tenga ningún indicador observado, en cuyo caso el indicador de calidad conjunto se declarará ausente. Este indicador jugará el papel de variable dependiente en los modelos de regresión estimados en la sección 5 del trabajo.

En la Tabla 3.3.1 se presenta la distribución de frecuencias de dicho indicador por tipologías. Se observa que la mayor parte de los puntos de muestreo (63,4%) verifican los requisitos mínimos de calidad.

**Tabla 3.3.1: Nivel de calidad del agua por tipologías según el indicador conjunto**

			¿Agua de baja calidad?		Total
			No	Si	No
<b>Tipología</b>	109	Recuento	148	160	308
		% de Tipología	48,1%	51,9%	100,0%
		Residuos tipificados	-3,4	4,4	
	112	Recuento	312	147	459
		% de Tipología	68,0%	32,0%	100,0%
		Residuos tipificados	1,2	-1,6	
	115	Recuento	116	156	272
		% de Tipología	42,6%	57,4%	100,0%
		Residuos tipificados	-4,3	5,6	
	116	Recuento	7	16	23
		% de Tipología	30,4%	69,6%	100,0%
		Residuos tipificados	-2,0	2,6	
	117	Recuento	24	97	121
		% de Tipología	19,8%	80,2%	100,0%
		Residuos tipificados	-6,0	7,9	
	126	Recuento	355	55	410
		% de Tipología	86,6%	13,4%	100,0%
		Residuos tipificados	5,9	-7,8	
127	Recuento	157	16	173	
	% de Tipología	90,8%	9,2%	100,0%	
	Residuos tipificados	4,5	-6,0		
Total	Recuento	1119	647	1766	
	% de Tipología	63,4%	36,6%	100,0%	

Si analizamos la distribución de dicho indicador tanto geográficamente (ver Figura 3.3.1) como por tipologías (ver Tabla 3.3.1) se puede apreciar, de nuevo que los puntos de muestreo que no verifican los requisitos de calidad tienden a situarse a lo largo del cauce del río Ebro. Por tipologías la que muestra los niveles más bajos de calidad es la 117 (80,2% de las muestras con agua de mala calidad, ver Tabla 3.3.1)) cuyos puntos de muestreo se encuentran situados a lo largo de la parte baja del cauce del río Ebro (ver Figura 3.3.1) y que muestra mal comportamiento en ambos indicadores. Posiciones intermedias ocupan las tipologías 116 (69,6% de muestras con agua de mala calidad), 115 (57,4) y 109 (51,9%) cuyos ríos tienden a estar situados en la parte media-baja del cauce (ver Figura 3.3.1) y que tienden a tener bajos niveles del indicador IBMWP (ver Tabla 3.1.3). Finalmente los niveles más altos se dan en las tipologías 112 (32% de muestras con agua de mala calidad) y, sobre todo, las tipologías 126 (13,4%) y 127 (9,2%) cuyos puntos de muestreo tienden a estar situados en la parte alta y/o al sur de la cuenca pero alejadas del cauce del río Ebro y en las zonas de montaña de los Pirineos.



**Figura 3.3.1: Distribución geográfica del indicador de calidad conjunto por tipologías**

#### 4. METODOLOGIA ESTADISTICA

Una vez realizado el análisis comparativo de los indicadores biológicos IBMWP e IPS por tipologías y rechazada, claramente, la hipótesis de homogeneidad de comportamiento de dichos indicadores por tipologías, nos planteamos, a continuación, si es posible determinar umbrales de calidad por tipologías para los parámetros físico-químicos detallados en la Tabla 2.3 que mejoren significativamente los resultados obtenidos si determinamos dichos umbrales sin tener en cuenta la tipología del río. A tal fin en esta sección se describe, brevemente, la metodología estadística utilizada para analizar estas cuestiones.

Para exponerla introducimos la siguiente notación. Sea  $Y$  = nivel de calidad conjunta codificada de forma binaria como:

$$Y = \begin{cases} 0 & \text{si la calidad del agua verifica los mínimos exigidos} \\ 1 & \text{si la calidad del agua no verifica los mínimos exigidos} \end{cases}$$

donde los mínimos exigidos vienen especificados por el indicador conjunto de calidad descrito en la sección 3.3.

Sea  $X$  la variable físico-química cuyos umbrales de calidad se quieren establecer. Supondremos, además, que es necesario realizar una transformación funcional  $g$  tal que  $g(X)$  es aproximadamente normal. La determinación de  $g$  fue llevada a cabo en un trabajo anterior y en este trabajo se tomará como dada.

Sea  $T$  la tipología del río del que procede la muestra.

La determinación de los umbrales de calidad se basa en la estimación de un modelo de regresión logística de la forma:

$$P(Y=1|X,T) = \frac{e^{\alpha_T + \beta_T g(X)}}{1 + e^{\alpha_T + \beta_T g(X)}} \quad (4.1)$$

donde  $P(Y=1|X,T)$  es la probabilidad de que si el valor del parámetro físico-químico de la muestra de agua es  $X$  y  $T$  la tipología del río de la que se extrajo, el agua de la zona considerada no verifique los estándares de calidad mínimo exigidos. Dicha probabilidad mide, por tanto, el riesgo de que dicha zona presente niveles de contaminación altos en el futuro. Observar, en particular, que la influencia de la tipología  $T$  en el modelo (4.1) se muestra a través de la dependencia de los parámetros del mismo,  $\alpha_T$  y  $\beta_T$ , de dicha tipología. Si, en particular,  $\alpha_T = \alpha$  y  $\beta_T = \beta$  se concluiría que la tipología no ejerce una influencia significativa

sobre el riesgo y que, por lo tanto, podrían utilizarse los mismos umbrales en los ríos de toda la cuenca.

Además, si  $X_0$  es tal que  $g(X_0) = 0$ , se verifica que  $P(Y=1|X_0,T) = \frac{e^{\alpha_T}}{1+e^{\alpha_T}}$  por lo que el parámetro  $\alpha_T$  mide el riesgo de que una zona de tipología T y con un valor del indicador  $X = X_0$ , no verifique los estándares de calidad mínimos. Así mismo, el parámetro físico-químico es relevante para medir los niveles de calidad del agua si  $\beta_T$  es distinto de 0 de forma que, al ser la función  $g(X)$  creciente en  $X$  en todos los casos analizados, se tiene que si  $\beta_T > 0$  (respectivamente, si  $\beta_T < 0$ ) valores altos de  $X$  (respectivamente valores bajos de  $X$ ) vendrán asociados a riesgos altos de contaminación en los ríos de tipología T.

Fijado un nivel de riesgo  $p$  de que el agua de la zona considerada no verifique los estándares de calidad mínimo exigidos, donde  $p = P(Y=1|X,T)$  con  $0 < p < 1$ , el umbral de calidad para el parámetro físico-químico  $X$  en los ríos de tipología T vendrá dado por la expresión:

$$\text{Umbral}_{p,T} = g^{-1} \left( \frac{\log\left(\frac{p}{1-p}\right) - \alpha_T}{\beta_T} \right) \quad (4.2)$$

donde  $g^{-1}$  denota a la función inversa de  $g$ . De esta forma se verifica que:

$$\text{a) Si } \beta_T > 0 \text{ y } X \geq \text{Umbral}_{p,T} \quad (4.3)$$

ó

$$\text{b) Si } \beta_T < 0 \text{ y } X \leq \text{Umbral}_{p,T} \quad (4.4)$$

el agua tiene un riesgo del 100p% de estar contaminada en el futuro. Si  $\beta_T = 0$ ,  $X$  no es relevante para predecir los niveles de contaminación del agua en los ríos de tipología T y no debería ser utilizado para ello.

Así mismo, se lleva a cabo una validación intramuestral de los umbrales propuestos. Dicha validación utiliza los datos de la muestra analizada y los clasifica como muestras de agua contaminada o no contaminada de acuerdo a las reglas (4.3) o (4.4) según proceda, para un valor fijo de  $p$ . Utilizando dicho procedimiento de clasificación se calculan las siguientes cantidades:

$$\% \text{ Aciertos cumplen} = 100 * \frac{n_{\text{bien clasificad os, cumplen}}}{n_{\text{cumplen}}} \quad (4.5)$$

$$\% \text{ Aciertos no cumplen} = 100 * \frac{n_{\text{bien clasificados, no cumplen}}}{n_{\text{no cumplen}}} \quad (4.6)$$

$$\% \text{ Aciertos totales} = 100 * \frac{n_{\text{bien clasificados}}}{n_{\text{total}}} \quad (4.7)$$

donde  $n_{\text{bien clasificados, cumplen}}$  y  $n_{\text{bien clasificados, no cumplen}}$  son el número de muestras bien clasificadas entre las muestras que cumplen y no cumplen los requisitos de calidad, respectivamente;  $n_{\text{cumplen}}$  y  $n_{\text{no cumplen}}$  son, respectivamente, el número de muestras que cumplen y no cumplen dichos requisitos de calidad;  $n_{\text{bien clasificados}} = n_{\text{bien clasificados, cumplen}} + n_{\text{bien clasificados, no cumplen}}$  es el número de muestras bien clasificadas y  $n_{\text{total}} = n_{\text{cumplen}} + n_{\text{no cumplen}}$  es el número total de muestras.

Estos indicadores nos proporcionan una evaluación, en términos prácticos, de la fiabilidad del indicador X a la hora de pronosticar el riesgo que corre una zona de no verificar los indicadores de calidad mínimos exigidos para sus aguas. Así, el **%Aciertos cumplen** y **%Aciertos no cumplen** evalúan la fiabilidad del umbral dentro de los casos que cumplen y que no cumplen con los requisitos de calidad, respectivamente. Por su parte el **%Aciertos totales** evalúa la fiabilidad global del umbral en el total de la muestra.

Dada la gran cantidad de posibles modelos de la forma (4.1) que sería necesario estimar se ha recurrido a procedimientos de selección de variables (denominados *step-wise* en inglés) tomando como posibles variables independientes del modelo las *dummies* correspondiente a cada tipología así como las interacciones de dichas *dummies* con las variables X correspondientes. Se han aplicado tanto procedimientos de selección hacia delante (*forward selection* en inglés) como de eliminación hacia atrás (*backward selection* en inglés) seleccionando el modelo con mejor comportamiento en términos del indicador (4.7).

Se ha realizado, además, un estudio comparativo, en términos de los indicadores (4.5) a (4.6) con el modelo que supone homogeneidad en todas las tipologías de forma que  $\alpha_T = \alpha$  y  $\beta_T = \beta$  el cual se ha tomado como referencia. Finalmente, y en aquéllos parámetros físico-químicos en los que la CHE ha estado utilizando umbrales, se han incluido dichos umbrales en las comparaciones anteriores con el fin de evaluar su eficacia.

Los resultados obtenidos se exponen en la siguiente sección.

## **5. DETERMINACIÓN DE LOS UMBRALES DE CALIDAD**

En esta sección se presentan los resultados obtenidos al aplicar la metodología descrita en la sección 4 a cada una de las variables físico-químicas consideradas en el estudio. Los resultados se presentan agrupando las variables por bloques según el tipo de indicador (ver Tabla 2.3). Para cada parámetro físico-químico se comienza realizando un estudio estadístico por tipologías de carácter exploratorio con el fin de analizar la posible existencia de sesgos que disminuyan la validez del análisis. Posteriormente se lleva a cabo el proceso de determinación de umbrales así como una validación estadística de los mismos aplicando la metodología descrita en la sección 4.

En todos los análisis se han utilizado el paquete estadístico SPSS 15.0 y la hoja de cálculo EXCEL 2003 como herramientas de cálculo.

### **5.1. UMBRALES DE CALIDAD PARA LOS INDICADORES GENERALES**

En este apartado presentamos los umbrales de calidad biológica para todos los indicadores generales PH, COND20 y O<sub>2</sub> cuyo significado puede verse en la Tabla 2.3.

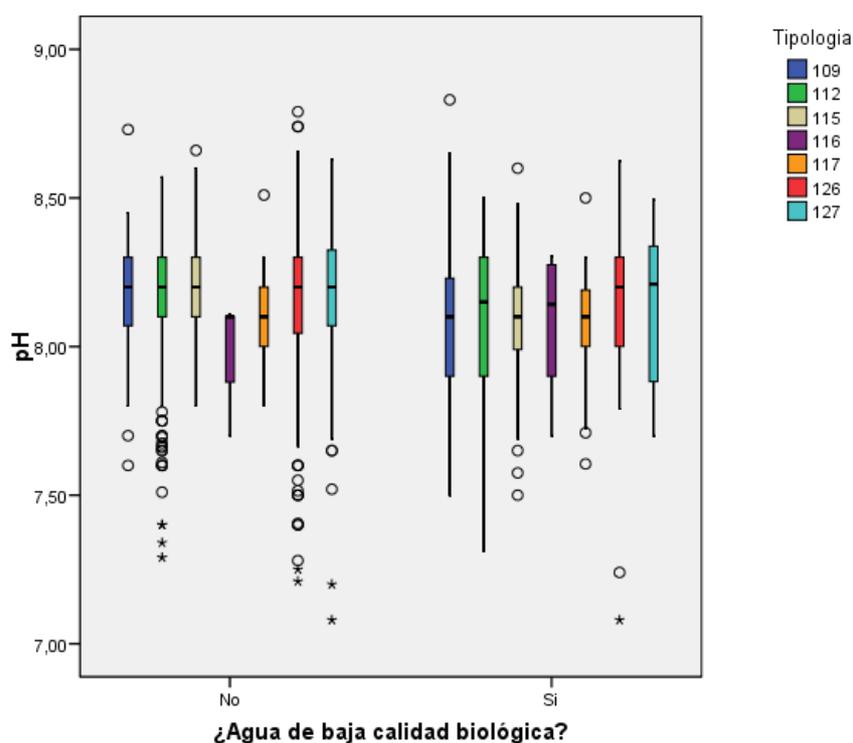
#### **5.1.1. Umbrales de calidad biológica para el índice pH**

De acuerdo con estudios realizados en proyectos anteriores se ha tomado  $g(X) = X$  al no mostrar el indicador analizado problemas serios de falta de normalidad.

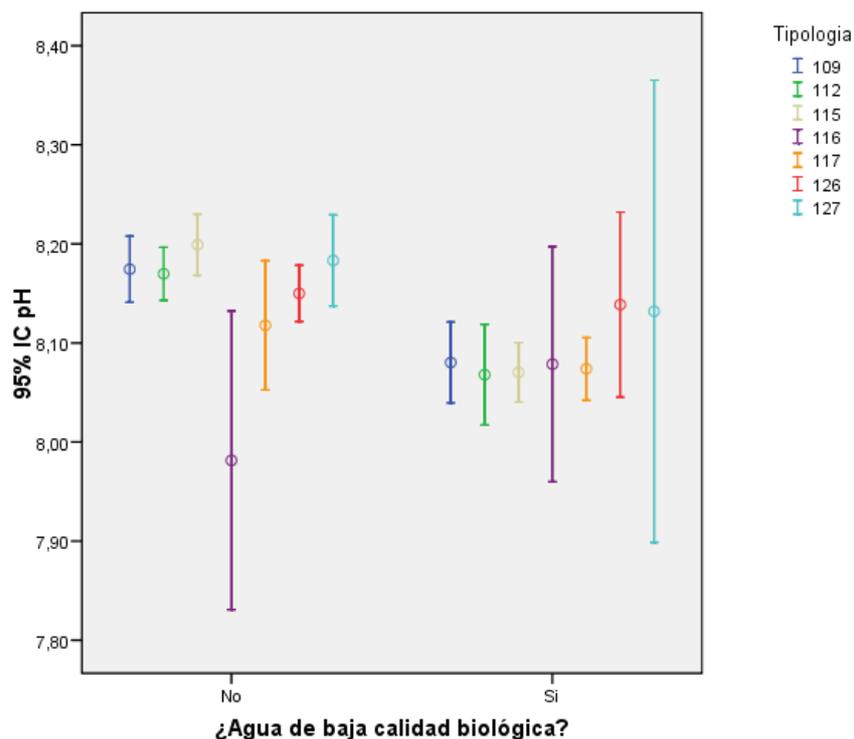
En la Tabla 5.1.1.1 se analiza el porcentaje de datos ausentes y válidos por tipologías y nivel de calidad biológica de las aguas. El porcentaje de datos ausentes oscila en torno a 16,5% llamando la atención el elevado porcentaje de datos ausentes en los ríos de la tipología 127 en las muestras de baja calidad (50%) así como el escaso número de muestras (8) disponibles de este tipo. Algo similar ocurre con los ríos de la tipología 116 de los que se dispone de un número escaso de observaciones. Todo ello hace que los resultados obtenidos respecto a estas dos tipologías hayan de tomarse con las debidas precauciones.

**Tabla 5.1.1.1: Datos ausentes y válidos del índice pH por tipologías y nivel de calidad de las aguas**

¿Aguas de baja calidad biológica?	Tipología	Válidos		Perdidos		Total	
		Casos	Porcentaje	Casos	Porcentaje	Casos	Porcentaje
No	109	116	78,4%	32	21,6%	148	100,0%
	112	266	85,3%	46	14,7%	312	100,0%
	115	108	93,1%	8	6,9%	116	100,0%
	116	7	100,0%	0	0,0%	7	100,0%
	117	23	95,8%	1	4,2%	24	100,0%
	126	298	83,9%	57	16,1%	355	100,0%
	127	119	75,8%	38	24,2%	157	100,0%
Si	109	129	80,6%	31	19,4%	160	100,0%
	112	116	78,9%	31	21,1%	147	100,0%
	115	138	88,5%	18	11,5%	156	100,0%
	116	14	87,5%	2	12,5%	16	100,0%
	117	90	92,8%	7	7,2%	97	100,0%
	126	42	76,4%	13	23,6%	55	100,0%
	127	8	50,0%	8	50,0%	16	100,0%
Total		1474	83,5%	292	16,5%	1766	100,0%



**Figura 5.1.1.1: Diagrama de cajas del índice pH por tipologías y nivel de calidad biológica de las aguas**



**Figura 5.1.1.2: Barras de error del índice pH medio por tipologías y nivel de calidad biológica de las aguas (nivel de confianza 95%)**

En las Figuras 5.1.1.1 y 5.1.1.2 así como en la Tabla 5.1.1.2 se presentan los resultados del estudio comparativo del índice pH entre las aguas de baja y alta calidad distinguiendo tipologías. Más concretamente se muestran los diagramas de cajas de dicho índice (Figura 5.1.1.1) y las barras de error del 95% de confianza para el índice pH medio (Figura 5.1.1.2) así como sus valores y los valores del índice pH mediano (Tabla 5.1.1.2). Finalmente, se dan los resultados obtenidos al realizar contrastes de Mann-Whitney presentando, para cada tipología y en total, los valores de los rangos medios de las muestras de alta y baja calidad así como el pvalor del contraste (Tabla 5.1.1.2).

Se observa la existencia de diferencias significativas en niveles medianos del índice pH del agua en las tipologías 109, 112 y 115 en las cuales dicho nivel tiende a ser significativamente superior en las aguas de alta calidad (ver Tabla 5.1.1.2). En el resto de las tipologías aunque, con la única excepción de la tipología 116, dicho patrón se conserva, no se aprecian diferencias que sean estadísticamente significativas.

**Tabla 5.1.1.2: Análisis descriptivo del índice pH por tipologías**

Tipología	Casos			Media			Mediana			Error típico de la media			Rangos promedio		Pvalor Mann-Whitney
	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		
No	Si	No		Si	No		Si	No		Si	No		Si	No	Si
<b>109</b>	116	129	245	8,17	8,08	8,12	8,20	8,10	8,17	0,02	0,02	0,01	139,33	108,32	0,001
<b>112</b>	266	116	382	8,17	8,07	8,14	8,20	8,15	8,20	0,01	0,03	0,01	203,68	163,58	0,001
<b>115</b>	108	138	246	8,20	8,07	8,13	8,20	8,10	8,10	0,02	0,02	0,01	151,99	101,20	0,000
<b>116</b>	7	14	21	7,98	8,08	8,05	8,10	8,14	8,10	0,06	0,05	0,04	9,00	12,00	0,322
<b>117</b>	23	88	111	8,12	8,07	8,08	8,10	8,10	8,10	0,03	0,02	0,01	62,80	54,22	0,252
<b>126</b>	298	42	340	8,15	8,14	8,15	8,20	8,20	8,20	0,01	0,05	0,01	170,54	170,24	0,985
<b>127</b>	119	8	127	8,18	8,13	8,18	8,20	8,21	8,20	0,02	0,10	0,02	64,48	56,81	0,567
<b>Total</b>	937	535	1472	8,17	8,08	8,13	8,20	8,10	8,19	0,01	0,01	0,01	805,74	615,24	0,000

**Tabla 5.1.1.3: Modelos estimados para el indicador pH distinguiendo y sin distinguir tipologías**

Variables seleccionadas	Distinguiendo tipologías			Sin distinguir tipologías		
	$\beta$	Error Estándar	Pvalor	$\beta$	Error Estándar	Pvalor
<b>Constante</b>	18,321	2,772	0,000	18,311	2,768	0,000
<b>pH</b>				-2,283	0,340	0,000
<b>pH_109</b>	-2,241	0,341	0,000			
<b>pH_112</b>	-2,358	0,342	0,000			
<b>pH_115</b>	-2,224	0,341	0,000			

**Tabla 5.1.1.4: Umbrales estimados para el indicador pH y validación predictiva de los mismos distinguiendo y sin distinguir tipologías  
(en negrita los umbrales utilizados en el proceso de comparación por tipologías descrito en la Tabla 5.1.1.5)**

Riesgo	Umbrales distinguiendo tipologías			%Aciertos			Umbrales sin distinguir tipología	%Aciertos		
	109	112	115	Cumplen	No cumplen	Totales		Cumplen	No cumplen	Totales
<b>25,00%</b>	8,67	8,23	8,73	22,00%	89,80%	51,80%	8,50	1,60%	99,20%	44,40%
<b>30,00%</b>	8,55	8,13	8,62	37,60%	83,80%	57,80%	8,39	13,30%	94,30%	48,80%
<b>35,00%</b>	8,45	8,03	8,52	44,70%	81,20%	60,70%	8,29	35,90%	80,40%	55,40%
<b>40,00%</b>	8,36	7,94	8,42	51,60%	74,20%	61,50%	8,20	59,80%	62,70%	61,10%
<b>45,00%</b>	<b>8,27</b>	<b>7,85</b>	<b>8,33</b>	<b>61,80%</b>	<b>66,10%</b>	<b>63,70%</b>	<b>8,11</b>	<b>68,40%</b>	<b>55,90%</b>	<b>62,90%</b>
<b>50,00%</b>	8,18	7,77	8,24	72,70%	54,60%	64,70%	8,02	82,00%	37,10%	62,30%
<b>55,00%</b>	8,09	7,68	8,15	83,50%	42,00%	65,30%	7,93	89,40%	25,30%	61,30%
<b>60,00%</b>	8,00	7,60	8,06	91,80%	37,70%	63,70%	7,84	93,70%	14,60%	59,00%
<b>65,00%</b>	7,90	7,51	7,96	95,10%	19,60%	62,00%	7,75	96,10%	8,60%	57,70%
<b>70,00%</b>	7,80	7,41	7,86	98,40%	8,10%	58,80%	7,65	98,00%	5,00%	57,20%
<b>75,00%</b>	7,69	7,30	7,74	99,60%	3,90%	57,60%	7,54	99,00%	2,90%	56,80%

**Tabla 5.1.1.5: Resultados obtenidos por tipologías utilizando los umbrales para el indicador pH mostrados en la Tabla 5.1.1.4 así como los umbrales proporcionados por la CHE**

Tipología	Calidad Observada	Calidad pronosticada						% Total aciertos		
		Distinguiendo tipología		Sin distinguir tipologías		CHE		Distinguiendo tipologías	Sin distinguir tipologías	CHE
		Alta	Baja	Alta	Baja	Alta	Baja			
109	Alta	66,38%	33,62%	37,93%	62,07%	100,00%	0,00%	59,18%	58,37%	47,35%
	Baja	47,29%	52,71%	23,26%	76,74%	100,00%	0,00%			
112	Alta	69,17%	30,83%	90,60%	9,40%	100,00%	0,00%	62,57%	69,11%	69,63%
	Baja	52,59%	47,41%	80,17%	19,83%	100,00%	0,00%			
115	Alta	68,52%	31,48%	16,67%	83,33%	100,00%	0,00%	67,07%	60,57%	43,90%
	Baja	34,06%	65,94%	5,07%	94,93%	100,00%	0,00%			
Total	Alta	68,37%	31,63%	61,84%	38,16%	100,00%	0,00%	62,89%	63,69%	56,13%
	Baja	44,13%	55,87%	33,94%	66,06%	100,00%	0,00%			

En la Tabla 5.1.1.3 se muestra el modelo seleccionado utilizando como variables independientes los indicadores de la tipología del río (denotados como TIP\_x donde x es el número de la tipología) y sus interacciones con el índice pH (denotados como pH\_x) y aplicando un procedimiento de selección de variables por eliminación hacia atrás tomando pvalor de entrada 0,01 y como pvalor de salida 0,02. Así mismo y, por razones comparativas, se muestra el modelo estimado utilizando como variables independientes una constante y el índice pH, el cual no tiene en cuenta, por tanto, la información proporcionada por la tipología del río. Los procesos de selección y estimación del modelo se han realizado utilizando, únicamente, las observaciones de las tipologías 109, 112 y 115 dado que, para el resto, los resultados obtenidos no fueron fiables debido al escaso tamaño muestral disponible en algunos de los grupos comparados.

Las estimaciones de los parámetros del modelo (ver Tabla 5.1.1.3) y del proceso de validación (ver Tabla 5.1.1.5) ponen de manifiesto que en los ríos de las tipologías 109, 112 y 115 el índice pH discrimina entre las muestras de agua de alta y baja calidad, siendo los coeficientes  $\beta$  estimados significativamente negativos.

Utilizando la expresión (4.2), el umbral para un riesgo  $0 < p < 1$  vendrá dado por la expresión:

$$\text{Umbral}_p = - \frac{\log\left(\frac{p}{1-p}\right) - 18,311}{2,283}$$

si no se distingue entre tipologías y

$$\text{Umbral}_p = - \frac{\log\left(\frac{p}{1-p}\right) - 18,321}{2,241 * I(\text{Tip} = 109) + 2,358 * I(\text{Tip} = 112) + 2,224 * I(\text{Tip} = 115)}$$

si se distinguen tipologías, donde  $I(\text{Tip}=x)$  es la función indicador de la tipología x que vale 1 si el río tiene tipología x y 0 en otro caso. En ambos casos se detectará riesgo alto de baja calidad biológica del agua si el índice pH es menor o igual que dicho umbral.

En la Tabla 5.1.1.4 se muestran los umbrales estimados para una red de valores del riesgo p, así como los porcentajes de éxito entre las muestras que cumplen los requisitos de calidad, entre las que no la cumplen y el porcentaje de éxitos totales calculados utilizando las expresiones (4.5) a (4.7). Así, por ejemplo, si el nivel mediano

del índice pH es inferior a 8,11 el modelo sin distinguir tipologías estima que existe un riesgo superior al 45% de que las aguas sean de baja calidad biológica. Distinguiendo por tipologías este umbral sube a 8,27 si el río es de la tipología 109; baja a 7,85 si es de tipología 112 y sube a 8,33 si es de la tipología 115. Utilizando estos umbrales los porcentajes de éxito para cada uno de los dos procedimientos fueron, respectivamente, 68,4% y 61,80% entre las muestras que cumplen, 55,9% y 66,10% entre las que no cumplen y 62,9% y 63,7% en total.

Finalmente, en la Tabla 5.1.1.5 se muestran los resultados obtenidos por tipologías utilizando los umbrales anteriores (distinguiendo y sin distinguir tipologías) junto a los correspondientes al procedimiento utilizado por la CHE que clasifica una muestra de agua como de buena calidad si el nivel de pH está entre 6,5 y 9. Así, por ejemplo, en la tipología 109, los porcentaje de éxito en la muestras que cumplen los estándares de calidad fueron 66,38%, 37,93% y 100% utilizando umbrales distinguiendo por tipologías, sin distinguir y el procedimiento de la CHE, respectivamente. Estos porcentajes fueron 52,71%, 76,74% y 0,00% entre las que no cumplen dichos estándares.

Se observa, en primer lugar, que los límites proporcionados por la CHE son poco operativos al clasificar todas las muestras como buenas. Por otro lado, aunque a nivel global no existen diferencias significativas en los resultados obtenidos en términos de porcentajes de aciertos (62,89% frente a 63,69%), si se analizan dichos resultados por tipologías se observa que los porcentajes de aciertos en las muestras de alta y baja calidad están más equilibrados cuando los umbrales se fijan distinguiendo tipologías que cuando no se distinguen, lo cual los hace más fiables al no saber, en una situación real, en qué situación están las aguas antes de ser analizadas.

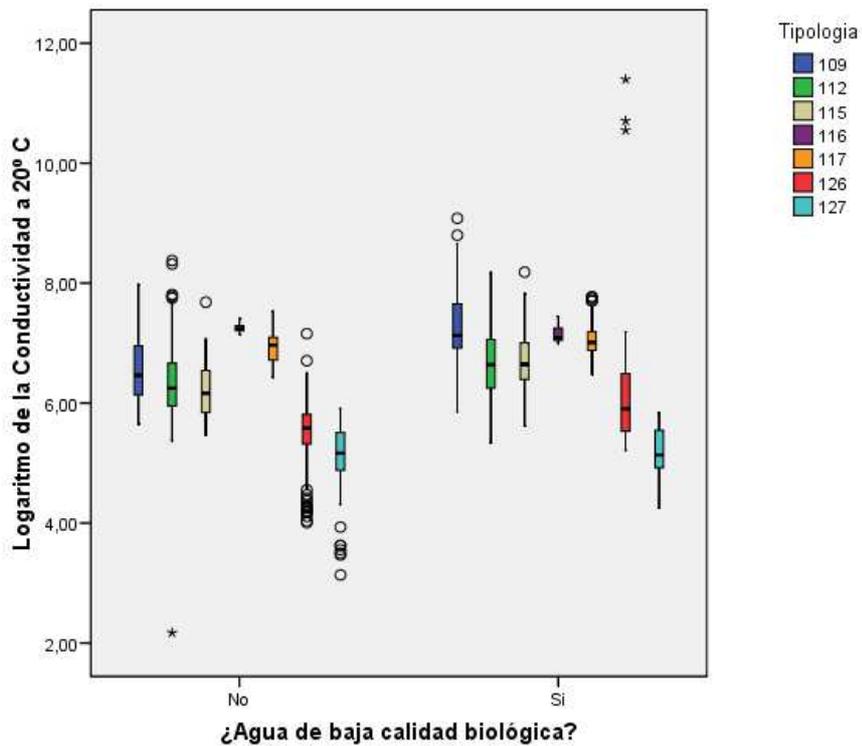
### 5.1.2. Umbrales de calidad biológica para la conductividad a 20° C

De acuerdo con estudios realizados en proyectos anteriores se ha tomado  $g(X) = \log(X)$  al mostrar el indicador analizado problemas serios de falta de normalidad por una elevada asimetría positiva.

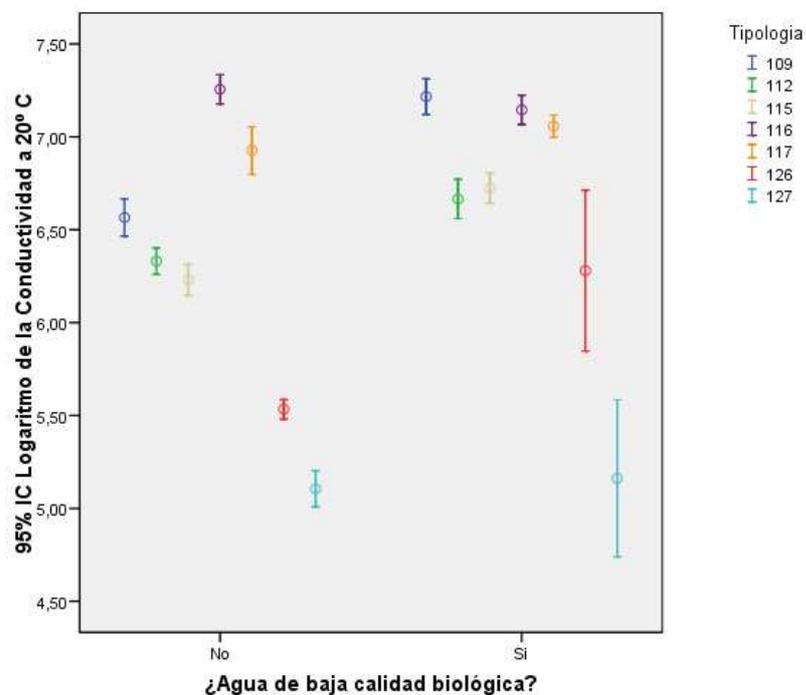
**Tabla 5.1.2.1: Datos ausentes y válidos de la conductividad a 20° C por tipologías y nivel de calidad de las aguas**

¿Aguas de baja calidad biológica?	Tipología	Válidos		Perdidos		Total	
		Casos	Porcentaje	Casos	Porcentaje	Casos	Porcentaje
No	109	116	78,4%	32	21,6%	148	100,0%
	112	266	85,3%	46	14,7%	312	100,0%
	115	108	93,1%	8	6,9%	116	100,0%
	116	7	100,0%	0	,0%	7	100,0%
	117	23	95,8%	1	4,2%	24	100,0%
	126	298	83,9%	57	16,1%	355	100,0%
	127	119	75,8%	38	24,2%	157	100,0%
Si	109	129	80,6%	31	19,4%	160	100,0%
	112	116	78,9%	31	21,1%	147	100,0%
	115	138	88,5%	18	11,5%	156	100,0%
	116	14	87,5%	2	12,5%	16	100,0%
	117	90	92,8%	7	7,2%	97	100,0%
	126	42	76,4%	13	23,6%	55	100,0%
	127	8	50,0%	8	50,0%	16	100,0%
<b>Total</b>		1474	83,47%	292	16,53%	1766	100,0%

En la Tabla 5.1.2.1 se analiza el porcentaje de datos ausentes y válidos por tipologías y nivel de calidad biológica de las aguas. El porcentaje de datos ausentes oscila en torno a 16,53% llamando la atención el elevado porcentaje de datos ausentes en los ríos de la tipología 127 en las muestras de baja calidad (50%) así como el escaso número de muestras (8) disponibles de este tipo. Algo similar ocurre con los ríos de la tipología 116 de los que se dispone de un número escaso de observaciones. Todo ello hace que los resultados obtenidos respecto a estas dos tipologías hayan de tomarse con las debidas precauciones.



**Figura 5.1.2.1: Diagrama de cajas del logaritmo de la conductividad a 20 ° C por tipologías y nivel de calidad biológica de las aguas**



**Figura 5.1.2.2: Barras de error de la media del logaritmo de la conductividad a 20° C por tipologías y nivel de calidad biológica de las aguas (nivel de confianza 95%)**

En las Figuras 5.1.2.1 y 5.1.2.2 así como en la Tabla 5.1.2.2 se presentan los resultados del estudio comparativo del logaritmo de la conductividad a 20° C entre las aguas de baja y alta calidad biológicas distinguiendo tipologías. Más concretamente, se muestran los diagramas de cajas del logaritmo del indicador (Figura 5.1.2.1) y las barras de error del 95% de confianza para sus valores medios (Figura 5.1.2.2) así como sus valores medios y medianos (Tabla 5.1.2.2). Finalmente se dan los resultados obtenidos al realizar contrastes de Mann-Whitney presentando, para cada tipología y en total, los valores de los rangos medios de las muestras de alta y baja calidad biológica así como el pvalor del contraste (Tabla 5.1.2.2).

Se observa la existencia de diferencias significativas en niveles medianos del logaritmo de la conductividad a 20° C del agua en las tipologías 109, 112, 115 y 126 en las cuales dicho nivel tiende a ser significativamente inferior en las aguas de alta calidad (ver Tabla 5.1.2.2). En el resto de las tipologías aunque, con la única excepción de la tipología 116, dicho patrón se conserva, no se aprecian diferencias que sean estadísticamente significativas.

**Tabla 5.1.2.2: Análisis descriptivo del logaritmo de la conductividad a 20° C por tipologías**

Tipología	Casos			Media			Mediana			Error típico de la media			Rangos promedio		Pvalor Mann-Whitney
	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		
No	Si	No		Si	No		Si	No		Si	No		Si	No	Si
<b>109</b>	116	129	245	6,57	7,22	6,91	6,46	7,13	6,95	0,05	0,05	0,04	84,25	157,84	0,000
<b>112</b>	266	116	382	6,33	6,67	6,43	6,25	6,64	6,34	0,04	0,05	0,03	171,22	238,00	0,000
<b>115</b>	108	138	246	6,23	6,72	6,51	6,16	6,65	6,49	0,04	0,04	0,03	85,90	152,92	0,000
<b>116</b>	7	14	21	7,26	7,14	7,18	7,24	7,09	7,16	0,03	0,04	0,03	14,71	9,14	0,056
<b>117</b>	23	90	113	6,93	7,06	7,03	6,97	7,01	7,01	0,06	0,03	0,03	45,72	59,88	0,064
<b>126</b>	298	42	340	5,53	6,28	5,63	5,58	5,91	5,60	0,03	0,21	0,04	161,43	234,88	0,000
<b>127</b>	119	8	127	5,11	5,16	5,11	5,17	5,13	5,17	0,05	0,18	0,05	63,91	65,31	0,917
<b>Total</b>	937	537	1474	5,96	6,84	6,28	5,91	6,88	6,27	0,02	0,03	0,02	566,11	1036,56	0,000

**Tabla 5.1.2.3: Modelos estimados para el logaritmo de la conductividad a 20° C distinguiendo y sin distinguir tipologías**

Variables seleccionadas	Distinguiendo tipologías			Sin distinguir tipologías		
	$\beta$	Error Estándar	Pvalor	$\beta$	Error Estándar	Pvalor
<b>Constante</b>	-14,7753	1,5403	0,0000	-10,6341	0,9373	0,0000
<b>LCOND20</b>				1,5728	0,1414	0,0000
<b>TIP_112</b>	7,3311	2,0543	0,0004			
<b>LCOND20_109</b>	2,1644	0,2248	0,0000			
<b>LCOND20_112</b>	1,0184	0,2070	0,0000			
<b>LCOND20_115</b>	2,3253	0,2400	0,0000			

**Tabla 5.1.2.4: Umbrales para la conductividad a 20 ° C y validación predictiva de los mismos distinguiendo y sin distinguir tipologías (en negrita señalados los umbrales utilizados en el proceso de comparación por tipologías descrito en la Tabla 5.1.2.5)**

Riesgo	Umbrales distinguiendo tipologías			% Aciertos			Umbrales sin distinguir tipologías	% Aciertos		
	109	112	115	Cumplen	No cumplen	Totales		Cumplen	No cumplen	Totales
<b>40,00%</b>	<b>764,457</b>	<b>1003,712</b>	<b>482,882</b>	<b>72,70%</b>	<b>71,80%</b>	<b>72,30%</b>	667,424	65,30%	70,80%	67,70%
<b>41,00%</b>	779,256	1045,457	491,578	73,50%	70,20%	72,10%	685,269	66,70%	69,20%	67,80%
<b>42,00%</b>	794,228	1088,609	500,363	75,30%	68,90%	72,50%	<b>703,453</b>	<b>68,00%</b>	<b>67,40%</b>	<b>67,70%</b>
<b>43,00%</b>	809,388	1133,242	509,247	77,10%	67,90%	73,10%	721,996	69,00%	65,50%	67,50%
<b>44,00%</b>	824,747	1179,436	518,236	79,80%	66,60%	73,40%	740,918	70,60%	64,50%	67,90%
<b>45,00%</b>	840,322	1227,273	527,340	79,60%	66,30%	73,80%	760,240	71,60%	62,40%	67,60%

**Tabla 5.1.2.5: Resultados obtenidos por tipologías utilizando los umbrales de la conductividad a 20° C de la Tabla 5.1.2.4 así como los umbrales proporcionados por la CHE**

Tipologia	Calidad Observada	Calidad pronosticada						% Total aciertos		
		Distinguiendo tipología		Sin distinguir tipologías		CHE		Distinguiendo tipologías	Sin distinguir tipologías	CHE
		Alta	Baja	Alta	Baja	Alta	Baja			
<b>109</b>	<b>Alta</b>	64,66%	35,34%	56,90%	43,10%	87,07%	12,93%	76,33%	74,29%	60,41%
	<b>Baja</b>	13,18%	86,82%	10,08%	89,92%	63,57%	36,43%			
<b>112</b>	<b>Alta</b>	84,21%	15,79%	69,55%	30,45%	95,11%	4,89%	69,11%	65,18%	70,94%
	<b>Baja</b>	65,52%	34,48%	44,83%	55,17%	84,48%	15,52%			
<b>115</b>	<b>Alta</b>	52,78%	47,22%	75,93%	24,07%	99,07%	0,93%	73,17%	65,04%	51,63%
	<b>Baja</b>	10,87%	89,13%	43,48%	56,52%	85,51%	14,49%			
<b>Total</b>	<b>Alta</b>	72,65%	27,35%	67,96%	32,04%	94,08%	5,92%	72,28%	67,70%	62,54%
	<b>Baja</b>	28,20%	71,80%	32,64%	67,36%	77,81%	22,19%			

En la Tabla 5.1.2.3 se muestra el modelo seleccionado utilizando como variables independientes los indicadores de la tipología del río (denotados como TIP\_x donde x es el número de la tipología) y sus interacciones con el logaritmo de la conductividad a 20° C (denotados como LCOND20\_x) y aplicando un procedimiento de selección de variables por eliminación hacia atrás tomando pvalor de entrada 0,01 y como pvalor de salida 0,02. Así mismo y, por razones comparativas, se muestra el modelo estimado utilizando como variables independientes una constante y dicho logaritmo, el cual no tiene en cuenta, por tanto, la información proporcionada por la tipología del río. Las estimaciones se han realizado utilizando, únicamente, las observaciones de las tipologías 109, 112 y 115 dado que, para el resto, los resultados obtenidos no fueron fiables debido al escaso tamaño muestral disponible en algunos de los grupos comparados.

Las estimaciones de los parámetros del modelo (ver Tabla 5.1.2.3) y del proceso de validación (ver Tabla 5.1.2.5) ponen de manifiesto que en los ríos de las tipologías 109, 112 y 115 el índice de conductividad a 20° discrimina entre las muestras de agua de alta y baja calidad, siendo los coeficientes  $\beta$  estimados significativamente positivo.

Utilizando la expresión (4.2), el umbral para un riesgo  $0 < p < 1$  vendrá dado por la expresión:

$$\text{Umbral}_p = \exp \left( \frac{\log \left( \frac{p}{1-p} \right) + 10,63}{1,57} \right)$$

si no se distingue entre tipologías y

$$\text{Umbral}_p = \exp \left( \frac{\log \left( \frac{p}{1-p} \right) + 14,78 - 7,33 * I(\text{Tip} = 112)}{2,16 * I(\text{Tip} = 109) + 1,02 * I(\text{Tip} = 112) + 2,33 * I(\text{Tip} = 115)} \right)$$

si se distinguen tipologías, donde  $I(\text{Tip}=x)$  es la función indicador de la tipología x que vale 1 si el río tiene tipología x y 0 en otro caso. En ambos casos se detectará riesgo alto de baja calidad biológica del agua si el nivel de conductividad a 20° C es mayor o igual que dicho umbral.

En la Tabla 5.1.2.4 se muestran los umbrales estimados para una red de valores del riesgo  $p$ , así como los porcentajes de éxito entre las muestras que cumplen los requisitos de calidad, entre las que no la cumplen y el porcentaje de éxitos totales calculados utilizando las expresiones (4.5) a (4.7). Así, por ejemplo, si el nivel mediano de la conductividad es superior a 667,42  $\mu\text{S}/\text{cm}$  el modelo estimado sin distinguir tipologías estima que existe un riesgo superior al 40% de que las aguas sean de baja calidad biológica. Distinguiendo por tipologías este umbral sube a 764,46  $\mu\text{S}/\text{cm}$  si el río es de la tipología 109; 1003,71  $\mu\text{S}/\text{cm}$  si es de tipología 112 y baja a 482,88  $\mu\text{S}/\text{cm}$  si es de la tipología 115. Utilizando estos umbrales los porcentajes de éxito para cada uno de los dos procedimientos fueron, respectivamente, 65,3% y 72,7% entre las muestras que cumplen, 70,8% y 71,8% entre las que no cumplen y 67,7% y 72,3% en total.

Finalmente, en la Tabla 5.1.2.5 se muestran los resultados obtenidos por tipologías utilizando los umbrales señalados en negrita (distinguiendo y sin distinguir tipologías) junto a los correspondientes al procedimiento utilizado por la CHE el cual clasifica el agua como de alta calidad si su nivel de conductividad es inferior a 1500  $\mu\text{S}/\text{cm}$  en las 3 tipologías analizadas. Así, por ejemplo, en la tipología 109, los porcentajes de éxito en la muestras que cumplen los estándares de calidad fueron 64,66%, 56,90% y 87,07% utilizando umbrales distinguiendo por tipologías, sin distinguir y el procedimiento de la CHE, respectivamente. Estos porcentajes fueron 86,82%, 89,92% y 36,43% entre las que no cumplen dichos estándares.

Se observa, en primer lugar, que los límites proporcionados por la CHE son poco operativos al clasificar un 77,81% aguas de baja calidad como buenas. Por otro lado, aunque a nivel global el procedimiento que distingue entre tipologías obtiene un mayor porcentaje de éxitos global que el que no distingue (72,28% frente a 67,7%), si analizamos los porcentajes de éxitos por tipologías, se observa que los porcentajes de aciertos en las muestras de alta y baja calidad están más equilibrados cuando los umbrales se fijan sin distinguir tipologías que cuando se distinguen, lo cual los hace más fiables al no saber, en una situación real, en qué situación están las aguas antes de ser analizadas.

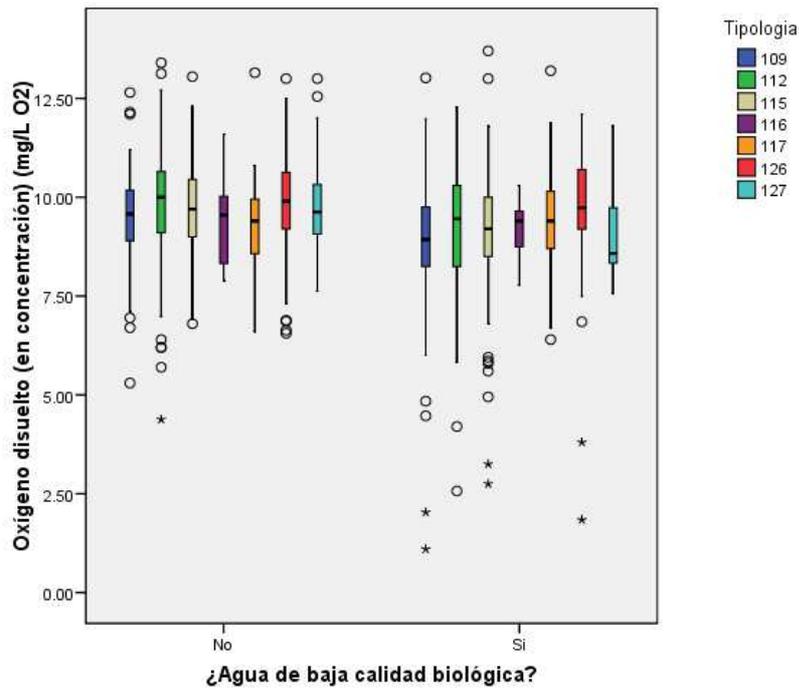
### 5.1.3. Umbrales de calidad biológica para el oxígeno disuelto (en concentración)

De acuerdo con estudios realizados en proyectos anteriores se ha tomado  $g(X) = X$  al no mostrar el indicador analizado problemas serios de falta de normalidad.

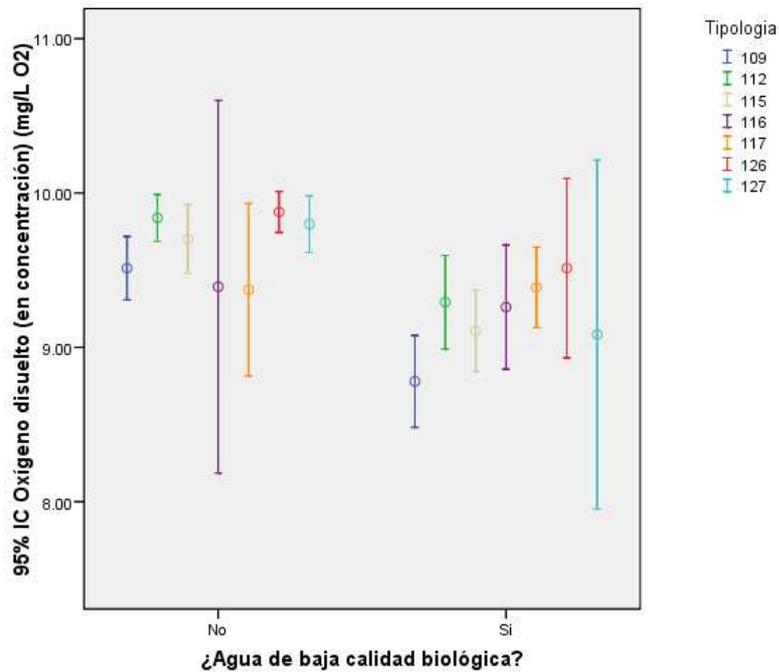
**Tabla 5.1.3.1: Datos ausentes y válidos del oxígeno disuelto por tipologías y nivel de calidad de las aguas**

¿Aguas de baja calidad biológica?	Tipología	Válidos		Perdidos		Total	
		Casos	Porcentaje	Casos	Porcentaje	Casos	Porcentaje
No	109	112	75,7%	36	24,3%	148	100,0%
	112	261	83,7%	51	16,3%	312	100,0%
	115	107	92,2%	9	7,8%	116	100,0%
	116	7	100,0%	0	,0%	7	100,0%
	117	23	95,8%	1	4,2%	24	100,0%
	126	295	83,1%	60	16,9%	355	100,0%
	127	119	75,8%	38	24,2%	157	100,0%
Si	109	127	79,4%	33	20,6%	160	100,0%
	112	112	76,2%	35	23,8%	147	100,0%
	115	134	85,9%	22	14,1%	156	100,0%
	116	14	87,5%	2	12,5%	16	100,0%
	117	90	92,8%	7	7,2%	97	100,0%
	126	42	76,4%	13	23,6%	55	100,0%
	127	8	50,0%	8	50,0%	16	100,0%
<b>Total</b>		1451	82,2%	315	17,8%	1766	100,0%

En la Tabla 5.1.3.1. se analiza el porcentaje de datos ausentes y válidos por tipologías y nivel de calidad biológica de las aguas. El porcentaje de datos ausentes oscila en torno a 17,8% llamando la atención el elevado porcentaje de datos ausentes en los ríos de la tipología 127 en las muestras de baja calidad (50%) lo cual, unido al escaso número de muestras (8) de este tipo hace que los resultados correspondientes a esta tipología deban tomarse con las debidas precauciones.



**Figura 5.1.3.1: Diagrama de cajas del oxígeno disuelto (en concentración) por tipologías y nivel de calidad biológica de las aguas**



**Figura 5.1.3.2: Barras de error de la media del oxígeno disuelto por tipologías y nivel de calidad biológica de las aguas (nivel de confianza 95%)**

En las Figuras 5.1.3.1 y 5.1.3.2 así como en la Tabla 5.1.3.2 se presentan los resultados del estudio comparativo del nivel de oxígeno disuelto (en concentración) entre las aguas de baja y alta calidad biológicas distinguiendo por tipologías. Más concretamente, se muestran los diagramas de cajas del nivel de oxígeno disuelto (Figura 5.1.3.1) y las barras de error del 95% de confianza para sus valores medios (Figura 5.1.3.2) así como sus valores medios y medianos (Tabla 5.1.3.2). Finalmente se dan los resultados obtenidos al realizar contrastes de Mann-Whitney presentando, para cada tipología y en total, los valores de los rangos medios de las muestras de alta y baja calidad biológica así como el pvalor del contraste (Tabla 5.1.3.2).

Se observa la existencia de diferencias significativas en los niveles medianos del oxígeno disuelto del agua en las tipologías 109, 112, 115 y 127 en las cuales dicho nivel tiende a ser significativamente superior en las aguas de alta calidad (ver Tabla 5.1.3.2). En el resto de las tipologías aunque, con la única excepción de la tipología 117, dicho patrón se conserva, no se aprecian diferencias que sean estadísticamente significativas.

**Tabla 5.1.3.2: Análisis descriptivo del oxígeno disuelto por tipologías**

Tipología	Casos			Media			Mediana			Error típico de la media			Rangos promedio		Pvalor Mann-Whitney
	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		
No	Si	No		Si	No		Si	No		Si	No		Si	No	Si
<b>109</b>	112	127	239	9,51	8,78	9,12	9,58	8,93	9,25	0,10	0,15	0,10	139,25	103,03	0,000
<b>112</b>	261	112	373	9,84	9,29	9,67	10,00	9,46	9,88	0,08	0,15	0,07	197,87	161,67	0,003
<b>115</b>	107	134	241	9,70	9,11	9,37	9,70	9,20	9,40	0,11	0,13	0,09	137,25	108,02	0,001
<b>116</b>	7	14	21	9,39	9,26	9,31	9,55	9,40	9,40	0,49	0,19	0,20	11,14	10,93	0,971
<b>117</b>	23	90	113	9,37	9,39	9,39	9,40	9,40	9,40	0,27	0,13	0,12	55,85	57,29	0,850
<b>126</b>	295	42	337	9,88	9,51	9,83	9,90	9,74	9,90	0,07	0,29	0,07	170,05	161,65	0,601
<b>127</b>	119	8	127	9,80	9,08	9,75	9,63	8,58	9,60	0,09	0,48	0,09	65,76	37,88	0,038
<b>Total</b>	924	527	1451	9,77	9,15	9,55	9,80	9,30	9,60	0,04	0,07	0,04	790,95	612,13	0,000

**Tabla 5.1.3.3: Modelos estimados para el nivel de oxígeno disuelto distinguiendo y sin distinguir tipologías**

Variables seleccionadas	Distinguiendo tipologías			Sin distinguir tipologías		
	$\beta$	Error Estándar	Pvalor	$\beta$	Error Estándar	Pvalor
<b>Constante</b>	2,8290	0,5256	0,0000	3,1206	0,5216	0,0000
<b>O2</b>				-0,3582	0,0549	0,0000
<b>O2_109</b>	-0,2969	0,0577	0,0000			
<b>O2_112</b>	-0,3824	0,0557	0,0000			
<b>O2_115</b>	-0,2777	0,0563	0,0000			

**Tabla 5.1.3.4: Umbrales del oxígeno disuelto y validación predictiva de los mismos distinguiendo y sin distinguir tipologías (en negrita señalados los umbrales utilizados en el proceso de comparación por tipologías descrito en la Tabla 5.1.3.5)**

Riesgo	Umbrales distinguiendo tipologías			% Aciertos			Umbrales sin distinguir tipologías	% Aciertos		
	109	112	115	Cumplen	No cumplen	Totales		Cumplen	No cumplen	Totales
<b>40,00%</b>	10,90	8,46	11,65	50,20%	75,30%	61,20%	9,84	48,10%	69,20%	57,30%
<b>41,00%</b>	10,76	8,35	11,50	51,30%	73,50%	61,00%	9,73	53,30%	66,80%	59,20%
<b>42,00%</b>	10,62	8,24	11,35	52,90%	72,70%	61,50%	9,61	56,30%	64,10%	59,70%
<b>43,00%</b>	10,48	8,14	11,20	54,60%	70,80%	61,70%	<b>9,50</b>	<b>61,70%</b>	<b>58,70%</b>	<b>60,40%</b>
<b>44,00%</b>	10,34	8,03	11,06	56,50%	68,40%	61,70%	9,39	64,40%	55,50%	60,50%
<b>45,00%</b>	10,21	7,92	10,91	58,50%	67,60%	62,50%	9,27	67,50%	53,10%	61,20%
<b>46,00%</b>	10,07	7,82	10,77	61,70%	66,50%	63,80%	9,16	71,30%	49,60%	61,80%
<b>47,00%</b>	<b>9,93</b>	<b>7,71</b>	<b>10,62</b>	<b>63,30%</b>	<b>63,50%</b>	<b>63,40%</b>	9,05	74,80%	46,40%	62,40%
<b>48,00%</b>	9,80	7,61	10,48	67,10%	60,30%	64,10%	8,94	77,70%	42,40%	62,30%
<b>49,00%</b>	9,66	7,50	10,33	70,80%	57,10%	64,80%	8,82	80,20%	38,60%	62,00%
<b>50,00%</b>	9,53	7,40	10,19	72,90%	54,70%	64,90%	8,71	82,30%	35,40%	61,80%

**Tabla 5.1.3.5: Resultados obtenidos por tipologías utilizando los umbrales para el oxígeno disuelto de la Tabla 5.1.3.4 así como los umbrales proporcionados por la CHE**

Tipología	Calidad Observada	Calidad pronosticada						% Total aciertos		
		Distinguiendo tipología		Sin distinguir tipologías		CHE		Distinguiendo tipologías	Sin distinguir tipologías	CHE
		Alta	Baja	Alta	Baja	Alta	Baja			
109	Alta	32,14%	67,86%	55,36%	44,64%	98,21%	1,79%	58,58%	61,51%	51,46%
	Baja	18,11%	81,89%	33,07%	66,93%	89,76%	10,24%			
112	Alta	95,79%	4,21%	66,28%	33,72%	97,32%	2,68%	70,78%	61,39%	71,05%
	Baja	87,50%	12,50%	50,00%	50,00%	90,18%	9,82%			
115	Alta	16,82%	83,18%	57,01%	42,99%	95,33%	4,67%	56,85%	57,68%	46,89%
	Baja	11,19%	88,81%	41,79%	58,21%	91,79%	8,21%			
Total	Alta	63,33%	36,67%	61,67%	38,33%	97,08%	2,92%	63,42%	60,38%	58,73%
	Baja	36,46%	63,54%	41,29%	58,71%	90,62%	9,38%			

En la Tabla 5.1.3.3 se muestra el modelo seleccionado utilizando como variables independientes los indicadores de la tipología del río (denotados como TIP\_x donde x es el número de la tipología) y sus interacciones con el nivel de oxígeno disuelto (denotados como O2\_x) y aplicando un procedimiento de selección de variables por eliminación hacia atrás tomando pvalor de entrada 0,01 y como pvalor de salida 0,02. Así mismo y, por razones comparativas, se muestra el modelo estimado utilizando como variables independientes una constante y dicho indicador, el cual no tiene en cuenta, por tanto, la información proporcionada por la tipología del río. Las estimaciones se han realizado utilizando, únicamente, las observaciones de las tipologías 109, 112 y 115 dado que, para el resto, los resultados obtenidos no fueron fiables debido al escaso tamaño muestral disponible en algunos de los grupos comparados.

Las estimaciones de los parámetros del modelo (ver Tabla 5.1.3.3) y del proceso de validación (ver Tabla 5.1.3.5) ponen de manifiesto que en los ríos de las tipologías 109, 112 y 115 el nivel de oxígeno disuelto en el agua discrimina entre las muestras de agua de alta y baja calidad, siendo los coeficiente  $\beta$  estimados significativamente negativos.

Utilizando la expresión (4.2), el umbral para un riesgo  $0 < p < 1$  vendrá dado por la expresión:

$$\text{Umbral}_p = - \frac{\log\left(\frac{p}{1-p}\right) - 3,12}{0,36}$$

si no se distingue entre tipologías y

$$\text{Umbral}_p = - \frac{\log\left(\frac{p}{1-p}\right) - 2,83}{0,30 * I(\text{Tip} = 109) + 0,38 * I(\text{Tip} = 112) + 0,28 * I(\text{Tip} = 115)}$$

si se distinguen tipologías, donde  $I(\text{Tip}=x)$  es la función indicador de la tipología x que vale 1 si el río tiene tipología x y 0 en otro caso. En ambos casos se detectará riesgo alto de baja calidad biológica del agua si el nivel de oxígeno disuelto en el agua es menor o igual que dicho umbral.

En la Tabla 5.1.3.4 se muestran los umbrales estimados para una red de valores del riesgo p, así como los porcentajes de éxito entre las muestras que cumplen los requisitos de calidad, entre las que no la cumplen y el porcentaje de éxitos totales

calculados utilizando las expresiones (4.5) a (4.7). Así, por ejemplo, si el nivel de oxígeno disuelto en el agua es inferior a 9,05 mg/L el modelo estimado sin distinguir tipologías estima que existe un riesgo superior al 47% de que las aguas sean de baja calidad biológica. Distinguiendo por tipologías este umbral cambia a 9,93 mg/L si el río es de la tipología 109; 7,71 mg/L si es de tipología 112 y 10,62 mg/L si es de la tipología 115. Utilizando estos umbrales los porcentajes de éxito para cada uno de los dos procedimientos fueron, respectivamente, 74,8% y 63,3% entre las muestras que cumplen, 46,4% y 63,5% entre las que no cumplen y 62,4% y 63,4% en total.

Finalmente, en la Tabla 5.1.3.5 se muestran los resultados obtenidos por tipologías utilizando los umbrales en negrita (distinguiendo y sin distinguir tipologías) junto a los correspondientes al procedimiento utilizado por la CHE, el cual clasifica el agua como de alta calidad si su nivel de oxígeno disuelto en agua es superior a 6,7 mg/L en la tipología 109 y 7,2 mg/L en las tipologías 112 y 115. Así, por ejemplo, en la tipología 109, los porcentaje de éxito en la muestras que cumplen los estándares de calidad fueron 32,14%, 55,36% y 98,21% utilizando umbrales distinguiendo por tipologías, sin distinguir y el procedimiento de la CHE, respectivamente. Estos porcentajes fueron 81,89%, 66,93% y 10,24% entre las que no cumplen dichos estándares.

Se observa, en primer lugar, que los límites proporcionados por la CHE son poco operativos al clasificar un 90,62% aguas de baja calidad como buenas. Por otro lado, aunque a nivel global se obtienen resultados ligeramente mejores, en términos de porcentajes de aciertos, distinguiendo tipologías que sin distinguir (63,42% frente a 60,38%), si analizamos los resultados por tipologías, se observa que los porcentajes de aciertos en las muestras de alta y baja calidad están más equilibrados cuando los umbrales se fijan sin distinguir tipologías que cuando se distinguen, lo cual los hace más fiables al no saber, en una situación real, en qué situación están las aguas antes de ser analizadas.

## 5.2. UMBRALES DE CALIDAD PARA LOS INDICADORES DE MATERIA ORGÁNICA

En este apartado presentamos los umbrales de calidad biológica para el indicador de materia orgánica DQO cuyo significado puede verse en la Tabla 2.3.

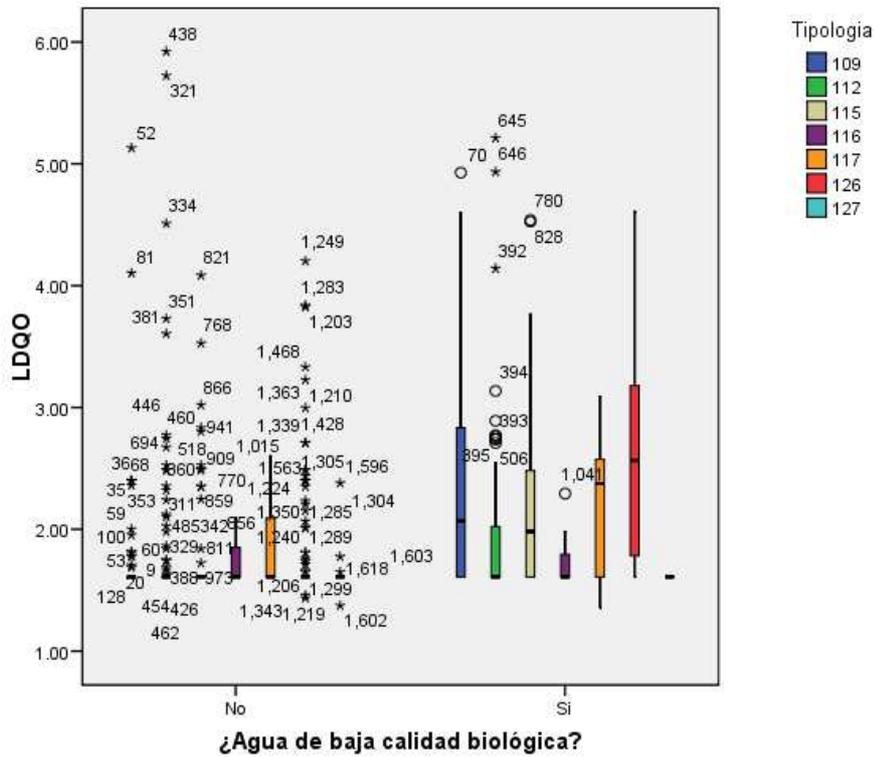
### 5.2.1. Umbrales de calidad biológica para la demanda química de oxígeno

De acuerdo con estudios realizados en proyectos anteriores se ha tomado  $g(X) = \log(X)$  al mostrar el indicador analizado problemas serios de falta de normalidad por una elevada asimetría positiva.

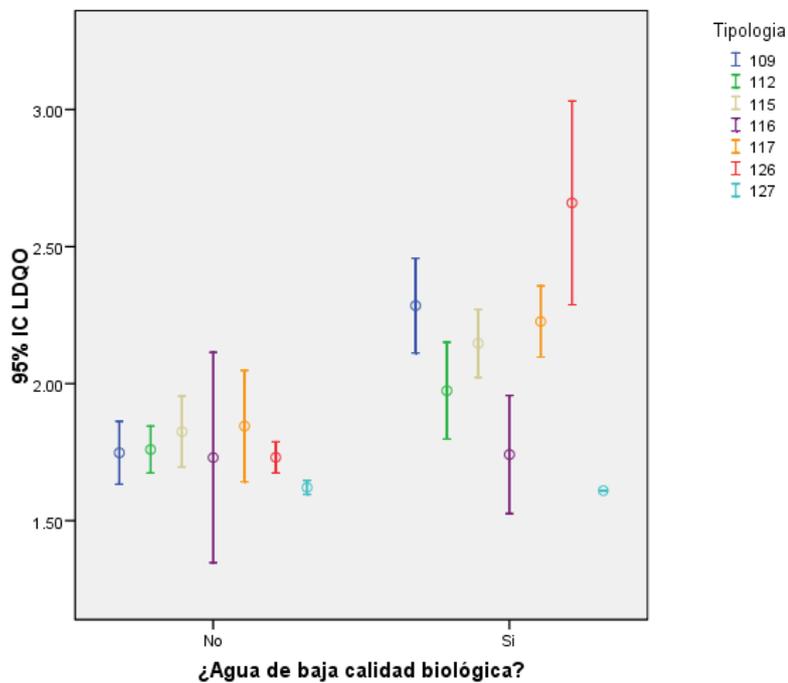
**Tabla 5.2.1.1: Datos ausentes y válidos del logaritmo de la demanda química de oxígeno por tipologías y nivel de calidad biológica de las aguas**

¿Aguas de baja calidad biológica?	Tipología	Válidos		Perdidos		Total	
		Casos	Porcentaje	Casos	Porcentaje	Casos	Porcentaje
No	109	78	52,7%	70	47,3%	148	100,0%
	112	175	56,1%	137	43,9%	312	100,0%
	115	62	53,4%	54	46,6%	116	100,0%
	116	4	57,1%	3	42,9%	7	100,0%
	117	15	62,5%	9	37,5%	24	100,0%
	126	191	53,8%	164	46,2%	355	100,0%
	127	64	40,8%	93	59,2%	157	100,0%
Si	109	86	53,8%	74	46,3%	160	100,0%
	112	69	46,9%	78	53,1%	147	100,0%
	115	100	64,1%	56	35,9%	156	100,0%
	116	8	50,0%	8	50,0%	16	100,0%
	117	57	58,8%	40	41,2%	97	100,0%
	126	25	45,5%	30	54,5%	55	100,0%
	127	2	12,5%	14	87,5%	16	100,0%
<b>Total</b>		936	53,0%	830	47,0%	1766	100,0%

En la Tabla 5.2.1.1 se analiza el porcentaje de datos ausentes y válidos por tipologías y nivel de calidad biológica de las aguas. El porcentaje de datos ausentes es muy elevado oscilando en torno a un 47%, siendo incluso mayor en los ríos de la tipología 127 en las muestras de baja calidad biológica (87,5% y un tamaño muestral igual a 2). Se aprecia, además, un número escaso de observaciones disponibles en la tipología 116. Todo ello hace que los resultados en general y, particularmente, los correspondientes a estas dos tipologías deban tomarse con las debidas precauciones.



**Figura 5.2.1.1: Diagrama de cajas del logaritmo de la demanda química de oxígeno por tipologías y nivel de calidad biológica de las aguas**



**Figura 5.2.1.2: Barras de error de la media del logaritmo de la demanda química media de oxígeno por tipologías y nivel de calidad biológica de las aguas (nivel de confianza 95%)**

En las Figuras 5.2.1.1 y 5.2.1.2 así como en la Tabla 5.2.1.2 se presentan los resultados del estudio comparativo del logaritmo de la demanda química de oxígeno entre las aguas de baja y alta calidad biológica distinguiendo tipologías. Más concretamente se muestran sus diagramas de cajas (Figura 5.2.1.1) y las barras de error del 95% de confianza (Figura 5.2.1.2) así como sus valores medios y medianos (Tabla 5.2.1.2). Finalmente, se dan los resultados obtenidos al realizar contrastes de Mann-Whitney presentando, para cada tipología y en total, los valores de los rangos medios de las muestras de alta y baja calidad biológicas así como el pvalor del contraste (Tabla 5.2.1.2).

Se observa que, con la única excepción de las tipologías 116 y 127, la mediana del logaritmo de la demanda química de oxígeno tiende a ser significativamente inferior en las aguas de alta calidad biológica (ver Tabla 5.2.1.2 y Figura 5.2.1.2).

**Tabla 5.2.1.2: Análisis descriptivo del logaritmo de la demanda química de oxígeno por tipologías**

Tipología	Casos			Media			Mediana			Error típico de la media			Rangos promedio		Pvalor Mann-Whitney
	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		
	No	Si		No	Si		No	Si		No	Si		No	Si	
<b>109</b>	78	86	164	1,75	2,28	2,03	1,61	2,07	1,61	0,06	0,09	0,06	64,82	98,53	0,000
<b>112</b>	175	69	244	1,76	1,97	1,82	1,61	1,61	1,61	0,04	0,09	0,04	115,20	141,02	0,000
<b>115</b>	62	100	162	1,83	2,15	2,02	1,61	1,98	1,61	0,06	0,06	0,05	62,45	93,31	0,000
<b>116</b>	4	8	12	1,73	1,74	1,74	1,61	1,61	1,61	0,12	0,09	0,07	6,50	6,50	1,000
<b>117</b>	15	57	72	1,85	2,23	2,15	1,61	2,37	2,22	0,09	0,06	0,06	23,90	39,82	0,007
<b>126</b>	191	25	216	1,73	2,66	1,84	1,61	2,56	1,61	0,03	0,18	0,04	99,30	178,80	0,000
<b>127</b>	64	2	66	1,62	1,61	1,62	1,61	1,61	1,61	0,01	0,00	0,01	33,53	32,50	0,955
<b>Total</b>	589	347	936	1,74	2,18	1,91	1,61	1,95	1,61	0,02	0,04	0,02	394,14	594,72	0,000

**Tabla 5.2.1.3: Modelos estimados para el logaritmo de la demanda química de oxígeno distinguiendo y sin distinguir tipologías**

Variables seleccionadas	Distinguiendo tipologías			Sin distinguir tipologías		
	$\beta$	Error Estándar	Pvalor	$\beta$	Error Estándar	Pvalor
<b>Constante</b>	-5,4534	0,8473	0,0000	-3,1166	0,3064	0,0000
<b>LDQO</b>				1,3986	0,1587	0,0000
<b>TIP_112</b>	4,7009	0,8532	0,0000			
<b>TIP_115</b>	5,7497	0,8561	0,0000			
<b>TIP_117</b>	6,8501	0,8774	0,0000			
<b>LDQO_109</b>	2,3159	0,3651	0,0000			
<b>LDQO_126</b>	1,4222	0,3353	0,0000			

**Tabla 5.2.1.4: Umbrales estimados para la demanda química de oxígeno y validación predictiva distinguiendo y sin distinguir tipologías (en negrita señalados los umbrales utilizados en el proceso de comparación por tipologías descrito en la Tabla 5.2.1.5)**

Riesgo	Umbrales distinguiendo tipologías		%Aciertos			Umbrales sin distinguir tipologías	%Aciertos		
	109	126	Cumplen	No cumplen	Totales		Cumplen	No cumplen	Totales
<b>26,00%</b>	6,71	22,18	0,40%	99,70%	39,40%	4,40	0,40%	99,70%	39,40%
<b>28,00%</b>	7,01	23,82	27,00%	96,10%	55,70%	4,73	0,40%	99,70%	39,40%
<b>30,00%</b>	7,31	25,50	27,30%	95,90%	55,80%	<b>5,07</b>	<b>83,70%</b>	<b>57,00%</b>	<b>73,20%</b>
<b>32,00%</b>	7,61	27,23	27,50%	95,30%	55,60%	5,42	84,80%	56,10%	73,50%
<b>34,00%</b>	<b>7,91</b>	<b>29,02</b>	<b>70,90%</b>	<b>66,30%</b>	<b>69,90%</b>	5,78	86,00%	55,80%	74,10%

**Tabla 5.2.1.5: Resultados obtenidos por tipologías utilizando los umbrales estimados para la demanda química de oxígeno de la Tabla 5.2.1.4 así como los umbrales proporcionados por la CHE**

		Calidad pronosticada						% Total aciertos		
		Distinguiendo tipología		Sin distinguir tipologías		CHE				
		Alta	Baja	Alta	Baja	Alta	Baja	Distinguiendo tipologías	Sin distinguir tipologías	CHE
<b>109</b>	<b>Alta</b>	14,10%	85,90%	82,05%	17,95%	97,44%	2,56%	56,71%	67,07%	60,98%
	<b>Baja</b>	4,65%	95,35%	46,51%	53,49%	72,09%	27,91%			
<b>112</b>	<b>Alta</b>	100,00%	0,00%	85,71%	14,29%	96,00%	4,00%	71,72%	71,31%	73,36%
	<b>Baja</b>	100,00%	0,00%	65,22%	34,78%	84,06%	15,94%			
<b>115</b>	<b>Alta</b>	0,00%	100,00%	79,03%	20,97%	91,94%	8,06%	61,73%	67,90%	42,59%
	<b>Baja</b>	0,00%	100,00%	39,00%	61,00%	88,00%	12,00%			
<b>117</b>	<b>Alta</b>	0,00%	100,00%	66,67%	33,33%	100,00%	0,00%	79,17%	69,44%	36,11%
	<b>Baja</b>	0,00%	100,00%	29,82%	70,18%	80,70%	19,30%			
<b>126</b>	<b>Alta</b>	98,43%	1,57%	85,34%	14,66%	95,81%	4,19%	88,89%	85,19%	89,35%
	<b>Baja</b>	84,00%	16,00%	16,00%	84,00%	60,00%	40,00%			
<b>Total</b>	<b>Alta</b>	71,79%	28,21%	83,69%	16,31%	95,78%	4,22%	71,91%	73,19%	66,08%
	<b>Baja</b>	27,89%	72,11%	43,03%	56,97%	79,82%	20,18%			

En la Tabla 5.2.1.3 se muestra el modelo seleccionado utilizando como variables independientes los indicadores de la tipología del río (denotados como TIP\_x donde x es el número de la tipología) y sus interacciones con el logaritmo de la demanda química de oxígeno (denotados como LDQO\_x) y aplicando un procedimiento de selección de variables por eliminación hacia atrás tomando pvalor de entrada 0,01 y como pvalor de salida 0,02. Así mismo y, por razones comparativas, se muestra el modelo estimado utilizando como variables independientes una constante y dicho logaritmo el cual no tiene en cuenta, por tanto, la información proporcionada por la tipología del río. Las estimaciones se han realizado utilizando las observaciones de las tipologías 109, 112, 115, 117 y 126.

Las estimaciones de los parámetros del modelo (ver Tabla 5.2.1.3) y del proceso de validación (ver Tabla 5.2.1.5) ponen de manifiesto que en los ríos de las tipologías 109, 112, 115, 117 y 126 la demanda química de oxígeno discrimina significativamente entre las muestras de agua de alta y baja calidad biológicas, siendo los coeficientes  $\beta$  estimados significativamente positivos.

Utilizando la expresión (4.2), el umbral para un riesgo  $0 < p < 1$  vendrá dado por la expresión:

$$\text{Umbral}_p = \exp \left( \frac{\log \left( \frac{p}{1-p} \right) + 3,12}{1,4} \right)$$

si no se distinguen tipologías y

$$\text{Umbral}_p = \exp \left( \frac{\log \left( \frac{p}{1-p} \right) + 5,45 - 4,7 * I(\text{Tip} = 112) - 5,75 * I(\text{Tip} = 115) - 6,85 * I(\text{Tip} = 117)}{2,32 * I(\text{Tip} = 109) + 1,42 * I(\text{Tip} = 126)} \right)$$

si se distinguen tipologías, donde  $I(\text{Tip}=x)$  es la función indicador de la tipología x que vale 1 si el río tiene tipología x y 0 en otro caso. En ambos casos se detectará riesgo alto de baja calidad biológica del agua si su demanda química de oxígeno es mayor o igual que dicho umbral.

En la Tabla 5.2.1.4 se muestran los umbrales estimados para una red de valores del riesgo p, así como los porcentajes de éxito entre las muestras que cumplen los

requisitos de calidad, entre las que no la cumplen y el porcentaje de éxitos totales calculados utilizando las expresiones (4.5) a (4.7). Así, por ejemplo, si el nivel mediano de la demanda química de oxígeno es superior a 5,78 mg/L O<sub>2</sub> el modelo estimado sin distinguir tipologías estima que existe un riesgo superior de al menos un 34% de que las aguas del río sean de baja calidad biológica. Distinguiendo tipologías este umbral aumenta a 7,91 mg/L O<sub>2</sub> si el río es de la tipología 109 y a 29,02 mg/L O<sub>2</sub> si es de la 126. Utilizando estos umbrales los porcentajes de éxito para cada uno de los dos procedimientos fueron, respectivamente, 86% y 70,9% entre las muestras que cumplen, 55,8% y 66,3% entre las que no cumplen y 74,1% y 69,9% en total.

Finalmente, en la Tabla 5.2.1.5 se muestran los resultados obtenidos por tipologías utilizando los umbrales en negrita (distinguiendo y sin distinguir tipologías) junto a los correspondientes al procedimiento utilizado por la CHE, el cual clasifica el agua como de alta calidad si su nivel de demanda química es inferior a 15 mg/L O<sub>2</sub>. Así, por ejemplo, en la tipología 109, los porcentaje de éxito en la muestras que cumplen los estándares de calidad fueron 14,10%, 82,05% y 97,44% utilizando umbrales distinguiendo por tipologías, sin distinguir y el procedimiento de la CHE, respectivamente. Estos porcentajes fueron 95,35%, 53,49% y 27,91% entre las que no cumplen dichos estándares.

Se observa, en primer lugar, que los límites proporcionados por la CHE son poco operativos al clasificar un 79,82% aguas de baja calidad como buenas. Por otro lado, aunque a nivel global no existen diferencias significativas en los resultados obtenidos, en términos de porcentajes de aciertos (71,91% frente a 73,19%), si los analizamos por tipologías los porcentajes de aciertos en las muestras de alta y baja calidad están más equilibrados cuando los umbrales se fijan sin distinguir tipologías que cuando se distinguen, lo cual los hace más fiables al no saber, en una situación real, en qué situación están las aguas antes de ser analizadas.

### 5.3. UMBRALES DE CALIDAD PARA LOS INDICADORES DE NUTRIENTES

En este apartado presentamos los umbrales de calidad biológica para todos los indicadores de nutrientes NH<sub>4</sub>, NO<sub>2</sub>, NO<sub>3</sub>, PO<sub>4</sub> y P\_TOT cuyo significado puede verse en la Tabla 2.3.

#### 5.3.1. Umbrales de calidad biológica para el amonio total

De acuerdo con estudios realizados en proyectos anteriores se ha tomado  $g(X) = \log(X)$  al mostrar el indicador analizado problemas serios de falta de normalidad por una elevada asimetría positiva.

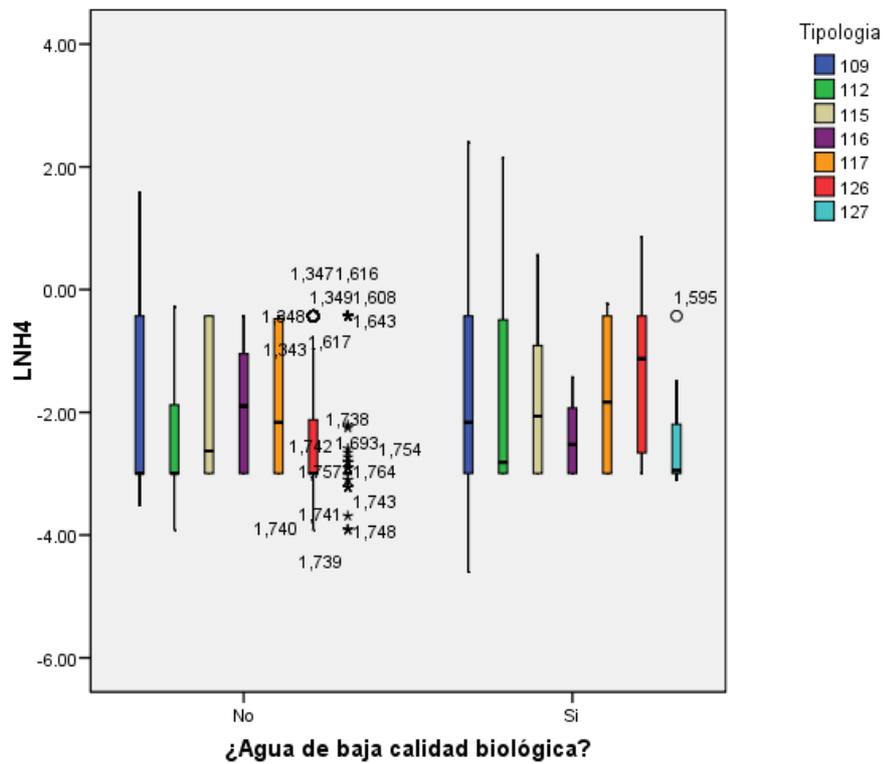
**Tabla 5.3.1.1: Datos ausentes y válidos del logaritmo del nivel de amonio total por tipologías y nivel de calidad biológica de las aguas**

¿Aguas de baja calidad biológica?	Tipología	Válidos		Perdidos		Total	
		Casos	Porcentaje	Casos	Porcentaje	Casos	Porcentaje
No	109	108	73,0%	40	27,0%	148	100.0%
	112	251	80,4%	61	19,6%	312	100.0%
	115	100	86,2%	16	13,8%	116	100.0%
	116	7	100,0%	0	,0%	7	100.0%
	117	23	95,8%	1	4,2%	24	100.0%
	126	269	75,8%	86	24,2%	355	100.0%
	127	105	66,9%	52	33,1%	157	100.0%
Si	109	105	65,6%	55	34,4%	160	100.0%
	112	88	59,9%	59	40,1%	147	100.0%
	115	120	76,9%	36	23,1%	156	100.0%
	116	13	81,3%	3	18,8%	16	100.0%
	117	84	86,6%	13	13,4%	97	100.0%
	126	30	54,5%	25	45,5%	55	100.0%
	127	8	50,0%	8	50,0%	16	100.0%
<b>Total</b>		1311	74,24%	455	25,76%	1766	100.0%

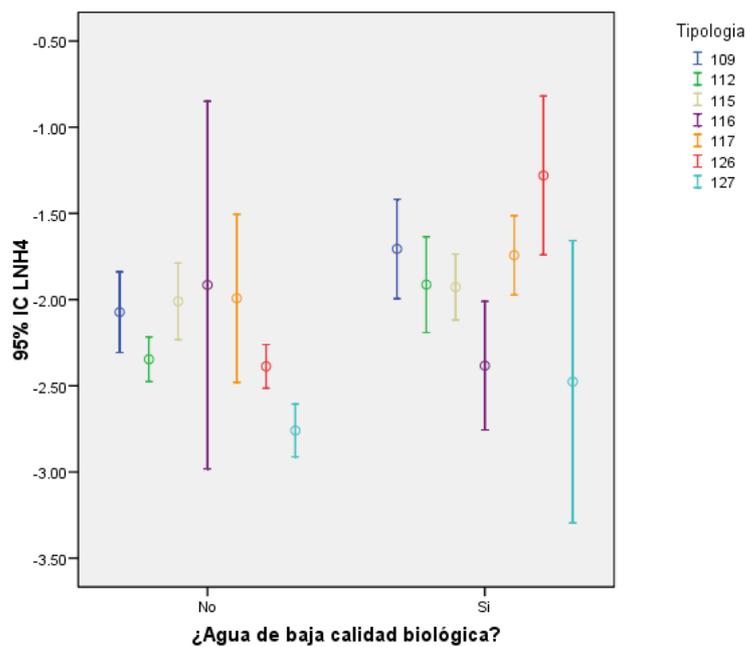
En la Tabla 5.3.1.1 se analiza el porcentaje de datos ausentes y válidos por tipologías y nivel de calidad biológica de las aguas. El porcentaje de datos ausentes oscila en torno a 25,76% siendo las tipologías más afectadas por este problema la 112 y la 126 en la que el porcentaje de datos ausentes está por encima del 40%, por lo que los resultados correspondientes a dichas tipologías deberían tomarse con precaución.

En este indicador también existe un alto porcentaje de datos censurados (82,84%) cuyo nivel de amonio es inferior a 0,13 mg/L NH<sub>4</sub> y cuyo valor se ha tomado

igual a 0,065, siendo la mayor parte de los mismos (72,65%) muestras de agua de alta calidad.



**Figura 5.3.1.1: Diagrama de cajas del logaritmo del nivel de amonio total por tipologías y nivel de calidad biológica de las aguas**



**Figura 5.3.1.2: Barras de error de la media del logaritmo del nivel de amonio total por tipologías y nivel de calidad biológica de las aguas (nivel de confianza 95%)**

En las Figuras 5.3.1.1 y 5.3.1.2 así como en la Tabla 5.3.1.3 se presentan los resultados del estudio comparativo del logaritmo del nivel de amonio total, para los datos no censurados, entre las aguas de baja y alta calidad biológica distinguiendo tipologías. Más concretamente se muestran sus diagramas de cajas (Figura 5.3.1.1) y las barras de error del 95% de confianza (Figura 5.3.1.2) así como sus valores medios y medianos (Tabla 5.3.1.3). Finalmente, se dan los resultados obtenidos al realizar contrastes de Mann-Whitney presentando, para cada tipología y en total, los valores de los rangos medios de las muestras de alta y baja calidad biológicas así como el pvalor del contraste (Tabla 5.3.1.3).

Se observa que únicamente para los ríos de las tipologías 109, 112 y 126 la mediana del logaritmo del nivel de amonio total tiende a ser significativamente inferior en las aguas de alta calidad biológica (ver Tabla 5.3.1.3). En el resto de tipologías, a excepción de la 116, dicho patrón se conserva, aunque no se aprecian diferencias que sean estadísticamente significativas (ver Figura 5.3.1.2).

**Tabla 5.3.1.2: Análisis descriptivo del logaritmo del nivel de amonio total por tipologías**

Tipología	Casos			Media			Mediana			Error típico de la media			Rangos promedio		Pvalor Mann-Whitney
	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		
	No	Si		No	Si		No	Si		No	Si		No	Si	
<b>109</b>	108	105	213	-2,07	-1,71	-1,89	-3,00	-2,16	-3,00	0,12	0,15	0,09	99,44	114,78	0,051
<b>112</b>	251	88	339	-2,35	-1,91	-2,23	-3,00	-2,81	-3,00	0,07	0,14	0,06	162,17	192,34	0,005
<b>115</b>	100	120	220	-2,01	-1,93	-1,96	-2,63	-2,06	-2,16	0,11	0,10	0,07	107,68	112,85	0,527
<b>116</b>	7	13	20	-1,92	-2,38	-2,22	-1,90	-2,53	-2,34	0,44	0,17	0,19	11,79	9,81	0,485
<b>117</b>	23	84	107	-1,99	-1,74	-1,80	-2,16	-1,83	-1,90	0,24	0,12	0,10	48,20	55,59	0,295
<b>126</b>	269	30	299	-2,39	-1,28	-2,28	-3,00	-1,13	-3,00	0,06	0,23	0,06	142,81	214,45	0,000
<b>127</b>	105	8	113	-2,76	-2,48	-2,74	-3,00	-2,95	-3,00	0,08	0,35	0,08	56,19	67,63	0,272
<b>Total</b>	863	448	1311	-2,32	-1,82	-2,15	-3,00	-2,00	-3,00	0,04	0,06	0,03	602,65	758,78	0,000

**Tabla 5.3.1.3: Modelos estimados para el logaritmo del nivel de amonio total distinguiendo y sin distinguir tipologías**

Variables seleccionadas	Distinguiendo tipologías			Sin distinguir tipologías		
	$\beta$	Error Estándar	Pvalor	$\beta$	Error Estándar	Pvalor
<b>Constante</b>	2,6341	0,5518	0,0000	0,1328	0,1204	0,2699
<b>LNH4</b>				0,3421	0,0518	0,0000
<b>TIP_109</b>	-1,5052	0,6590	0,0224			
<b>TIP_112</b>	-2,0124	0,7391	0,0065			
<b>LNH4_109</b>	0,6213	0,1711	0,0003			
<b>LNH4_112</b>	0,8012	0,2288	0,0005			
<b>LNH4_115</b>	1,1874	0,2655	0,0000			
<b>LNH4_117</b>	0,6762	0,2856	0,0179			
<b>LNH4_126</b>	2,4275	0,2863	0,0000			

**Tabla 5.3.1.4: Umbrales estimados para el amonio total y validación predictiva de los mismos distinguiendo y sin distinguir tipologías (en negrita señalados los umbrales utilizados en el proceso de comparación por tipologías descrito en la Tabla 5.3.1.5)**

Riesgo	Umbrales distinguiendo tipologías					%Aciertos			Umbrales sin distinguir tipologías	%Aciertos		
	109	112	115	117	126	Cumplen	No cumplen	Totales		Cumplen	No cumplen	Totales
26,00%	0.030	0.125	0.045	0.004	0.220	44,00%	91,00%	60,50%	0.032	1,20%	99,80%	36,90%
28,00%	0.036	0.142	0.049	0.005	0.229	44,00%	91,00%	60,50%	0.043	1,50%	99,80%	37,10%
30,00%	0.042	0.160	0.053	0.006	0.238	44,00%	91,00%	60,50%	0.057	63,80%	59,00%	62,10%
32,00%	0.048	0.180	0.058	0.007	0.248	44,10%	91,00%	60,60%	<b>0.075</b>	<b>64,70%</b>	<b>57,40%</b>	<b>62,10%</b>
34,00%	<b>0.056</b>	<b>0.201</b>	<b>0.062</b>	<b>0.008</b>	<b>0.257</b>	<b>76,30%</b>	<b>62,30%</b>	<b>71,40%</b>	0.098	65,50%	57,10%	62,50%
36,00%	0.064	0.224	0.067	0.009	0.267	76,00%	62,00%	71,20%	0.126	69,40%	51,10%	62,70%
38,00%	0.074	0.250	0.072	0.010	0.276	76,40%	61,60%	71,20%	0.162	72,40%	45,90%	62,80%
40,00%	0.085	0.277	0.077	0.011	0.286	76,50%	61,40%	71,20%	0.207	73,50%	40,50%	61,50%

**Tabla 5.3.1.5: Resultados obtenidos por tipologías utilizando los umbrales del nivel de amonio total de la Tabla 5.3.1.4 así como los umbrales proporcionados por la CHE**

		Calidad pronosticada						% Total aciertos		
		Distinguiendo tipología		Sin distinguir tipologías		CHE		Distinguiendo tipologías	Sin distinguir tipologías	CHE
		Alta	Baja	Alta	Baja	Alta	Baja			
109	Alta	3,70%	96,30%	60,19%	39,81%	69,44%	30,56%	48,36%	56,81%	52,11%
	Baja	5,71%	94,29%	46,67%	53,33%	65,71%	34,29%			
112	Alta	94,02%	5,98%	66,93%	33,07%	80,08%	19,92%	74,34%	61,65%	67,55%
	Baja	81,82%	18,18%	53,41%	46,59%	68,18%	31,82%			
115	Alta	0,00%	100,00%	50,00%	50,00%	70,00%	30,00%	54,09%	54,55%	45,91%
	Baja	0,83%	99,17%	41,67%	58,33%	74,17%	25,83%			
117	Alta	0,00%	100,00%	47,83%	52,17%	69,57%	30,43%	78,50%	62,62%	42,06%
	Baja	0,00%	100,00%	33,33%	66,67%	65,48%	34,52%			
126	Alta	98,51%	1,49%	71,38%	28,62%	79,55%	20,45%	93,31%	71,57%	75,92%
	Baja	53,33%	46,67%	26,67%	73,33%	56,67%	43,33%			
Total	Alta	67,24%	32,76%	64,71%	35,29%	76,70%	23,30%	71,05%	62,05%	60,53%
	Baja	22,25%	77,75%	42,62%	57,38%	67,92%	32,08%			

En la Tabla 5.3.1.3 se muestra el modelo seleccionado utilizando como variables independientes los indicadores de la tipología del río (denotados como TIP\_x donde x es el número de la tipología) y sus interacciones con el logaritmo del nivel de amonio total (denotados como LNH4\_x) y aplicando un procedimiento de selección de variables por eliminación hacia atrás tomando pvalor de entrada 0,01 y como pvalor de salida 0,02. Así mismo y, por razones comparativas, se muestra el modelo estimado utilizando como variables independientes una constante y dicho logaritmo, el cual no tiene en cuenta, por tanto, la información proporcionada por la tipología del río. Las estimaciones se han realizado utilizando las observaciones de las tipologías 109, 112, 115, 117 y 126.

Las estimaciones de los parámetros del modelo (ver Tabla 5.3.1.3) y del proceso de validación (ver Tabla 5.3.1.5) ponen de manifiesto que en los ríos de las tipologías 109, 112, 115, 117 y 126 el nivel de amonio total discrimina significativamente entre las muestras de agua de alta y baja calidad biológicas, siendo los coeficientes  $\beta$  estimados significativamente positivos.

Utilizando la expresión (4.2), el umbral para un riesgo  $0 < p < 1$  vendrá dado por la expresión:

$$\text{Umbral}_p = \exp\left(\frac{\log\left(\frac{p}{1-p}\right) - 0,133}{0,342}\right)$$

si no se distinguen tipologías y

$$\begin{aligned} \text{Umbral}_p &= \\ &= \exp\left(\frac{\log\left(\frac{p}{1-p}\right) - 2,634 + 1,505 * I(\text{Tip} = 109) + 2,012 * I(\text{Tip} = 112)}{0,621 * I(\text{Tip} = 109) + 0,801 * I(\text{Tip} = 112)) + 1,187 * I(\text{Tip} = 115) + 0,676 * I(\text{Tip} = 117) + 2,427 * I(\text{Tip} = 126)}\right) \end{aligned}$$

si se distinguen tipologías, donde  $I(\text{Tip}=x)$  es la función indicador de la tipología x que vale 1 si el río tiene tipología x y 0 en otro caso. En ambos casos se detectará riesgo alto de baja calidad biológica del agua si el nivel de amonio total de ésta es mayor o igual que dicho umbral.

En la Tabla 5.3.1.4 se muestran los umbrales estimados para una red de valores del riesgo  $p$ , así como los porcentajes de éxito entre las muestras que cumplen los requisitos de calidad, entre las que no la cumplen y el porcentaje de éxitos totales calculados utilizando las expresiones (4.5) a (4.7). Así, por ejemplo, si el nivel mediano de amonio total es superior a 0,075 mg/L  $\text{NH}_4$  el modelo estimado sin distinguir tipologías estima que existe un riesgo superior de al menos un 32% de que las aguas del río sean de baja calidad biológica. Distinguiendo tipologías este umbral cambia a 0,048 mg/L  $\text{NH}_4$  si el río es de la tipología 109, 0,180 mg/L  $\text{NH}_4$  si el río es de la tipología 112, 0,058 mg/L  $\text{NH}_4$  si es de la tipología 115, 0,007 mg/L  $\text{NH}_4$  si es de la 117 y 0,248 mg/L  $\text{NH}_4$  si es de la 126. Utilizando estos umbrales los porcentajes de éxito para cada uno de los dos procedimientos fueron, respectivamente, 64,70% y 44,10% entre las muestras que cumplen; 57,40% y 91% entre las que no cumplen y 62,10% y 60,6% en total.

Finalmente, en la Tabla 5.3.1.5 se muestran los resultados obtenidos por tipologías utilizando los umbrales señalados en negrita (distinguiendo y sin distinguir tipologías) junto a los correspondientes al procedimiento utilizado por la CHE que clasifica el agua como de alta calidad si su nivel de amonio total es inferior a 0,40 mg/L  $\text{NH}_4$ . Así, por ejemplo, en la tipología 109, los porcentaje de éxito en la muestras que cumplen los estándares de calidad fueron 3,70%, 60,19% y 69,44% utilizando umbrales distinguiendo por tipologías, sin distinguir y el procedimiento de la CHE, respectivamente. Estos porcentajes fueron 94,29%, 53,33% y 34,29% entre las que no cumplen dichos estándares.

Se observa, en primer lugar, que los límites proporcionados por la CHE son poco operativos al clasificar un 67,92% aguas de baja calidad como buenas. Por otro lado, aunque a nivel global distinguir por tipologías obtiene un porcentaje de aciertos superior a no distinguir (71,05% frente a 62,05%), si los analizamos por tipologías se observa que el procedimiento que las distingue tiende a clasificar las muestras siguiendo la regla de la mayoría obteniendo unos porcentajes de aciertos en las muestras de alta y baja calidad más desequilibrados que cuando los umbrales se fijan sin distinguirlas, lo cual los hace menos fiables al no saber, en una situación real, en qué situación están las aguas antes de ser analizadas.

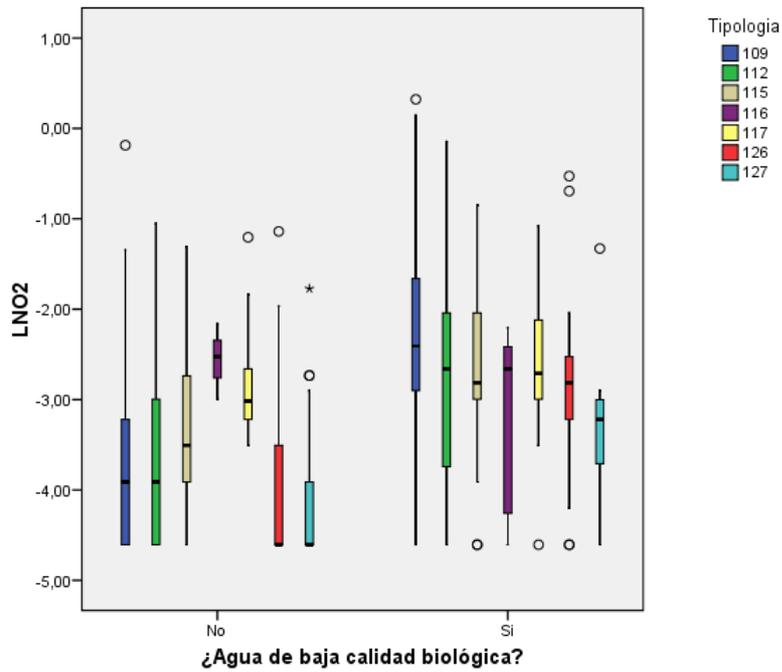
### 5.3.2. Umbrales de calidad biológica para los nitritos

De acuerdo con estudios realizados en proyectos anteriores se ha tomado  $g(X) = \log(X)$  al mostrar el indicador analizado problemas serios de falta de normalidad por una elevada asimetría positiva.

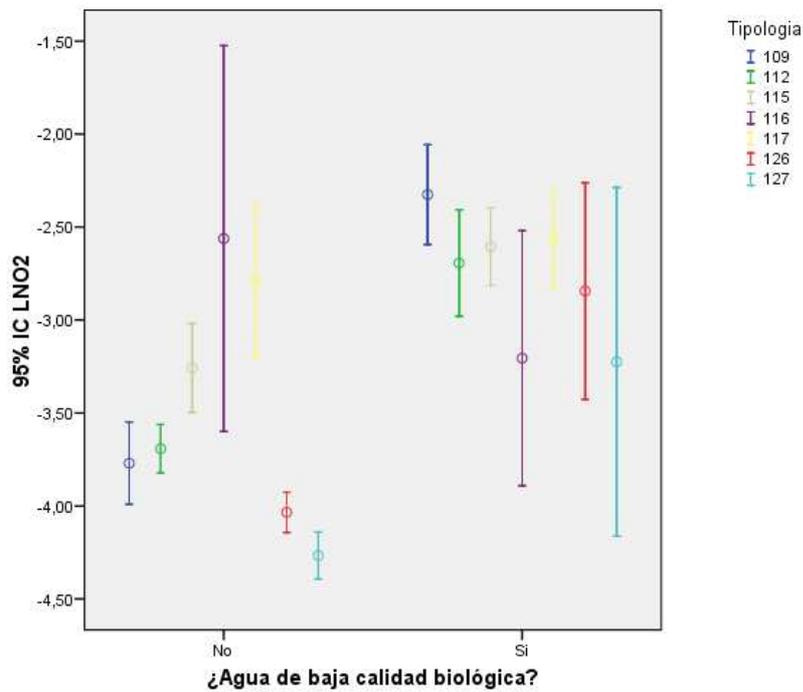
**Tabla 5.3.2.1: Datos ausentes y válidos del logaritmo del nivel de nitritos por tipologías y nivel de calidad biológica de las aguas**

¿Aguas de baja calidad biológica?	Tipología	Válidos		Perdidos		Total	
		Casos	Porcentaje	Casos	Porcentaje	Casos	Porcentaje
No	109	74	50,0%	74	50,0%	148	100,0%
	112	182	58,3%	130	41,7%	312	100,0%
	115	47	40,5%	69	59,5%	116	100,0%
	116	3	42,9%	4	57,1%	7	100,0%
	117	14	58,3%	10	41,7%	24	100,0%
	126	182	51,3%	173	48,7%	355	100,0%
	127	93	59,2%	64	40,8%	157	100,0%
Si	109	73	45,6%	87	54,4%	160	100,0%
	112	60	40,8%	87	59,2%	147	100,0%
	115	63	40,4%	93	59,6%	156	100,0%
	116	11	68,8%	5	31,3%	16	100,0%
	117	34	35,1%	63	64,9%	97	100,0%
	126	17	30,9%	38	69,1%	55	100,0%
	127	7	43,8%	9	56,3%	16	100,0%
<b>Total</b>		860	48,7%	906	51,3%	1766	100,0%

En la Tabla 5.3.2.1 se analiza el porcentaje de datos ausentes y válidos por tipologías y nivel de calidad biológica de las aguas. El porcentaje de datos ausentes es muy elevado y oscila en torno a 51,3% destacando los casos de las tipologías 117 y 126 cuyos porcentajes suben por encima del 60%.



**Figura 5.3.2.1: Diagrama de cajas del logaritmo del nivel de nitritos por tipologías y nivel de calidad biológica de las aguas**



**Figura 5.3.2.2: Barras de error de la media del logaritmo del nivel de nitritos por tipologías y nivel de calidad biológica de las aguas (nivel de confianza 95%)**

En las Figuras 5.3.2.1 y 5.3.2.2 así como en la Tabla 5.3.2.2. se presentan los resultados del estudio comparativo del logaritmo del nivel de nitritos entre las aguas de baja y alta calidad biológica distinguiendo tipologías. Más concretamente se muestran sus diagramas de cajas (Figura 5.3.2.1) y las barras de error del 95% de confianza (Figura 5.3.2.2) así como sus valores medios y medianos (Tabla 5.3.2.2). Finalmente, se dan los resultados obtenidos al realizar contrastes de Mann-Whitney presentando, para cada tipología y en total, los valores de los rangos medios de las muestras de alta y baja calidad biológicas así como el p-valor del contraste (Tabla 5.3.2.2).

Se observa que, con la excepción de los ríos de tipología 116, 117 y 127, la media del logaritmo del nivel de nitritos tiende a ser significativamente inferior en las aguas de alta calidad biológica (ver Tabla 5.3.2.2). En el resto de tipologías, con la única excepción de la 116, aunque dicho patrón se conserva, no se aprecian diferencias que sean estadísticamente significativas (ver Figura 5.3.2.2).

**Tabla 5.3.2.2: Análisis descriptivo del logaritmo del nivel de nitritos por tipologías**

Tipología	Casos			Media			Mediana			Error típico de la media			Rangos promedio		Pvalor Mann-Whitney
	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		
No	Si	No		Si	No		Si	No		Si	No		Si	No	Si
<b>109</b>	74	73	147	-3,77	-2,33	-3,05	-3,91	-2,41	-3,00	0,11	0,13	0,11	49,72	98,61	0,000
<b>112</b>	182	60	242	-3,69	-2,69	-3,44	-3,91	-2,66	-3,69	0,07	0,14	0,07	105,96	168,64	0,000
<b>115</b>	47	63	110	-3,26	-2,61	-2,88	-3,51	-2,81	-3,00	0,12	0,10	0,08	41,52	65,93	0,000
<b>116</b>	3	11	14	-2,56	-3,21	-3,07	-2,53	-2,66	-2,59	0,24	0,31	0,25	9,67	6,91	0,368
<b>117</b>	14	34	48	-2,79	-2,57	-2,63	-3,02	-2,71	-2,73	0,19	0,13	0,11	20,29	26,24	0,180
<b>126</b>	182	17	199	-4,03	-2,84	-3,93	-4,61	-2,81	-4,20	0,05	0,27	0,06	94,63	157,50	0,000
<b>127</b>	93	7	100	-4,27	-3,22	-4,19	-4,61	-3,22	-4,61	0,06	0,38	0,07	48,26	80,29	0,001
<b>Total</b>	595	265	860	-3,83	-2,60	-3,45	-3,91	-2,66	-3,54	0,04	0,06	0,04	345,73	620,83	0,000

**Tabla 5.3.2.3: Modelos estimados para el logaritmo del nivel de nitritos distinguiendo y sin distinguir tipologías**

Variables seleccionadas	Distinguiendo tipologías			Sin distinguir tipologías		
	$\beta$	Error Estándar	Pvalor	$\beta$	Error Estándar	Pvalor
<b>Constante</b>	2,8591	0,3680	0,0000	3,0203	0,3648	0,0000
<b>LN02</b>				1,1065	0,1134	0,0000
<b>LNO2_109</b>	0,9560	0,1207	0,0000			
<b>LNO2_112</b>	1,2166	0,1223	0,0000			
<b>LNO2_115</b>	0,8812	0,1334	0,0000			

**Tabla 5.3.2.4: Umbrales estimados para el nivel de nitritos y validación predictiva de los mismos distinguiendo y sin distinguir tipologías (en negrita señalados los umbrales utilizados en el proceso de comparación por tipologías descrito en la Tabla 5.3.2.5)**

Riesgo	Umbrales distinguiendo tipologías			% Aciertos			Umbrales sin distinguir tipologías	% Aciertos		
	109	112	115	Cumplen	No cumplen	Totales		Cumplen	No cumplen	Totales
<b>40,00%</b>	<b>0,033</b>	<b>0,068</b>	<b>0,025</b>	<b>75,20%</b>	<b>77,00%</b>	<b>76,00%</b>	<b>0,045</b>	<b>72,60%</b>	<b>73,50%</b>	<b>72,90%</b>
41,00%	0,034	0,071	0,026	76,20%	74,00%	75,40%	0,047	72,60%	73,50%	72,90%
42,00%	0,036	0,073	0,027	76,20%	74,00%	75,40%	0,049	72,60%	73,50%	72,90%
43,00%	0,037	0,076	0,028	76,20%	73,50%	75,20%	0,051	79,20%	63,80%	73,10%
44,00%	0,039	0,078	0,030	76,20%	73,50%	75,20%	0,052	79,20%	63,80%	73,10%
45,00%	0,041	0,081	0,031	80,20%	67,90%	75,40%	0,054	79,20%	63,80%	73,10%

**Tabla 5.3.2.5: Resultados obtenidos por tipologías utilizando los umbrales del nivel de nitritos de la Tabla 5.3.2.4 así como los umbrales proporcionados por la CHE**

Tipología	Calidad Observada	Calidad pronosticada						% Total aciertos		
		Distinguiendo tipología		Sin distinguir tipologías		CHE		Distinguiendo tipologías	Sin distinguir tipologías	CHE
		Alta	Baja	Alta	Baja	Alta	Baja			
109	Alta	74,32%	25,68%	78,38%	21,62%	95,95%	4,05%	78,91%	78,91%	66,67%
	Baja	16,44%	83,56%	20,55%	79,45%	63,01%	36,99%			
112	Alta	86,26%	13,74%	73,08%	26,92%	95,60%	4,40%	77,69%	70,66%	77,27%
	Baja	48,33%	51,67%	36,67%	63,33%	78,33%	21,67%			
115	Alta	34,04%	65,96%	61,70%	38,30%	91,49%	8,51%	68,18%	70,00%	48,18%
	Baja	6,35%	93,65%	23,81%	76,19%	84,13%	15,87%			
Total	Alta	75,25%	24,75%	72,61%	27,39%	95,05%	4,95%	75,95%	72,95%	67,74%
	Baja	22,96%	77,04%	26,53%	73,47%	74,49%	25,51%			

En la Tabla 5.3.2.3 se muestra el modelo seleccionado utilizando como variables independientes los indicadores de la tipología del río (denotados como TIP\_x donde x es el número de la tipología) y sus interacciones con el logaritmo del nivel de nitritos (denotados como LNO2\_x) y aplicando un procedimiento de selección de variables por eliminación hacia atrás tomando p-valor de entrada 0,01 y como p-valor de salida 0,02. Así mismo y, por razones comparativas, se muestra el modelo estimado utilizando como variables independientes una constante y dicho logaritmo, el cual no tiene en cuenta, por tanto, la información proporcionada por la tipología del río. Las estimaciones se han realizado utilizando, únicamente, las observaciones de las tipologías 109, 112 y 115 dado que, para el resto, los resultados obtenidos no fueron fiables debido al escaso tamaño muestral disponible en algunos de los grupos comparados.

Las estimaciones de los parámetros del modelo (ver Tabla 5.3.2.3) y del proceso de validación (ver Tabla 5.3.2.5) ponen de manifiesto que en los ríos de las tipologías 109, 112 y 115 el nivel de nitritos discrimina entre las muestras de agua de alta y baja calidad biológicas, siendo los coeficientes  $\beta$  estimados significativamente positivos.

Utilizando la expresión (4.2), el umbral para un riesgo  $0 < p < 1$  vendrá dado por la expresión:

$$\text{Umbral}_p = \exp\left(\frac{\log\left(\frac{p}{1-p}\right) - 3,020}{1,107}\right)$$

si no se distinguen tipologías y

$$\text{Umbral}_p = \exp\left(\frac{\log\left(\frac{p}{1-p}\right) - 2,859}{0,956 * I(\text{Tip} = 109) + 1,217 * I(\text{Tip} = 112) + 0,881 * I(\text{Tip} = 115)}\right)$$

si se distinguen tipologías, donde  $I(\text{Tip}=x)$  es la función indicador de la tipología x que vale 1 si el río tiene tipología x y 0 en otro caso. En ambos casos se detectará riesgo alto de baja calidad biológica del agua si el nivel de nitritos de ésta es mayor o igual que dicho umbral.

En la Tabla 5.3.2.4 se muestran los umbrales estimados para una red de valores del riesgo p, así como los porcentajes de éxito entre las muestras que cumplen los requisitos de calidad, entre las que no la cumplen y el porcentaje de éxitos totales

calculados utilizando las expresiones (4.5) a (4.7). Así, por ejemplo, si el nivel mediano de nitritos es superior a 0,045 mg/L NO<sub>2</sub> el modelo estimado sin distinguir tipologías estima que existe un riesgo superior de al menos un 40% de que las aguas del río sean de baja calidad biológica. Distinguiendo tipologías este umbral cambia a 0,033 mg/L NO<sub>2</sub> si el río es de la tipología 109; a 0,068 mg/L NO<sub>2</sub> si es de tipología 112 y a 0,025 mg/L NO<sub>2</sub> si es de la 115. Utilizando estos umbrales los porcentajes de éxito para cada uno de los dos procedimientos fueron, respectivamente, 72,60% y 75,20% entre las muestras que cumplen; 73,50% y 77% entre las que no cumplen y 72,90% y 76% en total.

Finalmente, en la Tabla 5.3.2.5 se muestran los resultados obtenidos por tipologías utilizando los umbrales anteriores (distinguiendo y sin distinguir tipologías) junto a los correspondientes al procedimiento utilizado por la CHE que clasifica el agua como de alta calidad si su nivel de nitritos es inferior a 0,15 mg/L NO<sub>2</sub>. Así, por ejemplo, en la tipología 109, los porcentaje de éxito en la muestras que cumplen los estándares de calidad fueron 74,32%, 78,38% y 95,95% utilizando umbrales distinguiendo por tipologías, sin distinguir y el procedimiento de la CHE, respectivamente. Estos porcentajes fueron 83,56%, 79,45% y 36,99% entre las que no cumplen dichos estándares.

Se observa, en primer lugar, que los límites proporcionados por la CHE son poco operativos al clasificar un 74,49% aguas de baja calidad como buenas. Por otro lado, aunque a nivel global el procedimiento que distingue tipologías obtiene mejores resultados, en términos de porcentajes de aciertos (75,95% frente a 72,95%), si analizamos por tipologías los porcentajes de aciertos en las muestras de alta y baja calidad se observa que están más equilibrados cuando los umbrales se fijan sin distinguir tipologías que cuando se distinguen, lo cual los hace menos fiables al no saber, en una situación real, en qué situación están las aguas antes de ser analizadas.

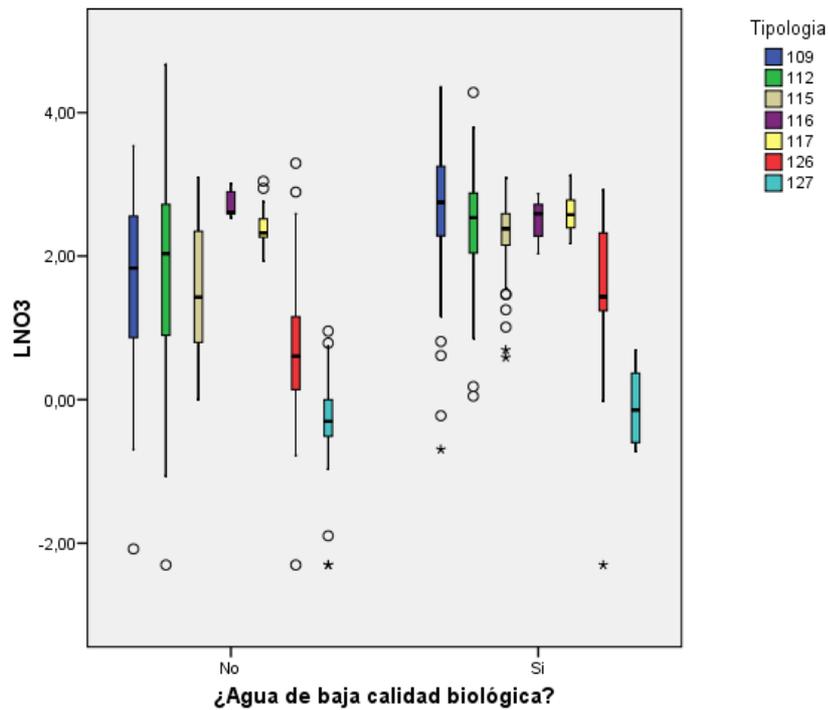
### 5.3.3. Umbrales de calidad biológica para los nitratos

De acuerdo con estudios realizados en proyectos anteriores se ha tomado  $g(X) = \log(X)$  al mostrar el indicador analizado problemas serios de falta de normalidad por una elevada asimetría positiva.

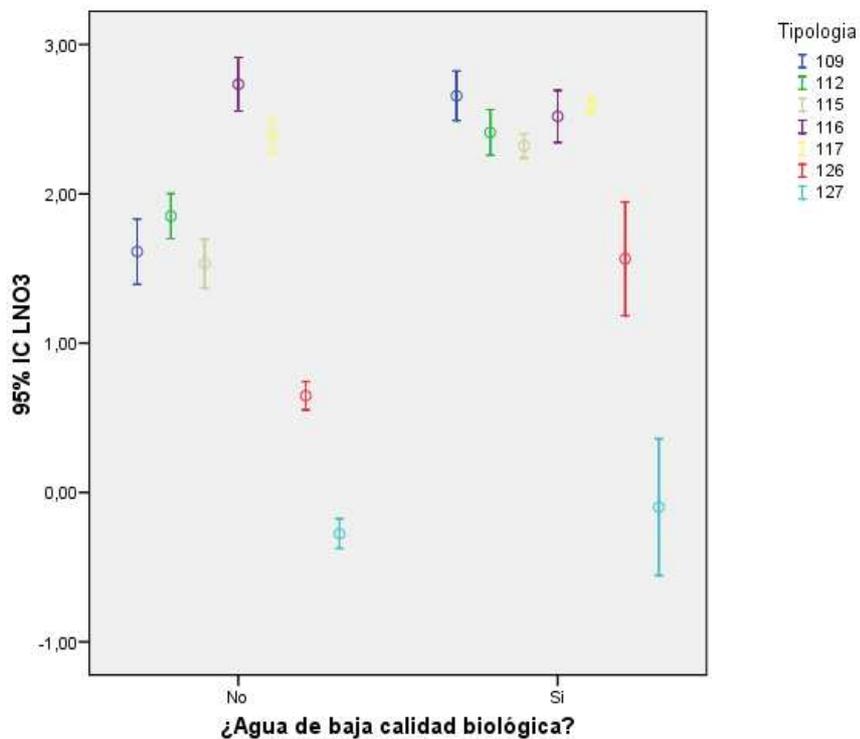
**Tabla 5.3.3.1: Datos ausentes y válidos del logaritmo del nivel de nitratos por tipologías y nivel de calidad biológica de las aguas**

¿Aguas de baja calidad biológica?	Tipología	Válidos		Perdidos		Total	
		Casos	Porcentaje	Casos	Porcentaje	Casos	Porcentaje
No	109	108	73,0%	40	27,0%	148	100,0%
	112	242	77,6%	70	22,4%	312	100,0%
	115	96	82,8%	20	17,2%	116	100,0%
	116	7	100,0%	0	,0%	7	100,0%
	117	23	95,8%	1	4,2%	24	100,0%
	126	263	74,1%	92	25,9%	355	100,0%
	127	102	65,0%	55	35,0%	157	100,0%
Si	109	105	65,6%	55	34,4%	160	100,0%
	112	88	59,9%	59	40,1%	147	100,0%
	115	116	74,4%	40	25,6%	156	100,0%
	116	13	81,3%	3	18,8%	16	100,0%
	117	84	86,6%	13	13,4%	97	100,0%
	126	30	54,5%	25	45,5%	55	100,0%
	127	8	50,0%	8	50,0%	16	100,0%
<b>Total</b>		1285	72,8%	481	27,2%	1766	100,0%

En la Tabla 5.3.3.1 se analiza el porcentaje de datos ausentes y válidos por tipologías y nivel de calidad biológica de las aguas. El porcentaje de datos ausentes oscila en torno a 27,2% llamando la atención el elevado porcentaje de datos ausentes en los ríos de las tipologías 126 y 127 en las muestras de baja calidad biológica (45,5% y 50%, respectivamente) lo cual hace que los resultados correspondientes a estas tipologías deban tomarse con las debidas precauciones.



**Figura 5.3.3.1: Diagrama de cajas del logaritmo del nivel de nitratos por tipologías y nivel de calidad biológica de las aguas**



**Figura 5.3.3.2: Barras de error de la media del logaritmo del nivel de nitratos por tipologías y nivel de calidad biológica de las aguas (nivel de confianza 95%)**

En las Figuras 5.3.3.1 y 5.3.3.2 así como en la Tabla 5.3.3.2 se presentan los resultados del estudio comparativo del logaritmo del nivel de nitratos entre las aguas de baja y alta calidad biológica distinguiendo tipologías. Más concretamente se muestran sus diagramas de cajas (Figura 5.3.3.1) y las barras de error del 95% de confianza (Figura 5.3.3.2) así como sus valores medios y medianos (Tabla 5.3.3.2). Finalmente, se dan los resultados obtenidos al realizar contrastes de Mann-Whitney presentando, para cada tipología y en total, los valores de los rangos medios de las muestras de alta y baja calidad biológicas así como el pvalor del contraste (Tabla 5.3.3.2).

Se observa que, con la excepción de los ríos de tipología 116 y 127, la mediana del logaritmo del nivel de nitratos tiende a ser significativamente inferior en las aguas de alta calidad biológica (ver Tabla 5.3.3.2). En el resto de tipologías, con la única excepción de la 116, aunque dicho patrón se conserva, no se aprecian diferencias que sean estadísticamente significativas (ver Figura 5.3.3.2).

**Tabla 5.3.3.2: Análisis descriptivo del logaritmo del logaritmo del nivel de nitratos por tipologías**

Tipología	Casos			Media			Mediana			Error típico de la media			Rangos promedio		Pvalor Mann-Whitney
	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		
	No	Si		No	Si		No	Si		No	Si		No	Si	
<b>109</b>	108	105	213	1,61	2,66	2,13	1,83	2,75	2,41	0,11	0,08	0,08	78,25	136,58	0,000
<b>112</b>	242	88	330	1,85	2,41	2,00	2,03	2,54	2,16	0,08	0,08	0,06	152,65	200,84	0,000
<b>115</b>	96	116	212	1,53	2,32	1,96	1,43	2,38	2,24	0,08	0,04	0,05	74,11	133,31	0,000
<b>116</b>	7	13	20	2,73	2,52	2,59	2,61	2,59	2,61	0,07	0,08	0,06	13,29	9,00	0,135
<b>117</b>	23	84	107	2,39	2,59	2,55	2,32	2,58	2,55	0,06	0,03	0,03	35,46	59,08	0,001
<b>126</b>	263	30	293	0,65	1,57	0,74	0,61	1,44	0,69	0,05	0,19	0,05	137,59	229,50	0,000
<b>127</b>	102	8	110	-0,28	-0,10	-0,26	-0,30	-0,15	-0,30	0,05	0,19	0,05	54,99	62,06	0,545
<b>Total</b>	841	444	1285	1,17	2,38	1,59	1,06	2,49	1,89	0,04	0,04	0,03	510,93	893,16	0,000

**Tabla 5.3.3.3: Modelos estimados para el logaritmo del nivel de nitratos distinguiendo y sin distinguir tipologías**

Variables seleccionadas	Distinguiendo tipologías			Sin distinguir tipologías		
	$\beta$	Error Estándar	Pvalor	$\beta$	Error Estándar	Pvalor
<b>Constante</b>	-2,4767	0,2432	0,0000	-2,1740	0,2239	0,0000
<b>LN03</b>				0,8511	0,0944	0,0000
<b>LNO3_109</b>	1,1045	0,1123	0,0000			
<b>LNO3_112</b>	0,6491	0,1035	0,0000			
<b>LNO3_115</b>	1,3918	0,1322	0,0000			

**Tabla 5.3.3.4: Umbrales estimados para el nivel de nitratos y validación predictiva de los mismos distinguiendo y sin distinguir tipologías (en negrita señalados los umbrales utilizados en el proceso de comparación por tipologías descrito en la Tabla 5.3.3.5)**

Riesgo	Umbrales distinguiendo tipologías			%Aciertos			Umbrales distinguiendo tipologías	%Aciertos		
	109	112	115	Cumplen	No cumplen	Totales		Cumplen	No cumplen	Totales
35,00%	5,38	17,49	3,80	64,80%	73,50%	68,30%	6,22	48,00%	87,10%	64,00%
36,00%	5,59	18,71	3,92	67,00%	71,80%	69,00%	6,54	49,80%	86,40%	64,80%
37,00%	5,82	20,00	4,04	68,20%	70,90%	69,30%	6,88	50,90%	83,80%	64,40%
38,00%	<b>6,04</b>	<b>21,36</b>	<b>4,17</b>	<b>69,10%</b>	<b>69,60%</b>	<b>69,30%</b>	7,24	53,10%	82,50%	65,20%
39,00%	6,28	22,79	4,30	70,40%	68,30%	69,50%	7,61	55,20%	80,60%	65,60%
40,00%	6,52	24,31	4,43	72,90%	66,70%	70,30%	7,99	56,70%	78,60%	65,70%
41,00%	6,77	25,91	4,56	73,30%	66,00%	70,30%	8,39	58,70%	76,70%	66,10%
42,00%	7,03	27,61	4,70	74,00%	65,70%	70,60%	8,80	61,00%	74,10%	66,40%
43,00%	7,30	29,41	4,84	75,00%	65,00%	71,00%	9,24	61,40%	71,50%	65,60%
44,00%	7,57	31,31	4,98	76,20%	64,40%	71,40%	9,69	63,70%	68,60%	65,70%
45,00%	7,85	33,33	5,13	77,10%	63,40%	71,50%	<b>10,16</b>	<b>65,00%</b>	<b>62,80%</b>	<b>64,10%</b>

**Tabla 5.3.3.5: Resultados obtenidos por tipologías utilizando los umbrales del nivel de nitratos de la Tabla 5.3.3.4 así como los umbrales proporcionados por la CHE**

Tipología	Calidad Observada	Calidad pronosticada						% Total aciertos		
		Distinguiendo tipología		Sin distinguir tipologías		CHE		Distinguiendo tipologías	Sin distinguir tipologías	CHE
		Alta	Baja	Alta	Baja	Alta	Baja			
109	Alta	48,15%	51,85%	66,67%	33,33%	94,44%	5,56%	66,20%	69,48%	66,20%
	Baja	15,24%	84,76%	27,62%	72,38%	62,86%	37,14%			
112	Alta	85,95%	14,05%	60,74%	39,26%	85,12%	14,88%	67,27%	60,00%	67,58%
	Baja	84,09%	15,91%	42,05%	57,95%	80,68%	19,32%			
115	Alta	50,00%	50,00%	73,96%	26,04%	98,96%	1,04%	75,47%	65,09%	46,70%
	Baja	3,45%	96,55%	42,24%	57,76%	96,55%	3,45%			
Total	Alta	69,06%	30,94%	65,02%	34,98%	90,36%	9,64%	69,27%	64,11%	61,32%
	Baja	30,42%	69,58%	37,22%	62,78%	80,58%	19,42%			

En la Tabla 5.3.3.3 se muestra el modelo seleccionado utilizando como variables independientes los indicadores de la tipología del río (denotados como TIP\_x donde x es el número de la tipología) y sus interacciones con el logaritmo del nivel de nitratos (denotados como LNO3\_x) y aplicando un procedimiento de selección de variables por eliminación hacia atrás tomando p-valor de entrada 0,01 y como p-valor de salida 0,02. Así mismo y, por razones comparativas, se muestra el modelo estimado utilizando como variables independientes una constante y dicho logaritmo, el cual no tiene en cuenta, por tanto, la información proporcionada por la tipología del río. Las estimaciones se han realizado utilizando, únicamente, las observaciones de las tipologías 109, 112 y 115 dado que, para el resto, los resultados obtenidos no fueron fiables debido al escaso tamaño muestral disponible en algunos de los grupos comparados.

Las estimaciones de los parámetros del modelo (ver Tabla 5.3.3.3) y del proceso de validación (ver Tabla 5.3.3.5) ponen de manifiesto que en los ríos de las tipologías 109, 112 y 115 el nivel de nitratos discrimina entre las muestras de agua de alta y baja calidad biológicas, siendo los coeficiente  $\beta$  estimados significativamente positivos.

Utilizando la expresión (4.2), el umbral para un riesgo  $0 < p < 1$  vendrá dado por la expresión:

$$\text{Umbral}_p = \exp \left( \frac{\log \left( \frac{p}{1-p} \right) + 2,17}{0,85} \right)$$

si no se distinguen tipologías y

$$\text{Umbral}_p = \exp \left( \frac{\log \left( \frac{p}{1-p} \right) + 2,48}{1,10 * I(\text{Tip} = 109) + 0,65 * I(\text{Tip} = 112) + 1,39 * I(\text{Tip} = 115)} \right)$$

si se distinguen tipologías, donde  $I(\text{Tip}=x)$  es la función indicador de la tipología x que vale 1 si el río tiene tipología x y 0 en otro caso. En ambos casos se detectará riesgo alto de baja calidad biológica del agua si el nivel de nitratos de ésta es mayor o igual que dicho umbral.

En la Tabla 5.3.3.4 se muestran los umbrales estimados para una red de valores del riesgo p, así como los porcentajes de éxito entre las muestras que cumplen los

requisitos de calidad, entre las que no la cumplen y el porcentaje de éxitos totales calculados utilizando las expresiones (4.5) a (4.7). Así, por ejemplo, si el nivel mediano de nitratos es superior a 7,24 mg/L NO<sub>3</sub> el modelo estimado sin distinguir tipologías estima que existe un riesgo superior de al menos un 38% de que las aguas del río sean de baja calidad biológica. Distinguiendo tipologías este umbral cambia a 6,04 mg/L NO<sub>3</sub> si el río es de la tipología 109; 21,36 mg/L NO<sub>3</sub> si es de tipología 112 y 4,17 mg/L NO<sub>3</sub> si es de la 115. Utilizando estos umbrales los porcentajes de éxito para cada uno de los dos procedimientos fueron, respectivamente, 53,1% y 69,10% entre las muestras que cumplen; 82,50% y 69,60% entre las que no cumplen y 65,20% y 69,30% en total.

Finalmente, en la Tabla 5.3.3.5 se muestran los resultados obtenidos por tipologías utilizando los umbrales en negrita (distinguiendo y sin distinguir tipologías) junto a los correspondientes al procedimiento utilizado por la CHE que clasifica las aguas como de alta calidad si el nivel de los nitratos es inferior a 20 mg/L NO<sub>3</sub>. Así, por ejemplo, en la tipología 109, los porcentaje de éxito en la muestras que cumplen los estándares de calidad fueron 48,15%, 66,67% y 94,44% utilizando umbrales distinguiendo por tipologías, sin distinguir y el procedimiento de la CHE, respectivamente. Estos porcentajes fueron 84,76%, 72,38% y 37,14% entre las que no cumplen dichos estándares.

Se observa, en primer lugar, que los límites proporcionados por la CHE son poco operativos al clasificar un 80,58% aguas de baja calidad como buenas. Por otro lado, aunque a nivel global el procedimiento que distingue entre tipologías obtiene mejores resultados en términos de porcentajes de aciertos que el que no distingue (69,27% frente a 64,11%), si los analizamos por tipologías los porcentajes de aciertos en las muestras de alta y baja calidad están más equilibrados cuando los umbrales se fijan sin distinguir tipologías que cuando se distinguen, lo cual los hace menos fiables al no saber, en una situación real, en qué situación están las aguas antes de ser analizadas.

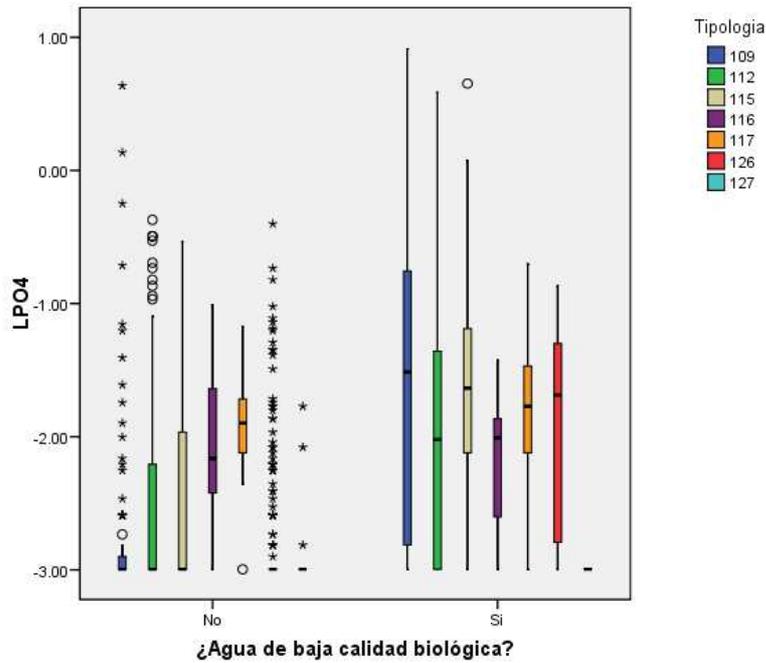
### 5.3.4. Umbrales de calidad biológica para los fosfatos

De acuerdo con estudios realizados en proyectos anteriores se ha tomado  $g(X) = \log(X)$  al mostrar el indicador analizado problemas serios de falta de normalidad por una elevada asimetría positiva.

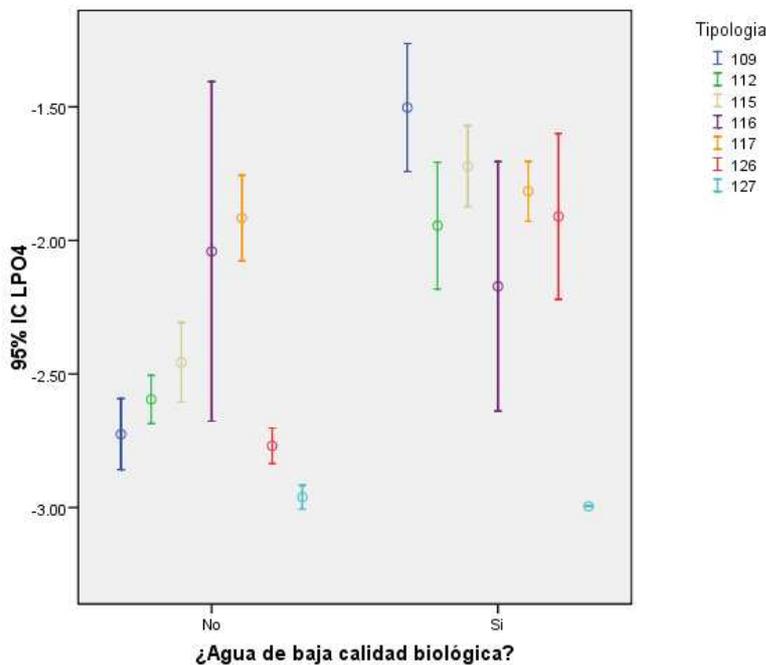
**Tabla 5.3.4.1: Datos ausentes y válidos del logaritmo del nivel de fosfatos por tipologías y nivel de calidad biológica de las aguas**

¿Aguas de baja calidad biológica?	Tipología	Válidos		Perdidos		Total	
		Casos	Porcentaje	Casos	Porcentaje	Casos	Porcentaje
No	109	102	68,9%	46	31,1%	148	100,0%
	112	207	66,3%	105	33,7%	312	100,0%
	115	90	77,6%	26	22,4%	116	100,0%
	116	7	100,0%	0	,0%	7	100,0%
	117	23	95,8%	1	4,2%	24	100,0%
	126	231	65,1%	124	34,9%	355	100,0%
	127	68	43,3%	89	56,7%	157	100,0%
Si	109	93	58,1%	67	41,9%	160	100,0%
	112	78	53,1%	69	46,9%	147	100,0%
	115	114	73,1%	42	26,9%	156	100,0%
	116	8	50,0%	8	50,0%	16	100,0%
	117	83	85,6%	14	14,4%	97	100,0%
	126	27	49,1%	28	50,9%	55	100,0%
	127	3	18,8%	13	81,3%	16	100,0%
<b>Total</b>		1134	64,2%	632	35,8%	1766	100,0%

En la Tabla 5.3.4.1 se analiza el porcentaje de datos ausentes y válidos por tipologías y nivel de calidad biológica de las aguas. El porcentaje de datos ausentes oscila en torno a 35,8% llamando la atención los elevados porcentaje de datos ausentes en los ríos de la tipología 127 tanto en las muestras de baja calidad biológica (81,3%) como en las de alta (56,7%) así como el de muestras de baja calidad en los ríos de la tipología 116 (50,9%) todo lo cual, unido al escaso número de muestras en algunos de estos grupos hace que los resultados correspondientes a estas tipologías deban tomarse con las debidas precauciones.



**Figura 5.3.4.1: Diagrama de cajas del logaritmo del nivel de fosfatos por tipologías y nivel de calidad biológica de las aguas**



**Figura 5.3.4.2: Barras de error de la media del logaritmo del nivel de fosfatos por tipologías y nivel de calidad biológica de las aguas (nivel de confianza 95%)**

En las Figuras 5.3.4.1 y 5.3.4.2 así como en la Tabla 5.3.4.2. se presentan los resultados del estudio comparativo del logaritmo del nivel de fosfatos entre las aguas de

baja y alta calidad biológica distinguiendo tipologías. Más concretamente, se muestran sus diagramas de cajas (Figura 5.3.4.1) y las barras de error del 95% de confianza (Figura 5.3.4.2) así como sus valores medios y medianos (Tabla 5.3.4.2). Finalmente, se dan los resultados obtenidos al realizar contrastes de Mann-Whitney presentando, para cada tipología y en total, los valores de los rangos medios de las muestras de alta y baja calidad biológicas así como el pvalor del contraste (Tabla 5.3.4.2).

Se observa que, con la excepción de los ríos de tipologías 116 y 127, la media del logaritmo del nivel de fosfatos tiende a ser significativamente inferior en las aguas de alta calidad biológica (ver Tabla 5.3.4.2). En el resto de tipologías aunque dicho patrón se conserva, no se aprecian diferencias que sean estadísticamente significativas (ver Figura 5.3.4.2).

**Tabla 5.3.4.2: Análisis descriptivo del logaritmo del nivel de fosfatos por tipologías**

Tipología	Casos			Media			Mediana			Error típico de la media			Rangos promedio		Pvalor Mann-Whitney
	¿Aguas de baja calidad?			¿Aguas de baja calidad?			¿Aguas de baja calidad?			¿Aguas de baja calidad?			¿Aguas de baja calidad?		
	No	Si	Total	No	Si	Total	No	Si	Total	No	Si	Total	No	Si	
<b>109</b>	102	93	195	-2,73	-1,50	-2,14	-3,00	-1,51	-3,00	0,07	0,12	0,08	69,52	129,24	0,000
<b>112</b>	207	78	285	-2,60	-1,94	-2,42	-3,00	-2,02	-3,00	0,05	0,12	0,05	128,21	182,26	0,000
<b>115</b>	90	114	204	-2,46	-1,72	-2,05	-3,00	-1,64	-1,98	0,07	0,08	0,06	74,78	124,39	0,000
<b>116</b>	7	8	15	-2,04	-2,17	-2,11	-2,16	-2,01	-2,12	0,26	0,20	0,16	8,00	8,00	1,000
<b>117</b>	23	83	106	-1,92	-1,82	-1,84	-1,90	-1,77	-1,80	0,08	0,06	0,05	46,54	55,43	0,220
<b>126</b>	231	27	258	-2,77	-1,91	-2,68	-3,00	-1,69	-3,00	0,03	0,15	0,04	121,56	197,46	0,000
<b>127</b>	68	3	71	-2,96	-3,00	-2,96	-3,00	-3,00	-3,00	0,02	0,00	0,02	36,07	34,50	0,904
<b>Total</b>	728	406	1134	-2,66	-1,76	-2,34	-3,00	-1,76	-3,00	0,02	0,05	0,03	450,69	776,95	0,000

**Tabla 5.3.4.3: Modelos estimados para el logaritmo del nivel de fosfatos distinguiendo y sin distinguir tipologías**

Variables seleccionadas	Distinguiendo tipologías			Sin distinguir tipologías		
	$\beta$	Error Estándar	Pvalor	$\beta$	Error Estándar	Pvalor
<b>Constante</b>	3,2096	0,4241	0,0000	2,2996	0,2626	0,0000
<b>LPO4</b>				1,1852	0,1092	0,0000
<b>TIP_112</b>	-2,1236	0,6224	0,0006			
<b>LPO4_109</b>	1,6850	0,2040	0,0000			
<b>LPO4_112</b>	1,0271	0,2178	0,0000			
<b>LPO4_115</b>	1,5407	0,2118	0,0000			

**Tabla 5.3.4.4: Umbrales estimados para el nivel de fosfatos y validación predictiva de los mismos distinguiendo y sin distinguir tipologías (en negrita señalados los umbrales utilizados en el proceso de comparación por tipologías descrito en la Tabla 5.3.4.5)**

Riesgo	Umbrales distinguiendo tipologías			%Aciertos			Umbrales sin distinguir tipologías	%Aciertos		
	109	112	115	Cumplen	No cumplen	Totales		Cumplen	No cumplen	Totales
<b>30,00%</b>	0,090	0,152	0,072	55,90%	80,40%	66,10%	0,070	67,70%	74,40%	70,50%
<b>31,00%</b>	0,093	0,159	0,074	55,90%	80,00%	65,90%	0,073	67,70%	74,40%	70,50%
<b>32,00%</b>	0,095	0,167	0,076	57,60%	78,60%	66,40%	0,076	71,40%	72,60%	71,90%
<b>33,00%</b>	0,098	0,174	0,079	58,10%	78,20%	66,50%	<b>0,079</b>	<b>71,40%</b>	<b>72,60%</b>	<b>71,90%</b>
<b>34,00%</b>	<b>0,100</b>	<b>0,182</b>	<b>0,081</b>	<b>74,20%</b>	<b>69,80%</b>	<b>72,40%</b>	0,082	72,40%	71,90%	72,20%
<b>35,00%</b>	0,103	0,190	0,083	74,90%	68,80%	72,40%	0,085	73,40%	70,90%	72,40%

**Tabla 5.3.4.5: Resultados obtenidos por tipologías utilizando los umbrales del nivel de fosfatos de la Tabla 5.3.4.4 así como los umbrales proporcionados por la CHE**

Tipología	Calidad Observada	Calidad pronosticada						% Total aciertos		
		Distinguiendo tipologías		Sin distinguir tipologías		CHE		Distinguiendo tipologías	Sin distinguir tipologías	CHE
		Alta	Baja	Alta	Baja	Alta	Baja			
<b>109</b>	<b>Alta</b>	86,27%	13,73%	85,29%	14,71%	94,12%	5,88%	78,97%	78,97%	68,72%
	<b>Baja</b>	29,03%	70,97%	27,96%	72,04%	59,14%	40,86%			
<b>112</b>	<b>Alta</b>	89,37%	10,63%	69,08%	30,92%	93,24%	6,76%	74,04%	67,37%	73,68%
	<b>Baja</b>	66,67%	33,33%	37,18%	62,82%	78,21%	21,79%			
<b>115</b>	<b>Alta</b>	25,56%	74,44%	61,11%	38,89%	91,11%	8,89%	63,73%	71,57%	54,41%
	<b>Baja</b>	6,14%	93,86%	20,18%	79,82%	74,56%	25,44%			
<b>Total</b>	<b>Alta</b>	74,19%	25,81%	71,43%	28,57%	92,98%	7,02%	72,37%	71,93%	66,52%
	<b>Baja</b>	30,18%	69,82%	27,37%	72,63%	70,53%	29,47%			

En la Tabla 5.3.4.3 se muestra el modelo seleccionado utilizando como variables independientes los indicadores de la tipología del río (denotados como TIP\_x donde x es el número de la tipología) y sus interacciones con el logaritmo del nivel de fosfatos (denotados como LPO4\_x) y aplicando un procedimiento de selección de variables por eliminación hacia atrás tomando pvalor de entrada 0,01 y como pvalor de salida 0,02. Así mismo y, por razones comparativas, se muestra el modelo estimado utilizando como variables independientes una constante y dicho logaritmo, el cual no tiene en cuenta, por tanto, la información proporcionada por la tipología del río. Las estimaciones se han realizado utilizando, únicamente, las observaciones de las tipologías 109, 112 y 115 dado que, para el resto, los resultados obtenidos no fueron fiables debido al escaso tamaño muestral disponible en algunos de los grupos comparados.

Las estimaciones de los parámetros del modelo (ver Tabla 5.3.4.3) y del proceso de validación (ver Tabla 5.3.4.5) ponen de manifiesto que en los ríos de las tipologías 109, 112 y 115 el nivel de fosfatos discrimina entre las muestras de agua de alta y baja calidad, siendo los coeficientes  $\beta$  estimados significativamente positivos.

Utilizando la expresión (4.2), el umbral para un riesgo  $0 < p < 1$  vendrá dado por la expresión:

$$\text{Umbral}_p = \exp\left(\frac{\log\left(\frac{p}{1-p}\right) - 2,30}{1,19}\right)$$

si no se distinguen tipologías y

$$\text{Umbral}_p = \exp\left(\frac{\log\left(\frac{p}{1-p}\right) - 3,21 + 2,12 * I(\text{Tip} = 112)}{1,69 * I(\text{Tip} = 109) + 1,03 * I(\text{Tip} = 112) + 1,54 * I(\text{Tip} = 115)}\right)$$

si se distinguen tipologías, donde  $I(\text{Tip}=x)$  es la función indicador de la tipología x que vale 1 si el río tiene tipología x y 0 en otro caso. En ambos casos se detectará riesgo alto de baja calidad biológica del agua si el nivel de fosfatos de ésta es mayor o igual que dicho umbral.

En la Tabla 5.3.4.4 se muestran los umbrales estimados para una red de valores del riesgo p, así como los porcentajes de éxito entre las muestras que cumplen los

requisitos de calidad, entre las que no la cumplen y el porcentaje de éxitos totales calculados utilizando las expresiones (4.5) a (4.7). Así, por ejemplo, si el nivel mediano de fosfatos es superior a 0,082 mg/L PO<sub>4</sub> el modelo estimado sin distinguir tipologías estima que existe un riesgo superior de al menos un 34% de que las aguas del río sean de baja calidad biológica. Distinguiendo tipologías este umbral cambia a 0,1 mg/L PO<sub>4</sub> si el río es de la tipología 109; 0,182 mg/L PO<sub>4</sub> si es de tipología 112 y 0,081 mg/L PO<sub>4</sub> si es de la 115. Utilizando estos umbrales los porcentajes de éxito para cada uno de los dos procedimientos fueron, respectivamente, 72,40% y 74,20% entre las muestras que cumplen; 71,90% y 69,80% entre las que no cumplen y 72,20% y 72,40% en total.

Finalmente, en la Tabla 5.3.4.5 se muestran los resultados obtenidos por tipologías utilizando los umbrales en negrita (distinguiendo y sin distinguir tipologías) junto a los correspondientes al procedimiento utilizado por la CHE que clasifica las aguas como de alta calidad si el nivel de fosfatos es inferior a 0,30 mg/L PO<sub>4</sub>. Así, por ejemplo, en la tipología 109, los porcentaje de éxito en la muestras que cumplen los estándares de calidad fueron 86,27%, 85,29% y 94,12% utilizando umbrales distinguiendo por tipologías, sin distinguir y el procedimiento de la CHE, respectivamente. Estos porcentajes fueron 70,97%, 72,04% y 40,86% entre las que no cumplen dichos estándares.

Se observa, en primer lugar, que los límites proporcionados por la CHE son poco operativos al clasificar un 70,53% aguas de baja calidad como buenas. Por otro lado, aunque a nivel global no existen diferencias significativas en los resultados obtenidos, en términos de porcentajes de aciertos (72,37% frente a 71,93%), si los analizamos por tipologías los porcentajes de aciertos en las muestras de alta y baja calidad están más equilibrados cuando los umbrales se fijan sin distinguir tipologías que cuando se distinguen, lo cual los hace menos fiables al no saber, en una situación real, en qué situación están las aguas antes de ser analizadas.

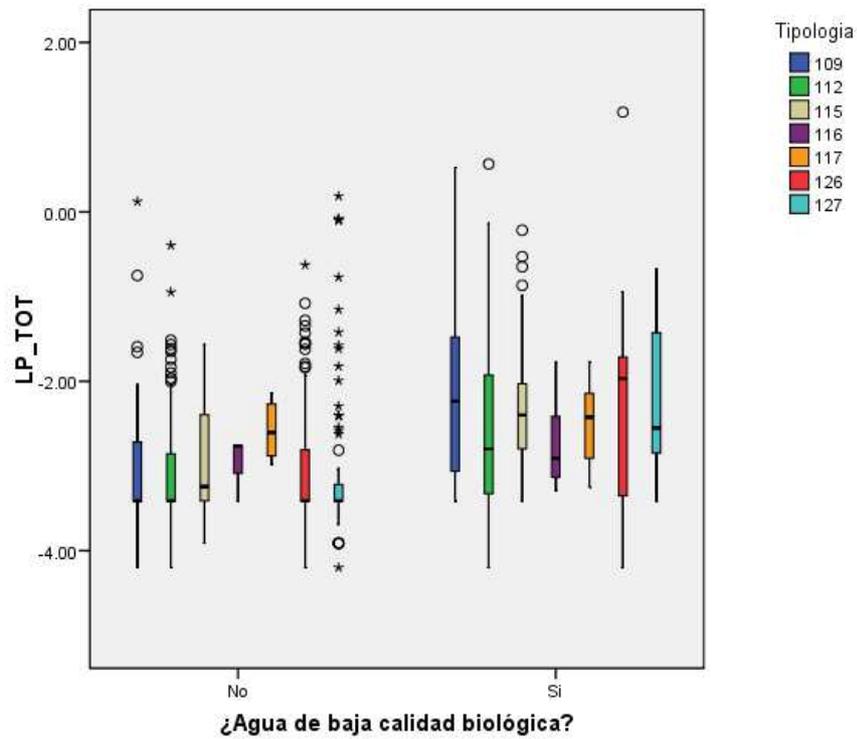
### 5.3.5. Umbrales de calidad biológica para el fósforo total

De acuerdo con estudios realizados en proyectos anteriores se ha tomado  $g(X) = \log(X)$  al mostrar el indicador analizado problemas serios de falta de normalidad por una elevada asimetría positiva.

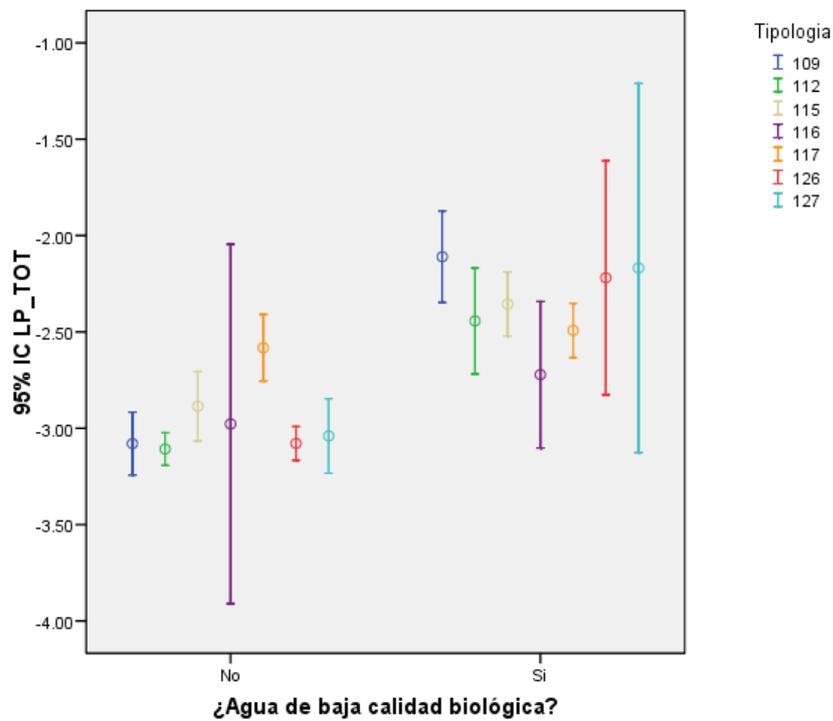
**Tabla 5.3.5.1: Datos ausentes y válidos del logaritmo del nivel de fósforo total por tipologías y nivel de calidad biológica de las aguas**

¿Aguas de baja calidad biológica?	Tipología	Válidos		Perdidos		Total	
		Casos	Porcentaje	Casos	Porcentaje	Casos	Porcentaje
No	109	74	50,0%	74	50,0%	148	100,0%
	112	186	59,6%	126	40,4%	312	100,0%
	115	51	44,0%	65	56,0%	116	100,0%
	116	3	42,9%	4	57,1%	7	100,0%
	117	14	58,3%	10	41,7%	24	100,0%
	126	184	51,8%	171	48,2%	355	100,0%
	127	93	59,2%	64	40,8%	157	100,0%
Si	109	79	49,4%	81	50,6%	160	100,0%
	112	63	42,9%	84	57,1%	147	100,0%
	115	73	46,8%	83	53,2%	156	100,0%
	116	11	68,8%	5	31,3%	16	100,0%
	117	39	40,2%	58	59,8%	97	100,0%
	126	18	32,7%	37	67,3%	55	100,0%
	127	7	43,8%	9	56,3%	16	100,0%
<b>Total</b>		895	50,68%	871	49,32%	1766	100,0%

En la Tabla 5.3.5.1 se analiza el porcentaje de datos ausentes y válidos por tipologías y nivel de calidad biológica de las aguas. El porcentaje de datos ausentes oscila en torno a 49,32% llamando la atención el elevado porcentaje de datos ausentes en los ríos de la tipología 126 en las muestras de baja calidad biológica (67,3%) lo cual hace que los resultados correspondientes a esta tipología deban tomarse con las debidas precauciones.



**Figura 5.3.5.1: Diagrama de cajas del logaritmo del nivel de fósforo total por tipologías y nivel de calidad biológica de las aguas**



**Figura 5.3.5.2: Barras de error de la media del logaritmo del nivel de fósforo total por tipologías y nivel de calidad biológica de las aguas (nivel de confianza 95%)**

En las Figuras 5.3.5.1 y 5.3.5.2 así como en la Tabla 5.3.5.2. se presentan los resultados del estudio comparativo del logaritmo del nivel de fósforo total entre las aguas de baja y alta calidad biológica distinguiendo tipologías. Más concretamente se muestran sus diagramas de cajas (Figura 5.3.5.1) y las barras de error del 95% de confianza (Figura 5.3.5.2) así como sus valores medios y medianos (Tabla 5.3.5.2). Finalmente, se dan los resultados obtenidos al realizar contrastes de Mann-Whitney presentando, para cada tipología y en total, los valores de los rangos medios de las muestras de alta y baja calidad biológicas así como el pvalor del contraste (Tabla 5.3.5.2).

Se observa que, con la excepción de los ríos de tipologías 116 y 117, la mediana del logaritmo del nivel de fósforo total tiende a ser significativamente inferior en las aguas de alta calidad biológica (ver Tabla 5.3.5.2). En el resto de tipologías aunque dicho patrón se conserva no se aprecian diferencias que sean estadísticamente significativas (ver Figura 5.3.5.2).

**Tabla 5.3.5.2: Análisis descriptivo del logaritmo del nivel de fósforo total por tipologías**

Tipología	Casos			Media			Mediana			Error típico de la media			Rangos promedio		Pvalor Mann-Whitney
	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		Total	¿Aguas de baja calidad?		
No	Si	No		Si	No		Si	No		Si	No		Si	No	Si
<b>109</b>	74	79	153	-3,08	-2,11	-2,58	-3,41	-2,23	-2,94	0,08	0,12	0,08	53,58	98,94	0,000
<b>112</b>	186	63	249	-3,11	-2,44	-2,94	-3,41	-2,80	-3,35	0,04	0,14	0,05	112,08	163,14	0,000
<b>115</b>	51	73	124	-2,89	-2,36	-2,57	-3,24	-2,40	-2,48	0,09	0,08	0,07	46,64	73,58	0,000
<b>116</b>	3	11	14	-2,98	-2,72	-2,78	-2,76	-2,91	-2,86	0,22	0,17	0,14	7,33	7,55	1,000
<b>117</b>	14	39	53	-2,58	-2,49	-2,52	-2,61	-2,42	-2,50	0,08	0,07	0,06	25,75	27,45	0,724
<b>126</b>	184	18	202	-3,08	-2,22	-3,00	-3,41	-1,97	-3,41	0,04	0,29	0,05	96,63	151,25	0,000
<b>127</b>	93	7	100	-3,04	-2,17	-2,98	-3,41	-2,55	-3,41	0,10	0,39	0,10	48,43	78,00	0,004
<b>Total</b>	605	290	895	-3,05	-2,33	-2,82	-3,41	-2,41	-3,21	0,03	0,05	0,03	369,18	612,44	0,000

**Tabla 5.3.5.3: Modelos estimados para el logaritmo del nivel de fósforo total distinguiendo y sin distinguir tipologías**

Variables seleccionadas	Distinguiendo tipologías			Sin distinguir tipologías		
	$\beta$	Error Estándar	Pvalor	$\beta$	Error Estándar	Pvalor
<b>Constante</b>	2,8333	0,4028	0,0000	3,0225	0,4001	0,0000
<b>LP_TOT</b>				1,2361	0,1394	0,0000
<b>LP_TOT_109</b>	1,0602	0,1488	0,0000			
<b>LP_TOT_112</b>	1,3660	0,1457	0,0000			
<b>LP_TOT_115</b>	0,9561	0,1586	0,0000			

**Tabla 5.3.5.4: Umbrales para el nivel de fósforo total y validación predictiva de los mismos distinguiendo y sin distinguir tipologías (en negrita señalados los umbrales utilizados en el proceso de comparación por tipologías descrito en la Tabla 5.3.5.5)**

Riesgo	Umbrales distinguiendo tipologías			%Aciertos			Umbrales sin distinguir tipologías	%Aciertos		
	109	112	115	Cumplen	No cumplen	Totales		Cumplen	No cumplen	Totales
<b>35,00%</b>	0,039	0,027	0,080	64,60%	75,30%	69,00%	<b>0,053</b>	<b>69,80%</b>	<b>70,70%</b>	<b>70,20%</b>
<b>36,00%</b>	0,040	0,028	0,082	66,20%	74,40%	69,60%	0,054	70,10%	69,30%	69,80%
<b>37,00%</b>	0,042	0,030	0,085	66,90%	74,00%	69,80%	0,056	71,70%	68,40%	70,30%
<b>38,00%</b>	0,044	0,031	0,088	67,20%	74,00%	70,00%	0,058	72,00%	66,50%	69,80%
<b>39,00%</b>	<b>0,045</b>	<b>0,032</b>	<b>0,091</b>	<b>71,10%</b>	<b>73,00%</b>	<b>71,90%</b>	0,060	73,30%	65,10%	70,00%
<b>40,00%</b>	0,047	0,034	0,093	79,10%	70,20%	75,50%	0,062	74,00%	62,80%	69,40%

**Tabla 5.3.5.5: Resultados obtenidos por tipologías utilizando los umbrales del nivel de fósforo total de la Tabla 5.3.5.4 así como los umbrales proporcionados por la CHE**

Tipología	Calidad Observada	Calidad pronosticada						% Total aciertos		
		Distinguiendo tipologías		Sin distinguir tipologías		CHE		Distinguiendo tipologías	Sin distinguir tipologías	CHE
		Alta	Baja	Alta	Baja	Alta	Baja			
<b>109</b>	<b>Alta</b>	70,27%	29,73%	72,97%	27,03%	93,24%	6,76%	73,20%	72,55%	68,63%
	<b>Baja</b>	24,05%	75,95%	27,85%	72,15%	54,43%	45,57%			
<b>112</b>	<b>Alta</b>	89,78%	10,22%	73,12%	26,88%	91,94%	8,06%	76,71%	69,48%	76,71%
	<b>Baja</b>	61,90%	38,10%	41,27%	58,73%	68,25%	31,75%			
<b>115</b>	<b>Alta</b>	3,92%	96,08%	52,94%	47,06%	84,31%	15,69%	60,48%	68,55%	54,03%
	<b>Baja</b>	0,00%	100,00%	20,55%	79,45%	67,12%	32,88%			
<b>Total</b>	<b>Alta</b>	71,06%	28,94%	69,77%	30,23%	91,00%	9,00%	71,86%	70,15%	69,01%
	<b>Baja</b>	26,98%	73,02%	29,30%	70,70%	62,79%	37,21%			

En la Tabla 5.3.5.3 se muestra el modelo seleccionado utilizando como variables independientes los indicadores de la tipología del río (denotados como TIP\_x donde x es el número de la tipología) y sus interacciones con el logaritmo del nivel de fósforo total (denotados como LP\_TOT\_x) y aplicando un procedimiento de selección de variables por eliminación hacia atrás tomando pvalor de entrada 0,01 y como pvalor de salida 0,02. Así mismo y, por razones comparativas, se muestra el modelo estimado utilizando como variables independientes una constante y dicho logaritmo, el cual no tiene en cuenta, por tanto, la información proporcionada por la tipología del río. Las estimaciones se han realizado utilizando, únicamente, las observaciones de las tipologías 109, 112 y 115 dado que, para el resto, los resultados obtenidos no fueron fiables debido al escaso tamaño muestral disponible en algunos de los grupos comparados.

Las estimaciones de los parámetros del modelo (ver Tabla 5.3.5.3) y del proceso de validación (ver Tabla 5.3.5.5) ponen de manifiesto que en los ríos de las tipologías 109, 112 y 115 el nivel de fósforo total discrimina significativamente entre las muestras de agua de alta y baja calidad biológicas, siendo los coeficientes  $\beta$  estimados significativamente positivos.

Utilizando la expresión (4.2), el umbral para un riesgo  $0 < p < 1$  vendrá dado por la expresión:

$$\text{Umbral}_p = \exp \left( \frac{\log \left( \frac{p}{1-p} \right) - 3,02}{1,24} \right)$$

si no se distinguen tipologías y

$$\text{Umbral}_p = \frac{\log \left( \frac{p}{1-p} \right) - 2,83}{1,06 * I(\text{Tip} = 109) + 1,37 * I(\text{Tip} = 112) + 0,96 * I(\text{Tip} = 115)}$$

si se distinguen tipologías, donde  $I(\text{Tip}=x)$  es la función indicador de la tipología x que vale 1 si el río tiene tipología x y 0 en otro caso. En ambos casos se detectará riesgo alto de baja calidad biológica del agua si el nivel de fósforo total de ésta es mayor o igual que dicho umbral.

En la Tabla 5.3.5.4 se muestran los umbrales estimados para una red de valores del riesgo p, así como los porcentajes de éxito entre las muestras que cumplen los requisitos de calidad, entre las que no la cumplen y el porcentaje de éxitos totales calculados utilizando las expresiones (4.5) a (4.7). Así, por ejemplo, si el nivel mediano de fósforo total es superior a 0,060 mg/L P el modelo estimado sin distinguir tipologías estima que existe un riesgo superior de al menos un 39% de que las aguas del río sean de baja calidad biológica. Distinguiendo tipologías este umbral cambia a 0,045 mg/L P si el río es de la tipología 109, 0,032 mg/L P si es de tipología 112 y 0,091 mg/L P si es de la 115. Utilizando estos umbrales los porcentajes de éxito para cada uno de los dos procedimientos fueron, respectivamente, 73,30% y 71,10% entre las muestras que cumplen; 65,10% y 73% entre las que no cumplen y 70,00% y 71,90% en total.

Finalmente, en la Tabla 5.3.5.5 se muestran los resultados obtenidos por tipologías utilizando los umbrales en negrita (distinguiendo y sin distinguir tipologías) junto a los correspondientes al procedimiento utilizado por la CHE el cual clasifica el agua como de alta calidad si su nivel de fósforo total es inferior a 0,12 mg/L P. Así, por ejemplo, en la tipología 109, los porcentaje de éxito en la muestras que cumplen los estándares de calidad fueron 70,27%, 72,97% y 93,24% utilizando umbrales distinguiendo por tipologías, sin distinguir y el procedimiento de la CHE, respectivamente. Estos porcentajes fueron 75,95%, 72,15% y 45,57% entre las que no cumplen dichos estándares.

Se observa, en primer lugar, que los límites proporcionados por la CHE son poco operativos al clasificar un 62,79% aguas de baja calidad como buenas. Por otro lado, aunque a nivel global el procedimiento que distingue entre tipologías obtiene mejores resultados que el que no, en términos de porcentajes de aciertos (71,86% frente a 70,15%), si analizamos por tipologías los porcentajes de aciertos en las muestras de alta y baja calidad están más equilibrados cuando los umbrales se fijan sin distinguir tipologías que cuando se distinguen, lo cual los hace menos fiables al no saber, en una situación real, en qué situación están las aguas antes de ser analizadas.

## 6. CONCLUSIONES

En este trabajo se ha llevado a cabo un proceso estadístico de estimación de umbrales para un conjunto de indicadores físico-químicos habitualmente utilizados por la CHE para determinar el estado de las masas de agua del tipo de río según la DMA en la cuenca del Ebro. El procedimiento seguido ha tenido en cuenta la tipología de la masa de agua, definida según la IPH, de la que se ha extraído la muestra y se ha analizado, en particular, si el tener en cuenta dicha información mejora significativamente el proceso de determinación del estado de la masa de agua respecto al procedimiento seguido habitualmente por la CHE. Así mismo, la determinación del estado biológico de cada muestra analizada se ha llevado a cabo a partir de los valores de los indicadores IBMWP e IPS, de manera que una muestra de agua se ha considerado de “baja calidad” (estado según los indicadores biológicos inferior a bueno) si al menos uno de dichos indicadores presenta un estado menor que bueno según los requisitos establecidos por la CHE.

Los umbrales se han determinado utilizando, para cada indicador, modelos de regresión logística binaria así como procedimientos de selección de variables que tienen en cuenta la tipología del río. Los niveles de riesgo se han tomado procurando que los porcentajes de éxitos totales, tanto en muestras de “alta calidad” como de “baja calidad”, fueran lo más elevados y similares posible, buscando, de esta forma, un equilibrio en ambos casos para el comportamiento del procedimiento de clasificación.

En la Tabla 6.1 se exponen, de forma resumida, los resultados obtenidos para cada indicador y cada tipología presentando, por un lado, las estimaciones de los umbrales y, por otro, las de los porcentajes de aciertos tanto en las muestras de agua de “alta calidad” como en las de “baja calidad” según los parámetros biológicos.

Las conclusiones extraídas del presente estudio son:

- 1) No se han podido estimar umbrales para los ríos de las **tipologías 111** (ríos de montaña silíceas), **116** (ríos del eje mediterráneo-continental mineralizado) y **127** (ríos de alta montaña). En el caso de las tipologías 111 y 127 debido a que la mayor parte de las muestras disponibles eran de alta calidad (en el caso de la tipología 111 todas las muestras) imposibilitando llevar a cabo el proceso de discriminación estadístico que exige tener un número mínimo de muestras de

ambos tipos. En el caso de la tipología 116 el número total de muestras resultó demasiado bajo para permitir este estudio.

- 2) Con la única excepción del indicador pH, **tener en cuenta la tipología no mejora sustancialmente los resultados obtenidos utilizando un umbral común** para todas ellas. Esta afirmación se puede realizar puesto que el procedimiento que distingue tipologías muestra un comportamiento mucho más desequilibrado en lo que al porcentaje de éxitos se refiere tendiendo, en muchos casos, a clasificar las muestras siguiendo, en mayor o menor grado, la regla de la mayoría. La posible causa podríamos encontrarla en unos **tamaños muestrales no muy elevados** y, en general, muy desequilibrados en cuanto a la composición por grupos se refiere.
- 3) En los ríos de las tipologías 117 (ríos de grandes ejes en ambiente mediterráneo) y 126 (ríos de montaña húmeda calcárea) tan sólo se calcularon umbrales para los indicadores DQO (demanda química de oxígeno) y NH<sub>4</sub> (nivel de amonio total) dado que para el resto de los indicadores la potencia discriminante fue muy escasa clasificando las muestras siguiendo la regla de la mayoría, es decir, todas las muestras de agua de ríos 117 de baja calidad y todas las de ríos de tipología 126 de buena calidad.
- 4) **Los porcentajes de éxito de los umbrales son estadísticamente significativos** oscilando en torno a un 68,52% en el caso de las muestras de alta calidad y un 62,82% en las muestras de baja calidad. Si distinguimos por indicadores, los mejores resultados corresponden a los **nitritos** y a los **fosfatos** con porcentajes de éxitos muy equilibrados que oscilan en torno a un 71.5%.
- 5) **Los procedimientos de clasificación del estado según los parámetros físico-químicos utilizados por la CHE tienden a ser muy permisivos** clasificando la mayor parte de las muestras como de alta calidad, provocando un comportamiento muy desequilibrado en cuanto al porcentaje de aciertos. Este valor es obviamente muy elevado en las muestras de alta calidad pero muy pequeño en las de baja calidad, lo cual no los haría demasiado recomendables si se quisiese calcular el estado de la masa solamente a partir de los parámetros físico- químicos .

- 6) Conviene hacer notar, finalmente, que el **no cumplimiento de los umbrales propuestos** debería utilizarse sólo como **signo de un posible problema** de baja calidad en las aguas, pero para confirmarlo necesitaría realizarse un examen más a fondo del estado del río mediante indicadores biológicos.

En este estudio el proceso de determinación de umbrales se ha realizado para cada indicador por separado, no obstante, sería interesante llevar a cabo procedimientos de determinación de umbrales simultáneos utilizando, por ejemplo, técnicas estadísticas de construcción de árboles de decisión. Este planteamiento queda propuesta para una futura línea de investigación.

**Tabla 6.1. Umbrales recomendados por tipologías y porcentaje de aciertos estimados en cada una de ellas**

Tipología	Regla de decisión y % aciertos	Indicador Físico-Químico								
		pH	Cond20 (µS/cm)	O2 (mg/L)	DQO (mg/L)	NH4 (mg/L)	NO2 (mg/L)	NO3 (mg/L)	PO4 (mg/L)	PTOT (mg/L)
109	Agua de baja calidad si	<8,27	>703,45	<9,50	>5,07	>0,075	>0,045	>10,16	>0,079	PTOT>0,053
	% Aciertos Baja Calidad	66,38%	56,90%	55,36%	82,05%	60,19%	78,38%	66,67%	85,29%	72,97%
	% Aciertos Alta Calidad	52,71%	89,92%	66,93%	53,49%	53,33%	79,45%	72,38%	72,04%	72,15%
112	Agua de baja calidad si	<7,85	>703,45	<9,50	>5,07	>0,075	>0,045	>10,16	>0,079	PTOT>0,053
	% Aciertos Baja Calidad	69,17%	69,55%	66,28%	85,71%	66,93%	73,08%	60,74%	69,08%	73,12%
	% Aciertos Alta Calidad	47,41%	55,17%	50,00%	34,78%	46,59%	63,33%	57,95%	62,82%	58,73%
115	Agua de baja calidad si	<8,33	>703,45	<9,50	>5,07	>0,075	>0,045	>10,16	>0,079	PTOT>0,053
	% Aciertos Baja Calidad	68,52%	75,93%	57,01%	79,03%	50,00%	61,70%	73,96%	61,11%	52,94%
	% Aciertos Alta Calidad	65,94%	56,52%	58,21%	61,00%	58,33%	76,19%	57,76%	79,82%	79,45%
117	Agua de baja calidad si				>5,07	>0,075				
	% Aciertos Baja Calidad				66,67%	47,83%				
	% Aciertos Alta Calidad				70,18%	66,67%				
126	Agua de baja calidad si				>5,07	>0,075				
	% Aciertos Baja Calidad				85,34%	71,38%				
	% Aciertos Alta Calidad				84,00%	73,33%				