ESCUELA TÉCNICA
SUPERIOR DE INGENIEROS
DE MINAS Y ENERGÍA

POLITÉCNICA

"Ingeniamos el futuro"

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE MINAS Y ENERGÍA

Propuesta metodológica para el análisis y protección de la
calidad del agua subterránea de acuerdo con los
requerimientos de la Directiva Marco y la Directiva derivada
de Aguas Subterráneas

TESIS DOCTORAL

Autor: Juan Grima Olmedo (Ingeniero de Minas)
Directora: María Teresa González Aguado

Madrid, 2016

Departamento de Ingeniería Geológica y Minera

**ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE MINAS Y
ENERGÍA**

**Propuesta metodológica para el análisis y protección de la
calidad del agua subterránea de acuerdo con los
requerimientos de la Directiva Marco y la Directiva derivada de
Aguas Subterráneas**

Autor: Juan Grima Olmedo (Ingeniero de Minas)

**Directora: Prof. Dra. María Teresa González Aguado. Ingeniera
de Minas**

2016

Agradecimientos

Con estas líneas me gustaría expresar mi gratitud a todas aquellas personas que han hecho posible esta Tesis. Aun así, muchas quedarán sin nombrar, por lo que pido disculpas de antemano. Especial mención merece Enrique Chacón, con quien comencé esta Tesis y que por culpa de un trágico accidente no pudo ver concluida.

A mi directora de tesis, que con su participación ha posibilitado llevar a buen puerto el esta investigación, sin cuyos atinados comentarios y correcciones hubiera sido imposible concluir.

Al Dr. Juan Antonio López Geta, cuyo apoyo personal y profesional desde el inicio de mi carrera ha sido clave en mi desarrollo en el apasionante mundo de la hidrogeología.

Al Dr. Juan Antonio Luque Espinar por su colaboración y ayuda prestada, que ha sido imprescindible para la realización de este trabajo.

A mis compañeros de la Oficina del IGME en Valencia y Granada, Bruno, José Antonio, Olga, Elisabeth, Juan Carlos, David, Julio, Rafael y Carlos por su participación en las discusiones científicas y técnicas con las que les he abrumado. Sus aportaciones han contribuido de manera muy notable a su mejora.

A mi familia, por inculcarme el deseo de superación e instarme a afrontar nuevos retos y en especial a mi padre, al que le habría gustado ver finalizado este trabajo.

Dedico esta Tesis a mi padre y a Enrique Chacón, in memóriam, a mi madre y a mis hermanos por su apoyo constante e incondicional

Resumen

La protección de las aguas subterráneas es una prioridad de la política medioambiental de la UE. Por ello ha establecido un marco de prevención y control de la contaminación, que incluye provisiones para evaluar el estado químico de las aguas y reducir la presencia de contaminantes en ellas.

Las herramientas fundamentales para el desarrollo de dichas políticas son la Directiva Marco del Agua y la Directiva Hija de Aguas Subterráneas. Según ellas, las aguas se consideran en buen estado químico si:

- la concentración medida o prevista de **nitratos** no supera los 50 mg/l y la de ingredientes activos de **plaguicidas**, de sus metabolitos y de los productos de reacción no supera el 0,1 µg/l (0,5 µg/l para el total de los plaguicidas medidos)
- la concentración de determinadas sustancias de riesgo es inferior al **valor umbral** fijado por los Estados miembros; se trata, como mínimo, del amonio, arsénico, cadmio, cloruro, plomo, mercurio, sulfatos, tricloroetileno y tetracloroetileno
- la concentración de cualquier otro contaminante se ajusta a la definición de buen estado químico enunciada en el anexo V de la Directiva marco sobre la política de aguas
- en caso de superarse el valor correspondiente a una **norma de calidad** o a un **valor umbral**, una investigación confirma, entre otros puntos, la falta de **riesgo significativo** para el medio ambiente.

Analizar el comportamiento estadístico de los datos procedentes de la red de seguimiento y control puede resultar considerablemente complejo, debido al sesgo positivo que suelen presentar dichos datos y a su distribución asimétrica, debido a la existencia de valores anómalos y diferentes tipos de suelos y mezclas de contaminantes.

Además, la distribución de determinados componentes en el agua subterránea puede presentar concentraciones por debajo del límite de detección o no ser estacionaria debida a la existencia de tendencias lineales o estacionales. En el

primer caso es necesario realizar estimaciones de esos valores desconocidos, mediante procedimientos que varían en función del porcentaje de valores por debajo del límite de detección y el número de límites de detección aplicables. En el segundo caso es necesario eliminar las tendencias de forma previa a la realización de contrastes de hipótesis sobre los residuos.

Con esta tesis se ha pretendido establecer las bases estadísticas para el análisis riguroso de los datos de las redes de calidad con objeto de realizar la evaluación del estado químico de las masas de agua subterránea para la determinación de tendencias al aumento en la concentración de contaminantes y para la detección de empeoramientos significativos, tanto en los casos que se ha fijado un estándar de calidad por el organismo medioambiental competente como en aquéllos que no ha sido así.

Para diseñar una metodología que permita contemplar la variedad de casos existentes, se han analizado los datos de la Red Oficial de Seguimiento y Control del Estado Químico de las Aguas Subterráneas del Ministerio de Agricultura, Alimentación y Medio Ambiente (Magrama). A continuación, y dado que los Planes Hidrológicos de Cuenca son la herramienta básica de las Directivas, se ha seleccionado la Cuenca del Júcar, dada su designación como cuenca piloto en la estrategia de implementación común (CIS) de la Comisión Europea. El objetivo principal de los grupos de trabajo creados para ello se dirigió a implementar la Directiva Derivada de Agua Subterráneas y los elementos de la Directiva Marco del Agua relacionadas, en especial la toma de datos en los puntos de control y la preparación del primer Plan de Gestión de Cuencas Hidrográficas.

Dada la extensión de la zona y con objeto de analizar una masa de agua subterránea (definida como la unidad de gestión en las Directivas), se ha seleccionado una zona piloto (Plana de Vinaroz Peñiscola) en la que se han aplicado los procedimientos desarrollados con objeto de determinar el estado químico de dicha masa.

Los datos examinados no contienen en general valores de concentración de contaminantes asociados a fuentes puntuales, por lo que para la realización

del estudio se han seleccionado valores de concentración de los datos más comunes, es decir, nitratos y cloruros.

La estrategia diseñada combina el análisis de tendencias con la elaboración de intervalos de confianza cuando existe un estándar de calidad e intervalos de predicción cuando no existe o se ha superado dicho estándar.

De forma análoga se ha procedido en el caso de los valores por debajo del límite de detección, tomando los valores disponibles en la zona piloto de la Plana de Sagunto y simulando diferentes grados de censura con objeto de comparar los resultados obtenidos con los intervalos producidos de los datos reales y verificar de esta forma la eficacia del método.

El resultado final es una metodología general que integra los casos existentes y permite definir el estado químico de una masa de agua subterránea, verificar la existencia de impactos significativos en la calidad del agua subterránea y evaluar la efectividad de los planes de medidas adoptados en el marco del Plan Hidrológico de Cuenca.

Abstract

Groundwater protection is a priority of the EU environmental policy. As a result, it has established a framework for prevention and control of pollution, which includes provisions for assessing the chemical status of waters and reducing the presence of contaminants in it. The measures include:

- criteria for assessing the chemical status of groundwater bodies
- criteria for identifying significant upward trends and sustained concentrations of contaminants and define starting points for reversal of such trends
- preventing and limiting indirect discharges of pollutants as a result of percolation through soil or subsoil.

The basic tools for the development of such policies are the Water Framework Directive and Groundwater Daughter Directive. According to them, the groundwater bodies are considered in good status if:

- measured or predicted concentration of nitrate does not exceed 50 mg / l and the active ingredients of pesticides, their metabolites and reaction products do not exceed 0.1 mg / l (0.5 mg / l for total of pesticides measured)
- the concentration of certain hazardous substances is below the threshold set by the Member States concerned, at least, of ammonium, arsenic, cadmium, chloride, lead, mercury, sulphates, trichloroethylene and tetrachlorethylene
- the concentration of other contaminants fits the definition of good chemical status set out in Annex V of the Framework Directive on water policy
- If the value corresponding to a quality standard or a threshold value is exceeded, an investigation confirms, among other things, the lack of significant risk to the environment.

Analyzing the statistical behaviour of the data from the monitoring networks may be considerably complex due to the positive bias which often presents such information and its asymmetrical distribution, due to the existence of outliers and different soil types and mixtures of pollutants.

Furthermore, the distribution of certain components in groundwater may have concentrations below the detection limit or may not be stationary due to the existence of linear or seasonal trends. In the first case it is necessary to estimate these unknown values, through procedures that vary according to the percentage of values below the limit of detection and the number of applicable limits of detection. In the second case removing trends is needed before conducting hypothesis tests on residuals.

This PhD thesis has intended to establish the statistical basis for the rigorous analysis of data quality networks in order to conduct the evaluation of the chemical status of groundwater bodies for determining upward and sustained trends in pollutant concentrations and for the detection of significant deterioration in cases in which an environmental standard has been set by the relevant environmental agency and those that have not.

Aiming to design a comprehensive methodology to include the whole range of cases, data from the Groundwater Official Monitoring and Control Network of the Ministry of Agriculture, Food and Environment (Magrama) have been analysed. Then, since River Basin Management Plans are the basic tool of the Directives, the Júcar river Basin has been selected. The main reason is its designation as a pilot basin in the common implementation strategy (CIS) of the European Commission. The main objective of the ad hoc working groups is to implement the Daughter Ground Water Directive and elements of the Water Framework Directive related to groundwater, especially the data collection at control stations and the preparation of the first River Basin Management Plan.

Given the size of the area and in order to analyze a groundwater body (defined as the management unit in the Directives), Plana de Vinaroz Peñíscola has been selected as pilot area. Procedures developed to determine the chemical status of that body have been then applied.

The data examined do not generally contain pollutant concentration values associated with point sources, so for the study concentration values of the most common data, i.e., nitrates and chlorides have been selected.

The designed strategy combines trend analysis with the development of confidence intervals when there is a standard of quality and prediction intervals when there is not or the standard has been exceeded.

Similarly we have proceeded in the case of values below the detection limit, taking the available values in Plana de Sagunto pilot area and simulating different degrees of censoring in order to compare the results obtained with the intervals achieved from the actual data and verify in this way the effectiveness of the method.

The end result is a general methodology that integrates existing cases to define the chemical status of a groundwater body, verify the existence of significant impacts on groundwater quality and evaluate the effectiveness of the action plans adopted in the framework of the River Basin Management Plan.

ÍNDICE GENERAL

Capítulo 1	INTRODUCCIÓN
Capítulo 2	ESTADO DEL ARTE
Capítulo 3	MATERIALES, MÉTODOS Y TÉCNICAS ESTADÍSTICAS
Capítulo 4	ANÁLISIS DE LA INFORMACIÓN
Capítulo 5	GESTIÓN DE VALORES POR DEBAJO DEL LÍMITE DE DETECCIÓN
Capítulo 6	RESULTADOS Y DISCUSIÓN
Capítulo 7	CONCLUSIONES
Capítulo 8	RECOMENDACIONES Y LÍNEAS DE INVESTIGACIÓN FUTURAS
Capítulo 9	BIBLIOGRAFÍA

ÍNDICE DE CONTENIDOS

AGRADECIMIENTOS	ii
RESUMEN	iv
ABSTRACT	v
ÍNDICE GENERAL	vii
ÍNDICE DE CONTENIDOS	viii
RELACIÓN DE FIGURAS	x
1. INTRODUCCIÓN	22
1.1. Organización del documento de tesis	22
1.2. Justificación de la investigación	22
1.3. Objetivos de la investigación	24
1.4. Marco normativo para la gestión de los recursos hídricos	26
1.5. Evaluación del riesgo en las aguas subterráneas	31
1.5.1 Impactos naturales y antrópicos	32
1.5.2 Ámbito de los aspectos relacionados con el agua subterránea	34
1.5.3 Agua subterránea como objetivo de protección	34
1.5.4 Sustentabilidad.....	34
1.6. Definiciones	35
2. ESTADO DEL ARTE	40
2.1. Baseline Project	40
2.2. The EU Water Framework Directive: Statistical aspects of the identification of groundwater pollution trends	42
2.2.1 Agregación de datos	43
2.2.2 Tendencias e inversión de las mismas	44
2.2.3 Punto de partida para la inversión de tendencias	44
2.2.4 Longitud de las series para la inversión de tendencias	45
2.3. Bridge Project	45
2.3.1 Determinación de los valores naturales de fondo	47
2.3.2 Selección del Estándar de Calidad de Referencia	49
2.3.3 Utilización de la metodología.....	49

2.4. Programa de control de agua subterránea de la US EPA en instalaciones de sustancias peligrosas	50
2.4.1 Resultados significativos en el muestreo de agua subterránea	51
2.4.2 Diseño del programa de control de detección.....	52
2.4.3 Diseño de los programas de control de tolerancia y recuperación.....	53
2.4.4 Establecimiento y actualización del valor de fondo.....	54
2.4.5 Gestión de valores por debajo del límite de detección.....	55
3. MATERIALES, MÉTODOS Y TÉCNICAS ESTADÍSTICAS.....	58
3.1. Modelo conceptual	60
3.2. Análisis exploratorio de datos.....	63
3.3. Aplicación de técnicas de suavizado	64
3.4. Selección del periodo de referencia.....	67
3.5. Análisis de tendencias	68
3.5.1. Parámetros a controlar	74
3.6. Actualización del nivel básico	75
3.7. Intervalos estadísticos	77
3.7.1. Pruebas de bondad de ajuste.....	79
3.7.2. Cálculo de estadísticos.....	80
3.7.3. Intervalos paramétricos	80
3.7.4. Intervalos no paramétricos	85
3.8. Diseño del programa de control con umbral definido.....	88
3.8.1. Intervalos de predicción.....	89
3.8.2. Tests de hipótesis	94
3.9. Diseño del programa de control sin umbral definido	95
3.9.1. Comparación del nivel básico con nuevos valores	96
3.9.2. Requerimientos y limitaciones	98
3.10. Tratamiento de valores anómalos	99
3.10.1. Tamaño mínimo de la muestra	102
3.11. Autocorrelación	102
3.12. Estacionalidad.....	104
4. ANÁLISIS DE LA INFORMACIÓN	107
4.1. Origen de los datos	107
4.2. Selección de una Demarcación Hidrográfica de referencia	110
4.3. Masas de agua subterránea analizadas	114
4.3.1. Plana de Vinaroz	116

4.3.2.	Plana de Oropesa-Torreblanca	118
4.3.3.	Plana de Castellón	118
4.3.4.	Plana de Sagunto.....	119
4.3.5.	Buñol - Cheste	120
4.3.6.	Plana de Valencia Norte.....	120
4.3.7.	Plana de Valencia Sur.....	121
4.3.8.	Sierra de las Agujas	121
4.3.9.	Plana de Jaraco	122
4.3.10.	Oliva – Pego.....	122
4.3.11.	Ondara - Denia.....	123
4.3.12.	Jávea	123
4.4.	Análisis de las series de datos	124
4.4.1.	Inspección visual	124
4.4.2.	Selección de niveles básicos y análisis de tendencias	125
4.5.	Aplicación de la metodología a una zona piloto	127
5.	GESTIÓN DE VALORES POR DEBAJO DEL LÍMITE DE DETECCIÓN	141
5.1.	Consideraciones generales	141
5.2.	Visualización de series con datos censurados	143
5.2.1.	Gráficos de caja	143
5.2.2.	Gráficos de dispersión.....	144
5.2.3.	Gráficos probabilísticos	145
5.3.	Técnicas de análisis de valores censurados. Cálculo de estadísticos descriptivos	146
5.3.1.	Sustitución.....	147
5.3.2.	Estimación por máxima verosimilitud (MLE)	149
5.3.3.	Regresión sobre estadísticos de orden (ROS)	151
5.3.4.	Técnicas no paramétricas	152
5.4.	Regresión no paramétrica y análisis de tendencias	153
5.5.	Cálculo de estadísticos descriptivos	155
5.5.1.	Estimación por máxima verosimilitud.....	155
5.5.2.	Regresión robusta sobre estadísticos de orden.....	157
5.5.3.	Kaplan-Meier	158
5.6.	Intervalos de confianza de las medidas de tendencia central	159
5.6.1.	Intervalos de confianza de la media y la mediana por máxima verosimilitud	159

5.6.2.	Intervalos de confianza de la media y la mediana mediante Regresión Robusta (ROS) y métodos no paramétricos	161
5.6.3.	Intervalos de confianza de la media y la mediana mediante técnicas de remuestreo (bootstrap).....	161
5.7.	Intervalos de predicción de las medidas de tendencia central.....	162
5.7.1.	Simulación de datos censurados en las masas de agua analizadas.....	162
5.8.	Simulación de datos censurados en la zona piloto seleccionada...	164
6.	RESULTADOS Y DISCUSIÓN	170
7.	CONCLUSIONES	175
8.	RECOMENDACIONES Y LÍNEAS DE INVESTIGACIÓN FUTURAS	184
9.	BIBLIOGRAFÍA.....	186
APÉNDICE A: ANÁLISIS EXPLORATORIO DATOS ESTACIONES CUENCA DEL JÚCAR		193
APÉNDICE B: ANÁLISIS DE VALORES CENSURADOS EN ESTACIONES SELECCIONADAS DE LA CUENCA DEL JÚCAR.....		200
APÉNDICE C: CÓDIGO R GENERADO PARA ANÁLISIS DE DATOS		205

Relación de figuras

Figura 1: Políticas sectoriales relacionadas con la Directiva Marco del Agua (European Communities, 2008).....	26
Figura 2: Etapas en el proceso de evaluación de riesgos.....	31
Figura 3: Definición del buen estado químico de las masas de agua subterránea (De Müller, D., UBA - Austria).....	46
Figura 4: Metodología en etapas para determinación del estado químico.....	50
Figura 5: Curva de potencia normal (n=10) para un límite de predicción al 99% (de USEPA, 2009)	53
Figura 6: Modelo conceptual y heterogeneidad del sistema agua-suelo en un emplazamiento de la costa Mediterránea (Ballesteros et al, 2001)	60
Figura 7: Resumen análisis exploratorio datos evolución de nitratos en punto de la red de control TAIG000591 (Plana de Sagunto).....	66
Figura 8: Elementos de la evaluación e inversión de tendencias (European Commission, 2009)	72
Figura 9: Posición relativa del intervalo de confianza de la media o mediana en relación con el estándar de calidad ambiental.....	78
Figura 10: Histograma de la media e intervalos de confianza para la media de una muestra mediante técnicas bootstrap	88
Figura 11: Posición relativa del valor umbral en relación con la norma de calidad.....	88
Figura 12: Histograma y gráfico Cuantil-Cuantil de los valores de concentración de nitratos en la estación JUIG002950	101
Figura 13: Detección de valores anómalos en la estación JUIG002950.....	101
Figura 14: Concentraciones de nitratos agrupadas en estación seca y húmeda	104
Figura 15: Red de cuencas piloto en Europa.....	108
Figura 16: Cuencas analizadas en color blanco.....	109
Figura 17: Histogramas y gráficos Q-Q Plot de concentraciones de Nitrato en puntos seleccionados de las cuencas del Ebro, Segura, Júcar, Duero, Tajo y Guadiana.....	110
Figura 18: Localización de la Demarcación Hidrográfica del Júcar	111
Figura 19: Evolución concentración nitratos en la Plana de Vinaroz.....	117
Figura 20: Evolución concentración nitratos en la Plana de Oropesa -Torreblanca.....	117
Figura 21: Evolución concentración nitratos en la Plana de Castellón	118
Figura 22: Evolución concentración nitratos en la Plana de Sagunto	119
Figura 23: Evolución concentración nitratos en Buñol-Cheste	120
Figura 24: Evolución concentración nitratos en Plana de Valencia Norte	120
Figura 25: Evolución concentración nitratos en Plana de Valencia Sur	121
Figura 26: Evolución concentración nitratos en Sierra de las Agujas	122
Figura 27: Evolución concentración nitratos en Plana de Jaraco.....	122
Figura 28: Evolución concentración nitratos en Oliva-Pego	122
Figura 29: Evolución concentración nitratos en Ondara-Denia	123
Figura 30: Evolución concentración nitratos en Ondara-Denia	123
Figura 31: Nivel básico estaciones seleccionadas	126
Figura 32: Localización de la masa de agua subterránea y estaciones de la red de control	127

Figura 33: Localización de la zona piloto y de las estaciones de la red de control	128
Figura 34: Evolución de parámetros seleccionados en la Masa de Agua Subterránea Plana de Vinaroz	129
Figura 35: Evolución de la concentración de cloruros en la estación 47 de la Plana de Vinaroz	130
Figura 36: Evolución de la concentración de nitratos en la estación 69 de la Plana de Vinaroz	133
Figura 37: Distribución de concentración cloruros en la red de intrusión de la Plana de Vinaroz-Peñíscola	135
Figura 38: Diagrama de dispersión de la concentración de cloruros en la Plana de Vinaroz-Peñíscola	136
Figura 39: Evolución de medianas en la red de intrusión de la Plana de Vinaroz-Peñíscola ...	137
Figura 40: Evolución temporal red operativa.....	137
Figura 41: Relación entre el límite de detección y el límite de cuantificación.....	142
Figura 42: Boxplot JUIG000213 suponiendo 15% censura en totalidad de datos y con datos separados en dos grupos	143
Figura 43: Gráfico de dispersión de la estación JUIG000213	144
Figura 44: Función de supervivencia de la estación JUIG000213	144
Figura 45: Función de supervivencia por grupos de la estación JUIG000213	145
Figura 46: Gráfica de probabilidad de la estación JUIG000213	145
Figura 47: Histogramas serie JUIG002334 con diferente % censura	149
Figura 48: Ajuste por máxima verosimilitud en la estación JUIG000213 (normal y lognormal)	150
Figura 49: Localización de la zona piloto.....	165
Figura 50: Series temporales en la Plana de Sagunto	166
Figura 51: Diagrama de flujo para la toma de decisiones en masas de agua subterránea.....	177
Figura 52: Tendencia general en la concentración de cloruros en la Plana de Vinaroz Peñíscola	179
Figura 53: Niveles Básicos en la Cuenca del Júcar (1)	194
Figura 54: Niveles Básicos en la Cuenca del Júcar (2)	195
Figura 55: Niveles Básicos en la Cuenca del Júcar (3)	196
Figura 56: Niveles Básicos en la Cuenca del Júcar (4)	197
Figura 57: Niveles Básicos en la Cuenca del Júcar (5)	198
Figura 58: Niveles Básicos en la Cuenca del Júcar (6)	199

Relación de ecuaciones

Ecuación 1: Intervalo de confianza de la media de un conjunto de datos normales	81
Ecuación 2: Función de densidad de probabilidad de una distribución gamma.....	82
Ecuación 3: Condición para rechazar la hipótesis nula $\mu_1 \geq C_s$ para el límite superior de confianza de la media en una distribución gamma	82
Ecuación 4: Probabilidad del límite superior de confianza de la media en una distribución gamma.....	82
Ecuación 5: Estimador de k corregido para el sesgo en una distribución gamma.....	83
Ecuación 6: Valor aproximado del límite superior de confianza para la media en una distribución gamma.....	83
Ecuación 7: Valor aproximado del límite inferior de confianza para la media en una distribución gamma.....	83
Ecuación 8: Valor ajustado del límite superior de confianza para la media en una distribución gamma.....	83
Ecuación 9: Límite inferior de confianza en una distribución lognormal	84
Ecuación 10: Estimadores insesgados de mínima varianza de los parámetros de una distribución lognormal	85
Ecuación 11: Desigualdad de Chebyshev	85
Ecuación 12: Intervalo de confianza de la media de Chebyshev	85
Ecuación 13: Ecuación general para el cálculo de un intervalo de predicción paramétrico	91
Ecuación 14: Intervalo de predicción paramétrico para m muestras futuras	91
Ecuación 15: Intervalo de predicción paramétrico para la media de m muestras futuras	92
Ecuación 16: Distribución de probabilidad acumulada de la densidad de probabilidad Beta ...	93
Ecuación 17: Error tipo 1 asociado al intervalo de predicción de la mediana de m valores futuros.....	93
Ecuación 18: Error tipo 1 asociado al intervalo de predicción de la mediana de tres valores futuros.....	94
Ecuación 19: Hipótesis nula y alternativa para verificar el estado químico de la MASub	95
Ecuación 20: Hipótesis nula y alternativa para verificar la efectividad de las medidas de recuperación en la MASub	95
Ecuación 21: Hipótesis nula y alternativa para la igualdad de medias entre la red y el nivel básico	97
Ecuación 22: Selección del estadístico para el test de Wilcoxon y de Mann-Whitney.....	97
Ecuación 23: Error cuadrático medio.....	146
Ecuación 24: Función de densidad conjunta de las observaciones de una m.a.s.....	149
Ecuación 25: Estimador de máxima verosimilitud de θ	150
Ecuación 26: Función de supervivencia	152
Ecuación 27: Estimador de Kaplan-Meier de la función de supervivencia	153
Ecuación 28: Estimación de la varianza de la función de supervivencia mediante la fórmula de Greenwood.....	153
Ecuación 29: Modelo de regresión lineal.....	154
Ecuación 30: Estimador robusto de Theil de la pendiente en regresión no paramétrica	154

Ecuación 31: Estimador robusto de Theil de la ordenada en el origen en regresión no paramétrica	154
Ecuación 32: Estimadores de la media y la desviación estándar para distribuciones lognormales	156
Ecuación 33: Percentiles de una distribución lognormal.....	156
Ecuación 34: Relación entre la media, la varianza y la función de distribución normal con datos censurados	157
Ecuación 35: Relación entre la media, la varianza y la función de distribución normal con datos censurados	157
Ecuación 36: Cálculo de la media mediante el procedimiento de Kaplan-Meier	158
Ecuación 37: . Error estándar de la media mediante Kaplan-Meier	159
Ecuación 38: Intervalo de confianza de la media.....	160
Ecuación 39: Método de Cox para el cálculo de intervalos de confianza de la media para una distribución lognormal	160
Ecuación 40: Intervalo de predicción paramétrico para m muestras futuras	162
Ecuación 41: Intervalo de predicción paramétrico para un nuevo valor (distribución lognormal)	162

Relación de tablas

Tabla 1: Evaluación de tendencias (artículo 5.1 y 5.2 de la DAS). Resumen de elementos y tests correspondientes	72
Tabla 2: Evaluación adicional de tendencias. Resumen de elementos y tests correspondientes	73
Tabla 3: Nivel ajustado de significación β	84
Tabla 4: Relación de masas de agua subterránea identificadas en riesgo en la Demarcación Hidrográfica del Júcar y motivo de su clasificación en este apartado	112
Tabla 5: Relación de masas de agua subterránea seleccionadas en la	114
Tabla 6: Relación de estaciones de control seleccionadas en la Demarcación Hidrográfica del Júcar	115
Tabla 7: Relación de estaciones de control seleccionadas en la Demarcación Hidrográfica del Júcar (continuación).....	116
Tabla 8: Bondad de ajuste de los puntos de control de la red operativa en Plana de Vinaroz y p-valores para las concentraciones de cloruros.....	131
Tabla 9: Intervalos de confianza de concentración de nitratos y cloruros y límites de predicción de concentración de cloruros en la Plana de Vinaroz Peñíscola.....	132
Tabla 10: p-valores para el test de Mann Kendal en la red operativa de la Plana de Vinaroz Peñíscola	139
Tabla 11: Estaciones seleccionadas en Plana de Sagunto y número de registros	166
Tabla 12: Intervalos de confianza de la media estimados en las estaciones seleccionadas.....	168
Tabla 13: Intervalos de confianza de la media en estaciones seleccionadas de la Cuenca del Júcar calculados por diferentes métodos y para diversos grados de censura (1).....	201
Tabla 14: Intervalos de confianza de la media en estaciones seleccionadas de la Cuenca del Júcar calculados por diferentes métodos y para diversos grados de censura (2).....	202
Tabla 15: Intervalos de confianza de la media en estaciones seleccionadas de la Cuenca del Júcar calculados por diferentes métodos y para diversos grados de censura (3).....	203
Tabla 16: Intervalos de confianza de la media en estaciones seleccionadas de la Cuenca del Júcar calculados por diferentes métodos y para diversos grados de censura (4).....	204

Capítulo 1:

INTRODUCCIÓN

1. INTRODUCCIÓN

1.1. Organización del documento de tesis

La presente Memoria de Tesis Doctoral está estructurada en 9 capítulos. De ellos, el capítulo dedicado a la presentación y discusión de resultados (VI) tiene entidad propia y puede ser leído de forma independiente. Se incluye además un capítulo donde se abordan los métodos y tratamientos estadísticos utilizados cuando no existen valores por debajo del límite de detección (III) y otro sobre gestión de valores censurados (V).

1.2. Justificación de la investigación

Los países de la Unión Europea afrontan el problema de una contaminación importante del suelo y los recursos hídricos (**Grima et al, 2002**). Aunque los especialistas técnicos hablan un lenguaje común y comprenden por igual las ciencias relacionadas con el medio subterráneo, existen diferencias en cuanto a la forma de determinar la calidad del agua subterránea. Dichas diferencias tienen su origen en las diferentes legislaciones, en la importancia a escala local del agua subterránea o en la consideración de los recursos suelo-agua subterránea de forma separada o conjunta.

No obstante, la entrada en vigor de la Directiva Marco del Agua y la Directiva Derivada de Aguas subterráneas han supuesto el establecimiento de una serie de requisitos que todos los estados miembros deben cumplir, entre ellos la clasificación del estado químico de las masas de agua subterránea.

A este respecto, la red de control debe ser diseñada (**Quevauvillier, 2005**) de forma que proporcione una visión coherente y completa del estado químico del agua subterránea en su correspondiente cuenca hidrográfica y sea capaz de detectar tendencias crecientes a largo plazo inducidas por la actividad antrópica.

Los problemas relacionados con la calidad del agua aumentan de forma sostenida, por lo que se hace necesaria una actuación a medio plazo. Al mismo tiempo, la lentitud y complejidad de los procesos burocráticos tiende a ralentizar el proceso de toma de decisiones, alejándolo de la realidad en muchas ocasiones.

La calidad del agua subterránea es un concepto que generalmente se ha relacionado con la idoneidad del recurso para sus usos en una zona determinada. Sin embargo, la Directiva de Aguas Subterráneas establece una distinción fundamental entre los acuíferos que presentan mala calidad debido a causas naturales, como puede ser el alto contenido en sulfatos en algunas cuencas detríticas terciarias, y los que presentan mala calidad debido a procesos contaminantes de origen antrópico. En el primer caso, aun estando clasificada el agua del acuífero como de mala calidad, el estado químico de la masa de agua subterránea puede ser bueno.

Por ello es necesario un marco hidrogeológico de referencia (**Wendland, 2008**) que evite la clasificación inadecuada de una masa de agua subterránea. Además, los estados miembros deben identificar los valores umbral de los contaminantes representativos de las masas de agua subterránea en riesgo de no cumplir los objetivos de calidad fijados por la Directiva de Aguas Subterráneas a más tardar en el año 2015.

Como se ha concluido en diferentes proyectos europeos (**BaSeLine, 2003**), es necesario estudiar los parámetros a controlar en las redes de control y la frecuencia de medida, alejándose de la forma tradicional de comparación con estándares de calidad, en la mayoría de los casos, del agua potable.

Por todo ello, se hace imprescindible investigar sobre el diseño de metodologías que permitan la clasificación de las masas de agua subterránea desde el punto de vista de la Directiva y que puedan ser utilizados a su vez como sistemas de alerta temprana con objeto de proteger los recursos hídricos y cumplir con otro de los requisitos fundamentales de la Directiva, como es la inversión de tendencias al aumento en la concentración de contaminantes.

1.3. Objetivos de la investigación

El objetivo fundamental de la investigación consiste en el diseño y desarrollo de un procedimiento para la toma de decisiones en cuanto a la clasificación del estado químico de una masa de agua subterránea y contrastar la existencia de incrementos significativos en la concentración de un parámetro determinado en la misma. Para ello es necesario considerar las características hidrogeológicas particulares de las redes de control de aguas subterráneas y sus condicionantes específicos.

Dada la subjetividad del término “significativo”, y con objeto de dotar al método de una base científica sólida, se aplicarán metodologías estadísticas basadas en tests de hipótesis, mediante los cuales se definirán los incrementos estadísticamente significativos.

El resultado de un contraste estadístico de hipótesis puede implicar la clasificación de una masa de agua subterránea como en mal estado químico desde el punto de vista de la Directiva de Aguas Subterráneas. Por ello es conveniente conocer el significado de las pruebas que se realicen sobre los datos. Los principales resultados de interés en el análisis del estado químico de la masa son los siguientes:

- cumplimiento de los estándares de protección establecidos
- empeoramiento significativo de su calidad
- existencia de tendencias significativas al aumento (o al descenso en el caso de acciones de recuperación)

Para la detección de empeoramientos significativos es necesario que se produzca un aumento en la concentración del parámetro analizado. El aumento debe ser superior a lo que podría ocurrir por azar, para lo cual es necesario conocer la variabilidad de los datos. La comparación puede realizarse tanto con valores de fondo calculados a partir de los datos como con estándares de calidad (fijados a partir de un subconjunto de los datos o establecidos por las autoridades ambientales).

En el caso de se haya fijado un valor umbral la comparación debe realizarse con él. El punto de partida para dicha comparación es el valor básico, definido como la media de los valores correspondientes a los años de referencia 2007 y 2008. Dado que el valor básico representa un valor medio o central, la comparación debe realizarse con medias o medianas.

Para ello es posible la elaboración de un intervalo de confianza para la media bien por técnicas tradicionales bien por técnicas tipo bootstrap. El valor relevante en caso de evaluar el estado ecológico de la masa de agua subterránea es el límite inferior de confianza, mientras que en caso de evaluar la inversión de tendencias el punto de comparación sería el límite superior.

Cuando el límite inferior de confianza no supere el valor umbral, no se clasificará la masa de agua subterránea en mal estado aunque exista una tendencia creciente. Sin embargo, cuando se supere el valor correspondiente al 75% de la norma de calidad y exista, además, una tendencia al aumento, será necesario iniciar la inversión de la misma.

En caso de que no se haya definido un valor umbral, la única comparación posible será la elaboración de intervalos de predicción para la detección de empeoramientos estadísticamente significativos. Cuando se constate un empeoramiento combinado con una tendencia al aumento deberá recomendarse la inversión de tendencias.

Por último, cuando los datos puedan ajustarse a una distribución normal es posible la utilización de métodos basados en intervalos de confianza paramétricos o bien métodos gráficos de Shewhart o CUSUM para la declaración de un empeoramiento significativo.

En el caso de que la población no siga una distribución normal, se utilizarán intervalos de confianza no paramétricos, no siendo posible la utilización de métodos gráficos de control, puesto que no existen alternativas no paramétricas en este caso.

Los tests utilizados para la detección de tendencias al aumento implican generalmente la comparación de un único conjunto de datos con un valor fijado o una concentración de fondo. En cambio, cuando el objetivo del análisis es la

verificación de inversión de tendencias, es necesario demostrar la existencia de una pendiente negativa en los valores de concentración.

1.4. Marco normativo para la gestión de los recursos hídricos

El 23 de octubre de 2000 fue aprobada por el Parlamento y la Comisión Europeos la **Directiva 2000/60/EC** por la que se establece un marco de acción comunitario en el campo de la política de los recursos hídricos. La Directiva Marco del Agua es, como consecuencia, el instrumento que proporciona el marco legislativo para la futura gestión de los acuíferos. La idea central de la Directiva Marco es el concepto de integración (**Del Moral, 2008**) del agua en las diferentes políticas sectoriales, en un sentido mucho más amplio que la idea clásica de gestión integrada del agua (Figura 1).

Figura 1: Políticas sectoriales relacionadas con la Directiva Marco del Agua (European Communities, 2008)

Esta idea involucraba simplemente la gestión conjunta de aguas superficiales y subterráneas, aunque es preciso aclarar que ha sufrido la lógica evolución conceptual a lo largo del tiempo, hasta contemplar la incorporación del mayor número posible de dimensiones y aspectos involucrados en la gestión y funcionamiento de un sistema hídrico determinado (**Molina y García Aróstegui, 2009**).

Uno de los ejes fundamentales sobre los que bascula la directiva es el uso sostenible del agua lo cual, dado el incremento constante de la demanda, plantea una serie de nuevos retos, tales como la gestión efectiva y sostenible de los acuíferos, la necesidad de conseguir un buen estado ecológico de todas las aguas a finales de 2015, la depuración de las aguas residuales y el mantenimiento de la funcionalidad de los ecosistemas con respecto al suelo y al agua subterránea entre otros.

La protección de las aguas subterráneas se establece, pues, como una de las prioridades de la política ambiental de la Unión Europea, lo cual, unida a la derogación de la Directiva 80/68/CEE (relativa a la protección de las aguas subterráneas contra la contaminación causada por determinadas sustancias peligrosas) en el año 2013, motivó la publicación de la **Directiva 2006/118/CE** del Parlamento Europeo y del Consejo, de 12 de diciembre de 2006, relativa a la protección de las aguas subterráneas contra la contaminación y el deterioro, con el foco puesto más en la prevención que en la limpieza de la contaminación **(López Geta et al, 2009)**.

En el Anexo IV (Parte A, apartado 3) de dicha Directiva se fijan los niveles básicos como el punto de partida para la determinación de tendencias significativas y sostenidas al aumento en las concentraciones de sustancias que se produzcan naturalmente y como resultado de actividades humanas.

Las masas de agua subterránea son sistemas complejos que presentan una variabilidad elevada, no solo en la geometría y características hidrodinámicas del medio geológico, sino en sus propiedades físico-químicas **(Ballesteros et al, 2001)**. Desde un punto de vista científico no tiene sentido referirse a la concentración de una sustancia en un acuífero como el resultado medido en un punto concreto del mismo **(Blum et al, 2009)**, si bien el uso de concentraciones medias es útil desde el punto de vista de la Directiva.

Desde una perspectiva regulatoria (Directiva de Aguas Subterráneas) el nivel de fondo se refiere a la concentración de una sustancia o al valor de un indicador en una masa de agua subterránea correspondiente a condiciones caracterizadas por una influencia antrópica mínima. En sentido estricto se refiere a condiciones inalteradas propias de la era preindustrial, lo cual no

parece una suposición muy realista (al menos para sustancias con origen tanto natural como antrópico). En cambio, para sustancias sintéticas el valor de fondo sí debe considerarse igual a cero (referencia el límite de detección), mientras que para sustancias con posible origen natural debe considerarse un rango de concentraciones **(Müller et al, 2006)**.

Por otra parte, en la Directiva se definen los niveles básicos como el valor medio medido por lo menos durante los años de referencia 2007 y 2008 de acuerdo con los programas de medidas implementados por los Estados Miembros con objeto de dar cumplimiento al artículo 8 de la misma o, en el caso de sustancias identificadas después de dichos años de referencia, durante el primer periodo en el que se disponga de datos representativos. Igualmente, se indica la posibilidad de utilización, cuando se disponga de ellos, de los datos recabados con anterioridad al comienzo del programa de control.

Su definición se realiza únicamente en las masas de agua subterránea en las que se han identificado riesgos relevantes durante el proceso de caracterización requerido por la Directiva Marco del Agua. Es decir, que se ha identificado la existencia de influencia antrópica, por lo que conseguir concentraciones de niveles naturales de fondo en dichas masas parece una tarea difícil cuanto menos.

Asimismo, los Estados Miembros podrán eximir de las medidas exigidas con objeto de prevenir o limitar la entrada de contaminantes en el agua subterránea, si se requieren costes económicos desproporcionados. Ello no significa que no deba tomarse ningún tipo de medida para evitar el deterioro de la calidad del agua subterránea, y en cualquier caso las exenciones previstas sólo podrán otorgarse cuando se realice el seguimiento de las masas de agua subterránea implicadas.

Las medidas obtenidas de las redes de control son una información valiosa que debe ser aprovechada para evaluar la existencia de impactos estadísticamente significativos desde el comienzo del periodo de control, incluso si no es posible revertir el acuífero a las concentraciones naturales de fondo. Con ese objetivo deben utilizarse los puntos disponibles en las redes de control para la elaboración de un valor básico actualizado o extendido.

La agregación de datos adicionales al valor básico existente tiene la ventaja de que se aumenta la potencia de los tests realizados, y pueden elaborarse intervalos estadísticos más precisos en los que se tenga en cuenta la variabilidad natural en la concentración de una sustancia.

En sintonía con el Ordenamiento Jurídico Comunitario se han desarrollado una serie de normas a escala nacional cuyo principal objetivo es integrar los mencionados ordenamientos. A continuación se mencionan brevemente las más relevantes:

- **Real Decreto 907/2007**, de 6 de julio, por el que se aprueba el Reglamento de la Planificación Hidrológica. Establece que los planes hidrológicos de cuenca deberán contener obligatoriamente una descripción general de las presiones e incidencias antrópicas significativas sobre las aguas, superficiales y subterráneas, incluyendo los usos del suelo y la contaminación puntual y difusa (que engloba las zonas mineras y los suelos contaminados). En el caso de aguas superficiales se consideran otras afecciones significativas de la actividad humana como los sedimentos contaminados.

También incluirán una lista de objetivos medioambientales para las aguas, superficiales y subterráneas, que deberán satisfacerse, como norma general, antes del 31 de diciembre de 2015. Entre estos objetivos medioambientales se encuentra el evitar o limitar la entrada de contaminantes a las aguas y el reducir la contaminación derivada de la actividad humana.

- **Real Decreto 140/2003**, de 7 de febrero, por el que se establecen los criterios sanitarios de la calidad del agua potable. Fija los estándares de calidad para las aguas destinadas al consumo humano. De este modo, en el caso de que la contaminación del suelo afecte a aguas destinadas a este uso deberán observarse los estándares de calidad fijados por la norma.
- **Real Decreto Legislativo 1/2001**, de 20 de julio, por el que se aprueba el texto refundido de la Ley de Aguas. Establece la protección del dominio público hidráulico contra su deterioro con el objetivo, entre otros, de impedir la acumulación de compuestos tóxicos o peligrosos en el subsuelo, capaces de contaminar las aguas subterráneas. La Autoridad de Cuenca se constituye

como el organismo responsable para delimitar zonas de protección para los recursos hídricos.

Asimismo, requiere un estudio hidrogeológico previo que demuestre la inocuidad de los vertidos cuando éstos puedan dar lugar a la infiltración de sustancias susceptibles de contaminar los acuíferos o las aguas subterráneas.

- **Decreto 2414/1961, de 30 de noviembre**

Por el que se aprueba el Reglamento de actividades molestas, insalubres, nocivas y peligrosas. Se administra localmente por los ayuntamientos. Requiere de permisos para toda actividad molesta, insalubre o peligrosa.

Una deficiencia de esta ley es que las autoridades locales a menudo no cuentan con los expertos técnicos requeridos para realizar las evaluaciones apropiadas que la protección del agua subterránea necesita. Algunas autoridades regionales han publicado leyes involucrándose en esta administración.

- **Real Decreto 849/1986, de 11 de abril**

Por el que se aprueba el Reglamento del Dominio Público Hidráulico que desarrolla los títulos preliminar, I, IV, V, VI, VII y VIII del texto refundido de la Ley de Aguas, aprobado por el Real Decreto Legislativo 1/2001, de 20 de julio.

- **Real Decreto 261/1996, de 16 de febrero**

Sobre protección de las aguas contra la contaminación producida por los nitratos procedentes de fuentes agrarias.

- **Real Decreto 1664/1998, de 24 de julio**

Por el que se aprueban los Planes Hidrológicos de Cuenca. Son los instrumentos teóricos para definir las zonas de protección dentro de las áreas de recarga de los campos de pozos. Se da prioridad a las ciudades más grandes.

1.5. Evaluación del riesgo en las aguas subterráneas

El análisis de riesgos se basa en el uso sistemático de la información existente para determinar la naturaleza y magnitud de dichos riesgos para la salud humana y los ecosistemas derivados de la existencia de sustancias contaminantes en el medio ambiente (Figura 2). Cuando se produce la liberación de un contaminante al medio, la posible afección a los diferentes compartimentos ambientales viene determinada por las características físico-químicas de la sustancia vertida y por las propiedades del suelo.

Para evaluar las posibles rutas de migración de un contaminante desde la fuente hasta el punto de exposición resulta imprescindible analizar de qué manera se distribuye dicho contaminante entre las distintas fases (fugacidad). Por otra parte, es posible realizar un análisis cuantitativo de riesgos mediante el análisis de escenarios concretos con estimaciones puntuales (determinístico) o mediante simulaciones de tipo Monte Carlo.

Este último método representa las incertidumbres de los parámetros del modelo con distribuciones de probabilidad. De esta forma es posible analizar los efectos simultáneos y acumulativos de la incertidumbre en los datos de un modelo y proporcionar información adicional tal como los porcentajes de población cuya exposición puede ser superior a los límites legales establecidos.

Figura 2: Etapas en el proceso de evaluación de riesgos

En el caso de los nitratos la Directiva 98/83/CE del Consejo (Comisión Europea, 1998), de 3 de noviembre de 1998, relativa a la calidad de las aguas destinadas al consumo humano, establece un valor máximo de 50 mg/l de nitratos. En España, la norma que transpone la citada Directiva (RD 140/2003) también establece un valor paramétrico de 50 mg/l.

Este valor se obtiene por medio de una evaluación de riesgos, mediante la cual se confirma que la dosis admisible de nitratos en adultos está comprendida entre 0-3,65 mg/kg de peso corporal y día. No obstante, en el año 2004 la Organización mundial de la salud confirmó un valor máximo de 50 mg/l de nitratos en el agua de consumo con objeto de prevenir la metahemoglobinemia en los bebés. De esta forma, aceptando dicho valor umbral se realizan los cálculos sobre los datos seleccionados.

1.5.1 Impactos naturales y antrópicos

El agua viaja mucho más lentamente en los acuíferos que en los sistemas de agua superficial y el tiempo de migración desde la superficie de la tierra al punto de descarga para el agua infiltrada y los contaminantes presentes en ella se mide en años, décadas o siglos, dependiendo del tipo de contaminante, las características del acuífero y la longitud de la trayectoria de flujo. En consecuencia, hay una gran cantidad de tiempo para el contacto íntimo entre el agua y material del suelo y roca, que se traduce en cambios en la calidad del agua con respecto a la composición mineral.

Como resultado, los metales se encuentran en el agua subterránea en prácticamente todos los países del mundo (**US EPA, 2008**). Dada la gran variedad en la geometría de los acuíferos, suelo y materiales de las formaciones atravesadas, tanto los tiempos de viaje como la calidad del agua subterránea (Concentración natural de fondo) es muy variable. Por tanto, pueden producirse riesgos inaceptables para la salud humana o de los ecosistemas por causas naturales. Un ejemplo de ello es el envenenamiento por arsénico de gran número de personas que beben agua poco profunda de origen subterráneo en Bangladesh y el oeste de Bengala (India), como

consecuencia de un fenómeno natural, no causado por la contaminación industrial (**Rahman et al., 2001**).

Además de los contaminantes procedentes de fuentes naturales, hay una amplia variedad de contaminantes que se originan a partir de fuentes puntuales, es decir, originadas a partir de fugas o vertidos, que se encuentran con frecuencia en las zonas más densamente pobladas del mundo. Lagunas de aguas residuales, rellenos sanitarios, fugas de los sistemas de alcantarillado (**Lerner 2002**), junto con instalaciones de eliminación de residuos, son ejemplos de cómo la calidad del agua subterránea puede verse negativamente afectada.

Igualmente, los sistemas de tratamiento de aguas residuales individuales (fosas sépticas) de las viviendas en zonas rurales y edificios que no están conectados al sistema de alcantarillado de las ciudades son otro factor a considerar en la evaluación de riesgos para los recursos de agua subterránea.

Un ejemplo de impacto antropogénico a gran escala sobre las aguas subterráneas se describe en **Santos et al. (2002)**, que midió zinc, cadmio, plomo y cobre en agua subterránea como resultado del accidente de Aznalcóllar. En la madrugada del 25 de abril de 1998 se produjo la rotura y desprendimiento de unos 50 metros del muro de contención de la balsa de decantación de los lodos piríticos que la empresa Boliden Apirsa, S.L. tenía en su explotación minera de Aznalcóllar (**Arenas, 2001**). Esta rotura ocasionó el vertido incontrolado de un volumen aproximado de 6 hm³ de aguas ácidas y lodos al río Agrio y desde éste al río Guadiamar. Como consecuencia de la avenida se afectó a una longitud de 62 km del cauce desde la balsa hasta el límite del Parque Nacional de Doñana, en la zona de Entremuros. La superficie total afectada se evaluó en 4630 ha pertenecientes a nueve municipios de la provincia de Sevilla y a una reducida extensión de áreas protegidas del Parque Natural de Doñana.

1.5.2 Ámbito de los aspectos relacionados con el agua subterránea

Con respecto a la zona saturada, en el esquema tradicional de la evaluación de riesgos se considera al agua subterránea desde la perspectiva de un medio de transporte para los contaminantes, esto es, como una vía, y no desde la perspectiva del agua subterránea en sí misma como un objetivo de protección.

1.5.3 Agua subterránea como objetivo de protección

Las aguas subterráneas desempeñan un papel importante en el abastecimiento de ciertas áreas, especialmente en países áridos o semiáridos. Por ejemplo, dos de los países más desarrollados del mundo consumen 425 (USA) y 326 (Canadá) litros por habitante y día respectivamente (**Baynes, 2004**). En algunos casos, especialmente en algunas zonas rurales de los países del tercer mundo, el consumo origina el agua directamente de pozos privados. Además, deben considerarse otros usos, entre los que se encuentran usos domésticos (limpieza y aseo), agrícolas e industriales. Por tanto, dado su papel el agua subterránea debe ser considerada como un objeto de protección. En la Unión Europea las masas de agua subterránea representan un almacén enorme de abastecimiento, con porcentajes que representan aproximadamente el 70% del agua de consumo doméstico proveniente del agua subterránea (**Foster and Chilton, 2003**). A escala mundial, este porcentaje es de al menos el 50% (**Zektser and Everett, 2004**). De hecho, con una tasa de extracción global de 600–700 km³/año, el agua subterránea es la materia prima más extraída del mundo (**Zektser and Everett, 2004**).

1.5.4 Sustentabilidad

Uno de los mayores desafíos con respecto a una política ambiental sostenible se encuentra en la protección de las aguas subterráneas. En general, el término sostenibilidad se refiere a un determinado estado de un proceso, o estado, que es capaz de continuar o ser mantenido indefinidamente. Existe una definición más amplia, debida a la Comisión Brundtland de la Asamblea

General de las Naciones Unidas (**United Nations, 1987**); un desarrollo que satisface las necesidades del presente sin comprometer la capacidad de las generaciones futuras para satisfacer sus propias necesidades o que afecten negativamente a un entorno más amplio. Esta expansión de la definición de sostenibilidad es especialmente relevante en relación con las aguas subterráneas como vía para el transporte de contaminantes, ya que dicho proceso puede conducir a riesgos en lugares alejados de la fuente.

En el nuevo marco definido por las Directivas Comunitarias para la gestión y el uso sostenible de los recursos hídricos se incluyen técnicas como la reutilización y la recarga artificial de acuíferos, bajo la premisa de que no deben comprometer los objetivos de calidad especificados. En concreto, en la parte B del Anexo VI de la Directiva Marco del Agua se considera la aplicación de dichas técnicas como una medida para conseguir el objetivo del buen estado, aunque no especifica metodologías concretas para su aplicación. No obstante, diversos autores (**Fernández Escalante, 2015**) han abordado la cuestión, incluyendo aspectos adicionales como la laminación de avenidas y el fomento de la biodiversidad en ecosistemas terrestres dependientes.

1.6. Definiciones

- A. Aguas continentales:** todas las aguas quietas o corrientes en la superficie del suelo y todas las aguas subterráneas situadas hacia tierra desde la línea que sirve de base para medir la anchura de las aguas territoriales.
- B. Aguas superficiales:** las aguas continentales -excepto las aguas subterráneas-, las aguas de transición y las aguas costeras, y, en lo que se refiere al estado químico, también las aguas territoriales.
- C. Aguas subterráneas:** todas las aguas que se encuentran bajo la superficie del suelo en la zona de saturación y en contacto directo con el suelo o el subsuelo.
- D. Aguas de transición:** masas de agua superficial próximas a la desembocadura de los ríos que son parcialmente salinas como

consecuencia de su proximidad a las aguas costeras, pero que reciben una notable influencia de flujos de agua dulce.

- E. Aguas costeras:** las aguas superficiales situadas hacia tierra desde una línea cuya totalidad de puntos se encuentra a una distancia de una milla náutica mar adentro desde el punto más próximo de la línea de base que sirve para medir la anchura de las aguas territoriales y que se extienden, en su caso, hasta el límite exterior de las aguas de transición.
- F. Cuenca hidrográfica:** la superficie de terreno cuya escorrentía superficial fluye en su totalidad a través de una serie de corrientes, ríos y, eventualmente, lagos hacia el mar por una única desembocadura, estuario o delta.
- G. Demarcación hidrográfica:** la zona marina y terrestre compuesta por una o varias cuencas hidrográficas vecinas y las aguas subterráneas y costeras asociadas, designada como principal unidad a efectos de la gestión de las cuencas hidrográficas.
- H. Norma de calidad de las aguas subterráneas:** Toda **norma de calidad medioambiental**, expresada como concentración de un contaminante concreto, un grupo de contaminantes o un indicador de contaminación en las aguas subterráneas, que no debe superarse en aras de la **protección de la salud humana y del medio ambiente**
- I. Valor umbral:** Una **norma de calidad** de las aguas subterráneas fijada por los Estados Miembros con arreglo al artículo 3
- J. Tendencia significativa y sostenida al aumento:** Cualquier **aumento significativo** desde el punto de vista estadístico y medioambiental de la concentración de un contaminante, grupo de contaminantes o indicador de contaminación en las aguas subterráneas para el que se haya determinado la necesidad de una inversión de la tendencia, de conformidad con el artículo 5
- K. Nivel de referencia**

La concentración de una sustancia o el valor de un indicador en una masa de agua subterránea correspondiente a condiciones no sometidas a alteraciones antropogénicas o sometidas a alteraciones mínimas, en relación con condiciones inalteradas;

Desde una perspectiva regulatoria (Directiva de Aguas Subterráneas) el **nivel de referencia** se refiere a la concentración de una sustancia o al valor de un indicador en una masa de agua subterránea correspondiente a condiciones caracterizadas por una influencia antrópica mínima.

L. Nivel básico

Se definen los **niveles básicos** como el valor medio medido por lo menos durante los años de referencia 2007 y 2008 de acuerdo con los programas de medidas implementados por los Estados Miembros con objeto de dar cumplimiento al artículo 8 de la misma o, en el caso de sustancias identificadas después de dichos años de referencia, durante el primer periodo en el que se disponga de datos representativos. Igualmente, se indica la posibilidad de utilización, cuando se disponga de ellos, de los datos recabados con anterioridad al comienzo del programa de control. La definición del nivel básico se realiza únicamente en las masas de agua subterránea en las que se han identificado riesgos relevantes durante el proceso de caracterización requerido por la Directiva Marco del Agua. Es decir, que se ha identificado la existencia de influencia antrópica, por lo que conseguir concentraciones de **niveles de referencia** en dichas masas parece una tarea difícil cuanto menos.

Asimismo, los Estados Miembros podrán eximir de las medidas exigidas con objeto de prevenir o limitar la entrada de contaminantes en el agua subterránea, si se requieren costes económicos desproporcionados. Ello no significa que no deba tomarse ningún tipo de medida para evitar el deterioro de la calidad del agua subterránea, y en cualquier caso las exenciones previstas sólo podrán otorgarse cuando se realice el seguimiento de las masas de agua subterránea implicadas.

M. Valor de fondo

Se define el valor de fondo como la concentración media en un periodo considerado.

Capítulo 2:

ESTADO DEL ARTE

2. ESTADO DEL ARTE

2.1. Baseline Project

El principal objetivo del proyecto “Natural Baseline Quality in European Aquifers: a basis for aquifer management (BaSeLine)”, fue el establecimiento de criterios para la definición de la calidad natural del agua subterránea y el desarrollo de una metodología de evaluación a escala europea en línea con los requerimientos de la Directiva Marco del Agua **(European Parliament and Council Directive, 2000; AIH-GE, 2003)**. Fue financiado por la Comisión Europea (proyectos EVK1-1999/0006 y EVK1-2002/ 40527) y en su desarrollo participaron instituciones de 12 países europeos. El proyecto se centró en los siguientes aspectos:

- Las diferentes escalas temporales que influyen los procesos naturales de las aguas subterráneas y la velocidad a que ocurren dichos procesos
- Variación de compuestos inorgánicos (principales y traza) en acuíferos representativos de referencia
- Carbono orgánico disuelto de forma natural en esos acuíferos
- Modelización del transporte reactivo para obtener indicaciones cuantitativas de la evolución de la calidad y de posibles cambios futuros
- Selección y aplicación de trazadores apropiados para las escalas temporales definidas
- Interpretación de tendencias en la calidad del agua
- Selección de indicadores óptimos en las redes de control

La definición de una metodología común apropiada para las necesidades de la comunidad en relación con la calidad de base del agua subterránea (tal como se define en el proyecto), implica la participación de todos los agentes implicados en la gestión de los recursos hídricos. Una aportación significativa del proyecto fue la creación de un foro de discusión compuesto por

investigadores, gestores y usuarios finales del recurso, con objeto de complementar e integrar las diferentes percepciones existentes.

La participación activa de los usuarios se fomentó en diferentes etapas. En la primera se realizaron contactos personales en cada uno de los países participantes y posteriormente se elaboró un cuestionario, en el cual se incluyeron aspectos relacionados con la calidad de base del agua subterránea.

Con objeto de evitar sesgos se analizaron los problemas de calidad del agua en acuíferos del sur de Europa con una explotación intensiva y en otros del Norte donde la extracción para regadío es menor pero existe mayor vulnerabilidad debido al poco espesor de la zona no saturada. Se consideró igualmente importante la inclusión de aspectos técnicos, económicos, sociales y administrativos **(Custodio, 2003)**.

No obstante, la definición de la calidad natural del agua subterránea es un asunto complejo que en ocasiones no es posible acotar con la precisión requerida. Es el caso de los acuíferos con altas tasas de renovación, en los que se hace necesario buscar alternativas para su definición, como puede ser la utilización de aguas fósiles en acuíferos de las mismas características o el agua en zonas de recarga con baja presión antrópica. Igualmente, es preciso señalar que en ocasiones el deterioro del agua subterránea se debe a la intensificación de las campañas de medida de la calidad conjuntamente con la mejora de las técnicas analíticas.

Como conclusiones del proyecto se señala la posibilidad de mejorar la calidad del agua subterránea por medio de una combinación de factores, entre los que tiene especial relevancia el uso del suelo y la utilización de las mejores técnicas agrarias. Igualmente, se reconoce la importancia de las redes de control como un elemento esencial en los programas de evaluación de la calidad del agua subterránea, orientadas a controlar la evolución de la calidad del agua subterránea y los impactos debidos a actividades antrópicas.

2.2. The EU Water Framework Directive: Statistical aspects of the identification of groundwater pollution trends

El principal objetivo del proyecto, financiado por la Dirección General de Medio Ambiente de la Comisión Europea y el Ministerio de Agricultura de Austria, consistió en el establecimiento de métodos para el cálculo de concentraciones medias representativas, y la evaluación de tendencias en relación con datos agregados a escala de masa de agua subterránea, considerando tanto impactos difusos como puntuales.

Se consideraron los siguientes aspectos:

Desarrollo de un método de agregación apropiado para la evaluación de la calidad del agua subterránea de una masa o grupo de masas, incluyendo los requisitos mínimos para su cálculo

Desarrollo de un método estadístico apropiado para la evaluación de tendencias y para la inversión de las mismas, así como los requerimientos para su cálculo

Consideración de concentraciones por debajo del límite de detección.

Se utilizó un conjunto de series temporales de calidad de 21 masas de agua subterránea procedente de 9 países. Como aspecto esencial se destacó la necesidad de contar con la descripción y caracterización de las masas de agua analizadas. Los parámetros disponibles para calibrar el procedimiento fueron nitratos, amonio, conductividad eléctrica, cloruros, pH, pesticidas, hidrocarburos clorados, metales, etc.

Los principales requerimientos en cuanto a las redes de control se refieren a la representatividad espacial de los puntos de control y a la existencia de valores por debajo del límite de detección. En cuanto al primero de ellos, se desarrolló un índice de representatividad con objeto de evaluar si la distribución de los emplazamientos de muestreo en una red de control es homogénea. En el caso de masas de agua subterránea heterogéneas o de la imposibilidad de establecer redes de control homogéneas, se subraya la necesidad de desarrollar redes de control hidrogeológicamente representativas.

Por otra parte, la existencia de valores por debajo del límite de detección y/o cuantificación constituye otra de las limitaciones de la metodología. Desde el punto de vista estadístico no es recomendable realizar la agregación de los datos y el análisis de tendencias cuando el límite de cuantificación excede en un 60% el valor límite (asimilable al valor umbral) en caso de que exista.

2.2.1 Agregación de datos

Para la agregación de los datos se consideran los puntos individuales de muestreo de una masa de agua subterránea en su conjunto y se calcula la media de los resultados de aquellos parámetros químicos para los que se han establecido valores umbral.

El cálculo de la media espacial se realiza por medio de un método simple de agregación espacial (**Grath et al, 2001**). Se considera el límite superior de confianza (con un nivel de significación del 95%) de la media aritmética. Dado que bajo ciertas condiciones no es posible el cálculo de la media aritmética, se propone la utilización de la media aritmética ponderada (límite superior de confianza) y la consideración de varias sub-masas de agua subterránea en caso necesario.

En caso de que las redes de control no cumplan los criterios fijados (índice de representatividad), es necesario replantear las redes o bien realizar subdivisiones que cumplan los criterios requeridos. Igualmente, es posible utilizar la media estimada por Kriging ordinario cuando los valores de intervención son superados por el límite superior del intervalo de confianza de la media seleccionada.

El intervalo de confianza depende de la variabilidad de la concentración y del número de estaciones de medida. A medida que se incrementa el número de puntos de medida disminuye el límite superior del intervalo, por lo que su utilización permite utilizar los recursos de forma efectiva, situando pocas estaciones en masas de agua subterránea con valores de concentración alejados (por defecto) de los valores umbral.

2.2.2 Tendencias e inversión de las mismas

La Directiva Marco del Agua requiere la utilización de las redes de vigilancia y operacionales para la identificación de tendencias antrópicas crecientes a largo plazo y la inversión de las mismas. El año de partida a partir del cual deben identificarse las tendencias debe ser identificado, y la inversión de tendencia debe realizarse con métodos estadísticos verificados en los que se haya establecido el nivel de confianza de forma previa.

Al igual que la agregación de datos, el análisis de tendencias debe realizarse a escala de masa de agua subterránea, con la misma metodología de agregación espacial y regularización. Dada la necesidad de garantizar la potencia y extensibilidad a otros factores de ajuste potencial, se optó por la utilización de métodos lineales (basados en un modelo lineal) en detrimento del test de Mann-Kendall, por lo que se propone el empleo de regresión lineal generalizada (ANOVA) en caso de tendencias crecientes monótonas y por un modelo en dos secciones en caso contrario, debido a su mayor simplicidad.

Para la evaluación de tendencias es preciso considerar un límite de cuantificación constante, evitando de esa forma la detección de tendencias inducidas.

2.2.3 Punto de partida para la inversión de tendencias

Se fijó como límite la detección de un incremento de concentración de un 30% con una potencia de un 90% o mayor. El punto de partida para la evaluación de tendencias es el mismo que para la red de control operacional, por lo que debe servir como alerta temprana para la detección de tendencias. El análisis de tendencias debe iniciarse cuando el límite superior del intervalo de confianza (con un nivel de significación del 95%) de la media calculada exceda el 75% del valor límite.

2.2.4 Longitud de las series para la inversión de tendencias

La Directiva Marco del Agua establece una frecuencia mínima de control de una vez por año para la evaluación de tendencias, y el requerimiento de que un incremento en la concentración de contaminante de un 30% debe ser detectada con una potencia del 90%. De acuerdo con dichos requerimientos, se deriva una longitud mínima de las series de 8 años (10 datos). En caso de frecuencia semestral o superior la longitud mínima se establece en cinco años (15 datos).

Para la inversión de las tendencias el procedimiento es análogo, resultando en longitudes mínimas de 14 años para el muestreo anual (18 datos) y en series de 10 años (30 datos) para muestreos con periodicidad mínima semestral.

2.3. Bridge Project

La Directiva sobre agua subterránea complementa las provisiones de la Directiva Marco del Agua (que en el artículo 17 insta a la Comisión y al Parlamento Europeos a adoptar medidas específicas para la protección del agua subterránea) y la Directiva existente de agua subterránea 80/68/EEC (relativa a la protección de las aguas subterráneas contra la contaminación causada por determinadas sustancias peligrosas), cuyos efectos dejaron de estar en vigor a partir de 2013.

Como consecuencia de la variabilidad natural en la composición química así como a la inexistencia de redes de observación y control adecuadas se desestimó la elaboración de una lista de estándares de calidad que pudieran ser aplicados a todas las masas de agua subterránea en Europa.

En este contexto, la Comisión lanzó el Proyecto BRIDGE (Background criteria for the identification of Groundwater Thresholds), centrado en el desarrollo de una metodología para el establecimiento de criterios de evaluación del estado químico de los acuíferos y el establecimiento de valores límite de contaminantes en el agua subterránea. Éstos últimos se definen como estándares de calidad establecidos por los Estados Miembros, y representan la

The diagram illustrates the relationship between groundwater quality and anthropogenic influence. It features a central plot with two vertical axes and a horizontal axis.

- Left Vertical Axis:** A green arrow pointing upwards, labeled "Increasing concentration".
- Right Vertical Axis:** An orange arrow pointing upwards, labeled "Increasing pollutant concentration".
- Horizontal Axis:** A dashed line labeled "Receptor-based standard".
- Diagonal Line:** A line separating the "POOR STATUS" (top) from the "GOOD STATUS" (bottom).
- Vertical Line:** A line separating "NATURAL QUALITY" (left) from "ANTHROGENICALLY INFLUENCED" (right).
- Quadrants and Regions:**
 - Top-Left (POOR STATUS, NATURAL QUALITY):** Contains a box stating "Naturally elevated concentrations giving rise to poor quality but good status groundwater".
 - Top-Right (POOR STATUS, ANTHROGENICALLY INFLUENCED):** Labeled "Polluted groundwater".
 - Middle-Right (GOOD STATUS, ANTHROGENICALLY INFLUENCED):** Labeled "Contaminated (altered) groundwater".
 - Bottom-Right (GOOD STATUS, ANTHROGENICALLY INFLUENCED):** Labeled "Pristine or high quality groundwater".
- Other Labels:**
 - Level of detection/quantification:** Indicated by a vertical double-headed arrow on the left side of the plot.
 - Vertical Axis Labels:** "Increasing concentration" (green arrow) and "Increasing pollutant concentration" (orange arrow).

El objetivo general del proyecto BRIDGE consistió en el desarrollo de una metodología para la obtención de valores umbral de contaminantes en masas de agua subterránea. Dichos valores umbral constituyen la base para la determinación del buen estado químico de las masas de agua subterránea y la evaluación de tendencias de contaminantes. Para su elaboración se fijaron los siguientes requisitos:

- Página | 46

La metodología propuesta se basa en las definiciones proporcionadas por la Directiva Marco del Agua, que se centran en los impactos en los posibles receptores de la contaminación y la importancia de dichos impactos. Por ello, cualquier valor umbral entre mal y buen estado ecológico debe obtenerse de una metodología basada en el riesgo de los posibles receptores.

Ello llevó a recomendar un proceso en etapas para derivar los umbrales medioambientales adecuados. La identificación de dichos valores debe estar basada en el conocimiento conceptual de la masa de agua subterránea y sus interacciones con los posibles receptores. Cada etapa supone un aumento de la complejidad en la toma de datos y análisis. En la primera etapa pueden utilizarse criterios conservativos, como por ejemplo la comparación directa de los datos obtenidos en los puntos de control con estándares de calidad definidos en las aguas superficiales asociadas. La segunda etapa implica una evaluación detallada de las características específicas de la masa de agua subterránea, y en etapas posteriores el requerimiento de datos aumenta aún más.

La adopción de una metodología en etapas permite mantener un balance entre la cantidad de datos necesaria y la complejidad del modelo conceptual, manteniendo proporcional el riesgo asumido con el esfuerzo realizado (a mayor riesgo mayor cantidad de datos requeridos). De esta forma se consigue una metodología práctica que permite minimizar costes manteniendo el nivel de protección a las aguas subterráneas.

2.3.1 Determinación de los valores naturales de fondo

La determinación de los valores naturales de fondo de una sustancia es importante, dado que un agua subterránea que contenga de forma natural elevadas concentraciones de una determinada sustancia no debe ser clasificada en mal estado únicamente por esta razón. Estos valores juegan un papel importante en el Proyecto BRIDGE, dado que son los puntos de partida para la definición de los valores umbral. En general no se definen para sustancias con origen antrópico exclusivamente.

Existen varias formas de determinar los valores naturales de fondo. Mientras que varios países han realizado estudios a escala nacional, otros se han centrado en estudios regionales o incluso locales. En todos los casos, aunque no se haya definido una metodología para la totalidad del país, se han realizado estudios al respecto y se dispone de indicaciones para la identificación de dichos valores.

La evaluación de muestras de agua subterránea libres de influencia antrópica ha sido utilizada en algunos de estos estudios. Este procedimiento presenta el inconveniente de que las muestras procedentes de acuíferos profundos pueden no ser representativas en muchos casos de los acuíferos someros, ya que presentan mayor mineralización en la mayoría de los casos.

La utilización de muestras procedentes de “reservas naturales”, también empleada en ciertos casos no puede ser considerada relevante a escala europea por su falta de representatividad espacial. Igualmente debe considerarse la simulación hidroquímica de los procesos de interacción entre las rocas y el agua subterránea. Sin embargo, este último procedimiento requiere un gran esfuerzo en tiempo y recursos, y es de aplicación en áreas con un gran número de puntos de control.

La utilización de métodos de separación, con objeto de separar los componentes naturales y antrópicos presentes en solución en el agua subterránea, se presenta como la técnica más utilizada y con mayores posibilidades de aplicación a escala europea. A su vez, existen dos técnicas para aplicar estos métodos, la separación de componentes por medio del análisis de la distribución de concentraciones y la utilización de métodos de preselección, mediante la exclusión de muestras con presencia de indicadores de influencia antrópica por encima de determinados valores.

Los métodos de preselección tienen varias ventajas. En primer lugar, no es necesario un conocimiento profundo del análisis estadístico, por lo que puede ser aplicado fácilmente por hidrogeólogos con poca experiencia en este campo. Igualmente puede ser aplicado en masas de agua subterránea con pocas muestras.

2.3.2 Selección del Estándar de Calidad de Referencia

Para la caracterización del estado de una masa de agua subterránea es necesario conocer la relación existente entre ésta y el receptor de la contaminación. En el caso de que el receptor sean los usos antrópicos es necesario determinar las necesidades que conllevan dichos usos y evaluar la eficacia de los tratamientos aplicados para determinar el estándar aplicable. Por ejemplo, si el agua subterránea es utilizada para abastecimiento a poblaciones sin ningún tipo de tratamiento, se aplicarán los valores establecidos en el Real Decreto 140/2003, por el que se establecen los criterios sanitarios de la calidad del agua de consumo humano.

2.3.3 Utilización de la metodología

Una vez identificados los receptores y el valor natural de fondo, pueden comenzar a aplicarse las etapas propuestas (Figura 4), en las que se evalúa el impacto de la contaminación del agua subterránea sobre los receptores con precisión creciente. En la primera etapa se determinan los valores naturales de fondo y se comparan con las concentraciones de los contaminantes existentes. En el caso de sustancias sintéticas, este valor debe ser cero.

En una segunda fase se realiza la comparación con los estándares de calidad seleccionados en función de los receptores. Esta fase se subdivide en otras dos. En la primera (2a), el valor umbral se basa en la relación existente entre el valor natural de fondo y un estándar de referencia. En la etapa siguiente (2b) se realiza la comparación directa con los valores de referencia ya existentes.

En la tercera etapa se determinará el impacto debido al agua subterránea. Por ejemplo, en el caso de un río cuyo caudal se deba tanto a escorrentía superficial como a recarga por parte de un acuífero, y cuya carga contaminante provenga básicamente de las aguas superficiales, es necesario calcular el porcentaje de impacto debido al agua subterránea con objeto de no penalizar innecesariamente el estado de la masa de agua subterránea. Se aplica el

concepto de factor de dilución con objeto de calcular dicho porcentaje de impacto.

En la cuarta y última etapa se consideran los procesos de atenuación natural que pueden ocurrir y cuya consecuencia final es la disminución del impacto en el receptor debido al

agua subterránea.

Por ejemplo, en el

caso de que se

produzcan reacciones químicas entre el contaminante y las capas subterráneas de roca o de otros estratos geológicos del acuífero, de forma que los contaminantes no alcancen los receptores, no es necesario penalizar innecesariamente el estado de la masa de agua subterránea.

Figura 4: Metodología en etapas para determinación del estado químico

2.4. Programa de control de agua subterránea de la US EPA en instalaciones de sustancias peligrosas

La Guía Unificada (USEPA, 2009) para el estudio de la calidad del agua subterránea en instalaciones industriales afectadas por Acciones de Recuperación RCRA (Resource Conservation and Recovery Act) proporciona un marco de actuación para el análisis estadístico de los datos procedentes de las redes de control. Dicha Ley para la Conservación y Recuperación de Recursos fue promulgada por el Congreso de los Estados Unidos con el objetivo de proteger la salud humana y el medio ambiente, reducir los residuos y conservar energía y recursos naturales y por último reducir o eliminar la producción de residuos peligrosos en el menor tiempo posible.

Las regulaciones del Subtítulo C de la Ley identifican los residuos sólidos que se consideran como peligrosos, y a continuación establecen diferentes requisitos administrativos para las tres categorías de entidades que manejan residuos: generadores, transportadores y dueños u operadores de instalaciones de tratamiento.

Los requerimientos de la Ley obligan a caracterizar de forma precisa la calidad del agua subterránea en dichas instalaciones, evaluar si se ha producido el vertido de compuestos peligrosos y determinar si los niveles medidos cumplen los estándares de calidad.

Para ello se han definido tres programas de control: detección, tolerancia y acción correctora. La detección es la primera etapa, cuando no se han identificado fugas (o en cantidades mínimas) de compuestos peligrosos, y su objetivo es identificar cambios significativos en la calidad del agua en relación con los niveles de fondo o con niveles de base establecidos. Las medidas tomadas en pozos situados aguas abajo de las instalaciones se comparan con los niveles de fondo, incluyendo valores de pozos situados aguas arriba. Si se encuentran evidencias estadísticamente significativas de la fuga de algún compuesto peligroso es necesario iniciar el programa de control de tolerancia, cuyo objetivo es comparar las medidas de la calidad del agua subterránea con un estándar de protección. Si un compuesto químico supera el estándar de calidad deben iniciarse acciones de remediación, cuya eficacia debe ser probada por medio de análisis estadísticos que demuestren que la concentración de los compuestos químicos peligrosos detectados se encuentra por debajo de los estándares de calidad.

2.4.1 Resultados significativos en el muestreo de agua subterránea

Es posible juzgar el resultado de un ensayo estadístico como significativo o no significativo. En el contexto de las redes de control de agua subterránea un resultado estadísticamente significativo puede obligar a cambiar el programa de control, incluso a iniciar actuaciones de recuperación, por lo que es fundamental comprender lo que representa y lo que no. En el contexto de las

redes de control de detección y tolerancia, un incremento estadísticamente significativo representa un empeoramiento de la calidad. Para determinar que dicho incremento se ha producido, es necesario que el cambio en la concentración del compuesto sea lo suficientemente grande, teniendo en cuenta la variabilidad natural del mismo, para que el cambio observado resulte improbable que haya ocurrido por azar.

2.4.2 Diseño del programa de control de detección

En él se consideran los elementos básicos que deben ser especificados en las autorizaciones administrativas, planes de control y revisiones periódicas. El programa debe ser capaz de detectar vertidos a las aguas subterráneas cuando se produzcan realmente, por lo que los ensayos deben tener la potencia estadística adecuada para detectar incrementos de concentración sobre el nivel de fondo. Se entiende por potencia efectiva de un protocolo de muestreo de una serie de contaminantes en una red de control como la probabilidad de detectar contaminación en la red de control cuando uno y solo uno de los pozos se encuentra contaminado por un único contaminante.

Otro objetivo fundamental es limitar el porcentaje de falsos positivos en el área controlada, ya que de otra forma se declararían como contaminados pozos cuya concentración es similar a la concentración de fondo.

Dado que no es posible mejorar la potencia de un test (para un número determinado de muestras utilizadas para establecer el valor de fondo) sin incrementar el riesgo asociado a los falsos positivos, es necesario desarrollar estrategias que permitan mantener bajo el riesgo total incluso aunque el número de ensayos sea muy elevado (múltiples constituyentes en gran número de pozos). Para abordar el problema de la comparación múltiple, se han desarrollado una serie de estrategias, entre las que merecen mención especial las siguientes:

Reducir el número de ensayos

- Bajar el error tipo I de cada uno de los ensayos individuales

- Cambiar el tipo de ensayo estadístico
- Aumentar el tamaño de la muestra empleada para obtener el valor de fondo

Una forma efectiva de describir la función de potencia asociada a un procedimiento particular de ensayo es por medio de gráficos. En la Figura 5 se muestra la potencia de un límite de predicción basado en una distribución normal estándar con 99% de confianza.

Figura 5: Curva de potencia normal (n=10) para un límite de predicción al 99% (de USEPA, 2009)

Las curvas de potencia identifican la probabilidad de que se produzca un incremento de la concentración en relación con el nivel de fondo superior a un valor comprendido entre 0 y 5 desviaciones estándar. Por ejemplo, la probabilidad de detectar un incremento superior a 3 desviaciones estándar es de un 55%.

2.4.3 Diseño de los programas de control de tolerancia y recuperación

La diferencia entre este tipo de programas y el anterior reside básicamente en que se evalúa la tolerancia por medio de un estándar de protección externo. Las comparaciones de efectúan por medio de ensayos de una sola muestra, sobre la que se calculan los intervalos de confianza de probabilidad asociados.

La hipótesis utilizada en el programa de control de tolerancia es la misma que en el de detección, es decir, no se ha producido contaminación, por lo que los ensayos deben detectar evidencias significativas de contaminación. En el caso de los programas de control de recuperación, la hipótesis se invierte, por lo que se considera probada la existencia de contaminación hasta que se encuentran evidencias significativas de lo contrario.

En ambos programas se realiza la distinción entre los estándares de calidad basados en concentraciones medias y aquéllos que representan concentraciones máximas (o percentiles superiores) de la distribución de concentraciones. La elección del intervalo de confianza debe estar basada en el tipo de estándar con el que se van a comparar los datos de la red de control. Por ejemplo, no deben utilizarse intervalos de confianza sobre la media aritmética con estándares basados en percentiles.

2.4.4 Establecimiento y actualización del valor de fondo

El valor de fondo proporciona información esencial en cualquier programa de control. La variabilidad natural es estimada a partir del valor de fondo, y la presencia de sustancias de origen antrópico puede ser determinada mediante la comparación de las concentraciones obtenidas aguas debajo de la instalación potencialmente peligrosa.

Para ello es preciso que el número de valores utilizado para calcular el valor de fondo sea suficiente para caracterizar su variabilidad natural, que no existan valores anómalos que distorsionen los resultados proporcionando valores anormalmente altos de la desviación estándar (con el consiguiente aumento en el intervalo de confianza del parámetro) y que las muestras sean estadísticamente independientes. Una implicación fundamental de este último requisito es que los valores deben ser estables en el tiempo, por lo que la existencia de tendencias en los valores utilizados para obtener el valor de fondo invalida esa asunción.

Por otra parte, dado que las condiciones naturales del agua subterránea tienen un carácter dinámico, no puede considerarse el nivel de fondo como un valor fijo, por lo que es necesario actualizarlo de forma periódica. Considerando que los programas de control de la EPA en instalaciones potencialmente peligrosas tienen una frecuencia cuatrimestral y que se precisan entre 4 y 8 nuevas medidas para realizar un ensayo de comparación estadística entre los valores existentes y las nuevas medidas, la actualización se realiza en general de forma anual o bianual.

Dicha actualización implica revisar las hipótesis de partida utilizadas en su construcción, como por ejemplo la hipótesis de normalidad (en caso de que se haya realizado), que debe ser verificada en el conjunto extendido de datos. La confirmación de impactos estadísticamente significativos implica que no es posible actualizar el valor de fondo, a menos que un remuestreo invalide los resultados iniciales.

2.4.5 Gestión de valores por debajo del límite de detección

En las redes de control se producen con frecuencia resultados por debajo del límite de detección, debido a las limitaciones del proceso de medida o a la técnica analítica utilizada, por lo que no es posible estimar la concentración del compuesto analizado. La única aseveración que puede ser realizada es que dicha concentración se encuentra en el intervalo comprendido entre 0 y el límite de cuantificación. Históricamente se ha constatado la utilización de técnicas fáciles de aplicar en comparación con los procedimientos específicos de gestión de valores censurados.

El principal motivo es la falta de familiaridad de los hidrogeólogos con dichas técnicas y la dificultad con el software capaz de implementarlas. Incluso en la actualidad la mayoría del software estadístico sólo incorpora rutinas para el análisis de datos censurados por la derecha (**Helsel, 2005**), por lo que para analizar los datos de concentración en agua subterránea es preciso convertir los valores censurados por la izquierda en valores censurados por la derecha y realizar posteriormente las transformaciones necesarias.

Helsel (2005) proporciona un resumen detallado de los métodos disponibles para analizar conjuntos de datos con valores censurados y concluye que los métodos de sustitución simple producen en general mayor sesgo e inexactitud que los métodos específicos en la estimación de parámetros como la media y la varianza, especialmente cuando el porcentaje de valores censurados es alto (**Guilliom and Helsel, 1986**).

Estudios posteriores (**Gibbons and Coleman, 2001**) concluyeron que la sustitución por cero producía mejores resultados que la sustitución por la mitad del límite de detección en relación con el porcentaje de falsos positivos, siempre y cuando se realizara un remuestreo de verificación. En estas circunstancias puede considerarse la sustitución simple como un método aceptable. No obstante, el estudio se realizó sin tener en cuenta la potencia de los ensayos estadísticos y mediante la utilización de límites paramétricos de predicción. En realidad, cuando los datos proceden de una distribución sesgada ninguno de los métodos de sustitución simple proporcionan resultados aceptables, con o sin remuestreo (**McNichols and Davis 1988**).

Se determinan cuatro estrategias para la gestión de valores censurados:

- Utilizar ensayos diseñados para incluir valores censurados, tales como el test de Tarone-Ware (similar a un ensayo de la t)
- Utilizar un ensayo no paramétrico basado en rangos, como el de Kruskal-Wallis
- Utilizar técnicas de estimación censuradas para obtener estimadores de la media y la desviación estándar
- Utilizar técnicas de regresión paramétrica sobre los estadísticos de orden para estimar la media y la desviación estándar

Las últimas son técnicas más antiguas que presentan la limitación de no funcionar con porcentajes de datos censurados cercanos al 50% y que además no son capaces de manejar varios límites de detección. Por ello la recomendación de la guía es utilizar las técnicas de estimación censurada para calcular la media y la desviación estándar, en particular el método de Kaplan-Meier y la regresión robusta sobre estadísticos de orden.

Capítulo 3:

***MATERIALES, MÉTODOS Y TÉCNICAS
ESTADÍSTICAS***

3. MATERIALES, MÉTODOS Y TÉCNICAS ESTADÍSTICAS

Los únicos requerimientos de la Directiva de Aguas Subterráneas en relación con la determinación de tendencias temporales en puntos de control concretos se refieren a la necesidad de que la evaluación se base en métodos estadísticos, como el análisis de regresión, de forma que se proporcionen datos de calidad científica equivalente que puedan compararse. Igualmente, deben definirse puntos de partida de inversión de dichas tendencias.

No obstante, de forma previa a la realización de cualquier análisis estadístico, incluida la regresión lineal, es necesario realizar una serie de comprobaciones, entre las que cabe destacar las siguientes:

- Verificación de normalidad en la distribución de los datos de concentración del parámetro químico analizado. En caso contrario es necesario realizar una transformación de los mismos o bien utilizar métodos no paramétricos.
- Realizar las modificaciones necesarias en los tests de hipótesis para tomar en consideración los valores por debajo del límite de detección y/o cuantificación
- Considerar las posibles causas de variaciones estacionales en los datos. Entre ellas se encuentran las fluctuaciones estacionales, la autocorrelación o la existencia de tendencias temporales. La eliminación de la variación estacional puede ser un requisito previo para la realización de un test de tendencia, que deberá ser realizado sobre los residuos del ajuste.

El artículo 8 de la Directiva Marco del Agua (DMA) establece que "los Estados miembros velarán por el establecimiento de programas de seguimiento del estado de las aguas, con el objeto de obtener una visión general coherente y completa del estado de las aguas de cada demarcación hidrográfica", indicando que en el caso de las aguas subterráneas "los programas incluirán el seguimiento del estado químico y cuantitativo".

En el anexo V la Directiva concreta los programas de seguimiento del estado químico de las masas de agua subterránea en el control de vigilancia y el control operativo. El primero se efectúa en todas las masas con objeto de

complementar y validar el procedimiento de **validación** del impacto y facilitar información para la evaluación de las tendencias prolongadas como consecuencia de modificaciones de las condiciones naturales y de la actividad antropogénica.

El control operativo se efectúa para las masas de agua subterránea identificadas en riesgo de no alcanzar el buen estado químico en 2015. Debe realizarse con la frecuencia suficiente y su objetivo, además de determinar el estado químico de las masas respecto de las cuales se haya establecido riesgo, es detectar la presencia de tendencias significativas y prolongadas al aumento de la concentración de cualquier contaminante.

Para ello se comparan los datos procedentes de la red de calidad con un valor numérico fijado por las autoridades ambientales (estándar de calidad o valor umbral). La elaboración de intervalos de confianza es el principal método utilizado en la mayoría de los programas existentes, ya que proporcionan una herramienta estadística fiable y flexible para comparar un parámetro estimado de una muestra con un límite numérico, teniendo en cuenta la variación e incertidumbre en los datos utilizados para su construcción.

Durante el control operativo la finalidad es determinar si la concentración de la sustancia se ha incrementado por encima del estándar. En cambio, la efectividad de los programas de medidas para invertir el mal estado químico debe realizarse comprobando si la concentración ha disminuido por debajo del estándar. En el primer caso el Límite Inferior de Confianza (LIC) es el criterio de interés, mientras que en el segundo el Límite Superior de Confianza es el valor de comparación. La ventaja de utilizar esta metodología radica en su posibilidad de aplicación a todo tipo de situaciones, tanto si existe un estándar de calidad definido como si no existe, o si la masa de agua de agua subterránea se encuentra en buen o mal estado químico. No obstante, el diseño de un programa estadístico depende de las características del punto concreto así como del número de observaciones disponibles y las distribuciones observadas.

Las masas de agua subterránea son sistemas complejos que presentan una alta variabilidad no sólo en su geometría y características hidrodinámicas del

medio geológico que los conforma, sino en sus propiedades físico-químicas (Ballesteros et al, 2001). Parece claro, pues, que no tiene sentido definir una concentración única para caracterizar la calidad de un acuífero en su conjunto (Blum et al, 2009). Sin embargo, la definición de valores medios de concentración es útil desde la perspectiva de la Directiva de Aguas Subterráneas.

3.1. Modelo conceptual

Para conseguir una gestión efectiva del agua subterránea de acuerdo con los

Figura 6: Modelo conceptual y heterogeneidad del sistema agua-suelo en un emplazamiento de la costa Mediterránea (Ballesteros et al, 2001)

principios establecidos en las Directivas Comunitarias, es necesario realizar, de forma previa a cualquier análisis, un modelo conceptual de la masa de agua subterránea en riesgo así como la selección de una red y un programa de control adecuado. Las conclusiones basadas en un modelo conceptual

erróneo pueden producir resultados en un sentido completamente diferente al

esperado.

La conceptualización de un sistema acuífero tiene como objetivo contribuir al mejor conocimiento de la masa de agua subterránea (Figura 6). Para ello es necesario integrar la información procedente de diversas fuentes. Para que un modelo sea fiable, debe incorporarse información obtenida de sondeos, campañas geofísicas y cartografía geológica entre otras.

Las masas de agua subterránea se encuentran en rocas permeables de diferente composición geológica, por lo que para convertir un modelo físico en un modelo funcional es preciso incorporar las características hidrodinámicas a escala regional y local. El flujo de agua subterránea se rige por las leyes de la dinámica, y las direcciones de flujo dependen del gradiente hidráulico. En algunas áreas la influencia de los sistemas de flujo regionales sobre determinados acuíferos locales es notable, por lo que la cuantificación de dicha influencia es uno de los objetivos del modelo. Igualmente, la identificación de zonas de recarga y las conexiones hidráulicas entre las masas de agua subterránea y las masas de agua superficial son aspectos relevantes a incluir en el mismo.

El modelo conceptual resulta, pues, de la integración de los modelos físico y funcional. Los modelos conceptuales no son necesariamente modelos numéricos (**European Commission, 2009**), pero deben basarse en la comprensión de las características geológicas e hidrogeológicas del sistema estudiado. No obstante, si el modelo conceptual se ha definido con rigor se puede elaborar un modelo matemático con la capacidad suficiente para estimar parámetros y condiciones de contorno, y confirmar las hipótesis del modelo así como obtener escenarios realistas para la gestión cualitativa y cuantitativa del recurso.

Tanto la evaluación del riesgo como la selección de la red de control deben basarse en el modelo conceptual. Los datos obtenidos de las redes de control definidas en los programas desarrollados para el cumplimiento de las Directivas Europeas deben ser utilizados para comprobar, validar y refinar las hipótesis del modelo. Los datos relativos a tiempos de tránsito, tasas de flujo y transporte y distribución de edades también son útiles para la validación.

Dado que una masa de agua subterránea es tridimensional, la concentración de contaminantes y los niveles de fondo en la concentración de sustancias de origen natural pueden variar de forma significativa en dirección vertical y horizontal, extremo que debe ser considerado cuando se establezcan valores umbral y se realicen los análisis relativos al estado cualitativo y a la existencia de tendencias en la masa de agua subterránea.

La Directiva Marco del Agua y la Directiva derivada de Aguas Subterráneas (Anexo II.A) proporcionan las siguientes indicaciones para el establecimiento de valores umbral en relación con el modelo conceptual definido:

- Deben basarse en el grado de interacción entre el agua subterránea y los ecosistemas acuáticos y terrestres dependientes (**López Geta et al, 2005**)
- Deben contemplar los usos actuales y potenciales del agua subterránea (por ejemplo abastecimiento o irrigación) y sus funciones (mantenimiento de ecosistemas, etc.)
- Debe determinarse valores umbral para todas las sustancias por la que el acuífero caracterizado se ha declarado en riesgo de no cumplir los objetivos fijados por la Directiva Marco del Agua en su artículo 4
- Deben incluir información basada en las concentraciones naturales, existentes como consecuencia de los procesos naturales hidrogeológicos e hidrogeoquímicos
- El proceso de derivación de los valores umbral debe incluir información relativa al origen de los contaminantes, su posible origen natural, sus propiedades toxicológicas y su potencial de dispersión, persistencia y bioacumulación
- Es necesario introducir las correcciones necesarias con respecto a la calidad de los datos y la precisión analítica en el proceso de cálculo de los valores umbral

En relación con la evaluación de tendencias (o de inversión de las mismas) el modelo conceptual juega un papel clave para la determinación de emplazamientos y frecuencias de muestreo, en el sentido de poder diferenciar las variaciones naturales de las tendencias sostenidas al aumento con un nivel adecuado de precisión y confianza

3.2. Análisis exploratorio de datos

De forma previa a la realización de cualquier análisis estadístico es necesario examinar los datos con objeto de investigar la calidad de los mismos y buscar posibles estructuras o patrones sin realizar ninguna hipótesis matemática acerca de la estructura de las observaciones. Los gráficos son una herramienta esencial para explorar y comprender pautas en cualquier conjunto de datos. La representación gráfica de los mismos proporciona información adicional a la que puede obtenerse mediante una prueba estadística formal. Por ejemplo, mediante la realización de un contraste de normalidad de Shapiro-Wilk puede concluirse que los datos no están normalmente distribuidos. Mediante la elaboración de un gráfico de probabilidad normal o de un histograma de los datos no sólo puede confirmarse esta conclusión sino que se obtiene información gráfica sobre los motivos por los cuales no puede aceptarse la hipótesis de normalidad (sesgo acusado, bimodalidad, un valor extremo aislado, etc.). Entre las principales técnicas de análisis cabe destacar las siguientes:

- *Gráficos de series temporales*

Se define una serie temporal como un conjunto de datos, correspondientes a la evolución de concentraciones de un determinado parámetro, ordenados en el tiempo. Los datos son de la forma (y_t, t) donde:

y_t : Variable endógena o dependiente

t : Variable exógena o independiente

Nota: realmente sólo hay una variable a estudiar que es y_t . En el análisis de regresión se analizan dos variables (se explica una variable a partir de la otra). Aquí sólo hay una variable, la cual se explica a partir de su pasado histórico.

La escala del eje vertical puede influir sobre la apreciación de las tendencias. En ese sentido, las escalas amplias pueden poner de manifiesto más fácilmente tendencias a largo plazo mientras que las escalas pequeñas enfatizan las tendencias a corto plazo.

- *Gráficos de caja*

Este tipo de gráficos estadísticos no es más que una representación gráfica de un conjunto de datos que brinda una impresión visual de la localización, dispersión, grado y dirección del sesgo. En el caso de una distribución que se aproxima a la forma de campana de Gauss (es decir, a la curva normal), el gráfico de caja también permite identificar los valores atípicos.

- *Histogramas*

Es una representación visual de los datos clasificados en una serie de grupos. Proporciona una herramienta para identificar la distribución subyacente de los datos. La impresión visual obtenida depende del número de grupos seleccionado. A mayor número de grupos, mayor detalle, mientras que un número pequeño aumenta el suavizado.

En el caso de medidas por debajo del límite de detección, deben utilizarse las cantidades indicadas por el laboratorio, ya que no pueden incluirse dichas medidas en el histograma.

3.3. Aplicación de técnicas de suavizado

La toma de datos de manera secuencial en el tiempo conlleva la existencia de alguna forma de variación aleatoria. En el caso de que las observaciones sean independientes y se encuentren idéntica y normalmente distribuidas se utiliza la regresión lineal para modelizar la relación existente entre una variable dependiente y una serie de variables explicativas. Los modelos así desarrollados son lineales, y permiten una explicación simplificada de la realidad. Si bien es posible modelar muchos fenómenos del mundo real de esta manera, la aplicación de modelos lineales a determinados fenómenos no es adecuada por las características intrínsecas de los mismos.

Con objeto de eliminar esa restricción se han desarrollado métodos para reducir o cancelar la variación aleatoria que realizan un número mínimo de supuestos sobre los modelos que describen las observaciones. La teoría de los modelos no paramétricos tiene por objeto el desarrollo de procedimientos de

inferencia estadística que no realizan una suposición explícita con respecto a la forma funcional de la distribución de probabilidad **(Fox, 2000)** de las observaciones de la muestra. Proporcionan una estimación suavizada de la relación para un conjunto de valores denominado ventana de la variable explicativa. El procedimiento se denomina suavizado, y la técnica más simple, los promedios móviles, fue una de las primeras en utilizarse, aunque con posterioridad se han desarrollado nuevas técnicas como la estimación kernel o la regresión local ponderada. Las medias móviles tienen el inconveniente de que el patrón presenta cambios muy bruscos, a menos que la ventana de datos sea muy grande, en cuyo caso los valores muy alejados tienen tanta influencia sobre un dato concreto como los valores más próximos.

Para evitar estos inconvenientes **(Helsel and Hirsch, 2002)** se utilizan las curvas LOWESS (Locally weighted Scatterplot Smoothing) o más genéricamente Loess (local regression) **(Cleveland, 1979)** y que se basan en ajustar modelos de regresión polinómicos locales para estimar cada punto y unir los puntos resultantes. En un primer paso se define el ancho de la ventana de datos a utilizar y se elige una función de ponderación con objeto de asignar mayor peso a los puntos más cercanos y menor peso a los más distantes. A continuación se utilizan mínimos cuadrados ponderados para ajustar una regresión polinómica dentro de la ventana y se estima el punto en cuestión. Repitiendo el proceso para cada observación en el conjunto de datos se obtiene una serie de valores que se representan en un diagrama de dispersión produciendo una curva de regresión no paramétrica.

Figura 7: Resumen análisis exploratorio datos evolución de nitratos en punto de la red de control TAIG000591 (Plana de Sagunto)

En definitiva, cuando se analizan posibles cambios en la calidad del agua, los gráficos de dispersión son una de las herramientas gráficas más utilizadas. De especial interés es la cuestión de si la relación entre las variables es lineal o curvada, si existen diferentes grupos de datos en sitios diferentes del gráfico y si la variabilidad es constante a lo largo del rango de variación de los datos. Por ello, el examen visual de la serie al que se añade un suavizado de datos tipo LOESS (Figura 7) proporciona mayor claridad en la interpretación y una indicación de cambios de tendencia a corto plazo, que pueden enmascarar de alguna forma la tendencia total de la serie.

3.4. Selección del periodo de referencia

Desde el punto de vista del marco regulatorio (DAS), nivel de fondo se refiere a la concentración de una sustancia o al valor de un indicador en una masa de agua subterránea correspondiente a la inexistencia (o a la existencia en grado mínimo) de alteraciones de origen antrópico con respecto a las condiciones naturales de dicha masa. En sentido estricto se refiere a las condiciones relativas a tiempos preindustriales, lo cual no parece muy realista (al menos para sustancias con un doble origen natural y antrópico). Para sustancias de origen no natural el nivel de fondo debe establecerse en cero (en relación con el límite de detección y/o cuantificación), mientras que para sustancias de origen natural debe ser considerado como un rango de valores (**Müller et al, 2006**).

Como se ha indicado anteriormente, nivel de base se refiere a la concentración media medida durante los años de referencia 2007 y 2008 como resultado de los programas de monitorización establecidos en consecuencia con el artículo 8 de la Directiva 2000/60/EC, o en el caso de sustancias identificadas con posterioridad a dichos años de referencia, durante el primer periodo para el cual se encuentren disponibles datos representativos. El nivel de base se define una vez que se han identificado riesgos relevantes durante el proceso de caracterización establecido por la Directiva Marco del Agua. Ello implica la existencia de presiones de origen antrópico, lo que comporta a su vez la dificultad de alcanzar concentraciones naturales de fondo en la mayoría de los casos, incluso con las medidas correctoras adoptadas.

Independientemente del punto de partida, la calidad de los datos es esencial para la elaboración de un programa de control, ya que todos los análisis estadísticos que se detallan a continuación se basan en la existencia de un nivel de referencia apropiado. Para ello es necesario que sea representativo, o en otras palabras, la distribución de sus datos debe reflejar la distribución de la población de la que se ha extraído la muestra.

Igualmente, el nivel básico debe ser apropiado para el test que se va a realizar. Por ejemplo, en el caso de que la calidad del agua subterránea haya variado en el tiempo (periodos de sequía, mejora en las técnicas analíticas utilizadas, etc.)

los datos de la red pueden no resultar adecuados para el análisis. El principal requerimiento del valor básico es que debe diseñarse de forma que refleje las condiciones actuales y las que pueden ocurrir en un futuro próximo.

Existe una serie de condiciones que debe cumplir cualquier nivel básico (o de referencia), entre las que se incluyen la independencia estadística, estacionariedad, inexistencia de valores anómalos y ajuste a las distribuciones requeridas en caso de pruebas paramétricas.

La independencia implica la no existencia de autocorrelación o tendencias en los datos, lo cual requiere generalmente una frecuencia de muestreo baja y un número mínimo de muestras comprendido entre 8 y 10 (**USEPA, 2009**). Dadas las frecuencias de muestreo empleadas en las redes de control no es probable la existencia de tales efectos, ya que dependiendo de la velocidad del agua subterránea es suficiente incluso con frecuencias trimestrales.

3.5. Análisis de tendencias

La identificación de tendencias sostenidas al aumento de la contaminación y el modo de invertir dichas tendencias es el segundo “pilar” de la Directiva de Aguas Subterráneas (**Comisión Europea, 2009**), que estipula que deberán identificarse las tendencias citadas con respecto a todos los contaminantes que caracterizan las aguas subterráneas en riesgo -aspecto relacionado con el análisis de presiones e impactos que se lleva a cabo de conformidad con la Directiva Marco del Agua. La cuestión del carácter “significativo” se puntualiza en el Anexo IV de la Directiva. Se refiere a la importancia estadística -puramente matemática- y a la importancia medioambiental, que a su vez remite a los riesgos reales que representan las tendencias al aumento identificadas.

Por tanto, con objeto de identificar la existencia o no de tendencias significativas y sostenidas al aumento de la contaminación, en cumplimiento de la Directiva 2006/118/CE, es necesario analizar la información proporcionada por las estaciones de seguimiento de calidad de aguas subterráneas en aquellas masas de agua subterránea que en función del análisis de presiones e impactos se encuentran en riesgo. Si bien la determinación de tendencias debe

realizarse únicamente en aquellas masas de agua subterránea en riesgo de no cumplir los objetivos de calidad fijados por la Directiva, los Estados Miembros pueden realizar evaluaciones en masas de agua subterránea que no se encuentren actualmente en riesgo pero en las cuales se hayan producido cambios en sus condiciones naturales o debido a la acción antrópica.

Los principios para la realización de dicho análisis se encuentran recogidos en el artículo 5 (criterios para la determinación de las tendencias significativas y sostenidas al aumento de las concentraciones de contaminantes, grupos de contaminantes o indicadores de contaminación, y definición de los puntos de partida de las inversiones de tendencia) y Anexo IV de la Directiva de Aguas Subterráneas, así como en los criterios y procedimientos descritos en el Documento Guía de la Comisión Europea nº 18 “*Guidance on Groundwater Chemical Status and Trend Assessment*”.

Asimismo, el anexo IV de la Directiva 2006/118/CE establece lo siguiente:

“PARTE A: DETERMINACIÓN DE TENDENCIAS SIGNIFICATIVAS Y SOSTENIDAS AL AUMENTO

Los Estados Miembros determinarán las tendencias significativas y sostenidas al aumento en todas las masas o grupos de masas de agua subterránea que presenten un riesgo con arreglo al Anexo II de la Directiva 2000/60/CE, teniendo en cuenta los siguientes aspectos:

1. De conformidad con el punto 2.4 del Anexo V de la Directiva 2000/60/CE, el programa de seguimiento deberá concebirse de forma tal que detecte las tendencias significativas y sostenidas al aumento de las concentraciones de los contaminantes que se hubieran determinado con arreglo al artículo 3 de la presente Directiva.

*2. El **procedimiento** de determinación de la aparición de tendencias significativas y sostenidas al aumento se llevará a cabo de la siguiente forma:*

- a) se elegirán **frecuencias y puntos de control** que sean **suficientes** para:*

i) proporcionar la información necesaria para garantizar que dichas tendencias al aumento puedan distinguirse de las variaciones naturales con un nivel adecuado de fiabilidad y precisión;

ii) permitir que dichas tendencias al aumento se determinen con tiempo suficiente para que puedan aplicarse medidas con objeto de impedir, o cuando menos mitigar en la medida de lo posible, cambios adversos en la calidad del agua que sean significativos para el medio ambiente. Esta determinación se llevará a cabo por primera vez en 2009, si es posible, y teniendo en cuenta los datos existentes, en el informe sobre determinación de tendencias del primer plan hidrológico de cuenca establecido en el artículo 13 de la Directiva 2000/60/CE, y después cada seis años como mínimo;

iii) tener en cuenta las características temporales, físicas y químicas, de la masa de agua subterránea, incluidas las condiciones de flujo y los índices de recarga del agua subterránea, así como el tiempo que ésta tarda en atravesar el suelo o el subsuelo;

b) se utilizarán métodos de control y análisis acordes con los principios internacionales de control de la calidad, entre ellos, si procede, las normas CEN o los métodos nacionales normalizados, para garantizar que se proporcionen datos de calidad científica equivalente que puedan compararse;

c) la evaluación se basará en un método estadístico, como el análisis de regresión, para analizar las tendencias en series temporales en puntos de control concretos;

d) con el fin de evitar sesgos en la determinación de las tendencias, todas las mediciones por debajo del límite de cuantificación se cifrarán en la mitad del valor del límite de cuantificación más alto registrado durante el período, con excepción de los plaguicidas totales.

3. Para la determinación de tendencias significativas y sostenidas al aumento en las concentraciones de sustancias que se produzcan naturalmente y como resultado de actividades humanas **se considerarán los niveles básicos y, cuando se disponga de ellos, los datos recabados con anterioridad al**

comienzo del programa de control, con objeto de informar acerca de la determinación de tendencias en el primer plan de gestión de cuenca hidrográfica a que hace referencia el artículo 13 de la Directiva 2000/60/CE.

PARTE B: PUNTO DE PARTIDA DE LAS INVERSIONES DE TENDENCIA

De acuerdo con el artículo 5, los Estados Miembros invertirán las tendencias significativas y sostenidas al aumento teniendo en cuenta los siguientes requisitos:

1. El **punto de partida** para aplicar medidas destinadas a invertir tendencias significativas y sostenidas al aumento será el momento en el cual la concentración del contaminante alcance el **75% de los valores paramétricos de las normas de calidad** (Figura 8) de las aguas subterráneas establecidas en el Anexo I y de los valores umbral establecidos con arreglo al artículo 3, **a menos que:**

a) sea necesario un punto de partida anterior para hacer posible que las medidas de inversión impidan del modo más rentable, o al menos mitiguen en la medida de lo posible, cualquier cambio adverso significativo en la calidad del agua subterránea;

b) se justifique un punto de partida distinto si el límite de detección no permite establecer la presencia de una tendencia cifrada en el 75% de los valores paramétricos; o

b) la tasa de aumento y la reversibilidad de la tendencia sean tales que, de tomarse un punto de partida posterior para aplicar medidas de inversión de la tendencia, éste seguiría haciendo posible que dichas medidas impidan del modo más rentable, o al menos mitiguen en la medida de lo posible, cualquier cambio adverso significativo desde el punto de vista medioambiental en la calidad del agua subterránea. Este punto de partida posterior no podrá suponer retraso alguno en el cumplimiento de los plazos para el logro de los objetivos medioambientales.

Figura 8: Elementos de la evaluación e inversión de tendencias (European Commission, 2009)

Para las actividades incluidas en el ámbito de aplicación de la Directiva 91/676/CEE, el punto de partida para la aplicación de medidas para invertir tendencias significativas y sostenidas se establecerá de conformidad con dicha Directiva y con la Directiva 2000/60/CE, en particular la adhesión a los objetivos medioambientales de protección de las aguas establecidos en su artículo 4.

Evaluación de tendencias (artículo 5.1 y 5.2 de la DAS)	Test	Test			
		Evaluación de tendencias	Evaluación de la inversión de tendencias	Determinación a nivel de la MAS	Puntos de control pertinentes
Determinar e invertir tendencias que supongan un riesgo significativo para los usos reales o potenciales del medio acuático	Ningún daño a los usos legítimos.	X	X	X	X
Determinar e invertir tendencias que supongan un riesgo significativo para la calidad de los ecosistemas acuáticos	Ningún daño a los ecosistemas acuáticos.	X	X	X	X
Determinar e invertir tendencias que supongan un riesgo significativo para los ecosistemas terrestres	Ningún daño a los ecosistemas terrestres.	X	X	X	X

Tabla 1: Evaluación de tendencias (artículo 5.1 y 5.2 de la DAS). Resumen de elementos y tests correspondientes

Nueva evaluación de tendencias	Test	Evaluación de tendencias	Evaluación de la inversión de tendencias	Determinación a nivel de la MAS	Puntos de control pertinentes
Se considerará una evaluación de tendencias para comprobar que los penachos resultantes de lugares contaminados no se expanden, no deterioran el estado químico de la masa o grupo de masas de agua subterránea y no suponen un riesgo para la salud humana ni para el medio ambiente (DAS, art. 5.5).	No hay expansión de penachos que supongan deterioro del estado químico ni riesgo para la salud humana ni el medio ambiente.	X			X
Evaluación del estado					
En la masa de agua subterránea no hay entradas ni conato de entrada de agua de mar ni de agua de una composición química sustancialmente diferente de otras masas de agua subterránea o aguas superficiales que pueda causar contaminación (DMA, anexo V 2.3.2).	No hay intrusión salina ni de otro tipo.	X			X
No hay deterioro de la calidad de las aguas para el consumo humano (DAS, artículo 4.2.c), iii) y anexo III 4)	Cumple los requisitos del artículo 7.3 de la DMA Zonas protegidas de captación de agua potable	X			X

Tabla 2: Evaluación adicional de tendencias. Resumen de elementos y tests correspondientes

1. *Una vez establecido un punto de partida para una masa de agua subterránea en riesgo de conformidad con el punto 2.4.4 del Anexo V de la Directiva 2000/60/CE y con arreglo al punto 1 anterior, aquél no se modificará durante el ciclo de seis años del Plan Hidrológico de Cuenca establecido de conformidad con el artículo 13 de la Directiva 2000/60/CE.”*

Para realizar la evaluación de tendencias significativas al aumento o al descenso es necesario considerar los siguientes puntos:

- Selección del método estadístico más apropiado para evaluar tendencias en cada punto de control
- Tratamiento de valores por debajo del límite de cuantificación
- Longitud de las series de datos

- Determinación de niveles de base para sustancias que ocurren simultáneamente de forma natural y debido a actividades antrópicas
- Niveles de confianza aceptables para la evaluación de tendencias
- Determinación de los puntos de partida para la inversión de tendencias
- Verificación estadística y nivel de confianza en la inversión de tendencias.

La longitud de la serie de datos necesaria para la evaluación de tendencias depende de dos factores diferenciados. Por una parte, de las características intrínsecas de la masa de agua subterránea y por tanto su velocidad de reacción ante actividades de carácter antrópico (cambio de usos del suelo, etc.), y por otra de la potencia de los ensayos estadísticos aplicados y de la calidad de los datos.

El número mínimo de medidas necesario para realizar la evaluación depende de la frecuencia de muestreo, del método de ensayo utilizado y de la potencia del mismo. El número máximo de medidas de la serie temporal viene condicionado por el modelo conceptual de la masa de agua subterránea y por la evolución temporal de concentraciones en dicho punto. El uso de series con gran cantidad de datos puede producir resultados sesgados por los primeros o los últimos años de medidas.

En el caso, por ejemplo, de oscilaciones en las tendencias, debe tomarse la última parte de los valores, una vez que se haya comprobado que éstos son consistentes con el modelo conceptual del acuífero (cambios estacionales, etc.)

3.5.1. Parámetros a controlar

La Directiva de Aguas Subterráneas especifica que deben identificarse las concentraciones de contaminantes, grupos de contaminantes o indicadores de contaminación en las masas de agua subterránea declaradas en riesgo, aunque no especifica los parámetros que deben ser sometidos a evaluación en su tendencia. Dado que el punto de partida para la inversión de tendencias debe ser establecido en relación con los estándares de calidad o los valores

umbral definidos para parámetros que hacen estar a la masa de agua subterránea en riesgo, parece evidente que las tendencias deben ser evaluadas para dichos parámetros.

Es posible, no obstante, la evaluación de tendencias en otras sustancias debidas a la actividad antrópica que puedan provocar tendencias ambientales significativas en el futuro, que de esta forma puedan servir como indicadores que puedan proporcionar alertas tempranas.

La identificación de tendencias para la identificación de sustancias que ocurren simultáneamente de forma natural y como resultado de actividades antrópicas debe tener en consideración los niveles de base, entendiendo como tales los valores medios medidos durante los años de referencia 2007 Y 2008, como resultado de los programas de medida y control establecidos al amparo del artículo 8 de la Directiva 200/60/EC. En el caso de sustancias identificadas con posterioridad a dicho periodo se tomará como referencia el primer periodo en el cual se encuentren datos representativos (Directiva de Aguas Subterráneas, Artículo 2 párrafo 6).

Los niveles de base son el punto de referencia mediante el cual se evaluarán los cambios posteriores (tendencias) en la concentración de contaminantes. Debe destacarse que, por su propia definición, los niveles de base son conceptualmente diferentes a los niveles de fondo, o concentración natural de la sustancia de que se trate en condiciones de no afección.

3.6. Actualización del nivel básico

Debido a la complejidad del comportamiento del agua subterránea y a la necesidad de contar con muestras de tamaño suficiente para la realización de los tests estadísticos, no debe considerarse el nivel básico como algo estático. Por ello debe ser revisado y actualizado periódicamente. Aunque su actualización puede parecer algo conceptualmente simple, es necesario comprobar la existencia de tendencias o cambios en la calidad antes de agregar nuevos valores. La adición de valores individuales al nivel básico

podría introducir tendencias de pequeña magnitud, que no alcanzarían a ser detectadas por los ensayos estadísticos.

En la Directiva de Aguas Subterráneas se menciona de forma explícita la posibilidad de incorporar datos procedentes de registros históricos para la elaboración del nivel básico. La agregación de medidas adicionales a las existentes tiene la ventaja de que se aumenta el poder estadístico de los tests y se pueden elaborar intervalos de predicción más precisos. Igualmente es posible incorporar la variabilidad natural del acuífero.

Por tanto, el paso siguiente a la realización del análisis preliminar de datos es la selección de los valores que se utilizarán para actualizar el nivel básico. En el proceso de toma de decisiones en relación con dicha selección es fundamental realizar un análisis detallado de las causas a las que se deben los patrones existentes en los datos o las tendencias observadas. Si los datos seleccionados son normales o susceptibles de ser normalizados mediante una transformación, es posible la utilización de técnicas paramétricas, mientras que en caso contrario la utilización de técnicas no paramétricas es la única alternativa viable. En ambos casos es necesario comprobar el cumplimiento de una serie de asunciones tales como la independencia estadística, estacionalidad espacial y temporal e inexistencia de valores anómalos entre otras. De forma adicional, en el caso de utilizar estadística paramétricas, es preciso comprobar las hipótesis realizadas sobre la distribución.

La existencia de tendencias en el nivel básico indica que el valor medio está cambiando, lo que contradice la hipótesis de variables aleatorias idénticamente distribuidas, que exige la necesidad de que la media y varianza de una serie de datos sean estables en el tiempo. Las fluctuaciones estacionales también vulneran esta hipótesis, puesto que en este caso es probable que tanto la media como la varianza oscilen.

La actualización no debe realizarse hasta que se hayan recogido valores suficientes para que pueda efectuarse una comparación estadística entre los valores existentes y el nuevo conjunto de datos. Dado que para ello se necesitan entre 4 y ocho nuevas medidas y asumiendo un muestreo semestral, no debería producirse antes de dos o tres años.

La determinación de tendencias se llevará a cabo por primera vez en 2009, si es posible, y en lo sucesivo, una vez cada seis años como mínimo (anexo IV A, 2. ii) de la DAS), teniendo en cuenta los datos existentes obtenidos en el control de vigilancia y el control operativo, así como los datos de seguimiento recopilados antes del comienzo del programa de seguimiento. Esto permitirá informar sobre las tendencias en el primer PHC (anexo IV A, 2. a), ii) y anexo IV A, 3). Habida cuenta de que el proyecto de Plan Hidrológico de Cuenca se someterá a la participación pública un año antes de su entrada en vigor, se recomienda que, si es posible, los Estados miembros evalúen las tendencias y la inversión de tendencias antes de presentar el proyecto de Plan.

Por otra parte, es necesario recordar que cada vez que se actualice el nivel básico es necesario recalcular todos los límites estadísticos basados el nivel básico, como los intervalos de predicción o control.

Cuando se dispone de un número reducido de datos es aconsejable utilizar técnicas no paramétricas (una excepción es el caso en el que se disponga de información precisa relativa a la distribución de los datos), puesto que los tests de normalidad no tienen la potencia necesaria en esos casos para decidir si los datos provienen de una distribución gaussiana, a menos que la desviación con respecto a la normalidad sea muy elevada (**Helsel and Hirsch, 2002**). En dicho supuesto la presencia de valores anómalos puede alterar completamente los resultados del análisis y producir como resultado la definición de intervalos de predicción de longitud desproporcionada que por lo tanto no conlleve la protección del medio ambiente.

3.7. Intervalos estadísticos

Se han utilizado diversas metodologías para la comparación de datos procedentes de las redes de control con un estándar de calidad determinado. La más aceptada es la elaboración de intervalos de confianza. En el caso de analizar la evolución de valores medios, el procedimiento consiste en estimar la media (o mediana) de la población por medio de la muestra (valor básico),

calculando un intervalo en el que previsiblemente se encuentre incluido su verdadero valor en vez de estimar el parámetro mediante un valor único.

Una de las ventajas de los intervalos de confianza frente a los tests de hipótesis tradicionales es la información adicional que proporcionan. Los límites superior e inferior indican las magnitudes máximas o mínimas del parámetro estimado, por lo que es posible descartar valores alejados de dichos

Figura 9: Posición relativa del intervalo de confianza de la media o mediana en relación con el estándar de calidad ambiental

límites. Por otra parte, la anchura del intervalo también aporta información valiosa. Un intervalo de confianza estrecho indica que el tamaño de la muestra es grande, por lo que la estimación del parámetro es precisa. Otra forma de expresar lo anterior es decir que el análisis tiene una potencia razonable. En el caso contrario, es decir, cuando el intervalo es muy amplio, la estimación es imprecisa y el estudio tiene poca potencia, lo cual suele suceder en muestras de pequeño tamaño.

A primera vista la media o la mediana parecen los valores indicados para efectuar la comparación con los valores procedentes de la red de control y comprobar la existencia de procesos de degradación en la calidad del agua subterránea. No obstante, ambos tienen la desventaja de que representan valores centrales, por lo que la probabilidad de que sean superados por datos procedentes de la misma distribución es del cincuenta por ciento en términos generales.

El límite inferior de confianza es utilizado en el control operativo, es decir, cuando el objetivo es verificar si los datos de concentración superan un determinado umbral, mientras que el límite superior se utiliza en las acciones de recuperación (Figura 9). En este último caso se supone afectada la masa de

agua subterránea y el objetivo consiste en evaluar la efectividad de las medidas de recuperación implementadas para la inversión de tendencias.

Los intervalos de confianza, como cualquier otro método estadístico, se basan en la suposición de que los datos de la muestra son representativos de la población de la cual han sido seleccionados, lo que significa que su distribución estadística puede considerarse similar a la de la mayoría de los datos de concentración de la población (que no han sido medidos).

Por tanto, un aspecto importante a la hora de seleccionar el método apropiado de ensayo es determinar si es posible modelizar los datos por medio de una distribución determinada (generalmente la normal) o no. En el primer caso se utilizan tests paramétricos, que presentan la ventaja de que proporcionan mayor potencia estadística (siempre que la hipótesis relativa a la distribución de los datos sea cierta). En el segundo se emplean tests no paramétricos, más apropiados para analizar datos que no siguen una distribución conocida o que presentan un porcentaje significativo de valores por debajo del límite de detección.

3.7.1. Pruebas de bondad de ajuste

Dado que la selección de un ensayo apropiado se basa en la distribución esperada de los datos, es preciso realizar contrastes de ajuste. Para ello es preciso comprobar si las distribuciones estadísticas más utilizadas, como la normal, simulan de forma apropiada los datos de la muestra. Existe una gran variedad de modelos en la literatura estadística, aunque la mayoría de ellos no son apropiados para analizar parámetros de concentración del agua subterránea, introduciendo, en cambio, una mayor complejidad matemática. Por ello, los modelos estadísticos se limitan a la distribución normal, lognormal y gamma.

3.7.2. Cálculo de estadísticos

En el caso de una distribución gamma $G(k, \theta)$, la media y la varianza de la población son funciones de los parámetros k y θ , por lo que para estimar la media previamente es necesario obtener estimadores de ambos. El cálculo del estimador de máxima verosimilitud de k es complejo, y requiere el procesamiento de funciones Digamma y Trigamma **(Choi and Wette, 1969)**. Aunque no existe una solución analítica, utilizando métodos numéricos es posible obtener un estimador de k mediante el método de Newton-Raphson **(Faires and Burden, 1993)**

3.7.3. Intervalos paramétricos

La distribución normal ha sido ampliamente utilizada, en parte porque ocurre a menudo y en parte porque es sencillo obtener resultados de aplicación inmediata, aunque este último punto ha perdido importancia con el desarrollo de los métodos informáticos.

Muchos datos ambientales son intrínsecamente positivos y a menudo presentan un sesgo acusado, por lo que la hipótesis de normalidad no es siempre adecuada. No obstante, en numerosas ocasiones es posible modelizar los datos por medio de distribuciones lognormales o gamma. En la práctica, suponiendo que los datos se ajustan a ambas, es difícil distinguir entre ellas cuando el tamaño de la muestra es inferior a 50.

Debido a su mayor facilidad computacional la distribución lognormal ha sido una elección común en el análisis estadístico de dichos datos. En este caso particular es conveniente recordar que aunque los intervalos de confianza se calculan de la misma manera sobre los valores transformados, la transformación inversa para volver a la escala original de concentraciones proporciona una estimación del intervalo de confianza de la mediana o la media geométrica.

En el caso de normalidad de los datos el intervalo de confianza (de una media) (basado en el estadístico de Student) viene dado por la siguiente fórmula:

$$LCL_{1-\alpha} = \bar{x} - t_{(1-\alpha, n-1)} \frac{s}{\sqrt{n}} \quad UCL_{1-\alpha} = \bar{x} + t_{(1-\alpha, n-1)} \frac{s}{\sqrt{n}}$$

Ecuación 1: Intervalo de confianza de la media de un conjunto de datos normales

Para una población distribuida normalmente (cuando el sesgo tiende a cero), la ecuación 3.1 proporciona el estimador óptimo de la media. También puede utilizarse cuando el sesgo es moderado ($|\text{sesgo}| < 0.5$). En caso de que sea superior, es posible que la Ecuación 1 no proporcione la cobertura deseada (0.95) a la media de la población, lo cual es evidente cuando el número de muestras es inferior a 20-25 (**Singh et al. (2002), Singh and Singh (2003)**).

El intervalo de confianza de la media en distribuciones lognormales se basa en el estadístico H de Land (**Land, 1971**) que, no obstante, proporciona valores inestables y anormalmente elevados del límite superior de confianza cuando la desviación estándar es alta, especialmente para muestras de pequeño tamaño (< 15-20 datos). Igualmente, el límite calculado de esta forma es muy sensible a valores anómalos. Por ejemplo, la adición de un valor por debajo del límite de detección puede provocar que el límite superior de confianza supere ampliamente el valor máximo, en cuyo caso la EPA recomienda la utilización de dicho valor máximo. El problema de utilización del valor máximo es que el valor de confianza es frecuentemente inferior al 95%.

Por ello es preferible comprobar en primer lugar si los datos se ajustan a una distribución gamma, ya que ésta proporciona valores estables y precisos de los límites. El intervalo de confianza de la media debe calcularse por medio de una distribución gamma ajustada cuando $0.1 \leq k < 0.5$ o por medio de una gamma aproximada cuando $k \geq 0.5$. Para valores de k inferiores a 0.1, el intervalo puede ser calculado por cualquiera de los dos cuando el número de muestras es superior a 15, pudiendo ser utilizados los métodos bootstrap de Hall o de la t en caso contrario.

Se dice que una variable aleatoria continua (concentración de contaminante en agua subterránea) sigue una distribución gamma, $G(k, \theta)$, con parámetros $k > 0$ (factor de forma) y $\theta > 0$ (factor de escala) cuando su función de densidad de probabilidad se encuentra dada por la siguiente ecuación:

$$f(x; k, \theta) = \frac{x^{k-1} e^{-x/\theta}}{\theta^k \Gamma(k)}; \text{ para } x > 0$$

$$= 0; \text{ en caso contrario}$$

Ecuación 2: Función de densidad de probabilidad de una distribución gamma

De acuerdo con **Singh et al. (2002)**, dada una muestra aleatoria de tamaño n procedente de una distribución gamma, $2nk/\theta$ sigue una distribución χ^2 con $2nk$ grados de libertad. Cuando el parámetro de forma es conocido, el test más potente de la hipótesis nula $H_0: \mu_1 \geq C_s$ frente a la hipótesis alternativa $H_1: \mu_1 < C_s$ consiste en rechazar la primera cuando:

$$\bar{x} / C_s < \chi^2_{2nk}(\alpha) / 2nk$$

Ecuación 3: Condición para rechazar la hipótesis nula $\mu_1 \geq C_s$ para el límite superior de confianza de la media en una distribución gamma

El límite superior de confianza más potente uniformemente para la media viene dado por la siguiente ecuación:

$$P(2nk\bar{x} / \chi^2_{2nk}(\alpha) \geq \mu_1) = 1 - \alpha$$

Ecuación 4: Probabilidad del límite superior de confianza de la media en una distribución gamma

En la práctica no se conoce k , por lo que es necesario estimarlo a partir de los datos. A partir de $\hat{k}$ y $\hat{\sigma}$ (estimadores de máxima verosimilitud de los parámetros de la distribución) una aproximación razonable consiste en sustituir

el valor de $\hat{k}$ por su estimador insesgado $\hat{k}^*$, obtenido de la siguiente ecuación (Johnson, Kotz, and Balakrishnan 1994):

$$\hat{k}^* = (n-3)\hat{k} / n + 2/(3n)$$

Ecuación 5: Estimador de k
corregido para el sesgo en una
distribución gamma

Por tanto, el límite superior de confianza aproximado viene descrito por la siguiente ecuación:

$$UCL_{aproximado} = 2n\hat{k}^* \bar{x} / \chi^2_{2n\hat{k}^*}(\alpha)$$

Ecuación 6: Valor aproximado del límite
superior de confianza para la media en una
distribución gamma

De forma análoga, el límite inferior de confianza de la media se obtiene de:

$$LCL_{aproximado} = 2n\hat{k}^* \bar{x} / \chi^2_{2n\hat{k}^*}(1-\alpha)$$

Ecuación 7: Valor aproximado del límite
inferior de confianza para la media en una
distribución gamma

Los límites así calculados son aproximados y no hay garantía de que se consiga el nivel de confianza $(1-\alpha)$, aunque proporciona un método de estimación de la media de una distribución gamma. Según los estudios de simulación de **Singh et al (2002)**, los límites de confianza proporcionan una cobertura del 95% cuando el factor de forma se aproxima a 0.5 incluso para muestras de pequeño tamaño ($n=5$).

Con objeto de conseguir el nivel de confianza deseado, **Grice and Bain (1980)** proporcionan un nivel de significación ajustado que puede ser usado en la ecuación 3.6 para conseguir dicho nivel de confianza. Los límites de confianza ajustados pueden ser expresados por:

$$UCL_{ajustado} = 2n\hat{k}^* \bar{x} / \chi^2_{2n\hat{k}^*}(\beta)$$

Ecuación 8: Valor ajustado del límite
superior de confianza para la media en
una distribución gamma

El nivel ajustado de significación β (**Grice and Bain, 1980**) se obtiene de la siguiente tabla:

n	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$
	probability level, β	probability level, β	probability level, β
5	0.0086	0.0432	0.0000
10	0.0267	0.0724	0.0015
20	0.0380	0.0866	0.0046
40	0.0440	0.0934	0.0070
--	0.0500	0.1000	0.0100

Tabla 3: Nivel ajustado de significación β

Los valores de tamaño de muestra diferentes a los que figuran en la Tabla 3 se obtienen por interpolación.

En cuanto al límite inferior de confianza de una cola de una distribución lognormal, según **Land (1971,1975)** se calcula a partir de la ecuación:

$$\text{LCL} = \exp(\bar{y} + 0.5s_y^2 - s_y H_{1-\alpha} / \sqrt{(n-1)})$$

Ecuación 9: Límite inferior de confianza en una distribución lognormal

Cuando la población es lognormal el valor proporcionado por la ecuación 3.8 es el límite de confianza insesgado más preciso de manera uniforme. No obstante, cuando existen valores extremos o datos procedentes de varias distribuciones, los resultados basados en el estadístico H pueden ser inaceptablemente elevados y de utilidad práctica, por tanto, muy escasa.

Los estimadores insesgados de mínima varianza de una distribución lognormal $\hat{\mu}_1$ y $\hat{\sigma}_1^2(\hat{\mu}_1)$ se obtienen de:

$$\begin{aligned}\hat{\mu}_1 &= \exp(\bar{y})g_n(s_y^2/2) \\ \sigma_1^2 &= \exp(2\hat{y})[g_n(2s_y^2) - g_n((n-2)s_y^2/(n-1))]\end{aligned}$$

Ecuación 10: Estimadores insesgados de mínima varianza de los parámetros de una distribución lognormal

Donde $g_n(u)$ es una función descrita en **Bradú and Mundlak (1970)**, posteriormente tabulada por **Gilbert(1987)**.

Igualmente es posible estimar el intervalo de confianza de la media en distribuciones lognormales por medio de estimadores insesgados de mínima varianza de la media y su desviación estándar.

Para ello se utiliza la desigualdad de Chebyshev:

$$P(-k\sigma_1 \leq x - \mu_1 \leq k\sigma_1) \geq 1 - \frac{1}{k^2}$$

Ecuación 11: Desigualdad de Chebyshev

Para el cálculo del intervalo de confianza se utiliza la siguiente ecuación:

$$CL = \hat{\mu}_1 \pm 4.359\hat{\sigma}_1(\mu_1)$$

Ecuación 12: Intervalo de confianza de la media de Chebyshev

3.7.4. Intervalos no paramétricos

La mayoría de los métodos no paramétricos de obtención de intervalos de confianza se basan en métodos de remuestreo (bootstrap) (**Efron y Tibshirani, 1993**). El procedimiento consiste en generar muestras de tamaño n de un conjunto de observaciones, con remplazamiento y repetir el proceso un número elevado de veces (1000-5000) como base para estudiar el comportamiento de

determinados estadísticos. A diferencia de los métodos estadísticos clásicos, la distribución del estadístico se determina simulando muestras aleatorias construidas directamente a partir de las observaciones, en lugar de realizar inferencias sobre la población.

La primera alusión a este método bajo tal denominación se debe a **Efron (1979)**, el cual concibió el método bootstrap como una ampliación a las técnicas jackknife, cuya mecánica consiste en producir muestras construidas al suprimir cada vez un solo elemento de la muestra original para valorar el efecto sobre determinados estadísticos (**Gil Flores, 2005**).

El método es más eficiente a medida que aumenta el número de datos, ya que con ello aumenta la información que proporciona la muestra sobre la población. No obstante, el método proporciona resultados correctos incluso con muestras pequeñas, pudiendo establecerse como inadecuados resultados obtenidos en muestras de tamaño inferior a cinco (**Chernick, 1999**).

Existen varias formas de conseguir intervalos de confianza de la media a través de técnicas bootstrap (Figura 10), entre las que figuran los siguientes:

- la metodología estándar utiliza la misma estructura que los procedimientos paramétricos en la construcción de intervalos de confianza. Si es posible asumir que el estadístico se distribuye según la curva normal pero el cálculo de la desviación estándar es complejo, la distribución muestral bootstrap proporciona un medio para su estimación. A continuación se introduce dicha desviación estándar en la expresión del intervalo de confianza paramétrico.
- El método del percentil hace uso literal de la idea del bootstrap. Suponiendo que $\hat{\theta}_i^*$ es el i-ésimo estimador bootstrap del parámetro seleccionado a partir de muestras de n elementos, se toma el intervalo que contiene el 90% de los $\hat{\theta}_i^*$ ordenados de menor a mayor como el intervalo de confianza de θ . El procedimiento consiste, pues, en utilizar los cuantiles empíricos de las replicaciones para obtener dicho intervalo. No obstante, en el caso de distribuciones asimétricas el método no proporciona resultados exactos.

- El tercer método (bootstrap accelerated bias-corrected percentile, BCA) intenta corregir la asimetría que puede presentar la distribución muestral del estadístico **(Efron y Tibshirani, 1993)**, para lo cual se modifica el método de cálculo de los percentiles. **Singh and Singh (2003)** evaluaron la efectividad del método en relación con los métodos paramétricos y otras técnicas bootstrap, apreciando que para datos no simétricos representa una ligera mejora comparado con el método del percentil simple. Aún así, la cobertura proporcionada es menor que la obtenida por aplicación de los métodos t y de Hall, especialmente en el caso de muestras de pequeño tamaño.
- Bootstrap t. Es otra variación de los métodos de remuestreo, que consiste en utilizar la metodología bootstrap para estimar los cuantiles del estadístico en lugar de utilizar los del estadístico de la t de Student.
- Método de **Hall, 1982**. Ajusta tanto para el sesgo como para la asimetría. En el caso de datos fuertemente asimétricos proporciona mejor cobertura que el bootstrap t, por lo que ha sido incluido por la EPA **(USEPA, 2002)** en el documento guía para el cálculo del límite superior de confianza.

En el caso de distribuciones gamma los dos últimos proporcionan coberturas similares, cercanas al 95% para muestras grandes. Este porcentaje disminuye a medida que se reduce el tamaño de la muestra. Para muestras distribuidas según la lognormal, las coberturas proporcionadas por ambos métodos son significativamente inferiores al 95%, incluso con muestras grandes (n=100) **(Singh and Singh, 2003)**. Por último, no es aconsejable su uso en presencia de valores anómalos, ya que proporcionan valores inestables, erráticos y anormalmente elevados **(Efron y Tibshirani, 1993)**.

Figura 10: Histograma de la media e intervalos de confianza para la media de una muestra mediante técnicas bootstrap

3.8. Diseño del programa de control con umbral definido

Cuando se quiere determinar el estado químico de una masa de agua subterránea y las autoridades ambientales han definido un valor umbral, los objetivos prioritarios son la identificación de tendencias sostenidas al

Figura 11: Posición relativa del valor umbral en relación con la norma de calidad

aumento en la concentración de

contaminantes y la verificación de conformidad con el umbral especificado. Las estrategias de control se basan frecuentemente en un análisis de tendencias seguido de la comparación de los datos procedentes de las redes de control con valores numéricos establecidos como estándares de protección de las aguas subterráneas.

En el Anexo II de la Directiva se definen los valores umbral como estándares de calidad del agua subterránea que deben ser fijados por los Estados Miembros en consonancia con el artículo 3. Deben ser establecidos para aquellos contaminantes e indicadores de contaminación (que determinan la inclusión de una masa o grupo de masas de agua subterránea en la categoría de en riesgo de no conseguir el buen estado cualitativo) definidos como resultado de la caracterización desarrollada de acuerdo con el artículo 5 de la Directiva.

Su definición debe basarse en el grado de interacción entre el agua subterránea y los ecosistemas terrestres y acuáticos dependientes asociados a dicha masa, los usos actuales y futuros del agua y sus características hidrogeológicas.

Pueden presentarse dos situaciones. En la primera de ellas el nivel básico se encuentra por debajo de la norma de calidad (Figura 11). La fijación del valor umbral depende de cada Estado Miembro, estando comprendido entre el límite superior del rango de concentraciones del periodo de referencia y la norma de calidad. En el segundo caso, el valor de fondo del periodo de referencia está por encima de la norma de calidad, por lo que la única alternativa es fijar el valor umbral como el límite superior del rango de concentraciones.

3.8.1. Intervalos de predicción

Representan el rango de concentraciones que tiene probabilidad $1-\alpha$ de contener un valor futuro de la variable aleatoria (en lugar de un parámetro, como los intervalos de confianza). Considera la variabilidad tanto en las medidas futuras como en el propio límite de predicción. Se obtienen de un número de muestras relativamente pequeño y presentan la ventaja de que son capaces de comprobar diferencias entre un conjunto de datos (nivel básico) y los datos de la red de control.

Son relativamente fáciles de construir y su interpretación es inmediata. El nivel básico es utilizado para construir un límite de concentración (LC), que se compara a posteriori con los datos de la red. En general son más anchos que

los intervalos de confianza equivalentes, ya que tratan con la variabilidad de las medidas individuales, que es inferior a la de estadísticos como la media.

Bajo la hipótesis nula (los datos de la red de control no muestran empeoramiento de la calidad en relación con el nivel básico). El rango aceptable de concentraciones comprende los valores inferiores (o iguales) a dicho límite. Su principal ventaja es la flexibilidad, ya que un solo valor es suficiente para el test.

Únicamente los valores del nivel básico deben ser empleados para obtener el intervalo de predicción, aunque la probabilidad de que el límite de concentración contenga m valores futuros (o estadísticos derivados de estos valores) depende del número de dichos valores (m) y de la forma de realizar las comparaciones.

En el caso más sencillo de comparación sucesiva de una serie de valores futuros todas las medidas deben estar por debajo del límite, por lo que en caso de que una de ellas lo supere se acepta la hipótesis alternativa y se declara un empeoramiento estadísticamente significativo. Otro tipo de test consiste en la comparación del límite de concentración con la media o la mediana de un conjunto futuro de datos.

Los intervalos de predicción obtenidos, independientemente del tipo de comparación realizada, deben cumplir una serie de requisitos para que el test tenga validez estadística, entre las que se encuentran las siguientes **(USEPA, 2009)**:

- Las medidas del nivel básico y red de control deben estar constituidas por variables aleatorias idénticamente distribuidas
- Es precisa la estacionariedad de los datos (es decir, no deben existir tendencias, autocorrelación o variaciones estacionales o cíclicas)
- El nivel básico no debe contener valores extremos

En relación con la distribución de los datos pueden clasificarse los intervalos de predicción en dos grandes grupos. Cuando los datos siguen una distribución normal o son susceptibles de ser normalizados por medio de una

transformación, el límite de concentración se obtiene por medio de intervalos de predicción paramétricos. En caso contrario deben utilizarse métodos no paramétricos. En el primer caso es posible realizar el cálculo con sólo cuatro valores, aunque diferentes autores recomiendan un mínimo de ocho (**Gibbons, 1987, USEPA, 2009**).

A) Intervalos de predicción paramétricos

El intervalo de predicción de m valores futuros se construye de forma que m valores de la red de control son analizados con objeto de determinar si se encuentran en un intervalo de predicción obtenido a partir del valor básico. Si alguno de dichos valores supera el límite superior del intervalo, existe evidencia estadística de que provienen de otra distribución con valores mayores que el nivel básico.

La ecuación general para el cálculo de un intervalo de predicción basado en poblaciones normales (o normalizables) es la siguiente:

$$LC = \bar{x} + kS$$

Ecuación 13: Ecuación general para el cálculo de un intervalo de predicción paramétrico

Donde x es la media del nivel básico, s es su desviación estándar y k es un multiplicador que depende del tipo de límite de predicción deseado.

Para obtener el límite superior de predicción con un error α para m comparaciones de muestras futuras se utiliza la siguiente expresión:

$$LP = \bar{x} + t_{1-\alpha/m, n-1} S \sqrt{1 + \frac{1}{n}}$$

Ecuación 14: Intervalo de predicción paramétrico para m muestras futuras

También es posible realizar la comparación con la media de un número m de valores futuros. Ello presenta la ventaja de que la potencia del límite de predicción en este caso es superior a la del test de m muestras futuras, debido a la menor variabilidad de la media en comparación con las medidas individuales. El límite se calcula en este caso por medio de la siguiente expresión:

$$LP = \bar{x} + t_{1-\alpha/m, n-1} s \sqrt{\frac{1}{p} + \frac{1}{n}}$$

Ecuación 15: Intervalo de predicción paramétrico para la media de m muestras futuras

B) Intervalos de predicción no paramétricos

Hay casos en los cuales los datos no tienen una distribución normal (o normalizable mediante las transformaciones oportunas), debido entre otras causas a la presencia de una fracción significativa de valores por debajo del límite de detección. Si éste es el caso, es posible utilizar gráficos de probabilidad censurados en conjunción con métodos tales como el de Kaplan-Meier o el de regresión robusta. Si aun así no es posible encontrar un ajuste razonable a la normal, es preciso emplear métodos no paramétricos. El límite superior de confianza se construye estableciendo como límite el mayor valor de la serie (o el segundo mayor). Por su propia naturaleza son muy vulnerables a la existencia de valores extremos, lo cual, en el caso de fracciones considerables de valores por debajo del límite de detección (donde es difícil obtener resultados fiables en los ensayos de detección), hace que cuando se sospecha la existencia de valores anómalos sea recomendable escoger el segundo mayor valor de la serie.

Al contrario que los intervalos de predicción paramétricos, los que no se basan en poblaciones normales no son ajustables, lo cual significa que en caso de que se desee aumentar el nivel de confianza asociado con la predicción, las

únicas opciones disponible son aumentar el tamaño de la muestra (nivel básico) o disminuir el número de futuros puntos a predecir.

Al igual que sucede cuando la distribución es normalizable, es posible construir intervalos de predicción para un conjunto de valores futuros o para un estadístico obtenido de ellos. En el primer caso, la fracción de la población contenida en el intervalo de predicción sigue una distribución de probabilidad Beta (**Guttman, 1970**).

$$I_t(j, n-j+1) = \int_{u=0}^t \frac{\Gamma(n+1)}{\Gamma(n-j+1)\Gamma(j)} u^{j-1} (1-u)^{n-j} du$$

Ecuación 16: Distribución de probabilidad acumulada de la densidad de probabilidad Beta

donde:

- n = tamaño muestra nivel básico
- j = rango del valor del límite de predicción
- $\Gamma(n) = (n-1)! = (n-1) \times (n-2) \times \dots \times 2 \times 1$

De acuerdo con **Gibbons, 1991**, la probabilidad de que un número de muestras m no supere dicho valor es de $n/(n+m)$. De forma más general, el intervalo de confianza asociado cuando el j^o mayor valor del nivel básico es seleccionado como intervalo de predicción está definido por la siguiente expresión (**Davis y McNichols, 1999**):

$$1 - \alpha = \frac{(j+m-1)(j+m-2)\dots(j+1)j}{(n+m)(n+m-1)\dots(n+2)(n+1)}$$

Ecuación 17: Error tipo 1 asociado al intervalo de predicción de la mediana de m valores futuros

De la misma forma, es posible construir intervalos de predicción para medianas de valores futuros. En este caso los valores más prácticos corresponden a tamaños de muestra (valor básico) impares. Por ejemplo, en el caso de muestras de tres valores, la no superación del límite de predicción en dos de los valores de la red de control permite asegurar que no existe un empeoramiento significativo. Ello hace innecesaria la realización de la tercera medición, ahorrando costes que pueden ser dedicados a emplazamientos más problemáticos sin aumentar el presupuesto. La probabilidad de que el intervalo de predicción contenga los valores futuros sigue una distribución de probabilidad Beta (**Davis y McNichols, 1999**). En particular, la probabilidad de que la mediana de tres valores futuros independientes no supere el límite superior del intervalo de predicción es:

$$1 - \alpha = \frac{(3n - 2j + 5)(j + 1)j}{(n + 3)(n + 2)(n + 1)}$$

Ecuación 18: Error tipo 1
asociado al intervalo de
predicción de la mediana de tres
valores futuros

3.8.2. Tests de hipótesis

La hipótesis estadística de partida debe ser que la concentración de la sustancia contaminante en la masa de agua subterránea no supera el estándar de calidad fijado a menos que los datos del muestreo demuestren lo contrario. La hipótesis nula, H_0 , asume que la concentración en agua subterránea es inferior o igual al estándar, mientras que la hipótesis alternativa, H_A , es aceptada cuando se supera el estándar de forma significativa. Formalmente, para un parámetro Θ , estimado a partir de la muestra, y representando un estándar E, la hipótesis del test se establece como:

$$H_0: \Theta \leq E \xrightarrow{vs} H_A: \Theta > E$$

Ecuación 19: Hipótesis nula y alternativa para verificar el estado químico de la MASub

Una vez que se ha superado el estándar fijado, se declara la masa de agua subterránea en mal estado y se inicia la inversión de tendencias, la hipótesis debe ser invertida. Se asume entonces que la concentración del contaminante supera el estándar de calidad hasta que se demuestre lo contrario. Es necesario demostrar, por tanto, que la concentración del contaminante se encuentra por debajo del estándar. La hipótesis del test se define como:

$$H_0: \Theta > E \xrightarrow{vs} H_A: \Theta \leq E$$

Ecuación 20: Hipótesis nula y alternativa para verificar la efectividad de las medidas de recuperación en la MASub

Independientemente del parámetro de la población (Θ) seleccionado como representativo del estándar de calidad, el contraste se basa en la elaboración de un intervalo de confianza a partir de los datos de la red de control con un determinado nivel de significación (α) y su posterior comparación con el estándar E.

El intervalo de confianza describe la distribución probable del estadístico de la muestra θ , empleado como estimador del parámetro de la población (Θ). Se utiliza entonces el límite inferior del intervalo para su comprobación con el estándar.

3.9. Diseño del programa de control sin umbral definido

De acuerdo con la Directiva de Aguas Subterráneas, los Estados Miembros pueden acogerse a la exención de implantar medidas para prevenir o limitar la entrada de contaminantes en una masa de agua subterránea cuando los costes

requeridos sean desproporcionados. Ello no significa que pueda permitirse el deterioro de la calidad del agua subterránea sin adoptar ningún tipo de medida. Las redes de agua subterránea son una herramienta útil para verificar la evolución de la calidad, de forma que aunque no sea posible revertir la calidad del acuífero a niveles naturales de fondo, es posible evaluar impactos estadísticamente significativos en la calidad del agua una vez que se inicia el periodo de control.

En determinados programas ambientales se recomienda diseñar tests estadísticos capaces de detectar incrementos (impactos estadísticamente significativos) en la concentración comprendidos entre tres y cuatro veces el valor de la desviación típica del nivel de referencia.

Los contrastes estadísticos utilizados en los programas de control, como se ha comentado en el apartado anterior, se basan en la comparación entre las medidas de la red de control y un estándar de calidad determinado por las autoridades ambientales. No obstante, en determinadas situaciones dicho estándar no se encuentra disponible, bien porque no ha sido fijado por el organismo competente, bien porque la concentración histórica supera el estándar. La única comparación posible es entonces, la que proporciona el fondo de valores en el punto de la red de calidad. Por ejemplo, arsénico y selenio superan en muchos casos los valores de calidad fijados en la reglamentación técnico-sanitaria, por lo que es necesario calcular el estándar a partir de los valores históricos.

3.9.1. Comparación del nivel básico con nuevos valores

Una vez elaborado el nivel básico las redes de control proporcionan nuevos datos de concentración con una periodicidad que depende de cada estación concreta. En caso de que no exista un estándar de calidad, es necesario comprobar si existe alguna diferencia estadísticamente significativa entre el nivel básico y los datos adicionales obtenidos de dichas redes. La comparación estadística entre ambos conjuntos de datos se denomina test de comparación

de dos muestras, y el test más utilizado, asumiendo la normalidad de ambas es el test de la t de Student.

La hipótesis nula es que la media de los valores de concentración de la estación analizada es igual o menor que la concentración media en el nivel básico.

$$H_0 : \mu_E \leq \mu_{NB} \xrightarrow{vs} H_1 : \mu_E \geq \mu_{NB}$$

Ecuación 21: Hipótesis nula y alternativa para la igualdad de medias entre la red y el nivel básico

Es frecuente que se presente el caso de no poder acreditar la normalidad de la población (debido al pequeño tamaño de las muestras, a la existencia de un porcentaje significativo de valores por debajo del límite de detección, etc.). En ese caso es posible recurrir a dos posibilidades. La primera es utilizar los mismos tests que en el caso normal, lo cual es factible únicamente en el caso de que el test sea robusto.

La segunda es utilizar un estadístico cuya distribución no dependa de la población (tests no paramétricos). Se trata de probar la hipótesis de igualdad de distribuciones (**Chacón, 1980**).

$$\begin{aligned} H_0 : F_1(x) &= F_2(x) \\ S_1 &= \sum_{i=1}^{n_1} R_{x_{1i}} \rightarrow S_1' = n_1(n_1 + n_2 + 1) - S_1 \\ P(S_1 < s_{1\alpha} \cup S_1' < s_1\alpha) &= \alpha \\ \binom{n_1 + n_2}{n_1} \alpha / 2 \end{aligned}$$

Ecuación 22: Selección del estadístico para el test de Wilcoxon y de Mann-Whitney

No obstante, si lo que se quiere es comparar dos muestras de tamaños n_1 y n_2 obtenidas de dos poblaciones P_1 y P_2 , es posible utilizar el test de la t de Student dada la robustez del mismo.

Los ensayos anteriormente descritos son utilizados igualmente cuando se desea actualizar el nivel básico, ya que cuando se rechaza la hipótesis nula se asume que no existen diferencias en la media, por lo que es posible añadir los nuevos valores al conjunto inicial de datos.

3.9.2. Requerimientos y limitaciones

El test de la t de Student ha sido ampliamente utilizado como procedimiento estadístico de referencia para la comparación de medias de dos muestras. No obstante, es necesario comprobar que se cumplen una serie de requisitos de forma previa a su aplicación:

- ♦ Los valores deben ser independientes, lo cual es complicado de demostrar, especialmente cuando sólo se dispone de unos pocos valores
- ♦ El test asume que la distribución de los datos es normal. Dado que es un test robusto, los resultados son generalmente válidos aunque no se cumpla la hipótesis de normalidad
- ♦ No obstante, cuando los datos son fuertemente sesgados. En el caso de distribuciones lognormales con sesgo positivo (u otro tipo de distribución), los resultados pueden ser incorrectos si no se realiza previamente la normalización de los mismos.

Muchas variables ambientales con sesgo positivo se corresponden con distribuciones tipo gamma, por lo que es conveniente comprobar esta posibilidad en los datos.

El test de Shapiro-Wilk se utiliza para comprobar si la serie sigue una distribución normal. Los parámetros de la población son generalmente desconocidos, y el tamaño máximo de muestra para el que se utiliza es $n=50$. En caso de muestras de mayor tamaño el test utilizado es el de Lilliefors.

En cuanto a los datos que siguen una distribución gamma, los estadísticos utilizados se basan en la función de distribución empírica. Los más empleados son los tests de Kolmogorov-Smirnov (K-S) y Anderson-Darling (A-D). Los valores críticos de ambos tipos de distribuciones empíricas no se encuentran fácilmente disponibles, especialmente para valores bajos del factor de forma ($k < 1$). Por ello, se utilizan experimentos de simulación tipo Monte Carlo. El código ProUCL, desarrollado por la EPA de Estados Unidos es de libre disposición y ha sido el software utilizado para la obtención de diversas figuras en el presente trabajo.

3.10. Tratamiento de valores anómalos

Los valores anómalos pueden ser clasificados como representativos y no representativos (**Chambers, 1986**). Los primeros están asociados con estructuras poblacionales complejas en las que los datos proceden de dos poblaciones diferentes. Una que genera la mayor parte de los datos y otra que genera valores extremos sistemáticamente distanciados de la mayoría. Los segundos son datos excepcionales que pueden aparecer de manera muy puntual en la población, a menudo atribuidos a errores de transcripción o de registro, y que por su naturaleza merecen ser analizados separadamente del resto.

En la práctica es difícil reconocer la naturaleza de los valores extremos presentes en una muestra, ni siquiera cuándo debemos juzgar un dato como tal. Si en la muestra se detecta algún valor extremo, cabe admitir, aunque con reservas, que se tratará de un dato anómalo no representativo, posiblemente único en toda la población. Si por el contrario, se detecta un número significativo de datos con valores separados del resto, puede interpretarse que éstos son anómalos y constituyen un reflejo de un colectivo más amplio existente en la población.

La presencia de valores anómalos (valores inusuales de concentración extrema) no es infrecuente en los datos de las redes de control (**Helsel and Hirsch, 2002; Gibbons and Coleman, 2001**). En ocasiones se deben

simplemente a un error tipográfico, mientras que en otras el problema es la calibración incorrecta de los aparatos de medida. Sin embargo, es posible que representen la variabilidad temporal de la calidad natural de la masa de agua subterránea.

En cada una de estas situaciones, los valores anómalos (en el contexto estadístico) representan valores inconsistentes con la distribución de las medidas restantes. Con ello se viola una de las asunciones básicas previas a la realización de cualquier análisis estadístico, como es la identificación de los datos con variables aleatorias idénticamente distribuidas.

Si no se elimina un valor anómalo (con una concentración muy elevada en relación con los valores utilizados para la elaboración del periodo de referencia) se producirá un incremento en los valores de la media y la desviación estándar. Por una parte, ello producirá un incremento en la magnitud de los límites paramétricos de predicción o de control, calculados a partir de la muestra. Será mucho más difícil identificar las posibles superaciones del estándar de calidad establecido. Lo mismo puede aplicarse en el caso de la elaboración de intervalos de predicción no paramétricos, especialmente cuando se selecciona el valor máximo como indicador. Si dicho valor máximo es un valor anómalo no representativo de la población utilizada para la elaboración del nivel de referencia, es improbable que un test de hipótesis sea capaz de distinguir aumentos en la concentración del contaminante analizado, bajando de esta forma la potencia del método.

Como ejemplo tomemos los datos de concentración de nitratos en la Plana de Sagunto (estación JUIG002950). Como puede observarse (Figura 12), existe un valor claramente diferenciado del resto, que con nivel de significación del 10% haría que rechazáramos la hipótesis de normalidad de los datos.

Figura 12: Histograma y gráfico Cuantil-Cuantil de los valores de concentración de nitratos en la estación JUIG002950

Para comprobar si el dato puede considerarse anómalo se aplican (Figura 13) los métodos de Dixon (Dixon, 1950) y Rosner (1975). En este caso ambos métodos detectan la existencia de un valor anómalo potencial al 5% de valor de significación.

Es destacable el hecho de que si el valor de significación seleccionado hubiera sido del 1% el valor no habría sido declarado anómalo.

User Selected Options				Outlier Tests for Selected Variables			
From File		WorkSheet_a.wst					
Full Precision		OFF					
Test for Suspected Outliers with Dixon test		1					
Test for Suspected Outliers with Rosner test		1					
Rosner's Outlier Test for C1							
Mean		157,0					
Standard Deviation		68,85					
Number of data		34					
Number of suspected outliers		1					
#	Mean	sd	Potential outlier	Obs. Number	Test value	Critical value (5%)	Critical value (1%)
1	157,0	67,83	370,0	11,00	3,140	2,970	3,300
For 5% Significance Level, there is 1 Potential Outlier							
Therefore, Observation 370,0 is a Potential Statistical Outlier							
For 1% Significance Level, there is no Potential Outlier							

Figura 13: Detección de valores anómalos en la estación JUIG002950

A continuación se realiza el mismo test eliminando el valor declarado anómalo de la serie.

Como puede observarse la variación es mínima, siendo en este caso inapreciable la eliminación del valor anómalo detectado.

Por tanto, puede parecer recomendable la eliminación de los valores anómalos de la serie, incluso aunque no se conozcan las causas reales de su existencia, ya que en este caso se reducen los valores de los intervalos de predicción y se protege el medio ambiente de forma más eficaz. No obstante, las estrategias de eliminación sistemática de valores anómalos pueden conducir a la

eliminación de cambios reales e importantes en las concentraciones naturales de referencia.

Éstas no son, por lo general, estáticas en el tiempo en una masa de agua subterránea. Incluso aunque no se haya producido un cambio en las condiciones naturales de concentración, es posible que un valor anormalmente alto no represente otra cosa que una porción de la distribución que aún no ha sido observada. También se da el caso contrario, cuando estaciones que históricamente han proporcionado valores positivos arrojan mediciones de cero. Es el caso de la estación JUIG003866, en la que para el cálculo del nivel básico se ha eliminado una medición (de valor nulo).

Es recomendable calcular los estadísticos con y sin los valores anómalos, de forma que sea posible comparar el impacto potencial de los mismos en el proceso de toma de decisiones.

3.10.1. Tamaño mínimo de la muestra

El número mínimo de muestras que deben tomarse para obtener estimaciones fiables es la cuestión más importante a resolver cuando existen valores por debajo del límite de detección. En caso de que sólo se encuentren disponibles unos pocos valores cuantificables (4-6), es recomendable tomar más datos hasta que se disponga de muestras de tamaño suficiente.

El número mínimo de valores por encima del valor de detección debe estar comprendido entre 8 y 10.

3.11. Autocorrelación

Con objeto de demostrar la independencia estadística es necesario verificar la estacionariedad de la media y la varianza y la ausencia de efectos tales como la autocorrelación o la existencia de tendencias. Puede definirse la autocorrelación entre miembros de series temporales como la correlación entre observaciones consecutivas. En el caso de los recursos hídricos subterráneos

dicha autocorrelación es positiva, es decir, a continuación de valores altos de concentración siguen valores altos y a continuación de valores bajos se presentan valores bajos. Sucede cuando los términos de error del modelo no son independientes, sino que están vinculados entre sí. En esta circunstancia los estimadores de mínimos cuadrados ordinarios dejan de ser eficientes.

Por ello es necesario realizar los tests correspondientes sobre los valores utilizados como nivel básico, para lo cual es preciso disponer de un número mínimo de medidas comprendido entre 8 y 10. La función de autocorrelación de la muestra es una herramienta efectiva **(Chatfield, 2004)** para evaluar diferentes tipos de autocorrelación. Por ejemplo, cuando existe una tendencia en la serie, la función de autocorrelación no disminuye a medida que aumenta el retardo, sino que se mantiene de forma persistente en el tiempo. Igualmente, la existencia de valores anómalos puede alterar los resultados del test, por lo que es necesario proceder a su eliminación de forma previa a su elaboración.

A menos que las medidas sigan de forma aproximada una distribución normal, los resultados pueden ser inexactos, por lo que para el tratamiento de datos no normales deben utilizarse alternativas no paramétricas como el test de Von Neumann (o el test de Bartels, que se basa en los rangos de dicho test). Se requiere un mínimo de entre 10 y 12 medidas para construir la función de autocorrelación. De otra manera, será difícil construirla, especialmente para valores altos del retardo. Es necesario que la toma de datos se realice a intervalos regulares, con objeto de que puntos separados por un determinado retardo representen el mismo periodo de tiempo.

En la práctica es difícil que existan medidas regularmente espaciadas en la red de control, por lo que el problema potencial de la autocorrelación puede ser minimizado por medio de una toma de muestras lo suficientemente espaciada en el tiempo. Diversos autores **(Gibbons, 1994 y ASTM, 2004)** han sugerido la toma de muestras trimestral como un periodo suficiente para eliminar los efectos debidos a la autocorrelación.

3.12. Estacionalidad

Las redes de control de calidad proporcionan una colección de observaciones de concentración de un parámetro recogidas secuencialmente en el tiempo. Muchas series temporales presentan cierta periodicidad de carácter mensual, anual, etc.

Por ejemplo, los valores de concentración de nitratos varían debido a la afección originada por variables exógenas, como la precipitación (se produce un aumento en la concentración de dicho ión tras una serie de precipitaciones). En realidad el proceso es más complejo, pues tras un aumento inicial, que disuelve los nitratos, se produce un descenso, con motivo del efecto de dilución.

Para comprobar en detalle dicha evolución sería necesario hacer un test de tendencias estacional, lo que requiere medidas mensuales. Las redes disponibles rara vez proporcionan un número tan grande de medidas, siendo la frecuencia de muestreo habitual de un par de veces al año. Para ilustrar dicho efecto se han agrupado los datos procedentes de la estación JUIG002950, agrupando los datos existentes en función de dos periodos principales, uno seco o de aguas bajas, y otro húmedo o de aguas altas.

En la estación seleccionada los meses más habituales de muestreo son mayo y noviembre, lo que concuerda a grandes rasgos con dichos periodos de aguas altas y bajas, dado que en la cuenca mediterránea las precipitaciones más torrenciales se producen en septiembre y octubre, meses en los que los episodios de gota fría son habituales.

De la Figura 14 se deduce que los valores de concentración parecen ser mayores en los meses de noviembre, lo que concuerda

Figura 14: Concentraciones de nitratos agrupadas en estación seca y húmeda

bastante bien con la teoría expuesta en el párrafo anterior. No obstante, con objeto de evaluar la componente estacional sería necesario disponer de datos mensuales de longitud suficiente para hacer el análisis de la serie temporal y elaborar el correlograma.

Capítulo 4:

ANÁLISIS DE LA INFORMACIÓN

4. ANÁLISIS DE LA INFORMACIÓN

4.1. Origen de los datos

La Directiva Marco del Agua (DMA) establece la **Demarcación Hidrográfica** (la cual debe ser definida en función de los límites naturales de las cuencas hidrográficas) como **unidad de gestión** de los recursos hídricos, incluyendo las aguas continentales (superficiales y subterráneas), las de transición y las costeras, independientemente de su tamaño y características. Se entienden por aguas subterráneas a efectos de dicha Directiva todas las aguas que se encuentran bajo la superficie del suelo en la zona de saturación y en contacto directo con el suelo o el subsuelo, y por acuífero una o más capas subterráneas de roca o de otros estratos geológicos que tienen la suficiente porosidad y permeabilidad para permitir ya sea un flujo significativo de aguas subterráneas o la extracción de cantidades significativas de aguas subterráneas.

La DMA establece la **masa de agua subterránea** (un volumen claramente diferenciado de aguas subterráneas en un acuífero o acuíferos) como **unidad geográfica de referencia para muchas de las obligaciones que establece**, entre las que pueden citarse el cumplimiento de los objetivos medioambientales, el control de la evolución del recurso y la adopción de medidas de protección y restauración masa de agua subterránea, por lo que se ha decidido seleccionar la masa de agua subterránea como el punto de partida para la realización de la investigación.

La implementación de la Directiva Marco de Aguas plantea una serie de retos de carácter técnico, científico y práctico, motivo por el cual, en la reunión de Directores de Agua de la UE + Noruega celebrada en París los días 23 y 24 de octubre de 2000, los Estados Miembros y la Comisión Europea acordaron desarrollar una Estrategia Común (CIS) para la implementación de la DMA. Los objetivos de la Estrategia Común se centran en:

- facilitar la coherencia y comparabilidad entre Estados Miembros
- mejorar la comprensión y enfoque común
- realización de esfuerzos y actividades conjuntas

- limitar los riesgos de una mala aplicación de la Directiva
- compartir experiencia e información
- desarrollar guías
- mejorar la gestión de la información

Para alcanzar estos objetivos se han establecido 10 grupos de trabajo y se han creado foros de consejo de expertos en materias como sustancias prioritarias, aguas subterráneas y emisión de informes de los Estados Miembros a la Comisión. La coordinación de los trabajos se ha llevado a cabo por el Grupo de Coordinación Estratégica, dentro del marco de las reuniones de los Directores del Agua de los Estados Miembros.

España ha asumido un alto nivel de compromiso en el desarrollo de la Estrategia Común, fruto de la cual ha sido el desarrollo de varios documentos guía cuya evaluación integrada se ha realizado de forma simultánea en cuencas hidrográficas de diversos lugares de la Unión Europea.

Figura 15: Red de cuencas piloto en Europa

Para la comprobación de la efectividad de las guías y su aplicación cruzada se ha creado una red de cuencas piloto que incluye 15 cuencas nacionales y transfronterizas (Figura 15). En España, a través del Ministerio de Medio Ambiente y Medio Rural y Marino, se propuso comprobar y evaluar los documentos guía y trabajar en el desarrollo de una plataforma de Sistema de Información Geográfica común la Cuenca Piloto del Júcar (CPJ).

Página | 109

Las cuencas estudiadas han sido Duero, Ebro, Guadalquivir, Guadiana, Júcar, Miño-Sil, Segura y Tajo. La base de datos proporcionada por el Ministerio presenta más de 50.000 registros, con 74 parámetros diferentes para su tratamiento y casi 1.300 estaciones de control repartidas entre las ocho cuencas estudiadas. El trabajo se ha centrado en las estaciones que presentan parámetros que en algún momento han superado el 75% de la norma de calidad o del valor umbral además de valorar su evolución temporal teniendo en cuenta para ello la representación gráfica de los datos (Figura 17). De esta manera se ha conseguido reducir la cantidad de información hasta los 13.774 registros. La revisión sistemática de toda esta información ha permitido establecer las bases para la determinación de los niveles básicos, como se detalla en los siguientes epígrafes.

Figura 17: Histogramas y gráficos Q-Q Plot de concentraciones de Nitrato en puntos seleccionados de las cuencas del Ebro, Segura, Júcar, Duero, Tajo y Guadiana

4.2. Selección de una Demarcación Hidrográfica de referencia

Como se ha citado en apartados anteriores, la designación de la cuenca hidrográfica del Júcar (Figura 18) como cuenca piloto (**Estrela et al, 2004**) la ha convertido en el centro de numerosos estudios cuyo propósito último es proporcionar herramientas para la toma de decisiones en el campo de la

gestión de los recursos hídricos. Dado que ese es precisamente el enfoque de esta tesis se ha seleccionado dicha cuenca para la aplicación de la metodología propuesta.

Figura 18: Localización de la Demarcación Hidrográfica del Júcar

Los estudios de investigación desarrollados por el IGME sobre los 23.100 Km² de la Comunidad Valenciana han permitido distinguir 12 Sistemas Acuíferos. La mayoría de los ríos de la Comunidad se integran en la Cuenca del Júcar, que abarca una extensión de 42.988,6 Km². En la Demarcación Hidrográfica del Júcar pueden distinguirse a grandes rasgos una zona interior montañosa, con altitudes que rebasan los 1500 metros, y una zona costera, constituida por llanuras litorales, conocidas comúnmente como planas.

La metodología empleada por la Confederación Hidrográfica del Júcar para identificar las masas de agua subterránea en riesgo de no alcanzar los objetivos medioambientales para el año 2015 es similar a la empleada por el Ministerio de Medio Ambiente para la identificación de las presiones y análisis del impacto en las masas de agua superficial **(MIMAM, 2005)**, y consiste básicamente en la determinación del riesgo como una combinación de las presiones y los impactos definidos.

CODIGO	NOMBRE MASA DE AGUA	CONT. PUNTUAL	CONT. DIFUSA	INTRUSIÓN	EXTRACCIÓN
080.106	Plana de Cenia				X
080.107	Plana de Vinaroz		X	X	X
080.110	Plana de Oropesa -Torreblanca		X	X	X
080.127	Plana de Castellón		X	X	X
080.128	Plana de Sagunto		X	X	X
080.129	Mancha Oriental				X
080.130	Medio Palancia				X
080.131	Liria - Casinos		X		X
080.140	Buñol - Cheste		X		X
080.141	Plana de Valencia Norte		X	X	
080.142	Plana de Valencia Sur		X	X	
080.143	La Contienda				X
080.144	Sierra del Ave				X
080.148	Hoya de Játiva		X		X
080.149	Sierra de las Agujas		X		X
080.151	Plana de Jaraco			X	X
080.152	Plana de Gandía		X	X	X
080.155	Valle de Albaida				X
080.157	Sierra de la Oliva				X
080.158	Cuchillo - Moratilla				X
080.159	Rocín				X
080.160	Villena - Benejama				X
080.161	Volcadores - Albaida				X
080.162	Almirante Mustalla		X		
080.163	Oliva - Pego		X	X	X
080.164	Ondara - Denia		X	X	X
080.165	Montgó		X		X
080.166	Peñón - Bernia		X		X
080.169	Muro de Alcoy				X
080.170	Salt San Cristobal				X
080.171	Sierra Mariola				X
080.172	Sierra Lácera				X
080.173	Sierra del Castellar				X
080.174	Peñarrubia				X
080.175	Hoya de Castalla				X
080.176	Barrancones - Carrasqueta				X
080.178	Serrella - Aixorta - Algar				X
080.179	Depresión de Benisa			X	X
080.180	Jávea			X	X
080.181	Sierra de Salinas				X
080.182	Argüeña - Maigó				X
080.184	San Juan - Benidorm		X		X
080.185	Agost - Monnegre				X
080.186	Sierra del Cid				X
080.187	Sierra del Reclot				X
080.188	Sierra de Argallet				X
080.189	Sierra de Crevillente				X
080.190	Bajo Vinalopó			X	X

Tabla 4: Relación de masas de agua subterránea identificadas en riesgo en la Demarcación Hidrográfica del Júcar y motivo de su clasificación en este apartado

En la tabla 4 se detallan las masas de agua subterránea identificadas en riesgo en la Demarcación Hidrográfica del Júcar.

Desde el punto de vista hidrogeológico es posible distinguir entre tres tipos de formaciones (**IGME, 1988**):

- ♦ Macizos calcáreos mesozoicos, con diferentes grados de carstificación, distribuidos por el interior de la Comunidad Valenciana, y que representan su mayor potencial hidráulico. En ocasiones (existencia de excedentes) son la fuente de alimentación lateral de los acuíferos costeros
- ♦ Acuíferos detríticos pliocuaternarios del litoral, intensamente explotados, y en los que se realiza la mayor parte de las actividades económicas existentes
- ♦ Acuíferos terciarios del interior

En cuanto a las presiones más significativas cabe destacar que se han realizado numerosos análisis por parte del CEDEX y de la Confederación Hidrográfica del Júcar con objeto de evaluar el riesgo de que las masas de agua subterránea no alcancen los objetivos de calidad fijados por la Directiva Marco del Agua. En base a dichos análisis se han definido las presiones significativas (**Estrela et al, 2004**):

- Extracciones significativas. Merece la pena destacar el hecho de que una parte importante de dicha abstracción se debe a la demanda urbana, por lo que se presentan picos estacionales asociados al turismo en las zonas costeras
- Obras de regulación de cauces
- Contaminación difusa y puntual. Como causas principales deben mencionarse la contaminación difusa debida a los nitratos de origen agrícola, identificados como el problema básico de las aguas subterráneas por varios autores (**De Stefano et al, 2014**) y la intrusión

marina. Las zonas con mayor contenido de nitratos se sitúan de forma general en las zonas costeras, en las provincias de Castellón y Valencia **(MARM, 2007)**, debido a la existencia de una gran zona de regadío y la consiguiente aplicación intensiva de fertilizantes. La existencia de zonas húmedas asociadas a esos acuíferos puede provocar problemas en los hábitats asociados.

4.3. Masas de agua subterránea analizadas

En función de los criterios anteriormente descritos y de la disponibilidad de datos de las redes de control, se han seleccionado las siguientes masas de agua subterránea para su análisis:

<i>Código identificativo</i>	<i>Nombre</i>
080.107	Plana de Vinaroz
080.110	Plana Oropesa-Torreblanca
080.127	Plana de Castellón
080.128	Plana de Sagunto
080.131	Liria-Casinos
080.140	Buñol-Cheste
080.141	Plana de Valencia Norte
080.142	Plana de Valencia Sur
080.149	Sierra de las Agujas
080.151	Plana de Jaraco
080.152	Plana de Gandía
080.163	Oliva-Pego
080.164	Ondara - Denia
080.165	Montgó
080.180	Jávea
080.184	San Juan - Benidorm
080.190	Bajo vinalopó

Tabla 5: Relación de masas de agua subterránea seleccionadas en la Demarcación Hidrográfica del Júcar

Los parámetros analizados han sido la concentración de nitratos en la mayor parte de los casos y la de cloruros cuando se encuentra disponible. Dentro de

las masas de agua subterránea sólo se han seleccionado aquéllas que presentan una serie de registros de al menos ocho valores, que se han marcado en color amarillo en las tablas 6 y 7.

Puntos de la red de control en la Cuenca del Júcar					
Código	Nº registros	Código	Nº registros	Código	Nº registros
Plana de Vinaroz		Plana de Castellón		Plana de Sagunto	
08.10.091	3	CA0812122	6	CA0820052	7
CA0810084	3	CA0812123	6	CA0821053	5
CA0810089	4	CA0812124	6	JUIG000542	4
JUIG002851	5	JUIG001746	36	JUIG002683	8
JUIG002973	4	JUIG002609	17	JUIG002924	8
JUIG003150	3	JUIG002642	3	JUIG002950	34
JUIG003420	8	JUIG002662	27	JUIG003137	5
JUIG003526	8	JUIG002955	3	JUIG003550	3
JUIG003764	8	JUIG003036	3	JUIG003594	17
Plana de Oropesa - Torreblanca		JUIG003060	14	JUIG003835	4
CA0807028	5	JUIG003174	4	JUIG003866	40
CA0811042	7	JUIG003204	4	JUIG003929	4
JUIG001311	4	JUIG003255	4	JUIG003940	8
JUIG001312	8	JUIG003259	4	Liria - Casinos	
JUIG001387	26	JUIG003266	4	CA083954	4
JUIG003314	9	JUIG003267	4	CA0820057	7
JUIG003328	5	JUIG003298	10	CA0822008	7
JUIG003624	8	JUIG003333	4	CA0822035	5
Buñol - Cheste		JUIG003353	4	CA0822038	5
CA0823034	8	JUIG003378	8	CA0825073	6
CA0827087	7	JUIG004046	8	Montgó	
JUIG000322	5	Plana de Gandía		JUIG000860	6
Sierra de las Agujas		C080045001	3	Oliva - Pego	
CA0826075	7	C080045002	3	CA0838146	4
JUIG000213	47	CA0832004	6	JUIG000440	5
JUIG002123	5	JUIG000438	4	JUIG003128	5
JUIG002708	5	JUIG000439	5	JUIG003489	8
JUIG003999	4	JUIG000621	5	JUIG003727	3

Tabla 6: Relación de estaciones de control seleccionadas en la Demarcación Hidrográfica del Júcar

Puntos de la red de control en la Cuenca del Júcar					
Código	Nº registros	Código	Nº registros	Código	Nº registros
Plana de Valencia Norte		Plana de Valencia Sur		Plana de Jaraco	
C080035001	3	CA0826002	9	JUIG000578	4
CA0825098	3	CA0826076	8	JUIG001505	8
CA0826074	5	CA0826077	4	JUIG002122	3
CA0831022	5	CA0826078	8	JUIG002129	5
JUIG000326	7	CA0826080	6	Ondara - Denia	
JUIG000377	4	CA0826102	5	C080056001	3
JUIG000413	6	JUIG000333	5	C080056002	3
JUIG000625	4	JUIG000412	46	CA0838149	4
JUIG002584	8	JUIG002558	5	JUIG000352	8
JUIG002617	4	JUIG002599	5	JUIG001011	18
JUIG002622	4	JUIG003121	7	JUIG002092	15
JUIG003768	8	JUIG003130	4	JUIG003098	14
JUIG003950	4	San Juan - Benidorm		Jávea	
JUIG003965	4	CA0848015	4	JUIG002334	22
		Bajo Vinalopó			
		CA0899005	4		

Tabla 7: Relación de estaciones de control seleccionadas en la Demarcación Hidrográfica del Júcar (continuación)

A continuación se describen de forma somera las masas de agua subterránea seleccionadas y se representa la evolución de concentración de la sustancia que origina el riesgo en las mismas.

4.3.1. Plana de Vinaroz

Desde el punto de vista hidrogeológico, en la Plana de Vinaroz-Peñíscola se diferencian tres formaciones diferentes (IGME, 1988 y Morell, 2001): formación detrítica superior, formación margoarcillosa intermedia y formación caliza profunda.

La formación detrítica superior corresponde al acuífero detrítico miocuaternario, que está constituido por arenas, gravas y conglomerados con escasa proporción de arcillas.

El nivel acuífero inferior está formado por calizas del Jurásico y Cretácico con

potencias de hasta 400 m. Entre ambos se encuentra una formación margo-arcillosa del Mioceno que separa ambos niveles acuíferos.

La Plana de Vinaroz es una de las nueve masas de agua subterránea con presión significativa debido a la intrusión marina en el ámbito territorial de la CHJ, lo que ha originado que en los sectores

Figura 19: Evolución concentración nitratos en la Plana de Vinaroz

dulce sea clorurada sódica. Igualmente, se han considerado como presiones significativas la contaminación difusa y la extracción de agua.

En la Figura 19 puede observarse la evolución de los puntos seleccionados. En general, aunque existen pocos datos en las series, se aprecia en todos ellos una clara tendencia creciente en la concentración de nitratos.

próximos a la línea de costa la composición química del agua

Figura 20: Evolución concentración nitratos en la Plana de Oropesa -Torreblanca

4.3.2. Plana de Oropesa-Torreblanca

El acuífero consiste en un extenso campo de derrame (**IGME, 1988**), cuyo principal nivel acuífero está definido por un conjunto de conglomerados con intercalaciones de gravas, arenas, limos y arcillas.

El impermeable de base está constituido por margas y arcillas miocenas. Existen cuatro puntos con más de 8 registros, cuyos gráficos se representan en la Figura 20.

4.3.3. Plana de Castellón

Figura 21: Evolución concentración nitratos en la Plana de Castellón

El acuífero superior está formado por sedimentos pliocuaternarios situados sobre materiales mesozoicos o sobre materiales terciarios de muy baja permeabilidad (**IGME, 1988**).

La masa se encuentra definida en riesgo tanto por contaminación difusa como por intrusión y extracción intensiva.

Asociadas a la masa se encuentran, en sus extremos meridional y septentrional, las marjalerías de Chilches-Almenara y Castellón, respectivamente.

Aunque en general la longitud de las series de concentración no es elevada (Figura 21), existen estaciones de control con puntos suficientes para realizar análisis estadísticos

4.3.4. Plana de Sagunto

Los materiales acuíferos que conforman el acuífero de la Plana de Sagunto se encuentran representados por niveles detríticos de gravas y/o arenas con espesores en general inferiores a 10-15 m, incluidos en una formación pliocuaternaria fundamentalmente limo-arcillosa **(Ballesteros et al, 2005)**.

En función de la información obtenida de sondeos mecánicos y geofísicos se ha podido determinar que los niveles de materiales permeables no son continuos ni lateral ni verticalmente y que su distribución en el espacio es también muy variable de unas zonas a otras.

El subsistema acuífero de la Plana de Sagunto (Figura 22) está constituido por una alternancia de gravas, arenas y conglomerados susceptibles de presentar una carstificación importante englobados dentro de una formación pliocuaternaria eminentemente limo-arcillosa **(IGME, 1988)**, que descansa sobre una formación arcillosa muy

Figura 22: Evolución concentración nitratos en la Plana de Sagunto

conductora, de espesor creciente hacia el mar. Subyacente a esta formación se detecta la presencia de materiales acuíferos mesozoicos, normalmente Muschelkalk. Estos materiales muestran un progresivo escalonamiento hacia la costa.

El acuífero detrítico está limitado hacia el interior por el afloramiento de estos materiales, en tanto que hacia el sur y norte el límite es relativamente convencional y está determinado por la presencia de materiales del Buntsandstein que lo individualizan parcialmente de las Planas de Valencia y Castellón respectivamente.

4.3.5. Buñol - Cheste

Figura 23: Evolución concentración nitratos en Buñol-Cheste

En la masa afloran materiales de edades comprendidas entre el Mesozoico y el Cuaternario. Estos últimos se encuentran en las inmediaciones de la Rambla de Poyo, y están constituidos por conglomerados situados en el interior de arcillas arenosas. Las calizas pontienses situadas a techo de la formación miocena conforman un acuífero de

elevada discontinuidad (IGME, 1988), mientras que los materiales infrayacentes del Serravalliense-Tortonense son delgados

paquetes de conglomerados calcáreos con características de acuífero multicapa. Los materiales mesozoicos, por otra parte se sitúan en las sierras de Perenchiza y La Rodana.

Los piezómetros representativos no muestran una tendencia clara al descenso continuado (Figura 23), por lo que no se considera la masa en riesgo de forma general, aunque pueden existir problemas de bombeo intensivo en alguna zona de borde (MIMAM, 2005).

4.3.6. Plana de Valencia Norte

La masa de agua subterránea de la Plana Norte puede asimilarse a efectos hidrogeológicos a un acuífero multicapa, en el que se diferencian dos niveles permeables principales. El tramo superior está constituido por una alternancia de materiales detríticos cuaternarios de carácter libre, mientras que el tramo inferior, de una gran heterogeneidad y anisotropía, está formado por un acuífero mioceno de carácter semiconfinado. El acuífero cuaternario supera los 200 metros de espesor en

Figura 24: Evolución concentración nitratos en Plana de Valencia Norte

las proximidades del cauce del río Turia y del barranco del Carraixet, disminuyendo de forma considerable en las proximidades de la Albufera. **(Pérez Gago et al, 2001).**

Existen numerosos pozos y sondeos (Figura 24) destinados a regadío que captan el acuífero cuaternario, el cual se encuentra afectado en gran parte de su superficie por el uso de abonos y fertilizantes y el reciclado de los excedentes de riego. Las principales presiones sobre la masa son la contaminación difusa y la intrusión.

4.3.7. Plana de Valencia Sur

Figura 25: Evolución concentración nitratos en Plana de Valencia Sur

La descripción hidrogeológica de la Plana de Valencia Sur es muy similar a la de la Plana de Valencia Norte. El acuífero cuaternario alcanza sus mayores espesores en puntos próximos al cauce del río Júcar, en Algemesí.

Al igual que en el caso anterior, la masa se ha definido en riesgo debido a presiones debidas a contaminación difusa (Figura 25) e intrusión.

4.3.8. Sierra de las Agujas

Tiene una extensión superficial cercana a unos 180 km², y está constituida por materiales acuíferos de edad Jurásico y Cretácico **(IGME, 1988)**. El impermeable de base está compuesto por arcillas en facies Keuper, las cuales constituyen el límite meridional excepto en el sector de Tabernes de la Valldigna, donde el límite es abierto, existiendo conexión hidráulica con la Plana de Gandía y posiblemente con Sierra Grossa **(IGME, 1988)**. Los pliegues

y fracturación existentes provocan la conexión hidráulica entre el acuífero Jurásico y el acuífero Cretácico.

Las presiones identificadas en la masa se deben a la contaminación difusa (Figura 26) y a la extracción.

Figura 26: Evolución concentración nitratos en Sierra de las Agujas

4.3.9. Plana de Jaraco

Figura 27: Evolución concentración nitratos en Plana de Jaraco

Es un acuífero detrítico formado por materiales cuaternarios y, eventualmente, pliocuaternarios. Está constituido esencialmente por gravas, limos y arcillas, correspondientes a depósitos aluviales, de pie de monte y sedimentos marinos o mixtos (IGME, 1988).

De los cuatro puntos de control existentes en la masa solamente uno de ellos (Figura 27) cuenta con ocho registros, que en líneas generales y a falta de un análisis de detalle, muestran una clara tendencia al ascenso.

De los cuatro puntos de control existentes en la masa solamente uno de ellos (Figura 27)

4.3.10. Oliva – Pego

El acuífero de Oliva Pego se define como de tipo detrítico litoral y desde el punto de vista ambiental presenta un interés notable debido a su asociación con la zona húmeda litoral de Pego-Oliva, que se encuentra incluida en la lista de la convención de

Figura 28: Evolución concentración nitratos en Oliva-Pego

RAMSAR como humedal de interés internacional.

Es un acuífero multicapa dispuesto sobre un impermeable de base integrado por margas miocenas o por margas con yesos del Trias. Sin embargo, en algunos sectores, el acuífero carbonatado mesozoico de Pego constituye la base del acuífero (**Domínguez et al, 2008**).

En la Figura 28 está representada la evolución en la concentración de nitratos del punto JUIG003489.

4.3.11. Ondara - Denia

La masa de agua subterránea Ondara-Denia (Figura 29) está formada por materiales cuaternarios y pliocuaternarios que corresponden a depósitos aluviales de llanura de inundación de piedemonte. Dichos materiales son muy heterométricos y están constituidos fundamentalmente por gravas, arenas, limos y arcillas.

Figura 29: Evolución concentración nitratos en Ondara-Denia

Figura 30: Evolución concentración nitratos en Ondara-Denia

4.3.12. Jávea

La Plana de Jávea es una masa de agua subterránea constituida por materiales detríticos del Cuaternario, fundamentalmente gravas, arenas y limos (**IGME, 1988**), que se extiende sobre el valle aluvial del río Gorgos en su tramo final (**Ballesteros et al, 2001**). El impermeable de base está constituido por margas del Mioceno. El acuífero ha sufrido

procesos de intrusión salina debido a la sobreexplotación de sus recursos, lo que ha provocado el desplazamiento de la interfase agua dulce agua salada hacia el interior. No existen registros de longitud suficiente en la masa para realizar análisis estadísticos fiables salvo en la estación JUIG002334 (Figura 30).

4.4. Análisis de las series de datos

En general los valores observados son muy altos, por encima de la norma de calidad y presentan fuertes oscilaciones. Este comportamiento produce una gran incertidumbre sobre cualquier intento de predicción del comportamiento de la variable estudiada. En consecuencia, no es posible la determinación de las tendencias en muchas de las estaciones de la Demarcación, por lo que se han seleccionado las estaciones resaltadas en amarillo en la Tabla 6 y la Tabla 7, con objeto de tener un tamaño mínimo de muestra para aplicar la metodología diseñada.

4.4.1. Inspección visual

En el apéndice A se muestran las figuras correspondientes al análisis exploratorio de datos. Los cambios en la calidad del agua pueden producirse de forma abrupta o bien pueden realizarse de forma gradual en el tiempo. Por ello, el examen visual de la serie, al que se añade un suavizado de los datos tipo LOESS, proporciona una indicación de cambios de tendencia a corto plazo, que pueden enmascarar de alguna forma la tendencia total de la serie. A ello se le han sumado los histogramas correspondientes y gráficos boxplot de la serie en la escala original y en escala logarítmica.

En líneas generales es posible distinguir tres grandes grupos. El primero de ellos está compuesto por las estaciones en las cuales se produce un aumento constante de la concentración, característica que se puede apreciar sin necesidad de realizar un análisis estadístico de detalle. El segundo grupo lo integran las estaciones en las que es posible distinguir dos tendencias claramente, por lo cual el nivel básico debe ser escogido entre los puntos del

segundo tramo, justificando de forma cuidadosa su elección. El último lo componen aquellas estaciones en las cuales se mantiene la tendencia con oscilaciones de pequeña amplitud.

4.4.2. Selección de niveles básicos y análisis de tendencias

De acuerdo con el apartado anterior, para la elaboración del nivel básico deben tenerse en cuenta las siguientes consideraciones:

1. En series claramente crecientes se ha utilizado la totalidad de los puntos, ya que para detectar incrementos estadísticamente significativos es condición necesaria la inexistencia de tendencias en el nivel básico. En caso de que la totalidad de los valores se encuentren por debajo del umbral la Directiva de Aguas Subterráneas obliga igualmente a la inversión de tendencias cuando se alcance el 75% de dicho umbral.
2. Cuando existe un único valor extremo que no supera los tests descritos se ha eliminado dicho valor, con objeto de no elevar los intervalos de confianza como se ha comentado con anterioridad. En la Figura 23 (Masa de agua subterránea de Buñol-Cheste) puede observarse un ejemplo claro en el que la no eliminación de dicho valor podría producir la declaración del mal estado de la masa incluso aunque todos los demás valores se encuentran por debajo del umbral
3. Cuando existen valores aislados de concentración tras los cuales existe un intervalo de tiempo muy grande sin medidas, se han eliminado los valores iniciales, al considerar que no dichos valores no son representativos de las condiciones actuales de la masa y que incluso las técnicas de medida y precisión de los datos han sufrido modificaciones de forma sustancial. Como ejemplo puede observarse en la Figura 20 (Oropesa-Torreblanca y Ondara-Denia). En esta última el valor eliminado tampoco supera el test de Dixon para la detección de valores anómalos

Figura 31: Nivel básico estaciones seleccionadas

Aplicación de la metodología a una zona piloto

La Plana de Vinaroz (Figura 32) es una de las nueve masas de agua subterránea (IGME, 1988) en el dominio territorial de la Demarcación Hidrográfica del Júcar en la que se ha declarado una presión significativa debido a la intrusión marina debido a la elevada concentración en nitratos en sectores próximos a la costa.

Figura 32: Localización de la masa de agua subterránea y estaciones de la red de control

Igualmente se han identificado presiones adicionales, en concreto contaminación difusa y extracción.

Las fuentes de información para la caracterización del agua subterránea en esta masa de agua subterránea han sido el Instituto Geológico y Minero de España y la Confederación

Hidrográfica del Júcar. Con objeto de generar los datos e información requerida para gestionar la intrusión de agua de mar en acuíferos costeros, la Confederación Hidrográfica del Júcar estableció una red de control a lo largo de la costa española del Mediterráneo, comenzando con unas 40 estaciones. Hoy, su tamaño ha aumentado hasta el centenar, las cuales son muestreadas con periodicidad semestral, analizando los cloruros, conductividad, nitratos, bicarbonatos y temperatura. La localización (Figura 32) de la masa de agua subterránea seleccionada se muestra en la figura conjuntamente con los puntos de la red operativa.

Área de estudio y Marco Hidrogeológico

Desde el punto de vista hidrogeológico existen tres formaciones diferentes en la masa de agua subterránea Plana de Vinaroz Peñíscola (Figura 33). La formación superior detrítica (Mioceno Superior a Cuaternario), que contiene arena, gravas y conglomerados con una pequeña proporción de arcilla. El nivel acuífero inferior (Jurásico Superior- Cretácico Inferior) tiene naturaleza caliza. Entre medias se sitúa una formación arcillosa del Mioceno, que desconecta ambas unidades.

Análisis exploratorio de datos

La información proporcionada por la red de control es muy variable, con estaciones que presentan más de treinta años de medidas con otras que fueron instaladas hace escasamente cinco años.

Figura 33: Localización de la zona piloto y de las estaciones de la red de control

En trabajos anteriores la Plana de Vinaroz Peñíscola ha sido definida como en mal estado por la Confederación Hidrográfica del Júcar. La razón es que el índice de explotación (el cociente entre el volumen extraído y el recurso disponible) es mayor que 0.8 (en realidad es superior a 1), causando la intrusión de agua de mar. Los valores umbral definidos por el Organismo de Cuenca son 1100 mg/L para los cloruros y 250 mg/L para los sulfatos.

Conformidad con estándares

Una vez realizado el análisis exploratorio de datos, se han realizado ensayos de bondad de ajuste para distribuciones de tipo normal, gamma y lognormal.

Los resultados, incluyendo los p-valores para el ajuste de los valores de concentración de cloruros se muestran en la Tabla 8.

Figura 34: Evolución de parámetros seleccionados en la Masa de Agua Subterránea Plana de Vinaroz

En una primera aproximación se observa (Figura 34) que los valores de cloruros tienden a permanecer estables o incluso a descender, lo que se observa con la representación en detalle de la estación 08.10.047 (Figura 35). En ella se observa un máximo en el año 2001, tras lo cual se produce una estabilización.

Dado que estos valores son la base para la elaboración del nivel de fondo, es preciso comprobar la existencia de valores anómalos o tendencias al aumento en la concentración de contaminantes.

En el caso de la estación 08.10.047, tras la aplicación del test de Rossner, se ha decidido descartar este valor, ya que su inclusión produciría un aumento del nivel calculado, lo que implicaría menor grado de protección de la masa de agua subterránea.

Figura 35: Evolución de la concentración de cloruros en la estación 47 de la Plana de Vinaroz

La independencia (o inexistencia de correlación) es otro de los requisitos para el análisis, lo que implica periodos de muestreo lo suficientemente espaciados. Dado que en la Demarcación Hidrográfica del Júcar la frecuencia de muestreo es semestral, la existencia de dichos efectos es muy improbable, ya que según la bibliografía mencionada, incluso con una frecuencia de muestreo cuatrimestral sería posible descartarlos.

Plana de Vinaroz - Cloruros					Bondad de ajuste											
Estación	Fecha inicio	p-valor			Cloruros			Conduct			Nitratos			Bicarbon		
		normal	gam	logn	N	G	L	N	G	L	N	G	L	N	G	L
08.10.010	15/04/1982	1.34e-11	0.012	5.52e-07	x	x	x	x	x	x	✓	✓	✓	x	x	x
08.10.026	01/02/1977	4.31e-04	0.950	0.579	x	✓	✓	x	x	x	✓	✓	x	✓	✓	✓
08.10.039	16/06/1977	1.26e-04	0.105	6.20e-06	x	x	x	x	x	x	x	x	x	x	x	x
08.10.044	10/06/1980	1.85e-03	0.208	1.06e-02	x	x	x	x	x	x	✓	✓	x	x	x	x
08.10.047	14/06/1977	1.70e-11	0.069	3.33e-05	x	x	x	x	✓	x	✓	✓	✓	✓	✓	✓
08.10.050	21/04/1982	5.27e-04	0.075	4.92e-05	x	x	x	✓	✓	x	✓	✓	✓	✓	✓	✓
08.10.062	13/06/1972	6.82e-08	0.524	0.098	x	✓	x	x	x	x	✓	✓	✓	x	x	x
08.10.069	16/06/1977	1.57e-04	0.004	5.13e-12	x	x	x	x	x	x	✓	x	x	✓	✓	✓
08.10.091	03/10/2007	7.88e-02	0.529	1.17e-02	✓	✓	x	✓	✓	✓	x	x	x	✓	✓	✓

N representa la Normal, G Gamma y L Lognormal

Tabla 8: Bondad de ajuste de los puntos de control de la red operativa en Plana de Vinaroz y p-valores para las concentraciones de cloruros

Por otra parte, se han calculado los intervalos de confianza de la concentración de cloruros bajo la hipótesis de que los datos no siguen una distribución discernible, excepto para las estaciones 08.10.026 y 08.10.062, que siguen una distribución gamma y la 08.10.091, que sigue una distribución normal.

En lo que se refiere al estado químico la concentración de nitratos es el otro parámetro investigado, a consecuencia de la definición de la contaminación difusa como una presión relevante. De hecho, esta masa de agua subterránea ha sido definida en mal (o pobre) estado en el Plan Hidrológico Nacional 2009-2015 a partir de las concentraciones medidas en el año 2010.

Con objeto de evaluar el cumplimiento de estándares y analizar la evolución de la calidad del agua se ha utilizado el valor máximo de concentración de nitratos en agua subterránea fijado por la Directiva de Nitratos (50 mg/L). Como puede comprobarse en la Tabla 9, únicamente dos estaciones se encuentran en conformidad con el estándar.

Los intervalos de confianza calculados muestran valores extremadamente altos en estaciones como la 08.10.050, donde se supera el umbral ampliamente, mientras que en otras como la 08.10.047 el límite inferior del intervalo de confianza se encuentra próximo al estándar. De acuerdo con los

requerimientos de la Directiva Marco del Agua, es necesaria la elaboración de un plan de medidas con objeto de mejorar la calidad del agua.

En relación con la concentración de cloruros debe destacarse el hecho de que la masa de agua subterránea ha sido clasificada en mal estado debido a la intrusión salina, y la Confederación Hidrográfica del Júcar ha establecido un estándar de calidad de 1100 mg/L. Se ha verificado la conformidad de los valores proporcionados por la red de control con dicho estándar.

Límites de predicción

Los datos disponibles en la segunda mitad de 2012 han sido utilizados para comprobar la conformidad con los límites de predicción elaborados. Dado que los valores reales no superan en ningún caso los límites de predicción calculados, no puede afirmarse que exista un deterioro en la calidad del agua subterránea.

En la estación 08.10.26 el límite de predicción ha sido calculado usando la aproximación de Wilson-Hilferty (**Krishnamoorthy et al, 2008**) que establece que la raíz cúbica de una variable aleatoria gamma sigue una distribución normal. Para ello se han realizado los cálculos correspondientes de acuerdo con la ecuación 14 y se ha realizado la transformación inversa para obtener la concentración original. El test de Shapiro ha sido utilizado para comprobar la hipótesis de normalidad. También se ha calculado el mismo límite utilizando la aproximación lognormal, pero ha sido descartado, dado que su valor está por encima del primero.

Estación	Intervalo de confianza		Lím. Pred Cloruros	
	Nitratos	Cloruros	Estimado	Medido
08.10.010	25.26 - 30.29	56.96 - 79.26	1199	-----
08.10.026	23.24 - 58.22	452.2 - 560.7	1301.82	342.51
08.10.039	84.0 - 159.1	1077.43 - 1337.61	2925	1116.26
08.10.044	105.6 - 148.1	58.73 - 66.04	117.42	59.27
08.10.047	66.22 - 110.98	241.18 - 314.04	1400	-----
08.10.050	178.7 - 202.3	58.27 - 64.75	77.91	-----
08.10.062	106.7 - 126.9	127.65 - 178.75	622.21	124.47
08.10.069	95.8 - 149.8	1221 - 1378	2087	1025.40
08.10.091	142.9 - 179.7	205.7 - 231.6	313.95	243.01

Tabla 9: Intervalos de confianza de concentración de nitratos y cloruros y límites de predicción de concentración de cloruros en la Plana de Vinaroz Peñíscola

Análisis de tendencias

La identificación de tendencias sostenidas al aumento es otro de los pilares de la Directiva de Aguas Subterránea (**European Commission, 2009**), por lo que es preciso identificar su existencia con respecto a los contaminantes responsables de la declaración en riesgo de la masa de agua subterránea. Se ha optado por utilizar ensayos que verifiquen la existencia de tendencias monótonas al aumento.

La herramienta para tal comprobación son la regresión lineal (**Grath et al, 2001**), de acuerdo con la Directiva, mientras que el ensayo de Mann Kendal y el método de Sen (para estimación de la pendiente) son los ensayos de referencia en caso de que los datos no puedan ser ajustados a una distribución paramétrica.

En la Figura 36 se muestra la evolución de nitratos en la estación 69 de la Plana de Vinaroz. Se adivina una tendencia creciente que se pone de manifiesto cuando se añade una línea de suavizado de los datos. La aplicación de un test formal, en este caso el de Mann Kendal confirma las suposiciones, con un p-valor de 0.0462

Figura 36: Evolución de la concentración de nitratos en la estación 69 de la Plana de Vinaroz

Análisis de la evolución espacial

La Directiva Marco del Agua establece la Masa de Agua Subterránea como la unidad de gestión a considerar en los Planes Hidrológicos de Cuenca. Por ello, es posible dividir áreas geográficas extensas con el objeto de gestionarlas adecuadamente.

Siguiendo el mismo espíritu de la Directiva, es posible subdividir una Masa de Agua Subterránea en dos Masas, una de ellas considerada en riesgo de no cumplir los objetivos ambientales y la otra no considerada en riesgo, por lo que el esfuerzo de caracterización y el grado de control son considerablemente menores.

Por ello es necesario incidir de nuevo en la necesidad de caracterizar adecuadamente las masas de agua subterránea mediante la descripción de sus aspectos más relevantes, que deben incluir el modelo conceptual de la misma.

A pesar de lo anterior, existen ocasiones en las que no es posible realizar tal subdivisión, bien porque no existe un modelo conceptual definido bien porque no existen datos suficientes para realizar la subdivisión. La primera consecuencia es la necesidad de recabar más datos, que debe ser puesta de manifiesto en los Planes Hidrológicos de Cuenca correspondientes.

En dichos casos, la evaluación de una tendencia sostenida y significativa al aumento en la concentración de contaminantes a escala de masa de agua subterránea requiere la integración de los resultados del análisis de tendencias en cada uno de los puntos de la red de control.

Esto quiere decir que para declarar una masa de agua subterránea en mal estado no basta la presencia de una tendencia significativa en una de las estaciones de control. En algunos casos es posible determinar el impacto ambiental significativo mediante el análisis de varias estaciones, que en caso de resultar negativo puede evitar tener que declarar la totalidad de la masa en mal estado.

Siguiendo con la Plana de Vinaroz Peñíscola, se han evaluado los cambios en la concentración de cloruros a escala de masa de agua subterránea, con el propósito de verificar la existencia de tendencias regionales en la misma. Para ello es preciso diseñar una metodología que permita confirmar de manera objetiva la existencia de impactos significativos en el contenido de cloruros en el agua subterránea y la existencia de tendencias a largo plazo, permitiendo además cuantificar las mismas.

En la Plana de Vinaroz-Peñíscola existen un total de 58 estaciones, cuyos diagramas de caja se muestran en la Figura 37, elaborada con los paquetes estadísticos `wq` (Jassby, A.D. and Cloern, J.E., 2011) y `reshape` (Wickham, 2007) desarrollados en el entorno de programación del lenguaje R (R Development Core Team, 2011) con objeto de comparar la distribución de la serie histórica (hasta el año 2011). Del análisis de los datos se deduce que los valores de concentración se distribuyen en un rango amplio, desde valores inferiores a 50 mg/L hasta concentraciones por encima de los 2000 mg/L. Igualmente se aprecia una mayor dispersión para los valores más altos. Por último, es visible la existencia de sesgos acusados, que hacen dudar de su normalidad.

Figura 37: Distribución de concentración cloruros en la red de intrusión de la Plana de Vinaroz-Peñíscola

En la Figura 38 se ha representado la evolución de la concentración de cloruros de la totalidad de los puntos que componen la red de intrusión, a la que se ha superpuesto una curva de suavizado.

Figura 38: Diagrama de dispersión de la concentración de cloruros en la Plana de Vinaroz-Peñíscola

existe ninguna. En concreto, en el periodo 2005-2010 los únicos registros activos son los correspondientes a las estaciones de la red operativa.

La primera característica reseñable de los datos es la existencia de una tendencia decreciente que se ha intentado poner de manifiesto mediante la representación en escala logarítmica. Igualmente se observa la ausencia de datos en el periodo 2002-2004 así como la disminución en su número a partir del año 2005. A partir de dicha fecha, los registros se corresponden con la red operativa, que está compuesta por nueve estaciones cuya medida continúa en el presente. La disminución en los valores de concentración en el tiempo se enfatiza mediante la representación de los valores de las medianas, según se refleja en la Figura 39. No obstante, la cantidad de información anual disponible es muy variable, desde años en los que todas o casi todas las estaciones tienen medidas hasta años en los que no

Figura 39: Evolución de medianas en la red de intrusión de la Plana de Vinaroz-Peñíscola

Dado que es necesario disponer de datos en la totalidad de las fechas en las cuales se desea realizar el análisis de tendencias, se han seleccionado los puntos de dicha red operativa (Figura 40).

Figura 40: Evolución temporal red operativa

Los valores obtenidos mediante la realización del test de Mann Kendall en los registros históricos muestran la presencia de tendencias en 4 de las 9 estaciones de la red operativa (Tabla 10), lo que corresponde a un porcentaje inferior al 50%. Parece obvio que no es posible determinar tendencias globales de esta forma.

Dado que no es posible comprobar de forma rigurosa dicha evolución por medio del test de Mann-Kendall es necesario acudir a otras alternativas que permitan detectar tendencias a largo plazo.

La primera opción consiste en utilizar las variaciones estacionales de forma que al eliminarse las diferencias debidas al periodo en el que se realiza la toma de datos aumente la potencia del test y se obtengan resultados concluyentes. No obstante, las medidas disponibles en la red operativa tienen periodicidad semestral, por lo que es preciso explorar otras opciones.

En este sentido se han realizado investigaciones por parte del Servicio Geológico de los Estados Unidos (USGS) para determinar tendencias en la concentración de contaminantes en Áreas de Gestión de Aguas Subterráneas, cuyo principal resultado ha sido la modificación del test estacional de Mann-Kendall (**Helsel et al, 2006**) para incluir emplazamientos (coordenadas) en lugar de estaciones (periodos de muestreo). El test así diseñado tiene capacidad para indicar cambios en las tendencias regionales en zonas concretas.

La aplicación del test regional de Mann-Kendall proporciona los resultados indicados en la Tabla 10 en el conjunto de estaciones de control de la Plana de Vinaroz para el periodo 1985-2011. Aunque el p-valor es altamente significativo, es preciso destacar el alto porcentaje de puntos de registro con la totalidad de valores ausentes, por lo que los resultados deben tomarse con precaución.

Aplicando el ensayo a las estaciones de la red operativa (excepto la s92, debido a la ausencia de medidas hasta 2004), en las cuales están presentes la totalidad de los datos, no es posible distinguir impactos significativos en el periodo 2007-2011.

<i>Código estación</i>	<i>p valor</i>
<i>s10</i>	0.000
<i>s39</i>	1.000
<i>s47</i>	0.422
<i>s62</i>	0.001
<i>s10</i>	0.000
<i>s39</i>	1.000
<i>s47</i>	0.422
<i>s62</i>	0.001
<i>S91</i>	0.763

Tabla 10: p-valores para el test de Mann Kendal en la red operativa de la Plana de Vinaroz Peñíscola

Para obtener resultados representativos es preciso incrementar el periodo temporal considerado, aunque la ausencia de valores en el periodo 2002-2004 no permite determinar el intervalo necesario para obtener valores aceptables. La aplicación del método entre los años 2000-2011 proporciona valores negativos para la pendiente media en el entorno de -7,4 mg/L por año del parámetro analizado, con un p-valor altamente relevante y un porcentaje de valores ausentes por debajo del 50% en todas las estaciones. No es necesario, por tanto, la realización de tests de homogeneidad, ya que la tendencia se ha identificado con claridad.

Capítulo 5:

GESTIÓN DE VALORES POR DEBAJO DEL LÍMITE DE DETECCIÓN

5. GESTIÓN DE VALORES POR DEBAJO DEL LÍMITE DE DETECCIÓN

5.1. Consideraciones generales

Se define el límite de detección como la cantidad o concentración mínima de sustancia que puede ser detectada con fiabilidad por un método analítico determinado. Se corresponde con la concentración mínima medida en una muestra que contenga el analito que el laboratorio es capaz de discriminar de la concentración proporcionada por la medida de un blanco. Es, por tanto, un límite cualitativo. Con objeto de establecer además límites cuantitativos se han definido una serie de límites en los análisis **(Currie, 1968)** basados en la teoría estadística de los contrastes de hipótesis:

- El nivel crítico, o la señal a partir de la cual puede decirse que se ha detectado de forma fiable el compuesto de interés en el agua subterránea
- El límite de detección, que es un umbral a partir del cual los valores medidos pueden considerarse diferentes de un blanco con un nivel especificado de probabilidad.
- El límite de determinación, o la concentración sobre la que es posible realizar una determinación cuantitativa con una incertidumbre relativa especificada.

Los valores de dichos límites dependen de una serie de factores, entre los cuales el más importante es la desviación estándar de un conjunto de muestras cuya concentración es exactamente cero. Como en la práctica esto es difícil de conseguir, suelen utilizarse muestras en las cuales el componente está presente en baja concentración. Para ello el analista supone que el error de medida es el mismo a concentración cero que para la disolución estándar. Seleccionando una probabilidad α (error tipo I, o probabilidad de detectar el compuesto cuando en realidad no está presente) es posible calcular el nivel crítico. Escogiendo el error β o de segunda clase (aceptar la ausencia del

contaminante cuando en realidad está presente), el límite de detección puede ser evaluado. Por último, el límite de cuantificación se obtiene especificando el máximo error estadístico tolerable en una medida cuantitativa.

Figura 41: Relación entre el límite de detección y el límite de cuantificación

En la Figura 41 se ilustra la relación entre el blanco, el límite de detección y el límite de cuantificación por medio de la función de densidad de probabilidad para medidas distribuidas normalmente en el blanco, en el límite de detección y en el límite de cuantificación.

Por motivos prácticos suele definirse el límite de detección como tres veces la desviación estándar del blanco, y el límite de cuantificación como diez veces dicha cantidad. Para una señal situada en el límite de detección, la probabilidad α (falso positivo) es pequeña (en torno al 1%). En cambio, el error β es de un 50%. Por ello, es frecuente trabajar con el límite de cuantificación, que presenta una probabilidad mucho menor de falsos negativos.

Los valores de concentración de un contaminante determinado (medidos en una red de control ambiental) por debajo del límite de detección (LDD) se conocen generalmente en estadística como datos censurados (o incompletos). Existen diferentes tipos de censura:

- Por el lado izquierdo de la distribución. Se sabe que una observación se encuentra en el intervalo comprendido entre cero y el LDD y que el valor real ha sido censurado por las limitaciones inherentes a la instrumentación empleada o al método de medida.
- Por la derecha. La observación está por encima de cierto valor, pero se ignora cuánto

- En un intervalo. El valor real está comprendido entre dos valores, generalmente el límite de detección y el límite de cuantificación

5.2. Visualización de series con datos censurados

En general los métodos de representación gráfica de datos censurados no se han utilizado de forma generalizada. Ello es debido, en parte, a que la información proporcionada para los valores por debajo del límite de detección es incompleta, lo que motiva que las gráficas no sean tan informativas como debieran.

No obstante, los procedimientos gráficos ayudan a determinar la distribución de los datos (si son simétricos o sesgados), identificar la presencia de valores censurados y su porcentaje, identificar tendencias en función del tiempo, comparar datos de dos o más poblaciones y confirmar las conclusiones obtenidas de los ensayos estadísticos. La representación gráfica de los datos es de vital importancia para la comprensión de los aspectos mencionados anteriormente y es un paso necesario antes de asumir cualquier distribución en un conjunto de datos.

5.2.1. Gráficos de caja

Figura 42: Boxplot JUIG000213 suponiendo 15% censura en totalidad de datos y con datos separados en dos grupos

Entre las más utilizadas figuran los gráficos de caja (boxplots) censurados (Figura 42), en los cuales se añade una línea horizontal en el límite superior de detección (en caso de que existan varios).

De esta forma es posible calcular los percentiles situados por encima de la línea de forma análoga a como se realiza en el caso de muestras con la totalidad de sus datos cuantificables, mientras que los percentiles situados por debajo deben ser estimados con algunos de los métodos descritos a continuación. En el caso de que existan varios límites de detección, la información mostrada corresponde al más alto de ellos. En el entorno de trabajo R los percentiles situados por debajo son estimados mediante regresión robusta (ROS).

En la Figura 43 se muestra el gráfico de valores del punto JUIG002950 suponiendo que se establece un único límite de detección en 175 mg/L. De esta forma, el porcentaje de valores censurados se encuentra en el entorno del 35%, lo que aún permite obtener estimaciones de la media y la mediana. En el segundo gráfico se han dividido los valores en dos grupos (los primeros diez datos y a continuación los restantes).

5.2.2. Gráficos de dispersión

Figura 43: Gráfico de dispersión de la estación JUIG000213

En ellos se representan los valores censurados mediante líneas discontinuas que abarcan un intervalo comprendido entre cero y el límite de detección. En la Figura 44 se han representado los datos de la estación JUIG000213.

Figura 44: Función de supervivencia de la estación JUIG000213

5.1.1. Gráficos de la función de supervivencia

La función de supervivencia se define, de forma genérica, como la probabilidad de que un suceso determinado se produzca después de un determinado

tiempo. En el eje de ordenadas se representan los percentiles de los datos obtenidos mediante el procedimiento de Kaplan-Meier (descrito en epígrafes posteriores), mientras que en el eje de abcisas se representan valores de concentración. En el gráfico únicamente se representan los valores no censurados, aunque su posición depende también de los valores censurados **(Helsel, 2012)**. En la Figura 45 se visualiza la función de supervivencia de los valores correspondientes al nivel básico en la estación JUIG000213. En ella

Figura 45: Función de supervivencia por grupos de la estación JUIG000213

deben estacarse las líneas punteadas, que comprenden el intervalo de confianza al 95% de dicha función.

En algunos casos interesa verificar si dos o más muestras proceden de la misma distribución, concretamente si la distribución de concentraciones

es diferente entre los grupos. Para ello es posible representar en la misma gráfica las funciones de supervivencia de dos o más grupos. Como ejemplo se observan en la las funciones de supervivencia

del nivel básico en la estación JUIG000213 agrupando los diez primeros y los diez últimos valores. Se ha supuesto un porcentaje de censura del 15%, el mismo que para la figura anterior.

5.2.3. Gráficos probabilísticos

Mediante esta técnica gráfica es posible contrastar la hipótesis de que los datos siguen una determinada distribución. Generalmente los datos de concentración son ajustados de forma más precisa mediante la distribución lognormal.

Ello se debe a que suelen tener sesgo positivo tanto por sus características intrínsecas (no son

Figura 46: Gráfica de probabilidad de la estación JUIG000213

posibles valores de concentración negativos) como por la existencia de posibles valores anómalos. En la Figura 46 se muestra el ajuste a una distribución lognormal de los valores de la estación JUIG000213nb con el porcentaje de censura del 15%.

5.3. Técnicas de análisis de valores censurados. Cálculo de estadísticos descriptivos

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{N}}$$

Ecuación 23: Error cuadrático medio

Los métodos para la estimación de estadísticos en series con valores por debajo del límite de detección pueden ser divididos en tres categorías: métodos basados en sustitución directa, en la asunción de distribuciones conocidas o métodos robustos. La habilidad de cada uno de ellos para estimar los estadísticos de la población puede medirse por medio de la raíz cuadrada del error cuadrático medio.

Cuando se analizan datos de concentración con valores por debajo del límite de detección no debe realizarse ningún tipo de análisis estadístico como si dichos datos no existieran, ya que ello produce una pérdida de información importante y dificulta la interpretación de los resultados. Las principales clases de tratamientos estadísticos utilizados en la literatura para la determinación de estadísticos básicos en datos ambientales censurados son los siguientes **(Antweiler, 2008)**: sustitución, máxima verosimilitud, regresión sobre los estadísticos de orden y técnicas no paramétricas, en concreto el método de Kaplan-Meier.

Únicamente los dos últimos permiten el análisis de conjuntos de datos en los que existen múltiples límites de detección. En este caso los cuantiles asociados con algunos valores censurados pueden superar los cuantiles de los valores detectados **(Helsel, 2005)**, por lo que es necesario utilizar métodos de cálculo diferentes.

5.3.1. Sustitución

La forma más sencilla de gestión de valores censurados es la sustitución simple por un valor concreto, de forma previa a la realización de cualquier análisis estadístico. Dicho valor suele ser una fracción del límite de cuantificación (0, 1/2 o 1). Si los valores por debajo del límite de cuantificación representan una ausencia de contaminación, la sustitución por cero es lógica.

	Mín.		1 st Qu.		Median		Mean		3 rd Qu.		Max.		sd	
Javea	14		49.2		89		105.6		139.6		378		87.2	
Javeasust0	0	0	49.2	0	89	0	102.6	74.5	139.6	139.6	378	378	90.6	106.1
Javeasustld0.5	15	50	49.2	50	89	50	108.1	104	139.6	139.6	378	378	87.5	85.3
Javeasustld	30	100	49.2	100	89	100	108.1	133.5	139.6	139.6	378	378	84.7	67.6
% censura	B	A	B	A	B	A	B	A	B	A	B	A	B	A

Tabla 11: Estadísticos obtenidos de la serie JUIG002334 por sustitución con diferentes porcentajes de censura (B=18, A=59)

Sin embargo, parece razonable pensar que los valores por debajo del límite de cuantificación representan una concentración comprendida entre cero y dicho límite, especialmente cuando existen valores por encima en la misma serie. En este caso, la sustitución por LC/2 representa el estimador de máxima verosimilitud de la media o la mediana en una población uniformemente distribuida en el intervalo (0, LC).

Son métodos baratos y muy atractivos por su sencillez. De hecho, en el apartado 2d de la Parte A del Anexo IV de la Directiva 2006/118/CE (Aguas Subterráneas) se dice literalmente que “con el fin de evitar sesgos en la determinación de las tendencias, todas las mediciones por debajo del límite de cuantificación se cifrarán en la mitad del límite de cuantificación más alto registrado durante el periodo, con la excepción de la totalidad de los plaguicidas”.

Sin embargo, la sustitución de los valores censurados por LC/2 proporciona resultados sesgados **(Singh, Maichle, and Lee, 2006)** incluso cuando el porcentaje de valores por debajo del límite de detección es sólo del 5-10%.

Como principal inconveniente del método debe mencionarse el hecho de que la elección del valor de sustitución es esencialmente arbitraria y puede provocar errores sustanciales. Por ello, no es recomendable la utilización de este tipo de métodos, salvo en los siguientes casos **(USEPA, 2009)**:

- La muestra es demasiado pequeña (<5)
- La proporción de valores censurados es inferior al 10-15%
- Existen procesos que generan distribuciones diferentes para los valores por debajo del límite de detección, tales como las variaciones estacionales en el nivel piezométrico o la composición del agua subterránea.

Incluso aunque las estimaciones de la media o la mediana sean relativamente buenas con bajos porcentajes de datos censurados, la estimación de la desviación estándar es generalmente pobre, lo que influirá en el cálculo de los intervalos estadísticos, que a su vez son la base para la toma de decisiones sobre el estado químico de una masa de agua subterránea.

Como ejemplo (Figura 47) se ha analizado la estación JUIG002334, situada en la masa de agua subterránea Javea, con objeto de obtener los estadísticos descriptivos de la serie completa y de la misma serie realizando la sustitución de los valores censurados por 0, la mitad del límite de cuantificación y por último por dicho límite. Para ello se han supuesto límites de 30 y 100 mg/L, lo que implica porcentajes de valores censurados del 18 y del 59% respectivamente.

Figura 47: Histogramas serie JUIG002334 con diferente % censura

Como puede observarse en la Tabla 11 la desviación estándar de la muestra es de 87.2, mientras que los procedimientos de sustitución producen intervalos para la misma magnitud comprendidos entre 84.7 y 90.6 mg/L para el caso de pequeña censura (18%). En el caso de censura severa (59%) el intervalo se encuentra entre 67.6 y 106.1.

5.3.2. Estimación por máxima verosimilitud (MLE)

Se basa en el ajuste de distribuciones conocidas para el cálculo de los estadísticos. Dada una muestra aleatoria simple la función de densidad conjunta de dicha muestra es la siguiente (**Chacón, 1980**):

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i, \theta)$$

Ecuación 24: Función de densidad conjunta de las observaciones de una m.a.s.

Esta función tomará valores diferentes en función del valor de θ . Considerando la expresión anterior como una función de θ se denomina a $L(\theta/x_1, x_2, \dots, x_n)$ función de verosimilitud. De entre los valores posibles de θ existe uno que tal que hace máxima la función de verosimilitud. En la práctica se suele tomar el logaritmo de esta función:

$$\hat{\theta}_{mle} = \arg \max \left(\frac{\partial L_n L(\theta / x_1, x_2, \dots, x_n)}{\partial \theta} \right)$$

Ecuación 25: Estimador de máxima verosimilitud de θ

Por tanto, suponiendo que los datos siguen una distribución determinada (generalmente normal o lognormal), se escogen los estadísticos que mejor ajusten los valores por encima del límite de cuantificación y el porcentaje de valores por debajo del mismo. Los estimadores de máxima verosimilitud aportan estimadores más precisos que las gráficas probabilísticas, ya que tiene menor RMSE, aunque ambos métodos consiguen estimaciones insesgadas cuando los datos presentan un ajuste perfecto a la distribución asumida y el número de datos es grande. En el caso de muestras de pequeño tamaño ($n=5$) la estimación es imprecisa y el sesgo grande (**Gleit, 1985**).

Figura 48: Ajuste por máxima verosimilitud en la estación JUIG000213 (normal y lognormal)

Igualmente, la presencia de valores anómalos puede distorsionar los estimadores de la media y la desviación estándar (**Singh and Nocerino, 2002**),

provocando que los valores censurados extrapolados sean incluso negativos.

Cuando el porcentaje de datos incompletos es alto (superior al 40%), no es recomendable en general el uso de métodos paramétricos. En la Figura 48 se muestra el ajuste por máxima verosimilitud de los datos de concentración de la estación JUIG000213 suponiendo que siguen una distribución normal (izquierda) o lognormal (derecha).

5.3.3. Regresión sobre estadísticos de orden (ROS)

Es un método para el cálculo de estadísticos que asume que los datos siguen una distribución conocida. En general suele utilizarse la distribución lognormal, en cuyo caso la gráfica de probabilidad de los logaritmos ordenados de los datos en función de los cuantiles de una distribución normal estandarizada debería proporcionar una línea recta.

Son métodos robustos que proporcionan resultados aceptables incluso en presencia de sesgos acusados (datos no normales), ya que combinan los valores medidos por encima del límite de cuantificación con valores extrapolados asumiendo una distribución determinada, que se fija a los primeros. Dicha distribución se utiliza para extrapolar un conjunto de valores por debajo del límite, aunque éstos no se utilizan como estimaciones de muestras concretas, sino como conjuntos de datos con los cuales se estiman los estadísticos.

Ello presenta dos ventajas frente a los métodos distribucionales. En primer lugar no son afectados por el ajuste de las distribuciones asumidas, ya que se utilizan los datos reales, y en segundo lugar las estimaciones de los valores extrapolados pueden reconvertirse a su escala original, por lo que se evitan sesgos debidos a la transformación, como en el caso de distribuciones lognormales, sesgos que en muchas ocasiones pueden ser inaceptables (especialmente en series muy asimétricas) y resultar en la toma de decisiones incorrectas **(Singh, Singh, and Iaci , 2002)**.

Otra ventaja es que no hay límite en el porcentaje de datos censurados, que pueden llegar a ser del 80%. Obviamente, en estos casos los resultados deben ser interpretados con precaución, especialmente en el caso de muestras de pequeño tamaño. Igualmente puede ser utilizado en presencia de múltiples límites de detección.

La regresión sobre los estadísticos de orden en muestras pequeñas es más precisa que la estimación por máxima verosimilitud, por lo que se utiliza generalmente como alternativa en estos casos.

5.3.4. Técnicas no paramétricas

El análisis de supervivencia es un conjunto de procedimientos estadísticos que se emplean para analizar información censurada. Su nombre se debe a que se diseñaron para analizar el tiempo que transcurre hasta que ocurre un suceso, tal como la muerte del paciente en estudios clínicos. Posteriormente se ha aplicado en el campo de la ingeniería para calcular fallos en sistemas mecánicos, adoptando el nombre de teoría de fiabilidad.

Existen diferentes tipos de censura: por la derecha, por la izquierda y por intervalo. En el primer caso se sabe que el tiempo de supervivencia es superior a un valor, mientras que en el caso de los datos ambientales su censura es por la izquierda, por lo que se sólo se sabe que ciertas concentraciones son menores que una cierta cantidad (generalmente el límite de detección). El tercer caso combina los dos anteriores, ya que se sabe que la supervivencia se encuentra comprendida entre dos valores.

La función de supervivencia $S(t)$ se define como **(Letón, 2002)**:

$$S(t) = \text{Pr } ob(T > t) = 1 - F(t)$$

donde

$$F(t) = \int_0^t f(u) du$$

Ecuación 26: Función de supervivencia

Los métodos no paramétricos se basan únicamente en los rangos de los datos **(Antweiler, 2008)** y no realizan hipótesis sobre la distribución estadística de la que proceden. Debido a ello no se ven afectados por la existencia de valores anómalos (frecuentes en datos ambientales), por lo que producen buenos resultados incluso en conjuntos pequeños de datos. Las técnicas más utilizadas para la estimación de la función de supervivencia son las de Kaplan-Meier y Turnbull.

Para una muestra de tamaño N de una población sean $t_1 < t_2 < \dots < t_N$ los tiempos transcurridos hasta la muerte de todos ellos. Para cada t_i se define:

- d_i el número de muertes en el momento i
- n_i el número de sujetos en riesgo justo antes de t_i

El estimador de Kaplan-Meier es el estimador de máxima verosimilitud de la función de supervivencia $S(t)$. Es un producto de la forma:

$$\hat{S}(t) = \prod_{i=1}^k \frac{b_i - d_i}{b_i}$$

Ecuación 27: Estimador de Kaplan-Meier de la función de supervivencia

donde: b = número total de observaciones (detectadas y censuradas)
 d = número de observaciones no censuradas

El estimador de Kaplan-Meier está implementado en el entorno de programación R por medio de la función `survfit` (paquete NADA). En realidad trata los valores como si fueran censurados por la derecha, aunque previamente selecciona de forma automática una constante de la cual resta los valores de concentración de la serie.

Como cualquier estadístico, el estimador de Kaplan-Meier tiene una varianza, que puede ser calculada mediante varios estimadores. Uno de los más comunes es la fórmula de Greenwood (**Collet, 2003**).

$$\hat{Var}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{i=1}^k \frac{d_i}{n_i(n_i - d_i)}$$

Ecuación 28: Estimación de la varianza de la función de supervivencia mediante la fórmula de Greenwood

5.4. Regresión no paramétrica y análisis de tendencias

La estimación de la pendiente en una serie temporal de datos de concentración se realiza generalmente mediante técnicas de regresión lineal. No obstante,

cuando no se cumplen las suposiciones en las cuales se sustenta dicho análisis es preciso utilizar alternativas no paramétricas, entre las cuales se encuentra el estimador de Theil-Sen. Dado el modelo de regresión clásico:

$$Y = \alpha + \beta X + \varepsilon$$

Ecuación 29: Modelo de regresión lineal

Theil propuso obtener el estimador de la pendiente (β) mediante la mediana de todas las posibles pendientes de la muestra.

$$S_{ij} = \frac{(Y_j - Y_i)}{(X_j - X_i)}$$

donde: $i < j$

Ecuación 30:
Estimador robusto de Theil de la pendiente en regresión no paramétrica

El estimador de la pendiente es la mediana de los valores S_{ij} . El estimador de α (ordenada en el origen) se obtiene mediante:

$$\alpha = \text{mediana}(\bar{Y} - \beta \bar{X})$$

Ecuación 31: Estimador robusto de Theil de la ordenada en el origen en regresión no paramétrica

De forma similar a como se realiza la estimación de la pendiente en regresión lineal simple el método de regresión no paramétrica más utilizado para el análisis de datos censurados es el método de Akritas-Theil-Sen, que consiste **(Helsel, 2012)** en realizar una estimación inicial de la pendiente y obtener con él los residuos en el eje de las Y. Con posterioridad se determina el estadístico S de Kendall y se analizan las pendientes que producen un estadístico S igual

a cero. Para finalizar se calcula la mediana de dichas pendientes y se toma ésta como la pendiente final.

5.5. Cálculo de estadísticos descriptivos

Cuando existe una mezcla de datos censurados y no censurados se plantea el problema de cómo calcular los estadísticos descriptivos (media, mediana y desviación estándar). En este sentido, ya a finales de los ochenta **Gilbert, (1980)** describió los métodos sencillos de sustitución y algunos estimadores de máxima verosimilitud como el de Cohen. Poco después, **Helsel y Guilliom (1986)** estimaron los parámetros relativos a la distribución en datos de calidad del agua, a continuación **Helsel y Cohn (1988)** estimaron sus estadísticos descriptivos y poco después **Helsel (1990)** comparó varios métodos de cálculo de dichos estadísticos.

En publicaciones posteriores (**Helsel, 2005 y Helsel, 2012**) se definieron los métodos de obtención de estadísticos de datos censurados, que han sido descritos en apartados anteriores. Los resultados obtenidos en el presente trabajo han sido obtenidos por medio del entorno de trabajo R, concretamente mediante el paquete NADA (**Lee, 2010**).

5.5.1. Estimación por máxima verosimilitud

La estimación de parámetros en poblaciones normales y lognormales es un procedimiento completamente analítico de maximización. Ha sido estudiada por gran número de investigadores, entre los que figuran **Cohen (1950 y 1951)** y **Guilliom y Helsel (1986)**. El primero derivó las ecuaciones de máxima verosimilitud para muestras censuradas mediante el método numérico iterativo de Newton-Raphson. Con objeto de facilitar la obtención de estimadores de la media y la desviación estándar elaboró unas tablas con las constantes necesarias para obtener dichos estimadores de máxima verosimilitud.

Como regla general los datos de concentración en agua subterránea se adaptan mejor a una distribución lognormal, debido a sus características

intrínsecas, en particular los sesgos positivos y su incompatibilidad con valores negativos. Las fórmulas para la obtención de la media y la desviación estándar (**Gilbert, 1987**) son las siguientes:

$$\hat{\mu} = \exp\left(\mu_{\log} + \frac{\hat{\sigma}_{\log}^2}{2}\right)$$

$$\sigma^2 = \hat{\mu}[\exp(\hat{\sigma}_{\log}^2) - 1]$$

Ecuación 32: Estimadores de la media y la desviación estándar para distribuciones lognormales

donde $\hat{\sigma}$ y $\hat{\mu}$ son las estimaciones de la media y la varianza, respectivamente, de los logaritmos neperianos de los datos.

En cambio, los percentiles pueden ser obtenidos directamente mediante la transformación inversa de los límites de confianza para la media en escala logarítmica mediante:

$$p_k = \exp[\mu_{\log} + z_k \sigma_{\log}]$$

Ecuación 33: Percentiles de una distribución lognormal

donde z_k es el percentil k-ésimo en la escala original y z_k es el percentil k-ésimo de la distribución normal estándar.

Uno de los principales inconvenientes del método es la imposibilidad de utilizar más de un límite de detección, a lo que debe sumarse la posibilidad de obtener soluciones más exactas mediante software más moderno tanto comercial como de libre distribución, como por ejemplo el paquete NADA (**Lee, 2010**).

5.5.2. Regresión robusta sobre estadísticos de orden

La utilización de puntuaciones normales sobre los estadísticos de orden proporciona métodos potencialmente más robustos. Suponiendo que una transformación logarítmica (o de otro tipo) proporciona n_0 valores por debajo del límite de detección y n_1 observaciones por encima, suponiendo que el conjunto de datos es independiente e idénticamente distribuido con la misma media y desviación estándar, se cumple que la media y la varianza satisfacen la siguiente ecuación **(Shumway, 2002)** :

$$y_i = \mu_y + \sigma_y \phi^{-1}(P_i)$$

Ecuación 34: Relación entre la media, la varianza y la función de distribución normal con datos censurados

donde:

$$P_i = \text{Prob}(Y_i < y_i)$$

$$\Phi^{-1}(\bullet) = \text{Inversa de la función de distribución normal}$$

El procedimiento comúnmente aceptado para su resolución consiste en sustituir las probabilidades por los rangos ajustados, de forma que la ecuación de regresión se transforma de la siguiente manera **(Shumway, 2002)**:

$$y_i = \mu_y + \sigma_y \phi^{-1}\left(\frac{i - 3/8}{n + 1/4}\right) + \varepsilon_i$$

Ecuación 35: Relación entre la media, la varianza y la función de distribución normal con datos censurados

donde:

$$i = 1, 2, \dots, n \text{ (valores no censurados)}$$

Ello implica que los errores residuales ε_i tienen igual varianza y su correlación es cero. Una vez ajustada la distribución más apropiada a los datos (por encima del límite de detección) sobre un gráfico probabilístico se realiza una regresión lineal sobre los mismos. A continuación se estima el valor de los datos censurados a partir del modelo de regresión y se combinan con los datos no censurados como si existiese ningún tipo de censura. La media es estimada mediante el punto de corte de dicha recta con el eje de ordenadas mientras que la pendiente de la recta proporciona la desviación estándar.

5.5.3. Kaplan-Meier

La media de una serie con datos censurados se obtiene calculando el área bajo la curva de Kaplan-Meier **(Kleind and Moeschberger, 2003)**.

$$\hat{\mu} = \int_0^{\infty} \hat{S}(t) dt$$

Ecuación 36: Cálculo de la media mediante el procedimiento de Kaplan-Meier

Para evaluar esta integral **(Datta, 2005)** es necesario hacer una corrección al estimador de Kaplan-Meier si la observación mayor se corresponde con un valor censurado. Una corrección de este tipo es utilizada por **Efron, 1967**, y consiste en tratar dicha observación como un valor a partir del cual el estimador vale cero.

En el caso de valores censurados por la izquierda la situación es análoga. Cuando la observación menor está por debajo del límite de detección se desconoce el final de la función de supervivencia **(Helsel, 2012)**. Ello no representa un inconveniente para el cálculo de percentiles pero impide el cálculo de la media, ya que la última área a integrar es desconocida. La corrección de Efron en este caso consiste en suponer que el valor censurado

coincide con el límite de detección, lo que equivale a estimar la media con un sesgo positivo.

La varianza de la media (para m valores cuantificados y n datos) es calculada por medio de la siguiente fórmula (Lee, 2003):

$$Error_estandar = \sqrt{\frac{m}{m-1} \sum_{i=1}^m \frac{A_r^2}{(n-r)(n-r+1)}}$$

donde: $A_r = \text{Áreas_Kaplan_Meier}$

Ecuación 37: Error estándar de la media mediante Kaplan-Meier

Por tanto, multiplicando este error estándar por la raíz cuadrada de n se obtiene la desviación estándar.

5.6. Intervalos de confianza de las medidas de tendencia central

En muchos casos resulta de interés calcular un rango estimado de valores que tenga una probabilidad dada de contener un parámetro de la población. Al igual que en el caso de series con la totalidad de sus valores por encima del límite de detección, la determinación del estado químico de una masa de agua subterránea puede ser realizada mediante la elaboración de intervalos de confianza, de predicción o de tolerancia. Los procedimientos básicos para calcular dichos intervalos han sido descritos en apartados anteriores, y pueden ser divididos en procedimientos de máxima verosimilitud, Estimación Robusta (ROS) y Kaplan-Meier.

5.6.1. Intervalos de confianza de la media y la mediana por máxima verosimilitud

En general se asume que los valores de concentración en agua subterránea siguen una distribución normal o lognormal (por las razones antes

mencionadas). En el primer caso el intervalo de confianza viene definido por la ecuación 3.1:

$$LCL_{1-\alpha} = \bar{x} - t_{(1-\alpha, n-1)} \frac{s}{\sqrt{n}} \quad UCL_{1-\alpha} = \bar{x} + t_{(1-\alpha, n-1)} \frac{s}{\sqrt{n}}$$

Ecuación 38: Intervalo de confianza de la media

El intervalo de confianza de la mediana es el mismo, ya que en una distribución normal la media es idéntica a la mediana.

En el caso de distribuciones lognormales el intervalo de confianza es más complicado de calcular, ya que debido a los sesgos de la transformación inversa no es posible calcular el intervalo en escala logarítmica y después simplemente exponenciarlos a fin de obtener el intervalo de confianza en su escala original. La ecuación de Cox es utilizada para el cálculo de los límites de confianza:

$$LIC = \exp\left[\bar{y} + \frac{s_y^2}{2} - z \sqrt{\frac{s_y^2}{n} + \frac{s_y^4}{2(n-1)}}\right]$$

$$LSC = \exp\left[\bar{y} + \frac{s_y^2}{2} + z \sqrt{\frac{s_y^2}{n} + \frac{s_y^4}{2(n-1)}}\right]$$

Ecuación 39: Método de Cox para el cálculo de intervalos de confianza de la media para una distribución lognormal

En cambio, el intervalo de confianza de la mediana se obtiene mediante la transformación inversa de los valores del intervalo de confianza en escala logarítmica.

5.6.2. Intervalos de confianza de la media y la mediana mediante Regresión Robusta (ROS) y métodos no paramétricos

Mediante las técnicas de regresión sobre estadísticos de orden no es posible construir un intervalo de confianza que contenga con una determinada probabilidad el verdadero valor de la media o la mediana. En cambio, mediante técnicas no paramétricas, como el procedimiento de Kaplan-Meier si que es posible calcular intervalos de confianza para la mediana (o cualquier otro percentil). El estimador de Kaplan-Meier del k -ésimo percentil con la mayor probabilidad de supervivencia igual o menor que k .

5.6.3. Intervalos de confianza de la media y la mediana mediante técnicas de remuestreo (bootstrap)

Proporcionan una alternativa cada vez más extendida para el cálculo de intervalos de confianza de los estadísticos centrales mediante cualquiera de los métodos descritos anteriormente.

El procedimiento es conceptualmente muy sencillo y puede resumirse en los pasos siguientes (**Helsel, 2012**):

- Dada una muestra de tamaño n , tomar una muestra con reemplazamiento también de tamaño n
- Calcular (mediante máxima verosimilitud, regresión robusta o Kaplan-Meier) el estadístico objetivo
- Salvar la estimación del parámetro
- Seleccionar una nueva muestra de tamaño n con reemplazamiento y repetir el proceso
- Calcular la media de las estimaciones
- Obtener los percentiles 2.5 y 97.5% de las estimaciones

5.7. Intervalos de predicción de las medidas de tendencia central

Cubren el rango de valores entre los cuales se encuentran los valores de m observaciones futuras con una probabilidad determinada (generalmente el 95%). En el caso de asumir una distribución normal se aplican las fórmulas analizadas en el capítulo 3. El límite inferior se obtiene sustituyendo el signo más por menos después de la media muestral. Los límites de predicción de la mediana son los mismos.

$$LP = \bar{x} + t_{1-\alpha/m, n-1} s \sqrt{1 + \frac{1}{n}}$$

Ecuación 40: Intervalo de predicción paramétrico para m muestras futuras

El intervalo de predicción que cubre el rango probable de valores en el caso de una distribución lognormal es:

$$IP = \exp(\bar{y} \pm t_{(1-\alpha/2), n-1} \sqrt{1 + \frac{1}{n}} s_y)$$

Ecuación 41: Intervalo de predicción paramétrico para un nuevo valor (distribución lognormal)

5.7.1. Simulación de datos censurados en las masas de agua analizadas

Dado que en las masas de agua subterránea investigadas en la cuenca del Júcar no se proporcionan datos sobre valores de concentración por debajo del límite de detección es necesario simular dichos valores con objeto de extraer conclusiones referentes a los métodos existentes.

Para ello se han impuesto límites de detección arbitrarios en cada una de las masas de forma que se establezcan tres niveles en el porcentaje de datos censurados. El primero corresponde a una censura ligera, en torno al 10% de valores por debajo del límite de detección. A continuación se fija otro límite cuyo resultado es una censura media, cercana al 25%, y para finalizar se determina el valor que produce una censura severa, variando este último supuesto entre el 40 y el 60%.

En la tabla 5.1 se muestran los resultados de la simulación con los correspondientes intervalos de confianza de la media mediante la utilización de seis métodos diferentes. Los tres primeros se corresponden con técnicas de sustitución directa mientras que los siguientes se han obtenido por medio de imputación, entendiendo ésta como el resultado de un proceso de modelización estadística. Como puede observarse en la tabla se han obtenido los intervalos sustituyendo los valores por debajo del límite de detección por cero, por la mitad del límite de detección y por último por dicho límite.

En cuanto a los métodos de imputación, se han seleccionado la regresión sobre los estadísticos de orden (en su variante robusta, con objeto de evitar los sesgos debidos a la transformación a la escala original), el método de Kaplan-Meier (dado que es un estándar en la literatura científica de análisis de supervivencia) y el método de estimación por máxima verosimilitud, como representante de la estadística paramétrica. Para cada uno de ellos se muestra el porcentaje real de valores censurados y el número de medidas por encima del límite de detección. De esta forma es posible comparar los resultados obtenidos mediante simulación con los valores reales, que se muestran en la primera fila para cada uno de los puntos de agua estudiados.

Pueden realizarse las siguientes consideraciones:

- Los métodos de sustitución proporcionan resultados aproximados en ocasiones, pudiendo variar la sustitución óptima entre cero y el límite de detección
- Analizados en detalle proporcionan estimaciones sesgadas de la desviación estándar, por lo que los intervalos de confianza obtenidos son más amplios que los correspondientes mediante técnicas específicas
- Las técnicas de imputación proporcionan valores aceptables en la mayoría de los casos, en concreto las de Kaplan-Meier y ROS.
- Las técnicas de estimación por medio de estimadores de máxima verosimilitud proporcionan resultados poco precisos, especialmente cuando

los datos no se ajustan a la distribución supuesta o el número de datos es pequeño. Por tanto, no deben seleccionarse a menos que existan gran cantidad de datos por encima del límite de detección, siendo necesario en cualquier caso verificar las hipótesis de normalidad (o ajuste a cualquier otra distribución seleccionada)

- A pesar de lo anterior, cuando la muestra es pequeña (5 o menos valores) no es posible aplicar de forma precisa ninguna técnica de análisis de datos censurados. En este caso la sustitución por la mitad del límite de detección es la única alternativa viable, al menos hasta que existan suficientes valores para realizar un análisis apropiado

5.8. Simulación de datos censurados en la zona piloto seleccionada

Como se ha comentado en capítulos anteriores, los documentos elaborados en el marco de la estrategia común de implementación (CIS), fueron evaluados en el ámbito territorial de la cuenca del Júcar (Demarcación Hidrográfica del Júcar, 2013).

Debe destacarse el hecho de que no se han identificado contaminación debida a fuentes puntuales. Por esa razón no hay conjuntos de datos adecuados para diseñar métodos estadísticos eficaces para hacer frente a los datos censurados por la izquierda. Como consecuencia, para evaluar el desempeño de los métodos disponibles se han seleccionado series de datos disponibles, en este caso los nitratos, y a posteriori se han censurado artificialmente.

Para ello se ha seleccionado la Plana de Sagunto (Figura 22), ya que se dispone de un conjunto de estaciones con registros históricos.

Los materiales acuíferos que conforman el acuífero de la Plana de Sagunto se encuentran representados por niveles detríticos de gravas y/o arenas con espesores en general inferiores a 10-15 m, incluidos en una formación pliocuaternaria fundamentalmente limo-arcillosa (Ballesteros et al, 2005). En la Plana de Sagunto existen 13 estaciones con valores por encima del valor

umbral para los nitratos (50 mg / L). Los datos históricos se han comparado hasta 2010.

Figura 49: Localización de la zona piloto

El primer paso a dar en el proceso de análisis de series temporales es la representación visual de los datos, con el objeto de identificar patrones, tendencias o variaciones estacionales. Si las observaciones son variables aleatorias independientes e idénticamente distribuidas suele usarse la regresión lineal para modelizar la relación existente entre una variable dependiente y varias variables explicativas.

Sin embargo, la aplicación de modelos lineales a datos de concentración de contaminantes en agua subterránea no suele ser adecuada, debido a sus características intrínsecas. De hecho, muchas variables de este tipo no están distribuidas normalmente y ni siquiera presentan una distribución simétrica (**Montgomery et al, 1986**). Para superar esa dificultad (**Helsel and Hirsch, 2002**) se utiliza la regresión local (LOESS) o la regresión local ponderada (LOWESS) (**Cleveland, 1979**).

En la Tabla 12 y Figura 50 se muestran las estaciones seleccionadas (con un mínimo de ocho valores), el número de registros existentes y las series temporales de concentración de nitratos correspondientes en la Plana de Sagunto junto con la curva procedente del suavizado.

Masa de Agua Subterránea Plana de Sagunto			
Código estación	Longitud	Código estación	Longitud
CA0820052	7	JUIG003550	3
CA0821053	5	JUIG003594	17
JUIG000542	4	JUIG003835	4
JUIG002683	8	JUIG003866	40
JUIG002924	8	JUIG003929	4
JUIG002950	34	JUIG003940	8
JUIG003137	5		

Tabla 12: Estaciones seleccionadas en Plana de Sagunto y número de registros

Figura 50: Series temporales en la Plana de Sagunto

En la Tabla 13 se muestran los resultados obtenidos. De izquierda a derecha se muestran los datos originales y los intervalos de confianza del valor medio calculado por diferentes métodos. En la primera columna se han incluido las

seis estaciones de las cuales se dispone de al menos ocho valores, identificadas con el código de la estación. En la segunda columna se indica la longitud de las series temporales originales y los límites de confianza de la media. En tercer lugar se indica el porcentaje de censura realizado de forma artificial con objeto de realizar la simulación. Para ello se ha supuesto un límite de detección arbitrario que produce un porcentaje de valores censurados similar en todas las estaciones. Se han seleccionado tres grados de censura, fijados en el 12, el 25 y el 50%.

En la columna situada bajo el encabezado “Valores sobre el límite de detección” se indica el número de valores sin censura una vez aplicado el grado de censura correspondiente. En las siguientes columnas se incluyen los intervalos de confianza calculados por cada uno de los métodos analizados, que se dividen en métodos de sustitución y de imputación. Entre los primeros hay que diferenciar entre la sustitución por 0, por la mitad del límite de detección y por el valor de dicho límite. La estimación por máxima verosimilitud, la regresión robusta y el método de Kaplan-Meier son los métodos de imputación analizados.

Como puede observarse en los resultados, los métodos de regresión robusta y de Kaplan-Meier son los que proporcionan resultados más aproximados a los datos reales y con longitudes menores del intervalo de confianza correspondiente, lo que implica un mayor grado de protección al medio ambiente, en el sentido de que dan indicaciones de falta de verificación del estándar ambiental o de detección de fugas de compuestos peligrosos antes que los otros métodos contemplados.

Intervalo de confianza del valor central										
LCI- LSC= Límite de confianza Inferior-Superior Sd=desviación estándar Código de estación	Datos originales		% datos censurados	Valores sobre el límite de detección	Substitución			Estimación máxima verosimilitud	Regresión robusta	Kaplan-Meier
					0	0.5 límite detección	Límite detección			
JUIG002683 media=133 sd=139.3	N=8 LIC=47.5 LSC=230.3		12.5	7	50.1-230.8	49.1-232.2	53.2-229.7	39.8-1455.2	49.51-231.2	39.6-231.2
			25	6	46.2-228.1	50.6-231.7	52.9-231.2	33.7-2629.5	47.7-236.23	40.7-233.1
			50	4	27.6-223.5	62.1-234.3	80.1-238.4	17.6-29791.1	-----	136.5-265.4
JUIG002924 media=129.4 sd=106.1	N=8 LIC=67.7 LSC=203.9		12.5	7	63.3-202.5	67.3-202.3	71.6-203.8	64.3-317.4	73.07-202.03	64.2-205.1
			25	6	50.9-199.7	65.4-200.2	77.9-204.9	57.3-366.7	71.71-207.31	69.1-207.7
			50	4	26.6-192	59-198.4	94.7-213.9	37.3-604.1	-----	73.8-214
JUIG002950 media=157 sd=68.9	N=34 LIC=135.1 LSC=180		11.7	32	129-178.2	132.7-180.4	135.5-180.4	133.2-187.3	135.7-180.3	135.2-180.6
			23.5	16	64.3-139.1	123-171.5	179.9-206	101.9-191.1	143-193.1	177.7-204.9
			47.1	12	43.5-119.8	120-166.5	194.6-216.7	84.4-181.7	132.3-198.1	192.5-215.7
JUIG003594 media=79.5 sd=40.2	N=17 LIC=61.6 LSC=98.7		15.3	12	44.1-92.8	60.8-97.6	76.1-102.9	50.5-147.5	69-100.9	74.4-102.6
			23.1	11	39.4-89.8	59.7-96	78.1-104	47.7-149.8	67.4-101.4	75.7-103.3
			46.1	7	18.6-75.5	54.7-91.6	88.8-110.4	32.8-163.2	49.8-101.8	85.3-109.1
JUIG003866 media=122.4 sd=75.1	N=40 LIC=99.8 LSC=145.5		12.8	33	94-143.1	99.2-145.9	105.1-148.1	148-1012.7	107-148.9	108.3-149.8
			25.6	29	85.6-139.1	99.5-145.7	113.1-152.1	143.4-1276.6	106.9-152.6	111.9-151.8
			48.7	20	81.2-136.3	100.1-145.8	118.3-155	140.1-1456	110.7-162.5	116.9-154.6
JUIG003940 media=85.9 sd=66.6	N=8 LIC=42.9 LSC=130		12.5	7	39.9-128.1	43.9-130.1	46.3-130.5	38.9-256.3	42.4-130.6	41.6-132.8
			37.5	5	30.5-127.9	38-129.2	48.5-130.3	29.1-508.1	50.9-145.9	64.5-137.6
			50	4	19.5-124	50.1-130.5	78.9-136.9	21.6-802.7	-----	101.7-147.1

Tabla 13: Intervalos de confianza de la media estimados en las estaciones seleccionadas

Capítulo 6:

RESULTADOS Y DISCUSIÓN

6. RESULTADOS Y DISCUSIÓN

La evaluación del estado químico de una masa de agua subterránea debe basarse en el diseño de estrategias para la detección de aumentos estadísticamente significativos en la concentración de aquellas sustancias (a partir del análisis de presiones e impactos) que pueden provocar su clasificación como en riesgo de no conseguir buen estado cualitativo y la posterior declaración del mal estado químico de dicha masa. Ello se consigue fundamentalmente por medio de comparaciones entre datos pasados y presentes. El valor básico (definido en la Directiva de Aguas Subterráneas) es el punto de partida para dicha comparación.

La calidad de los datos utilizados para el establecimiento del valor básico es, por tanto, un criterio fundamental para el éxito de un programa de control, por lo que es necesario aplicar técnicas para su análisis. Las técnicas exploratorias son una herramienta clave para mejorar el conocimiento sobre un conjunto de valores de concentración. Se basan fundamentalmente en la visualización de los datos y deben incluir técnicas de suavizado para la exploración y comprensión de los patrones y tendencias existentes.

Una vez realizados los análisis descritos es preciso seleccionar un periodo de referencia. El diseño de métodos estadísticos para determinar la existencia de incrementos sobre dicha concentración de referencia es la base de dicha comparación. Dado que el valor básico es el valor medio medido por lo menos durante los años de referencia 2007 y 2008 y que la frecuencia de la toma de datos en muchos programas de control es bianual, es deseable incorporar los datos históricos existentes con objeto de aumentar el tamaño de la muestra e incorporar la variabilidad natural, que no puede ser observada en unos pocos valores.

A continuación es necesario plantear tests estadísticos formales con objeto de obtener resultados comparables (tal como exige la Directiva). La existencia de tendencias debe ser comprobada en primer lugar, y las técnicas utilizadas para detectar incrementos de concentración están diseñadas para detectar

conformidad con los valores umbral y detectar impactos significativos en la calidad del agua subterránea.

El valor básico representa un valor de tendencia central, por lo que la evaluación debe basarse en la media o en la mediana tanto si se ha definido un valor umbral como si no. En el primer caso se estima la media por medio de intervalos de confianza y se compara el límite superior del intervalo con el valor umbral. Su cálculo requiere generalmente la asunción de normalidad de la población de origen. Si no es posible normalizar los datos o el porcentaje de valores por debajo del límite de detección es muy alto deben utilizarse métodos estadísticos específicos para el manejo de este tipo de datos.

En el caso de que no se haya definido un valor umbral, la metodología debe basarse en la construcción de intervalos de predicción y en el análisis de tendencias. Los primeros son apropiados para la detección de impactos estadísticamente significativos en la calidad del agua, y deben utilizar como valor de comparación valores futuros individuales o preferiblemente medias de valores futuros para su comparación con el valor básico. El análisis de tendencias (Figura 52) proporciona información sobre la existencia de tendencias sostenidas crecientes en la concentración de contaminantes.

Un aspecto relevante, considerado en la propia Directiva de Nitratos, es el intervalo temporal entre las medidas encaminadas a mejorar la calidad del recurso y las respuestas en la calidad de una masa de agua subterránea. La elaboración de un modelo conceptual adecuado de la masa de agua subterránea en riesgo es una cuestión esencial.

La selección de una red de control adecuada conjuntamente con un programa de control apropiado son elementos fundamentales en cualquier estrategia de control de la calidad del agua. En este sentido, la geoestadística proporciona una herramienta óptima para el diseño de redes.

Los únicos métodos de análisis de datos con valores censurados que pueden manejar series con múltiples límites de detección son Kaplan-Meier y estimación robusta (ROS), por lo que cuando se ha reportado más de un límite (debido a cambios en el método de medida, actualización de la

instrumentación, etc.), son los únicos métodos que es posible utilizar. En caso de que quiera utilizarse cualquier otro es preciso censurar por debajo del límite de detección más alto utilizado.

Cuando el porcentaje de valores por debajo del límite de detección es bajo (10-15%) es factible utilizar métodos basados en la sustitución simple. Este límite, aunque en cierto modo arbitrario, proporciona resultados aceptables en la mayoría de las situaciones (**USEPA, 2009**). Sin embargo, cuando el porcentaje es mayor (<50%) es preferible utilizar métodos paramétricos. Teniendo en cuenta que dichas técnicas asumen que no existen valores censurados, es posible utilizar métodos para evaluar la normalidad aproximada en presencia de este tipo de valores. Por último, es posible utilizar técnicas no paramétricas cuando no sea posible asumir normalidad. Ello presenta inconvenientes, especialmente en el caso de que el nivel básico esté compuesto de un número reducido de valores.

La visualización de los datos es un paso previo que debe realizarse de forma similar al caso de valores por encima del límite de detección.

La aplicación de métodos paramétricos (máxima verosimilitud) proporciona resultados poco precisos, que pueden llegar a ser inaceptables en el caso de un número pequeño de muestras. En este caso, la presencia de uno o dos valores anómalos pueden alterar totalmente los resultados. Además, si el cálculo de la media se realiza mediante transformación de los valores obtenidos mediante ajuste a una distribución lognormal, por ejemplo, se introduce un sesgo debido a dicha transformación.

Cuando el porcentaje de datos censurados es sensiblemente inferior al 50% tanto los métodos de regresión robustos sobre los estadísticos de orden (ROS) como el método de Kaplan-Meier proporcionan resultados satisfactorios en la estimación de la media y la desviación estándar.

A medida que la proporción de valores censurados se acerca al 50% el método de Kaplan-Meier no proporciona estimadores de la media o la mediana, por lo que es preciso utilizar métodos alternativos, como la regresión robusta (ROS).

Si el porcentaje de valores censurados es superior al 80% no es posible obtener estimaciones de la media o la mediana, aunque es posible calcular el porcentaje de datos por encima de un valor umbral. Otros autores (**Helsel, 2005**), sin embargo, han sugerido que en series con gran número de valores (100-300) es posible obtener estimaciones aceptables de los estadísticos de orden superior (percentiles 90 y 95).

Cuando todos los valores se encuentran por debajo del límite de detección, es posible fijar este último como estándar y aplicar la regla de la doble cuantificación, que implica que dos mediciones consecutivas por encima del límite de detección equivalen automáticamente a un incremento estadísticamente significativo.

El cálculo de intervalos de confianza para la media y la desviación estándar puede ser realizado de varias formas. En el caso de poder asumir distribuciones normales o lognormales la distribución t de Student o la ecuación de Cox, respectivamente, proporcionan el intervalo de confianza deseado.

En el caso de la regresión robusta (ROS) no existe un método directo para calcular dicho intervalo de confianza. El método de Kaplan-Maier, sin embargo, sí que proporciona dicha estimación por varias vías, entre las cuales cabe citar la fórmula de Greenwood o el método del signo B-C.

En todos los casos es posible aplicar técnicas de bootstrap que permiten obtener intervalos de confianza de la media. Son métodos más precisos mediante los cuales se obtienen los mejores resultados. Otra consideración adicional que cabe realizar es que como los conjuntos de datos en las estaciones de medida van aumentando de forma progresiva, la escasez de datos no va a ser una dificultad en el futuro próximo.

Capítulo 7:

CONCLUSIONES

7. CONCLUSIONES

En primer lugar, dados los resultados generados en el transcurso de esta investigación, cabe afirmar que se han alcanzado los objetivos generales especificados en el primer capítulo de esta tesis.

En concreto, y en línea con las prioridades políticas definidas por la Comisión Europea mediante la Directiva Marco del Agua y la Directiva derivada de Aguas Subterráneas, se ha diseñado y elaborado una metodología que permite la toma de decisiones sobre el estado químico de una masa de agua subterránea, basada en los datos proporcionados por las redes de control.

Igualmente, debe tenerse en cuenta que dichas Directivas establecen una serie de objetivos ambientales, entre los que se encuentran la adopción de medidas para prevenir o limitar la entrada de contaminantes en el agua subterránea y prevenir su deterioro, revertir la tendencia al aumento en la concentración de contaminantes o asegurar el cumplimiento de los objetivos de protección y estándares específicos de calidad de las áreas protegidas. En este sentido, la metodología propuesta permite detectar incrementos significativos en la concentración de contaminantes, evitando de esta forma el deterioro del recurso y proporcionando un sistema de alerta temprana tanto para el control de fuentes puntuales de contaminación como para el estudio de las fuentes difusas, causantes de gran parte de la contaminación presente en los acuíferos.

Para evaluar el estado de los recursos hídricos es necesario compararlos con un estándar predefinido, cuantitativo o cualitativo. El último se denomina valor umbral y debe ser determinado para contaminantes que causan la clasificación del estado de una masa de agua subterránea (o de una masa de agua superficial asociada) como pobre, siendo los Organismos de Cuenca, por medio de los Planes Hidrológicos de Cuenca, los responsables de su determinación así como de la evaluación de la existencia de tendencias al aumento en la concentración de contaminantes (o su inversión).

Mientras que existen estándares de calidad de agua subterránea para nitratos y pesticidas fijados por la legislación comunitaria, es necesario establecer valores umbral para los demás contaminantes. Dichos valores son estándares de Calidad Medioambiental que deben ser establecidos por los Estados Miembros, como parte de los esquemas de protección que aseguren la prevención del deterioro de las aguas subterráneas a escala regional, contribuyendo de ese modo a la política comunitaria en materia de aguas.

Estos valores umbral deben ser determinados para todos los contaminantes que han contribuido a la clasificación de una masa de agua subterránea como “en riesgo”, de acuerdo con el análisis de presiones e impactos realizado en función del artículo 5 de la Directiva Marco del Agua. Su determinación se basa, por tanto, en un análisis de riesgos, de acuerdo con el esquema tradicional: fuente→ vías de exposición→ receptores. La evaluación cuantitativa de riesgos puede definirse como el proceso que tiene por objeto caracterizar la naturaleza y magnitud de los riesgos que comporta la presencia de sustancias peligrosas en el medio ambiente para la salud humana o los ecosistemas. El modelo conceptual es una herramienta fundamental en el proceso de valoración y gestión de riesgos. Debe referirse al emplazamiento objeto de estudio y, en caso necesario, incluir un modelo físico conceptual del sistema suelo-agua.

Para ello es preciso diseñar contrastes estadísticos que comparen las medidas de la red de control y un estándar de calidad determinado por las autoridades ambientales. Debe destacarse el hecho de que la normativa comunitaria está basada en prevenir y limitar la entrada de contaminantes en el agua subterránea, por lo que el establecimiento de un estándar de calidad es condición necesaria pero no suficiente para preservar la calidad del recurso hídrico, ya que transmite la falsa sensación de que los estados miembros están habilitados para contaminar hasta el nivel definido por el estándar de calidad. A este respecto, la metodología permite detectar deterioros en la calidad del recurso o mejoras debidas a la implementación de un plan de medidas.

A ello debe sumarse que las Directivas incluyen un nuevo concepto, que es el valor intrínseco del agua, expresado en la necesidad de protección de sus usos

legítimos, reales o potenciales. Asimismo, en determinadas situaciones dicho estándar no se encuentra disponible, bien porque no ha sido fijado por el organismo competente, bien porque la concentración histórica supera el estándar. La única comparación posible es entonces, la que proporciona el fondo de valores en el punto de la red de calidad.

Por tanto, el objetivo fundamental de un programa de control para la evaluación del estado de una masa de agua subterránea debe basarse en el diseño de estrategias para la detección de aumentos estadísticamente significativos en la concentración de aquellas sustancias que han ocasionado la declaración de dicha masa como en riesgo de no conseguir buen estado químico.

Ello se consigue fundamentalmente por medio de comparaciones entre datos pasados y presentes. El valor básico (definido en la Directiva de Aguas Subterráneas) es el punto de partida para dicha comparación. La calidad de los datos utilizados para el establecimiento del valor básico es, por tanto, un criterio fundamental para el éxito de un programa de control.

Figura 51: Diagrama de flujo para la toma de decisiones en masas de agua subterránea

El diseño de métodos estadísticos para determinar la existencia de incrementos en concentración sobre el valor básico es el procedimiento de referencia. Dado que el valor básico se utiliza en los ensayos estadísticos, es deseable incorporar los datos históricos existentes con objeto de aumentar el tamaño de la muestra e incorporar la variabilidad natural, que no puede ser observada en unos pocos valores.

El valor básico representa un valor de tendencia central. La media se estima por medio de intervalos de confianza, y su cálculo requiere generalmente la asunción de normalidad de la población de origen. Si no es posible normalizar los datos o el porcentaje de valores por debajo del límite de detección es muy alto deben utilizarse métodos estadísticos específicos para el manejo de este tipo de datos.

La existencia de tendencias debe ser comprobada en primer lugar, siendo preciso a continuación plantear tests estadísticos formales mediante la elaboración de intervalos de confianza de la media o la mediana, con objeto de obtener resultados comparables (tal como exige la Directiva), y detectar conformidad con los valores umbral o impactos significativos en la calidad del agua subterránea.

En el caso de que no se haya definido un valor umbral, la metodología debe basarse en la construcción de intervalos de predicción y en el análisis de tendencias. Los primeros son apropiados para la detección de impactos estadísticamente significativos en la calidad del agua, y deben utilizar como valor de comparación valores futuros individuales o preferiblemente medias de valores futuros para su comparación con el valor básico. El análisis de tendencias proporciona información sobre la existencia de tendencias sostenidas crecientes en la concentración de contaminantes.

Un aspecto relevante, considerado en la propia Directiva de Nitratos, es el intervalo temporal entre las medidas encaminadas a mejorar la calidad del agua subterránea y las respuestas en la calidad de una masa de agua subterránea.

Existen ocasiones en las que no es posible realizar la división de la masa de agua subterránea, bien por falta de información (no existe un modelo conceptual definido, etc.), bien por motivos de gestión. En tales casos, la existencia de una tendencia sostenida al aumento en la concentración de un contaminante no es motivo suficiente para declarar la masa en mal estado. Es necesaria entonces la aplicación de técnicas que permitan comprobar la presencia de tendencias regionales y verificar de esta forma la existencia de un impacto ambiental significativo. El método más eficaz para determinar tales tendencias es el desarrollado por el Servicio Geológico de los Estados Unidos (USGS). Se conoce como el test regional de Mann-Kendall y es una modificación del test estacional original de Mann Kendal para análisis de tendencias, sustituyendo las estaciones (periodos de muestreo) por emplazamientos (coordenadas). Ha sido verificado en la zona piloto Plana de Vinaroz Peñíscola.

Figura 52: Tendencia general en la concentración de cloruros en la Plana de Vinaroz Peñíscola

Con frecuencia, los resultados proporcionados por los laboratorios en relación con la concentración de contaminantes en suelo y agua subterránea incluyen valores clasificados por debajo del límite de detección (o cuantificación), es

decir, con concentraciones de las que únicamente se sabe que su valor está comprendido entre 0 y el límite correspondiente. La interpretación de series con este tipo de valores es compleja, aunque en los últimos años se han producido avances significativos para su interpretación, entre los que merece especial mención el desarrollo de software libre por diferentes organismos, como la Agencia de Medio Ambiente de los Estados Unidos.

En el caso de muchos de los datos disponibles sobre concentración de contaminantes, el tamaño de la muestra es un factor limitante, ya que normalmente no es posible disponer de conjuntos de datos con 10 o más valores. Ello tiene un gran impacto en los intervalos de confianza elaborados con objeto de verificar el cumplimiento de estándares o detectar fugas de contaminantes en emplazamientos en los que se realizan actividades potencialmente contaminantes. Con tamaños de muestra pequeños (menos de 10) la estimación de máxima verosimilitud en datos censurados funciona de forma poco satisfactoria, como consecuencia de asumir hipótesis distribucionales.

En consecuencia, proporciona intervalos de confianza más grandes en relación con otros procedimientos. La utilización de métodos de sustitución (la mitad del límite de detección) puede proporcionar una indicación de cumplimiento de dichos estándares, aunque el aumento del número de muestras es un requisito indispensable para realizar una evaluación precisa. Por otro lado, cuando hay más de un límite de detección (o se espera como consecuencia del cambio de precisión analítica en el tiempo) es necesario utilizar técnicas particulares para cumplir estos requisitos. Las técnicas de estimación por medio de estimadores de máxima verosimilitud proporcionan resultados poco precisos, especialmente cuando los datos no se ajustan a la distribución supuesta o el número de datos es pequeño. Por tanto, no deben seleccionarse a menos que existan gran cantidad de datos por encima del límite de detección, siendo necesario en cualquier caso verificar las hipótesis de normalidad (o ajuste a cualquier otra distribución seleccionada).

Las técnicas de regresión robusta y el método de Kaplan-Meier son los métodos recomendados, ya que proporcionan una mayor precisión y son

capaces de hacer frente a varios límites de detección. Además, se recomienda un mínimo de 10-15 muestras por encima del límite de detección para obtener resultados precisos. Por último, es necesario comprobar la existencia de tendencias al aumento en la concentración de contaminantes de forma previa a la realización de cualquier análisis estadístico, con objeto de eliminar su influencia.

A pesar de lo anterior, cuando la muestra es pequeña (5 o menos valores) no es posible aplicar de forma precisa ninguna técnica de análisis de datos censurados. En este caso la sustitución por la mitad del límite de detección es la única alternativa viable, al menos hasta que existan suficientes valores para realizar un análisis apropiado

Debe destacarse el hecho de que el uso de estrategias basadas en el análisis de tendencias en combinación con el cálculo de intervalos de confianza (cuando se han definido valores umbral) o de intervalos de predicción (cuando no se han definido) es una herramienta potente a la vez que flexible para la toma de decisiones en relación con el estado químico de una masa de agua subterránea.

Teniendo en cuenta la necesidad de evaluar el estado químico de las masas de agua subterránea en riesgo, y considerando que no existe un ensayo estadístico único apropiado para cualquier situación, la metodología propuesta permite la integración de las técnicas individuales existentes en una metodología general. Combina las etapas requeridas en dicha evaluación mediante la aplicación de técnicas estándar con indicaciones del tamaño de muestra adecuado. De esta forma se superan las limitaciones de los ensayos individuales incluso en aquellos casos en los que la autoridad ambiental competente no ha fijado un valor umbral. Cabe agregar que incluso cuando se ha declarado una masa de agua subterránea en mal estado y existen razones (como la inviabilidad técnica o económica), que impidan la aplicación de un plan de medidas, la investigación realizada proporciona una herramienta para prevenir el deterioro mediante la aplicación secuencial de las siguientes técnicas estadísticas:

Objetivo	Técnica estadística
Elaboración del Nivel Básico	Representación gráfica Regresión local (LOESS)
Pruebas de bondad de ajuste	Shapiro Wilk Chi cuadrado Gráficos de probabilidad
Detección de valores anómalos	Dixon Rossner
Análisis de tendencias	Regresión lineal Mann Kendall Theil-Senn
Verificación de conformidad con estándar	Intervalo de confianza de la media de una distribución normal, gamma o lognormal Límites de confianza mediante bootstrap
Análisis de tendencias regionales	Mann Kendall estacional
Valores censurados. Conformidad con valor umbral	Regresión robusta (ROS) Kaplan-Meier Simulación de Montecarlo con las dos anteriores
Actualización nivel básico	Prueba t de Student Test de Wilcoxon

Como consideración final, se subraya el hecho de que es posible determinar el estado químico de una masa de agua subterránea por medio de software libre. En concreto, para la realización de todos los análisis estadísticos expuestos en esta tesis se ha utilizado el software estadístico R como herramienta de referencia. R está disponible como software libre bajo los términos de la fundación de software libre GNU, es de fácil implementación y no requiere recursos informáticos fuera de lo habitual.

Capítulo 8:

***RECOMENDACIONES Y LÍNEAS DE
INVESTIGACIÓN FUTURAS***

8. RECOMENDACIONES Y LÍNEAS DE INVESTIGACIÓN FUTURAS

En este capítulo se presentan algunas líneas de investigación que pueden ser objeto de interés y que han surgido de los resultados de este trabajo de investigación.

En primer lugar es necesario desarrollar valores umbral para los diferentes usos del agua subterránea, basado en la evaluación de riesgos, que es la herramienta que determina la legislación española para la declaración y gestión de suelos contaminados. Además, es necesario proporcionar bases sólidas para el desarrollo de dichos valores umbral en ecosistemas terrestres dependientes del agua subterránea.

Es preciso estudiar el impacto del cambio climático sobre la calidad de los recursos hídricos, y especialmente su afección sobre la evolución de los niveles básicos de concentración de contaminantes y el incremento de valores anómalos.

Una de las limitaciones del trabajo se debe al análisis de series de valores de concentración de contaminantes con valores por debajo del límite de detección. Debido a que no existen series con el suficiente número de registros, se han simulado artificialmente los datos con objeto de obtener diferentes grados de censura. Por tanto, es preciso reforzar las líneas de investigación relacionadas con objeto de validar el análisis realizado.

Capítulo 9:

BIBLIOGRAFÍA

9. BIBLIOGRAFÍA

- AIH-GE, 2003. Presente y futuro del agua subterránea en España y la Directiva Marco Europea (Present and future of groundwater in Spain and the European Framework Directive). Intern. Assoc. Hydrogeologists—Spanish Chapter/IGME, Madrid, pp. 1–540.
- American Society for Testing and Materials, 2005. Standard Guide for Developing Appropriate Statistical Approaches for Groundwater Detection Monitoring Programs (D6312-98[2005])
- American Society for Testing and Materials, 2004. D7048-04: Standard Guide for Applying Statistical Methods for Assessment and Corrective Action Environmental Monitoring Programs. West Conshohocken, PA: ASTM International
- Antweiler, R.C. and Taylor, H.E., 2008. Evaluation of Statistical Treatments of Left-Censored Environmental Data using Coincident Uncensored Data. Sets: I. Summary Statistics. Environ. Sci. Technol. 2008, 42, 3732–3738
- Arenas, J. M.^a, Carrero, G., Galache, J., Mediavilla, C., Silgado, A. y Vázquez, E. M. (2001). Actuaciones realizadas tras el accidente de Aznalcóllar. Boletín Geológico y Minero. Vol. Especial, 35-56. ISSN 0366-0176
- Baynes S (2004) Water efficient toilets: a canadian perspective. Canada Mortgage and Housing Corporation (electronic version). Retrieved 16 Oct 2009
- Blum, A., Pauwels, H., Wenland, F. And Grifioen, J., 2009. Background levels under the Water Framework Directive. Groundwater Monitoring (Water quality measurement series). John Wiley and Sons, Chichester, Wiltshire, p.p. 145-153. ISBN: 978-0470-77809-8.
- Ballesteros, B. J., Domínguez Sánchez, J.A., Díaz-Losada, E. y García Menéndez, O., 2009. Zonas húmedas mediterráneas y acuíferos asociados. Condicionantes hidrogeológicos del Marjal de Pego-Oliva (Alicante-Valencia). Boletín Geológico y Minero, 120 (3): 459-478. ISSN: 0366-0176
- Ballesteros, B.J., Barba-Romero, Muñoz, J. y Oriol Navarro Odriozola, J., 2005. Caracterización adicional Plana de Sagunto. Memoria. Trabajos técnicos para la aplicación de la Directiva Marco del Agua en materia de aguas subterráneas. Convenio de colaboración entre la Dirección General del Agua y el Instituto geológico y Minero de España. Base documental del IGME.
- Ballesteros Navarro, B.J., López Gutiérrez, J., Grima Olmedo, J., García Menéndez, O. y M. Pérez Gago, 2001. Recuperación de la calidad del acuífero costero de la plana de Jávea mediante la explotación sostenible de sus recursos. Investigación, gestión y recuperación de acuíferos contaminados, Alicante, 183-194. ISBN: 84-607-3216-9.
- Ballesteros, B.J., Rodríguez, L., López, J., Grima, J. García, O., Cobos, J.R., Gómez J.D. and De la Orden, J.A., 2001. Análisis y ordenación de recursos hídricos de la Marina Alta (Alicante). Alternativas y Directrices. Vol II. Evaluación de recursos subterráneos. Instituto Geológico y Minero de España-Diputación Provincial de Alicante. Fondo documental IGME. 179 pp. Alicante.
- Bartels R, 1982. The rank version of von Neumann's ratio test for randomness. J. Amer. Stat. Assn, 77,40-46.
- Bradu, D. and Mundlak, Y., 1970. *Estimation in Lognormal Linear Models*. Journal of the American Statistical Association, 65, 198-211.
- Chacón, E. y Míguez, F., 1980. Estadística aplicada. p.p. 428-429. Depósito legal: M-33461-1980. ISBN: 84-600-2023-1.
- Chatfield C., 2004. The Analysis of Time Series: An Introduction (6th Edition). Boca Raton, FL: Chapman and Hall.
- Chambers, R.L., 1986: "Outlier Robust Finite Population Estimation", Journal of the American Statistical Association, 81, pp. 1063-1069.
- Chambers, John, William Cleveland, Beat Kleiner, and Paul Tukey, 1983. Graphical Methods for Data Analysis, Wadsworth
- Chernick, M. R., 1999. *Bootstrap methods: a practitioner's guide*. Nueva York, Wiley & Sons.
- CHJ, 1998. Plan Hidrológico de la cuenca del Júcar. Ministerio de Medio Ambiente. Confederación Hidrográfica del Júcar.

- Choi, S. C., and Wette, R., 1969. Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias. *Technometrics*, Vol II, pp 383-690.
- Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatter-plots", *Journal of the American Statistical Association*, V. 74, pp. 829-836.
- Cohen, A. C. Jr., 1950. Estimating the Mean and Variance of Normal Populations from Singly Truncated and Double Truncated Samples. *Ann. Math. Statist.*, Vol. 21, pp. 557-569.
- Cohen, A. C. Jr., 1959. Simplified Estimators for the Normal Distribution When Samples Are Singly Censored or Truncated. *Technometrics*, Vol. 1, No. 3, pp. 217-237.
- Comisión Europea, 2008. Protección de las aguas subterráneas en Europa. Luxembourg: Office for Official Publications of the European Communities. 35 pp. ISBN 978-92-79-09818-5
- Comisión Europea, 1998. Directiva 98/83/CE del Consejo, de 3 de noviembre de 1998, relativa a la calidad de las aguas destinadas al consumo humano.
- Comisión Europea, 1979. Directiva 80/68/CEE del Consejo, de 17 de diciembre de 1979, relativa a la protección de las aguas subterráneas contra la contaminación causada por determinadas sustancias peligrosas. *Diario Oficial n° L 020 de 26/01/1980 p. 0043- 0048. Edición especial en español: Capítulo 15 Tomo 2 p. 162.*
- Currie, L.A., 1968. Limits for Qualitative Detection and Quantitative Determination: Application to Radiochemistry, *Anal. Chem.* 40, 586-593 (1968).
- Custodio, E., 2003. Groundwater role in world water resources: basics and facts. XI World Water Congress, Madrid. Intern. Water Resour. Assoc., pp. 1–17 (CD edited by CEDEX, Madrid).
- Datta, S., 2005. Estimating the mean life time using right censored data. *Statistical Methodology* 2, 65-69.
- Davis CB & McNichols R.B., 1999. Simultaneous nonparametric prediction limits. *Technometrics*, 41, 48.
- Del Moral Ituarte, L. (2008): "Integración de políticas sectoriales: agua y territorio", en Panel científico-técnico de seguimiento de la política de aguas. Sevilla, Convenio Universidad de Sevilla-Ministerio de Medio Ambiente.
- De Stefano, L., J.M. Fornés, J.M., López-Geta, J.A. and Villarroya, F. 2014. Groundwater use in Spain: An overview in light of the EU Water Framework Directive. *The International Journal of Water Resources Development*. DOI: 10.1080/07900627.2014.938260.
- Dixon, W.J., 1950. Analysis of extreme values. *Ann. Math. Stat.* 21, 4, 488-506.
- Domínguez J. A. y Ballesteros B. J. (2008). Análisis hidrogeológico de manantiales salinos asociados al acuífero carbonatado de Albuera-Gallinera-Mustalla en el entorno de la zona húmeda de Pego-Oliva (Valencia-Alicante). *Hidrogeología y Recursos Hidráulicos*. Vol. XXVIII. Publicaciones del Instituto Geológico y Minero de España, 209-222.
- Efron, B.; Tibshirani, R. J., 1993: An introduction to the Bootstrap. Nueva York, Chapman & Hall/CRC.
- Efron, B., 1987. "Better Bootstrap Confidence Intervals". *Journal of the American Statistical Association* 82: 171–185. JSTOR 2289144
- Efron, B., 1979: «Bootstrap methods: another look at the jackknife». *The Annals of Statistics*, 7, 1-26.
- Efron, B., 1967. The two sample problem with censored data, in: *Proc. Fifth Berkeley Symp. Math. Statist. Probab.*, vol. 4, Prentice-Hall, New York, 831–853.
- Estrela, T., Maestu, J., Fidalgo, A., Pérez, M.A., Fullana, J. y Pujante, A.M., 2004. Jucar Pilot River Basin. Provisional Article 5 Report. Pursuant to the Water Framework Directive – 208 pages. ISBN: 84-922007-2-3
- European Commission, 2009. Guidance on Groundwater Status and Trend Assessment. Guidance Document No 18. Technical Report - 2009- 026. ISBN 978-92-79-11374-1. European Communities, Luxembourg.
- European Commission, 2007. Guidance on Preventing or Limiting Direct and Indirect Inputs in the context of the Groundwater Directive 2006/118/EC, Guidance Document No 17. Technical Report - 2007 - 012. ISBN 978-92-79-06277-3. European Communities, Luxembourg.
- European Commission, 2001 – Statistical aspects of the identification of groundwater pollution trends and aggregation of monitoring results – Working Group 2.8 Statistics. Technical Report No. 1. ISBN 92-894-5639-6. European Communities, Luxembourg.
- European Parliament and Council Directive 2000/60/EC of 23 October 2000 establishing a framework for Community action in the field of water policy, OJ L 327, 22.12.2000, pp1-72.

European Parliament and Council Directive 2006/118/EC of 12 December 2006 on the protection of groundwater against pollution and deterioration, OJ L372, 27.12.2006, pp19-31.

Faires, J. D., and Burden, R. L., 1993. Numerical Methods. PWS-Kent Publishing Company, Boston, USA.

Fernández Escalante, A.E., González Herrarte, F.B. y San Sebastián Sauto, J., 2015. Recarga gestionada de acuíferos: multifuncionalidad en la zona regable de Santiuste. XXXIII Congreso Nacional de Riegos Universitat Politècnica de València, Valencia 2015. DOI:<http://dx.doi.org/10.4995/CNRIegos.2015.1445>

Fox, J., 2000. Nonparametric Simple Regression: Smoothing Scatterplots. Sage University Paper. Series: Quantitative applications in the Social Sciences, v 130. ISBN: 0-7619-1585-0.

Foster SSD, Chilton PJ (2003) Groundwater: the processes and global significance of aquifer degradation. Phil Trans R Soc Lond B 358:1957–1972. doi:10.1098/rstb.2003.1380

Gibbons, R.D. and Coleman, D.E., 2001. Statistical Methods for Detection and Quantification of Environmental Contamination, New York: Wiley.

Gibbons, R.D., 1994. The folly of Subtitle D statistics: when greenfield sites fail. Paper presented at the Proceedings of Waste Technology '94, Charleston, SC.

Gibbons RD, 1991. Some additional nonparametric prediction limits for ground-water detection monitoring at waste disposal facilities. *Ground Water*, 29(5), 729-736.

Gibbons, R.D., 1987. Statistical prediction intervals for the evaluation of ground-water quality. *Ground Water*, v 25, pp. 455-465.

Gilbert R.O., 1987. Statistical Methods for Environmental Pollution Monitoring. New York: Van Nostrand Reinhold.

Gil Flores, J., 2005. Aplicación del método bootstrap al contraste de hipótesis en la investigación educativa. *Revista de Educación* núm 336, pp 251-265.

Gleit, A., 1985. Estimation for small normal data sets with detection limits, *Environ. Science and Technol.*, 19, 1201-1206.

Grath, J., Scheidleder, A., Uhlig, S., Weber, K., Kralik, M., Keimel, T., Gruber, D., 2001: "The EU Water Framework Directive: Statistical aspects of the identification of groundwater pollution trends, and aggregation of monitoring results". Annex to the Final Report. Austrian Federal Ministry of Agriculture and Forestry, Environment and Water Management (Ref.: 41.046/01-IV1/00 and GZ 16 2500/2-I/6/00), European Commission (Grant Agreement Ref.: Subv 99/130794), in kind contributions by project partners. Vienna. December, 2001.

Grice, J.V., and Bain, L. J., 1980. *Inferences Concerning the Mean of the Gamma Distribution*. Journal of the American Statistical Association. Vol. 75, Number 372, 929-933.

Grima, J., Luque, J.A., Mejía, J.A. and Rodríguez, R. (2015). Methodological approach for the analysis of groundwater quality in the framework of the Groundwater Directive. *Journal of Environmental Earth Sciences*, 1-13. DOI 10.1007/s12665-015-4472-x. Springer, Berlin, Heidelberg.

Grima, J.; Luque, J.A.; Varela, M.; Sánchez, I., 2010. Trend of contaminants in groundwater bodies of intercommunity Spanish basins. European Groundwater Conference, Madrid, 20-21 may, poster session and proceedings.

Grima, J., Chacón, E., Ballesteros, B., Rodríguez, R. and Mejía, J.A., 2009. Communication of groundwater realities base don assessment and monitoring data. Groundwater Monitoring (Water quality measurement series). John Wiley and Sons, Chipelham, Wiltshire, p.p. 395-407. ISBN: 978-0470-77809-8.

Grima Olmedo, J., Ballesteros Navarro, B., López Gutiérrez, J. y García Menéndez, O., 2002. Sustentabilidad de las técnicas de recuperación de acuíferos y suelos contaminados y necesidades de investigación en la Unión Europea. Investigación, Gestión y recuperación de acuíferos contaminados. Ponencias. Editores: J.A. López Geta, J. Grima Olmedo y L. Rodríguez Hernández. ISBN: 84-607-6700-0.

Gilliom R.J. and Helsel D.R., 1986. Estimation of distributional parameters for censored trace level water quality data: I. Estimation techniques. *Water Resources Research*, 22, 135-146.

Guttman, I., 1970. *Statistical Tolerance Regions: Classical and Bayesian*. Darien, CT: Hafner Publishing.

Hall, P., 1988. *Theoretical comparison of bootstrap confidence intervals*. *Annals of Statistics*, 16, 927-953.

- Hall, P., 1992. *On the Removal of Skewness by Transformation*. Journal of Royal Statistical Society, B 54, 221-228.
- Helsel, D.R., 2012. Statistics for censored environmental data using Minitab and R. Wiley, 343 p.p. ISBN: 978-0-470-47988-9.
- Helsel, D.R., Mueller, D.K., and Slack, J.R.(2006). Computer program for the Kendall family of trend tests :U.S. Geological Survey Scientific Investigations Report 2005–5275, 4 p.
- Helsel, D.R., 2005. Nondetects and Data Analysis: Statistics for censored environmental data. Wiley, 268 p.p. ISBN: 978-0-471-67173-2.
- Helsel, D.R. and Hirsch, R.M., 2002. Techniques of Water-Resources Investigations of the United States Geological Survey. Book 4, Hydrologic Analysis and Interpretation. Statistical Methods in Water Resources. Environmental science and pollution research international, 14(5):297-307.
- Helsel, D.R., 1990. Less than obvious: "Statistical treatment of data below the detection limit." Environmental Science and Technology. Vol. 24, 1766-1774.
- Helsel, D.R. and Cohn, T.A., 1988. "Estimation of descriptive statistics for multiply censored water quality data." Water Resources Research. Vol. 24, 1997-2004.
- Hornik, K., 2008. Frequently Asked Questions on R. {ISBN} 3-900051-08-9. <http://CRAN.R-project.org/doc/FAQ/R-FAQ.html>
- Huston, C. and Juarez-Colunga, E., 2009. Guidelines for computing summary statistics for data-sets containing non-detects. http://www.bvcentre.ca/files/research_reports/08-03GuidanceDocument.pdf
- Ihaka, R. and Gentleman, R., 1996. "R: A language for data analysis and graphics," Journal of Computational and Graphical Statistics 5, 299-314.
- Instituto Geológico y Minero de España, 1988. Las Aguas Subterráneas en la Comunidad Valenciana. Uso, calidad y perspectivas de utilización. Colección Informe. ISBN: 84-7840-004-4
- Johnson, N.L., Kotz, S., and Balakrishnan, N., 1994. Continuous Univariate Distributions, Vol. 1. Second Edition. John Wiley, New York.
- Kaplan, E.L. & Meier, P., 1958. Non-parametric estimation from incomplete observations. *J. Amer. Stat. Assn*, 53, 457-481.
- Klein, J.P., and Moeschberger, M.L., 2003. Survival Analysis: Techniques for Censored and Truncated Data, second ed., Springer Verlag, New York.
- Land, C. E., 1971. *Confidence Intervals for Linear Functions of the Normal Mean and Variance*. Annals of Mathematical Statistics, 42, pp. 1187-1205.
- Land, C. E., 1975. *Tables of Confidence Limits for Linear Functions of the Normal Mean and Variance*. In Selected Tables in Mathematical Statistics, Vol. III, American Mathematical Society, Providence, R.I., pp. 385-419.
- Lerner DN (2002) Identifying and quantifying urban recharge: a review. *Hydrogeol J* 10(1):143–152
- Letón Molina, E., 2002. Generalización y equivalencias de tests no paramétricos para el análisis de datos censurados. Memoria para optar al grado de doctor presentada por Emilio Letón Molina bajo la supervisión del Doctor María Pilar Zuluaga Arias. ISBN: 84-669-1800-0.
- Lopaka Lee (2010). NADA: Nondetects And Data Analysis for environmental data. R package version 1.5-3. <http://CRAN.R-project.org/package=NADA>
- López Geta, J.A., Fornés Azcoiti, J.M., Ramos González, G., Villarroya Gil, F., 2009. Las Aguas Subterráneas: un recurso natural del subsuelo. auts.- 4ª ed.- Madrid: Instituto Geológico y Minero de España. 4º Edición 2009. ISBN: 978-84-7840-806-1. NIPO: 474-09-013-4
- López Geta, J.A., del Barrio Beato, V. y Vega Martín, L. 2005. Explotación de las aguas subterráneas en el Duero: Los retos de la cuenca. Instituto Geológico y Minero de España. /C Ríos Rosas, 23 Madrid –28003.
- Luque Espinar, J.A.; Grima Olmedo, J., Jiménez Sánchez, J., Jiménez Escamilla, M.C., Hueso Quesada, L.M., Gómez-Escalonilla Sánchez, M.D., 2010. Encomienda de gestión para la realización de trabajos científico-técnicos de apoyo a la sostenibilidad y protección de las Aguas Subterráneas (Ministerio de Medio Ambiente y Medio Rural y Marino – Instituto Geológico y Minero de España). Actividad 10: Apoyo a la implementación de la Directiva de protección de aguas subterráneas. Determinación de tendencias y de puntos de partida para la inversión de tendencias. Ocho volúmenes, 1180 pp. Informe interno.

- Luque, J.A; Grima, J; Varela, M; Sánchez, I., 2010. Searching for trends in data from groundwater monitoring networks through exploratory data analysis. European Groundwater Conference, Madrid, 20-21 may, poster session and proceedings.
- McNichols R.J. and Davis C.B., 1988. Statistical issues and problems in ground water detection monitoring at hazardous waste facilities. *Ground Water Monitoring Review*, 8, 135-150.
- Manly, B.F.J., 1997. *Randomization, Bootstrap, and Monte Carlo Methods in Biology*. Second Edition. Chapman Hall, London.
- Meier, P., Karrison, T., Chappell, R. and Xie, H., 2004. The price of Kaplan-Meier. *Journal of the American Statistical Association* 99, 890-897
- MIMAM, 2005. Manual para la identificación de las presiones y análisis del impacto en aguas superficiales. Dirección General del Agua. 14 de febrero de 2005. Madrid.
- Ministerio de Medio Ambiente y Medio Rural y Marino, 2007. Estudio General sobre la Demarcación Hidrográfica del Júcar.
http://www.phjucar.com/docs/cons_publica/iniciales/Est_Gral_CHJ.pdf
- Molina, J. L. y García Aróstegui, J.L., 2009. Nuevas metodologías para el análisis integrado de la gestión del agua subterránea. Aplicación al caso de estudio del Altiplano (Murcia, SE España). *Boletín Geológico y Minero*, 120 (1): 17-30 ISSN: 0366-0176
- Morell, I., Hernández, F., 2001. El Agua en Castellón: un reto para el siglo XXI". *Ed. Universidad Jaume I, Athenea*, 538 páginas (2000). ISBN: 84-8021-333-7
- Müller, D., Blum A., Hart A., Hookey J., Kunkel R., Scheidleder A., Tomlin C., Wendland F., 2006 – Final proposal for a methodology to set up groundwater threshold values in Europe, Deliverable D18, BRIDGE project, 63 p, www.wfd-bridge.net.
- Nieto, P., Custodio, E. y Manzano, M., 2005. Baseline Groundwater Quality: a European Approach. *Environmental Science and Policy* 8, 399-409
- Pérez Gago, M., Ballesteros Navarro, B., García Menéndez, O., Grima Olmedo, J. y López Gutiérrez, J., 2001. Contenido en nitratos de las aguas subterráneas en la plana de Valencia. Situación actual, evolución temporal y repercusión sobre La Albufera. Investigación, gestión y recuperación de acuíferos contaminados, Alicante, 119-129. ISBN: 84-607-3216-9.
- Pauwels et al., 2006 – Impact of hydrogeological conditions on pollutant behaviour in groundwater and related ecosystems, Deliverable D10, BRIDGE project, www.wfd-bridge.net.
- Quevauvillier, P, 2005. Groundwater monitoring in the context of EU legislation. Reality and integration needs. *Journal of environmental monitoring*. Vol 7 pp 89-102
- R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rahman MM, Chowdhury UK, Mukherjee SC, Mondal BK, Paul K, Lodh D, Biswas BK, Chanda CR, Basu GK, Saha KC, Roy S, Das R, Palit SK, Quamruzzaman Q, Chakrabort D (2001). Chronic arsenic toxicity in Bangladesh and West Bengal, India – a review and commentary. *Clin Toxicol* 39(7):683–700. doi:10.1081/CLT-100108509
- Real Decreto 140/2003, por el que se establecen los criterios sanitarios de la calidad del agua de consumo humano. BOE nº 45, pp 7228-7245.
http://www.msc.es/ciudadanos/saludAmbLaboral/docs/rd_140_2003.pdf
- Rosner, B., 1975. On the detection of many outliers. *Technometrics*, 17, 221-227.
- Sánchez, A., 2005. Estudio inicial para la identificación y caracterización de las masas de agua subterráneas de las cuencas intercomunitarias. Tomo I. Memorias y Anejos 1,2 y 3. Secretaría General para el territorio y la biodiversidad. Dirección General del Agua. Ministerio de Medio Ambiente y Medio Rural y Marino.
- Santos A, Alonso E, Callejón M, Jiménez JC (2002) Distribution of Zn, Cd, Pb and Cu metals in groundwater of the Guadiamar river basin. *Water Air Soil Pollut* 134(104):273–283
[http://www.phjucar.com/docs/otros_docs/Est_caract_mas_agua_subt\(memoriayfichasjucar\).pdf](http://www.phjucar.com/docs/otros_docs/Est_caract_mas_agua_subt(memoriayfichasjucar).pdf)
- Shumway, R.H. Azari, R.S. and Kayhanian, M., 2002. Statistical Approaches to Estimating Mean Water Quality Concentrations with Detection Limits. *Environ. Sci. Technol.* 2002, 36, 3345-3353
- Singh, A., y Singh, A. K., 2007. ProUCL Version 4.0 Technical Guide. EPA EPA/600/R-07/041.
- Singh, A., Maichle, R., and Lee, S., 2006. On the Computation of a 95% Upper Confidence Limit of the Unknown Population Mean Based Upon Data Sets with Below Detection Limit Observations. EPA/600/R-06/022, March 2006.

- Singh, A. and Singh, A.K., 2003. *Estimation of the Exposure Point Concentration Term (95% UCL) Using Bias-Corrected Accelerated (BCA) Bootstrap Method and Several other methods for Normal, Lognormal, and Gamma Distributions*. Draft EPA Internal Report.
- Singh, A. and Nocerino, J.M., 2002. Robust Estimation of Mean and Variance Using Environmental Data Sets with Below Detection Limit Observations. Vol. 60, pp. 69-86.
- Singh, A., Singh, A. K., Engelhardt, M., and Nocerino, J.M., 2002. On the Computation of the Upper Confidence Limit of the Mean of Contaminant Data Distributions. Under EPA Review.
- Singh, A., Singh, A. K., and Iaci, R. J., 2002. Estimation of the Exposure Point Concentration Term Using a Gamma Distribution. EPA/600/R-02/084.
- Tukey, J. W., 1977. *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.
- United Nations (1987) Report of the world commission on environment and development, A/RES/42/187, 96th plenary meeting, 11 December 1987
- U.S. Environmental Protection Agency, 2009. Office of Resource Conservation and Recovery. Unified guidance for statistical analysis of groundwater monitoring data at RCRA facilities
- US EPA (2008) U.S. EPA's 2008 report on the environment (final report), U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-07/045F (NTIS PB2008-112484)
- U.S. Environmental Protection Agency, 2002. Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites. OSWER 9285.6-10. December 2002.
- US Environmental Protection Agency, 1986. Resource Conservation and Recovery Act (RCRA) Ground-Water Monitoring Technical Enforcement Guidance Document. Washington, DC: USEPA. OSWER-9950.1.
- Wendland, F., Blum, A., Coetsiers, M., Gorova, R., Griffioen, J., Grima, J., Kinsby, K., Kunkel, R., Marandi, A., Melo, T., Panagopoulos, A., Pauwels, H., Ruisi, M., Traversa, P., Vermooten, J.S.A. and Walraevens, K., 2008. European aquifer typology: a practical framework for an overview of major groundwater composition at European scale. *Environmental Geology*. Vol 55, pp 77-85.
- Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12).
- Zektser IS, Everet LG (eds) (2004) Groundwater resources of the world and their use. UNESCO. IHP-VI Series on Groundwater No. 6

APÉNDICES

APÉNDICE A: ANÁLISIS EXPLORATORIO DATOS ESTACIONES CUENCA DEL JÚCAR

Figura 53: Niveles Básicos en la Cuenca del Júcar (1)

Figura 54: Niveles Básicos en la Cuenca del Júcar (2)

Figura 55: Niveles Básicos en la Cuenca del Júcar (3)

Figura 56: Niveles Básicos en la Cuenca del Júcar (4)

Figura 57: Niveles Básicos en la Cuenca del Júcar (5)

Figura 58: Niveles Básicos en la Cuenca del Júcar (6)

APÉNDICE B: ANÁLISIS DE VALORES CENSURADOS EN ESTACIONES SELECCIONADAS DE LA CUENCA DEL JÚCAR

	% cen	n>Id	Sust 0	Sust 0.5	Sust Id	KM	ROS	MLE
EN1	0	8	38.11 - 99.76					
	12.5	7	32.34-99.62	35.3-99.62	38.95-99.93	40.61-100.09	37.32-100.88	37.46-123.25
	25	6	26.46-99.3	33.96-99.77	39.99-100.64	42.09-106.97	34.6-104.44	34.15-138.07
	50	4	13.74-93.29	32.46-97.52	45.59-102.19	58.67-138.05	-----	23.7-207.72
EN2	0	8	57.9-137.95					
	12.5	7	54.04-137.81	58.54-137.37	62.27-137.56	66.39-143.14	59.98-140.79	51.31-238.37
	25	6	43.14-136.37	56.27-136.75	67.63-138.48	68.78-150.48	53.7-154.15	45.87-275.81
	50	5	34.65-132.87	54.21-135.79	73.53-139.8	75.89-165.91	53.45-163.23	38.89-338.07
EN3	0	8	46.14-170.18					
	12.5	7	44.75-170.21	45.55-169.48	45.43-170.87	47.36-178.44	42.08-184.77	40.19-500.58
	25	6	41.62-169.37	44.05-170.54	49.73-171.14	51.72-192.38	44.84-189.82	35.58-786.52
	50	4	23.98-168	45.85-170.76	67.72-173.59	-----	44.84-189.82	35.58-3883.96
EN4	0	8	14.99-27.16					
	12.5	7	11.85-26.83	13.71-27.02	15.24-26.96	15.37-27.44	13.82-26.96	14.53-28.94
	25	6	9.25-26.41	13.08-26.68	16.49-27.44	17.65-28.67	14.56-27.46	13.61-30.31
	50	4	3.75-24.42	12.5-25.96	20.25-28.46	-----	-----	10.54-32.51
EN5	0	10	10.15-23.23					
	10	9	9.43-23.33	9.88-23.36	10.3-23.22	10.35-23.47	9.61-23.38	9.44-33.31
	20	8	8.74-22.94	9.64-23.16	10.56-23.49	10.87-24.2	9.27-24.04	8.9-37.55
	50	6	6.58-22.54	12.07-23.58	17.15-25.13	17.15-27.21	11.89-25.71	7.21-53.9
EN6	0	7	8.78-53.68					
	14.3	6	8.54-53.44	8.69-53.58	8.47-53.06	8.51-56.83	8.96-57.11	3.83-2004.04
	28.6	5	8.08-53.44	8.31-53.68	8.66-53.68	8.95-64.21	8.61-62.89	2.91-10862.78
	42.9	4	7.59-53.47	10.13-53.83	11.92-54.18	-----	-----	1.78-191465.5
EN7	0	8	47.49-90.8					
	12.5	7	36.36-89.24	42.15-89.45	47.65-90.08	48.67-89.95	46.46-91.61	46.56-88.99
	25	6	26.49-85.45	38.35-89.16	49.29-91.02	49.66-91.17	43.02-89.71	43.59-91.21
	50	4	11.8-78.48	33.99-86.23	54.23-91.55	-----	-----	34.2-97.81
EN8	0	8	12.64-47.99					
	12.5	7	10.17-47.48	11.52-47.64	12.7-48.09	12.74-47.98	12.28-47.44	12.78-43.86
	25	6	7.62-46.85	10.66-47.5	13.04-47.75	13.29-48.81	11.86-47.75	11.58-48.45
	50	4	3.36-44.76	10.05-46.85	16.07-48.85	-----	-----	8-67.96

Tabla 14: Intervalos de confianza de la media en estaciones seleccionadas de la Cuenca del Júcar calculados por diferentes métodos y para diversos grados de censura (1)

	% cen	n>ld	Sust 0	Sust 0.5	Sust ld	KM	ROS	MLE
EN9	0	20	52.94-60.63					
	15	17	39.95-59.1	47.43-59.79	54.59-60.92	54.77-61.01	53.04-60.75	51.73-61.31
	20	16	36.19-57.78	46.03-59.12	54.83-61.01	54.98-61.22	53.07-60.83	51.22-61.23
	50	10	18.4-45.71	38.5-54.11	58.5-62.63	58.6-63.76	52.63-62.41	45.95-59.12
EN10	0	22	31.61-52.66					
	13.2	19	30.51-52.15	35.47-53.85	32.34-52.79	32.83-54.38	32.68-53.67	29.6-84.34
	22.7	17	28.9-51.45	31.24-52.35	33.25-52.84	39.79-55.72	37.33-55	28.79-95.39
	50	11	17-44.59	31.59-51.55	45.68-58.55	46.5-59.82	34.31-55.86	20.86-115.47
EN11	0	6	237.3-320.38					
	16.6	5	138.53-320.38	193.85-318.35	239.35-321.23	240.01-326.75	220.12-325.15	229.75-323.25
	33.33	4	79.03-311	163.35-311	241.35-320.72	243.41-342.07	-----	216.28-326.57
	50	3	45.83-285.2	145.83-305.2	245.83-321.72	-----	-----	193.93-331.14
EN12	0	27	35.57-61.85					
	11.1	24	33.63-61.48	34.65-61.97	35.89-62.07	35.78-61.52	35.62-62.23	34.65-70.53
	25.9	20	29.96-59.84	33.93-61.35	37.52-63.65	37.87-64.06	35.81-63.03	32.89-78.19
	48.1	14	21.74-54.65	35.39-62	48.62-69.97	49.18-70.79	35.84-62.96	27.17-93.46
EN13	0	14	89.87-162.83					
	14.2	12	78.85-161.65	87.7-162.21	95.94-165.23	97.65-165.87	91.91-165.24	83.26-202.79
	21.4	11	70.42-156.73	84.5-161.79	97.01-165.63	98.23-169.04	87.22-165.47	79.73-208.57
	50	7	39.42-142.2	75.8-156.8	110.8-171.79	-----	74.57-170.76	58.7-252.73
EN15	0	8	42.31-145.32					
	12.5	7	43.49-144.47	42.64-145.69	44.39-145.63	46.67-148.79	42.38-147.96	21.32-61426.08
	25	6	41.2-145.34	53.92-146.82	65.67-148.73	69.42-155.44	58.46-153.98	17.46-10 ⁵
	50	4	20.42-137.04	55.14-144.6	85.14-153.96	-----	-----	6.13-10 ⁷
EN16	0	8	53.84-151.3					
	12.5	7	50.17-149.87	52.46-150.99	54.84-151.99	54.65-162.31	53.52-157.21	51.89-234.3
	25	6	49.59-149.87	55.96-151.06	56.15-151.59	57.82-169.89	47.82-164.45	46.76-278.88
	50	4	21.59-145.3	47.84-150.43	74.09-153.53	-----	-----	30.86-578.16

Tabla 15: Intervalos de confianza de la media en estaciones seleccionadas de la Cuenca del Júcar calculados por diferentes métodos y para diversos grados de censura (2)

	% cen	n>ld	Sust 0	Sust 0.5	Sust Id	KM	ROS	MLE
EN18	0	8			67.8-204.09			
	12.5	7	64.05-203.76	67.4-204.39	71.55-202.85	78.6-206.05	70.66-207.21	64.34-317.41
	25	6	50.32-200.16	65.38-204.2	77.45-205	85.81-212	69.68-206.55	57.27-366.68
	50	4	26.62-191	57.51-197.94	94.7-210.97	-----	-----	37.28-604.15
EN19	0	17			147.41-199.05			
	11.7	15	131.35-198.31	140.87-198.15	149.7-199.33	153.09-203.18	153.29-201.94	139.41-230.06
	23.5	13	109.68-191.56	147.29-197.12	182.27-205.06	182.82-205.57	174.39-205.05	132.42-232.59
	47.1	9	61.8-162.5	128.39-185.21	194.35-210.15	194.86-212.47	169.68-209.11	109.59-222.1
EN20	0	13			78.5-112.35			
	15.3	11	59.41-108.69	70.52-109.88	79.97-112.3	80.59-112.62	76.08-112.04	76.05-111.28
	23.1	10	50.7-106.28	66.69-108.77	81.45-113.29	82.26-114.23	74.89-110.72	73.98-111.85
	46.1	7	30.31-94.08	60.08-104.54	89.85-117	90.15-120.92	66.96-112.11	67.18-111.6
EN21	0	39			103.62-148.48			
	12.8	34	99.06-147.15	102.55-148.9	107.38-150.58	109.57-151.91	107.49-150.66	101.39-189.5
	25.6	29	87.72-141.8	101.74-147.5	114.67-153.86	114.94-155.75	109.5-153.32	96.64-200.94
	48.7	20	83.03-138.65	101.79-147.64	119.26-156.92	135.67-170.6	112.23-164.45	80.34-225.53
EN22	0	8			44.17-128.92			
	12.5	7	40.59-127.51	43.19-129.25	46.55-129.34	46.99-133.25	43.75-130.62	38.89-256.27
	37.5	5	30.51-127.33	39.51-128.8	48.51-130.38	73.18-155.39	50.89-149.36	29.12-508.11
	50	4	19.48-124.21	52.67-130.22	80.73-137.94	-----	-----	21.61-802.7
EN23	0	8			32.48-104.62			
	12.5	7	28.96-103.92	31.73-103.98	33.23-105.12	32.98-105.89	31.36-105.39	32.44-153.73
	25	6	25.31-102.31	30.92-103.42	34.11-105.18	33.73-113.36	27.43-109.41	29.14-183.06
	50	4	13.73-101.18	26.86-103.05	39.98-106.8	-----	-----	19.14-412.82
EN26	0	7			20.43-24			
	14.3	6	12.66-23.97	16.77-24.01	20.46-24	21-24.46	19.45-24.39	20.16-24.41
	28.5	5	9.16-23.9	14.87-23.93	20.73-24.04	20.73-24.6	18.76-24.49	19.58-24.51
	42.8	4	3.59-21.21	13.87-22.94	24.04-24.57	-----	-----	18.67-24.59

Tabla 16: Intervalos de confianza de la media en estaciones seleccionadas de la Cuenca del Júcar calculados por diferentes métodos y para diversos grados de censura (3)

	% cen	n>ld	Sust 0	Sust 0.5	Sust ld	KM	ROS	MLE
EN27	0	5	50.08-68.96					
	20	4	24.46-68.96	39.46-68.96	51-68.96	51.16-72.5	46.3-72.24	48.27-70.9
	40	3	10.84-68.96	31.64-68.96	52.44-68.96	54.2-73.4	-----	44.1-72.2
	60	2	0-58.42	29-64.22	58-70.02	-----	-----	36.35-73.58
EN28	0	45	22.56-28.7					
	11.1	40	21.36-28.38	22.35-28.64	23.28-29.08	23.35-28.95	23.4-29.33	21.96-33.61
	26.6	33	18.44-26.97	22-28.42	25.62-30.03	25.83-30.39	25.24-30.44	20.85-34.79
	48.8	23	12.62-22.66	21.26-27.55	30.04-32.57	30.15-32.85	27.02-32.27	17.36-34.07
EN29	0	8	6.66-37.83					
	12.5	7	6.53-38.83	6.62-38.48	6.82-38.51	6.97-39.18	6.41-39.14	6.71-90.96
	25	6	6.05-38.31	6.8-38.48	7.53-38.85	8.13-39.95	6.56-39.53	5.92-143.68
	50	4	3.75-37.68	7.5-38.27	11.28-39.52	-----	-----	3.37-722.12
EN31	0	18	58.75-84.98					
	11.1	16	51.61-82.78	57.44-84.22	62.67-86	63.11-85.83	60.67-84.17	57.08-88.54
	33.3	12	34.72-75.56	51.11-81.11	65.28-87.5	65-87.5	55.14-83.08	51.23-88.83
	50	9	23.89-68.61	46.39-78.06	68.33-88.89	68.61-90.83	48.43-83.49	44.26-88.53
EN32	0	14	13.94-23.83					
	14.2	12	12.76-23.76	13.29-23.69	14-23.87	13.6-24.74	13.7-24.72	12.39-36.16
	21.4	11	12-23.56	13.23-23.8	14.42-23.99	14.39-25.51	13.8-25.26	11.98-39.21
	50	7	6-20.57	15.43-23.31	24.38-26.57	24.43-27.43	19.63-26.39	8.43-46.22
EN33	0	13	57.54-64.77					
	15.3	11	38.69-63.85	49.54-64	58.92-65.15	59.08-65.08	57.47-64.7	56.66-64.92
	23	10	33.31-62.46	47-63.46	59.69-65.31	59.69-65.46	57.15-64.85	55.98-64.83
	38.4	8	23.38-56.16	42-60.77	60.62-65.85	60.77-66.15	54.4-64.87	54.07-64.26
EN34	0	22	72.94-143.62					
	13.6	19	70.32-142.13	71.55-143.88	73.04-144.4	75.46-146.32	74.35-143.4	70.37-198.52
	27.2	16	62.87-138.93	72.31-143.32	82.41-148.41	84.51-150.69	75.98-146.87	66.13-231.94
	50	11	43.13-128.2	73.78-143.31	103.33-160.23	103.62-162.35	71.32-147.21	51.66-319.29

Tabla 17: Intervalos de confianza de la media en estaciones seleccionadas de la Cuenca del Júcar calculados por diferentes métodos y para diversos grados de censura (4)

APÉNDICE C: CÓDIGO R GENERADO PARA ANÁLISIS DE DATOS

Totalidad datos históricos

Establecer directorio de trabajo

Crear el directorio y copiar archivos a "C:/SistemaTomaDecisionesMAS"

Archivos ejemplo: JUIG002950.txt y JUIG002950nb.txt

Formato archivo: separador decimal punto y fecha dd-mm-aaaa

```
setwd("C:\\SistemaTomaDecisionesMAS")
```

Lectura archivo a analizar y representación gráfica

```
JUIG002950<-read.table("JUIG002950.txt",header=T)
EN<-data.frame(strptime(as.character(JUIG002950$Fecha), "%d-%m-%Y"),
JUIG002950$ N)
par(layout(matrix(c(1,2,3,4),2,2), c(1.5,1),c(1.5,1), respect = FALSE), bg ="gray80",
col.axis="blue")
plot(EN,pch=22,type="b", col="darkgreen",bg="yellow",main="Concentración nitratos
JUIG002950",xlab="tiempo",ylab=" mg/L N")
lines(lowess(EN, f=4),col = 3) # Añade el suavizado
hist(JUIG002950$N, probability=TRUE, breaks=20,main="histograma N",
xlab="concentración N", ylab="Frecuencia", axes=T,col="lightgreen")
lines(density(JUIG002950$ N), col="red", lwd=1)
boxplot(JUIG002950$ N, boxwex=0.3,main="boxplot N", horizontal =
F,axes=T,col="lightgreen")
points(mean(JUIG002950$N), pch=19,col="red")
abline(h=median(JUIG002950$N))
boxplot(log(JUIG002950$ N), boxwex=0.5, main="boxplot logN", horizontal =
TRUE,axes=T,col="lightgreen")
```

Grabación en directorio de trabajo de las representaciones gráficas

generadas en ficheros pdf

```
pdf("evolN.pdf", width=10,height=7)
par(bg="gray80", col.axis="blue")
plot(EN,pch=22,type="b", col="darkgreen",bg="yellow",main="Evolución concentración
nitratos estación JUIG002950",xlab="tiempo",ylab=" mg/L N")
dev.off()
pdf("boxplotN.pdf", width=6,height=3)
boxplot(JUIG002950$ N, main="boxplot N", horizontal =
TRUE,axes=T,col="lightgreen")
dev.off()
pdf("histN.pdf")
hist(JUIG002950$ N, main="histograma N", xlab="concentración N",
ylab="Frecuencia", axes=T,col="lightgreen")
dev.off()
pdf("suavizadoevolN.pdf",width=10,height=7)
par(bg="cornsilk", col.axis="blue")
plot(EN,pch=22,type="b",col="darkgreen",bg="yellow",main="Evolución concentración
nitratos estación JUIG002950 y suavizado",xlab="tiempo",ylab=" mg/L N")
lines(lowess(EN, f=4),col = 3) # Añade el suavizado
dev.off()
```

Representación gráfica alta resolución

```
install.packages("Cairo")
library(Cairo)
Cairo(600, 600, file="histNTAIG000591.png", type="png", bg="white")
hist(TAIG000591$ NTAIG000591, main="histograma NTAIG000591",
xlab="concentración N", ylab="Frecuencia", axes=T,col="#ff000018",pch=19,cex=2)
# semi-transparent red
dev.off()
```

Nivel Básico

Lectura archivo a analizar

```
JUIG002950nb<-read.table("JUIG002950nb.txt",header=T)
ENnb<-data.frame(strptime(as.character(JUIG002950nb$Fecha), "%d-%m-%Y"),
JUIG002950nb$N)
par(layout(matrix(c(1,2,3,4),2,2), c(1.5,1),c(1.5,1), respect = FALSE), bg = "gray80",
col.axis="blue")
plot(ENnb,pch=22,type="p", col="darkgreen",bg="yellow",main="Evolución nivel básico
nitratos estación JUIG002950",ylim=c(min(range(JUIG002950nb$N)/2),
max(range(JUIG002950nb$N)*2)), xlab="tiempo",ylab=" mg/L N")
hist(JUIG002950nb$N, probability=TRUE, breaks=20,main="histograma
JUIG002950nb", xlab="concentración N JUIG002950nb", ylab="Frecuencia",
axes=T,col="lightgreen")
lines(density(JUIG002950nb$N), col="red", lwd=1)
boxplot(JUIG002950nb$N, boxwex=0.3,main="boxplot Nnb", horizontal =
F,axes=T,col="lightgreen")
points(mean(JUIG002950nb$N), pch=19,col="red")
abline(h=median(JUIG002950nb$N))
boxplot(log(JUIG002950nb$N), boxwex=0.5, main="boxplot logNnb", horizontal =
TRUE,axes=T,col="lightgreen")
```

Representación gráfica nivel básico

```
par(bg="cornsilk", col.axis="blue")
plot(ENnb,pch=22,type="b", col="darkgreen",bg="yellow",main="Evolución nivel básico
nitratos estación TAIG000591",xlab="tiempo",ylab=" mg/L N")
hist(JUIG002950nb$N, main="histograma JUIG002950nb",breaks=25,
xlab="concentración N TAIG000591nb ", ylab="Frecuencia", axes=T,col="lightgreen")
```

Detección outliers

```
install.packages("outliers")
library(outliers)
dixon.test(JUIG002950nb$N)
grubbs.test(JUIG002950nb$N)
```

Autocorrelación

```
Install.packages("lawstat")
library(lawstat)
bartels.test(JUIG002950nb$N, alternative = "positive.correlated")
```

Verificación normalidad y ajuste a distribuciones gamma y lognormal

```
install.packages("MASS")
library(MASS)
shapiro.test(JUIG002950nb$N)
fitdistr(JUIG002950nb$N,"gamma")
fitdistr(JUIG002950nb$N,"lognormal")
```

Detección tendencias**# Regresión lineal**

```
install.packages("Epi")
library(Epi)
library(RColorBrewer)
JUIG002950nbaños<-cal.yr(JUIG002950nb$Fecha,format="%d-%m-%Y")
lmJUIG002950nb<-lm(JUIG002950nb$N~JUIG002950nbaños)
par(bg= "#ECB176FF")
plot(JUIG002950nb$N~ JUIG002950nbaños,col="blue",col.axis="brown",font.axis=2,
font.main=2, fg="red",main="Regresión lineal sobre conjunto datos
JUIG002950nb",xlab="",ylab=" concentración N")
abline(lmJUIG002950nb)
summary(lmJUIG002950nb)
```

Transformación Box-cox nivel básico

```
par(bg= "whitesmoke", col.axis="blue")
library(MASS)
boxcox(lmJUIG002950nb)
boxcox(lmJUIG002950nb, lambda = seq(-2, 4, 1/10)) # sustituir y por  $(y^{**\lambda}-1)/\lambda$  salvo
si  $\lambda=0$ 
pdf("bclmJUIG002950nb.pdf", width=10,height=7)
par(bg= "whitesmoke", col.axis="blue")
library(MASS)
boxcox(lm lmJUIG002950nb)
dev.off()
```

Grabación en directorio de trabajo de la representación gráfica del ajuste lineal

```
pdf("lmJUIG002950nb.pdf", width=10,height=7)
par(bg= "#ECB176FF")
plot(lmJUIG002950nb$N~ JUIG002950años,col="blue",col.axis="brown",font.axis=2,
font.main=2, fg="red",main="Regresión lineal sobre conjunto datos
JUIG002950nb",xlab="",ylab=" concentración N")
abline(lmJUIG002950nb)
dev.off()
```

Distribución no discernible

```
install.packages("Kendall")
library(Kendall)
MannKendall(JUIG0950nb$N)
```


Representación gráfica residuos ajuste modelo lineal

```
par(bg="blue", col.lab="white", col.main="white", fg="white")
plot(resid(lmJUIG002950nb), col.axis="white", xlab="", ylab="", main="Residuos del
ajuste lineal", col="white")
abline(h=0)
```

Grabación en fichero representación gráfica residuos ajuste modelo lineal

```
pdf("Resid lmJUIG002950nb.pdf", width=10, height=7)
par(bg="blue", col.lab="white", col.main="white", fg="white")
plot(resid(lmJUIG002950nb), col.axis="white", xlab="", ylab="", main="Residuos del
ajuste lineal estación JUIG002950nb", col="white")
dev.off()
```

Probability plot e intervalo confianza mediana

```
install.packages("e1071")
library(e1071)
p <- probplot(JUIG002950nb$N, line=FALSE)
lines(p, col="green", lty=2, lwd=2)
lines(p, h=.5, col="red", lwd=3, bend=TRUE)
lines(p, h=c(.05, .95), col="red", lwd=2, lty=3, bend=TRUE)
```

Representación valores censurados

```
par(bg="cornsilk", col.axis="blue")
JUIG002950nbcen<-read.table("JUIG002950nbcen.txt", header=T)
mean.sav <- function(JUIG002950nbcen){
  mean(ros(JUIG002950nbcen$obs, JUIG002950nbcen$cen))
}
res = censboot(JUIG002950nbcen, mean.sav, R=1000)
quantile(res$t, .025)
quantile(res$t, .975)
cenboxplot(JUIG002950nbcen$N, JUIG002950nbcen$Ncen, main="Boxplot censurado
estación JUIG002950", ylab="mg/L N", boxwex=0.4)
text(0.6, 175, "límite de detección")
JUIG002950km<-cenfit(JUIG002950nbcen$N, JUIG002950nbcen$Ncen)
JUIG002950ros<-cenros(JUIG002950nbcen$N, JUIG002950nbcen$Ncen)
JUIG002950mle<-cenmle(JUIG002950nbcen$N, JUIG002950nbcen$Ncen)
plot(ros(JUIG002950nbcen$N, JUIG002950nbcen$Ncen))
mean(ros(JUIG002950nbcen$N, JUIG002950nbcen$Ncen))
sd(ros(JUIG002950nbcen$N, JUIG002950nbcen$Ncen))

meaNnbcen<- function(JUIG002950nbcen){
  mean(ros(JUIG002950nb$N, JUIG002950nb$Ncen))
}
res = censboot(JUIG002950nbcen, meaNnbcen, R=300)
```

Intervalos confianza media basados en la t de Student

```
JUIG002950nb<-read.table("JUIG002950nb.txt", header=T)
x<-JUIG002950nb$N
mean(x)+qt(0.95, (length(x)-1))*sd(x)/sqrt(length(x))
```

Determinación intervalos confianza media bootstrap y representación gráfica

```
install.packages("plotrix")
library(plotrix)
JUIG002950nb<-read.table("JUIG002950nb.txt",header=T)
x<- JUIG002950nb$N
par(bg ="gray80", col.axis="blue")
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias bootstrap", xlab = "concentración de nitratos", ylab ="densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.95),col="red", lty=4,lw=3)
abline(v= quantile(boot,0.05) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.05)-6), 0.0012, pch=62, cex=0.75, round(quantile(boot,0.05),1))
legend((quantile(boot,0.5)-6), 0.0012, pch=181, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.95)-6), 0.0012, pch=60, cex=0.75, round(quantile(boot,0.95),1))
quantile(boot,0.95)
quantile(boot,0.05)
```

Grabación en fichero de la representación gráfica intervalo confianza bootstrap

```
pdf("meanbNTAIG000591.pdf", width=10,height=7)
hist(boot,freq=FALSE, col = 'grey80',breaks=15, main= "Histograma de medias bootstrap", xlab = "concentración de nitratos", ylab ="densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
dev.off()
```

Determinación intervalos confianza mediana bootstrap y representación gráfica

```
x<- TAI000591$ NTAIG000591
perc50<-function(x) quantile(x,0.50)
perc50(x)
boot <-numeric(1000)
for (i in 1:10000) boot[i] <- perc50(sample(x,replace=T))
hist(boot, probability=TRUE)
lines(density(boot), col="red", lwd=3)
quantile(boot,0.95)
quantile(boot,0.05)
```

Determinación intervalo predicción 95% m valores futuros

```
m<-4          ## Número de muestras futuras
α<-0.05       ## Falsos positivos
PL<-mean(x)+qt(1-( α /m),length(x)-1)*sd(x)* (1+1/length(x))**0.5
```

```
quantile(TAIG000591$ NTAIG000591, type=6, 0.9)
```

Cálculo intervalos confianza pendiente Theil-Senn

```
# R script for Theil-Sen Confidence band
# Compute bootstrapped confidence band around Theil-Sen trend line
# user inputs: list of x-values, list of y-values, desired confidence level
# Note: replace numbers in parentheses below with specific x and y values
# corresponding to data-specific ordered pairs
# x-values should be numeric values representing sampling dates or events
# y-values should be concentration values corresponding to these dates or events
# Script produces a plot of the Theil-Sen trend line, the confidence band around the
trend,
# and an overlay of the actual data values
x= c(TAIG000591$Fecha)
y= c(TAIG000591$NTAIG000591)
conf = .90
elimna= function(m){
#
# remove any rows of data having missing values
m= as.matrix(m)
ikeep= c(1:nrow(m))
for(i in 1:nrow(m)) if (sum(is.na(m[i,]))>=1)) ikeep[i]= 0
elimna= m[ikeep[ikeep>=1],]
elimna
}
theilsen2= function(x,y){
#
# Compute the Theil-Sen regression estimator
# Do not compute residuals in this version
# Assumes missing pairs already removed
#
ord= order(x)
xs= x[ord]
ys= y[ord]
vec1= outer(ys,ys,"-")
vec2= outer(xs,xs,"-")
v1= vec1[vec2>0]
v2= vec2[vec2>0]
slope= median(v1/v2)
coef= 0
coef[1]= median(y)-slope*median(x)
coef[2]= slope
list(coef=coef)
}
nb= 1000
temp= matrix(c(x,y),ncol=2)
temp= elimna(temp) #remove any pairs with missing values
x= temp[,1]
y= temp[,2]
n= length(x)
ord= order(x)
cut= min(x) + (0:100)*(max(x)-min(x))/100 #compute 101 cut pts
```

```
t0= theilsen2(x,y) #compute trend line on original data
tmp= matrix(nrow=nb,ncol=101)
for (i in 1:nb) {
  idx= sample(ord,n,rep=T)
  xboot= x[idx]
  yboot= y[idx]
  tboot= theilsen2(xboot,yboot)
  tmp[i,]= tboot$coef[1] + cut*tboot$coef[2]
}
lb= 0; ub= 0
for (i in 1:101){
  lb[i]= quantile(tmp[,i],c((1-conf)/2))
  ub[i]= quantile(tmp[,i],c((1+conf)/2))
}
tband= list(xcut=cut,lo=lb,hi=ub,ths0=t0)
yt= tband$ths0$coef[1] + tband$ths0$coef[2]*tband$xcut
plot(yt~tband$xcut,type='l',xlim=range(x),ylim=c(min(tband$lo),max(tband$hi)),main='Intervalos de confianza de la pendiente por Theil-Senn. Estación TAIG000591',xlab='Date',ylab='Conc')
points(x,y,pch=16)
lines(tband$hi~tband$xcut,type='l',lty=2)
lines(tband$lo~tband$xcut,type='l',lty=2)
t0 # valor de los coeficientes
```

Representación gráfica intervalos confianza pendiente Theil-Senn

```
pdf("TSNTAIG000591.pdf", width=10,height=7)
plot(yt~tband$xcut,type='l',xlim=range(x),ylim=c(min(tband$lo),max(tband$hi)),main='Intervalos de confianza de la pendiente por Theil-Senn. Estación TAIG000591',xlab='Date',ylab='Conc')
points(x,y,pch=16)
lines(tband$hi~tband$xcut,type='l',lty=2)
lines(tband$lo~tband$xcut,type='l',lty=2)
dev.off()
```

Intervalos tolerancia

```
install.packages("tolerance")
library(tolerance)
x<- TAIG000591$NTAIG000591
out <- nptol.int(x = x, alpha = 0.10, P = 0.95, side = 1,method = "WILKS", upper = NULL, lower = NULL)
plottol(out, x = x, plot.type = "both", side = "upper", x.lab = "Data", lty = "dashed", col = "gray")
JUIG002950nb<-read.table("JUIG002950nb.txt",header=T)
x<- JUIG002950nb$N
out <- nptol.int(x = x, alpha = 0.10, P = 0.95, side = 1,method = "WILKS", upper = NULL, lower = NULL)
plottol(out, x = x, plot.type = "both", side = "upper", x.lab = "Data", lty = "dashed", col = "gray")
```

Datos Red de Control **## Lectura archivo a analizar**

```
JUIG002950rc<-read.table("JUIG002950rc.txt",header=T)
```

```
plot(JUIG002950rc$N)
MannKendall(JUIG002950rc$N)
mean(JUIG002950rc$N)
```

Análisis Cuenca Júcar datos censurados

```
library (NADA)
library (boot)
library (xlsReadWrite)
library (ggplot2)
library (zoo)
```

Plana de Sagunto

```
JUIG002683<-read.table("JUIG002683.txt",header=T)
EN17<-data.frame(strptime(as.character(JUIG002683$Fecha),"%d-%m-%Y"),
JUIG002683$NJUIG002683)
JUIG002924<-read.table("JUIG002924.txt",header=T)
EN18<-data.frame(strptime(as.character(JUIG002924$Fecha),"%d-%m-%Y"),
JUIG002924$NJUIG002924)
JUIG002950<-read.table("JUIG002950.txt",header=T)
EN19<-data.frame(strptime(as.character(JUIG002950$Fecha),"%d-%m-%Y"),
JUIG002950$NJUIG002950)
JUIG003594<-read.table("JUIG003594.txt",header=T)
EN20<-data.frame(strptime(as.character(JUIG003594$Fecha),"%d-%m-%Y"),
JUIG003594$NJUIG003594)
JUIG003866<-read.table("JUIG003866.txt",header=T, dec=",")
EN21<-data.frame(strptime(as.character(JUIG003866$Fecha),"%d-%m-%Y"),
JUIG003866$NJUIG003866)
JUIG003940<-read.table("JUIG003940.txt",header=T)
EN22<-data.frame(strptime(as.character(JUIG003940$Fecha),"%d-%m-%Y"),
JUIG003940$NJUIG003940)
```

Intervalos confianza mle y K-M

EN17

```
# table(EN17[,2]<20)["TRUE"]/length(EN17[,2])
# 0.125
# table(EN17[,2]<25)["TRUE"]/length(EN17[,2])
# 0.25
# table(EN17[,2]<60)["TRUE"]/length(EN17[,2])
# 0.5
```

censura leve

```
JUIG002683nada<- read.table("JUIG002683nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG002683nada[,2], as.logical( JUIG002683nada[,3]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG002683nada[,2], as.logical(
JUIG002683nada[,3]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG002683nada[,2], as.logical( JUIG002683nada[,3]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG002683nada[,2], as.logical(
JUIG002683nada[,3]), conf.int=0.95)),1)[[4]], sep="-")
```

```
# censura media
JUIG002683nada<- read.table("JUIG002683nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG002683nada[,2], as.logical( JUIG002683nada[,4]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG002683nada[,2], as.logical(
JUIG002683nada[,4]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG002683nada[,2], as.logical( JUIG002683nada[,4]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG002683nada[,2], as.logical(
JUIG002683nada[,4]), conf.int=0.95)),1)[[4]], sep="-")

# censura severa
JUIG002683nada<- read.table("JUIG002683nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG002683nada[,2], as.logical( JUIG002683nada[,5]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG002683nada[,2], as.logical(
JUIG002683nada[,5]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG002683nada[,2], as.logical( JUIG002683nada[,5]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG002683nada[,2], as.logical(
JUIG002683nada[,5]), conf.int=0.95)),1)[[4]], sep="-")

round(mean(JUIG002683nada[,2]),1)
round(sd(JUIG002683nada[,2]),1)

EN17.20.0<-replace(EN17[,2], EN17[,2]<20, 0)
EN17.20.5<-replace(EN17[,2], EN17[,2]<20, 10)
EN17.20.ld<-replace(EN17[,2], EN17[,2]<20, 20)
EN17.25.0<-replace(EN17[,2], EN17[,2]<25, 0)
EN17.25.5<-replace(EN17[,2], EN17[,2]<25, 12.5)
EN17.25.ld<-replace(EN17[,2], EN17[,2]<25, 25)
EN17.60.0<-replace(EN17[,2], EN17[,2]<60, 0)
EN17.60.5<-replace(EN17[,2], EN17[,2]<60, 30)
EN17.60.ld<-replace(EN17[,2], EN17[,2]<60, 60)

length(EN17[,2])
length(EN17[,2][EN17[,2]>=20])
length(EN17[,2][EN17[,2]>=25])
length(EN17[,2][EN17[,2]>=60])

x<- EN17[,2]
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias
bootstrap", xlab = "concentración de nitratos", ylab ="densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))
legend((quantile(boot,0.5)-6), 0.0012, pch=171, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")
```

```
x<- EN17.20.0
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias
bootstrap", xlab = "concentración de nitratos", ylab ="densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))
legend((quantile(boot,0.5)-6), 0.0012, pch=171, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")
```

```
x<- EN17.20.5
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias
bootstrap", xlab = "concentración de nitratos", ylab ="densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))
legend((quantile(boot,0.5)-6), 0.0012, pch=171, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")
```

```
x<- EN17.20.ld
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias
bootstrap", xlab = "concentración de nitratos", ylab ="densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))
legend((quantile(boot,0.5)-6), 0.0012, pch=171, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
```



```

paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")

x<- EN17.25.0
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias
bootstrap", xlab = "concentración de nitratos", ylab ="densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))
legend((quantile(boot,0.5)-6), 0.0012, pch=171, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")

x<- EN17.25.5
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias
bootstrap", xlab = "concentración de nitratos", ylab ="densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))
legend((quantile(boot,0.5)-6), 0.0012, pch=171, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")

x<- EN17.25.Id
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias
bootstrap", xlab = "concentración de nitratos", ylab ="densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))

```



```
legend((quantile(boot,0.5)-6), 0.0012, pch=171, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")
```

```
x<- EN17.60.0
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias
bootstrap", xlab = "concentración de nitratos", ylab ="densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))
legend((quantile(boot,0.5)-6), 0.0012, pch=171, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")
```

```
x<- EN17.60.5
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias
bootstrap", xlab = "concentración de nitratos", ylab ="densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))
legend((quantile(boot,0.5)-6), 0.0012, pch=171, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")
```

```
x<- EN17.60.ld
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias
bootstrap", xlab = "concentración de nitratos", ylab ="densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
```

```
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))
legend((quantile(boot,0.5)-6), 0.0012, pch=171, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")
```

```
sustmediasEN17<-round(rbind(c(mean(EN17[,2]), mean(EN17[,2]), mean(EN17[,2])),
c(mean (EN17.20.0), mean(EN17.20.5), mean(EN17.20.ld)), c(mean(EN17.25.0),
mean(EN17.25.5), mean(EN17.25.ld)), c(mean(EN17.60.0), mean(EN17.60.5),
mean(EN17.60.ld))),2)
nadamediasEN17<-round(rbind(censtats(EN17[,2], (EN17[,2]<=0))[,2],
censtats(EN17[,2], (EN17[,2]<=20))[,2], censtats(EN17[,2], (EN17[,2]<=25))[,2],
censtats(EN17[,2], (EN17[,2]<=60))[,2]),2)
```

```
sustsdEN17<-round(rbind(c(sd(EN17[,2]), sd(EN17[,2]), sd(EN17[,2])), c(sd
(EN17.20.0), sd(EN17.20.5), sd(EN17.20.ld)), c(sd(EN17.25.0), sd(EN17.25.5),
sd(EN17.25.ld)), c(sd(EN17.60.0), sd(EN17.60.5), sd(EN17.60.ld))),2)
nadasdEN17<-round(rbind(censtats(EN17[,2], (EN17[,2]<=0))[,3], censtats(EN17[,2],
(EN17[,2]<=20))[,3], censtats(EN17[,2], (EN17[,2]<=25))[,3], censtats(EN17[,2],
(EN17[,2]<=60))[,3]),2)
```

```
valEN17<-cbind(sustmediasEN17, nadamediasEN17)
valEN17sd<-cbind(sustsdEN17, nadasdEN17)
```

```
valEN17<mean(EN17[,2])+0.125*mean(EN17[,2]) & valEN17>mean(EN17[,2])-
0.125*mean(EN17[,2])
valEN17sd<sd(EN17[,2])+0.125*sd(EN17[,2]) & valEN17sd>sd(EN17[,2])-
0.125*sd(EN17[,2])
```

```
write.xls(cbind(serie=EN17[,2], D_serie=EN17[,2]>0, seriep=EN17[,2],
D_seriep=EN17[,2]>20, seriem=EN17[,2], D_seriem=EN17[,2]>25, series=EN17[,2],
D_series=EN17[,2]>60, EN17.20.0, D_EN17.20.0=EN17[,2]>20, EN17.20.5,
D_EN17.20.5=EN17[,2]>20, EN17.20.ld, D_EN17.20.ld=EN17[,2]>20, EN17.25.0,
D_EN17.25.0=EN17[,2]>25, EN17.25.5, D_EN17.25.5=EN17[,2]>25, EN17.25.ld,
D_EN17.25.ld=EN17[,2]>25, EN17.60.0, D_EN17.60.0=EN17[,2]>60, EN17.60.5,
D_EN17.60.5=EN17[,2]>60, EN17.60.ld, D_EN17.60.ld=EN17[,2]>60), "EN17.xls")
```

```
# x<- EN17[,2]
# x<- EN17.20.0
# x<- EN17.20.5
# x<- EN17.20.ld
# x<- EN17.25.0
# x<- EN17.25.5
# x<- EN17.25.ld
# x<- EN17.60.0
# x<- EN17.60.5
# x<- EN17.60.ld
```

```
x<- EN17.20.0
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
```

```
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias
bootstrap", xlab = "concentración de nitratos", ylab ="densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))
legend((quantile(boot,0.5)-6), 0.0012, pch=171, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),2), round(quantile(boot,0.975),2), sep="-")
```

```
bEN17<-data.frame(obs=EN17[,2],cen=EN17[,2]<20)
mean.sav <- function(bEN17){
mean(ros(bEN17$obs,bEN17$cen))
}
resEN17 = censboot(bEN17,mean.sav,R=2000)
cimeanEN17<-c(quantile(resEN17$t,.025), quantile(resEN17$t,.975))
paste(round(cimeanEN17[1],2), round(cimeanEN17[2],2), sep="-")
```

```
bEN17<-data.frame(obs=EN17[,2],cen=EN17[,2]<25)
mean.sav <- function(bEN17){
mean(ros(bEN17$obs,bEN17$cen))
}
resEN17 = censboot(bEN17,mean.sav,R=2000)
cimeanEN17<-c(quantile(resEN17$t,.025), quantile(resEN17$t,.975))
paste(round(cimeanEN17[1],2), round(cimeanEN17[2],2), sep="-")
```

```
bEN17<-data.frame(obs=EN17[,2],cen=EN17[,2]<60)
mean.sav <- function(bEN17){
mean(ros(bEN17$obs,bEN17$cen))
}
resEN17 = censboot(bEN17,mean.sav,R=2000)
cimeanEN17<-c(quantile(resEN17$t,.025), quantile(resEN17$t,.975))
paste(round(cimeanEN17[1],2), round(cimeanEN17[2],2), sep="-")
```

```
paste(round(mean(cenmle(EN17[,2],EN17[,2]<85)),2)[3],
round(mean(cenmle(EN17[,2],EN17[,2]<85)),2)[4], sep="-")
```

```
bEN17<-data.frame(obs=EN17[,2],cen=EN17[,2]<20)
mean.sav <- function(bEN17){
mean(cenfit(bEN17$obs,bEN17$cen))[1]
}
resEN17 = censboot(bEN17,mean.sav,R=2000)
cimeanEN17<-c(quantile(resEN17$t,.025), quantile(resEN17$t,.975))
paste(round(cimeanEN17[1],2), round(cimeanEN17[2],2), sep="-")
```

EN18

```
# censura leve
JUIG002924nada<- read.table("JUIG002924nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG002924nada[,2], as.logical( JUIG002924nada[,3]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG002924nada[,2], as.logical(
JUIG002924nada[,3]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG002924nada[,2], as.logical( JUIG002924nada[,3]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG002924nada[,2], as.logical(
JUIG002924nada[,3]), conf.int=0.95)),1)[[4]], sep="-")

# censura media
JUIG002924nada<- read.table("JUIG002924nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG002924nada[,2], as.logical( JUIG002924nada[,4]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG002924nada[,2], as.logical(
JUIG002924nada[,4]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG002924nada[,2], as.logical( JUIG002924nada[,4]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG002924nada[,2], as.logical(
JUIG002924nada[,4]), conf.int=0.95)),1)[[4]], sep="-")

# censura severa
JUIG002924nada<- read.table("JUIG002924nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG002924nada[,2], as.logical( JUIG002924nada[,5]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG002924nada[,2], as.logical(
JUIG002924nada[,5]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG002924nada[,2], as.logical( JUIG002924nada[,5]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG002924nada[,2], as.logical(
JUIG002924nada[,5]), conf.int=0.95)),1)[[4]], sep="-")

round(mean(JUIG002924nada[,2]),1)
round(sd(JUIG002924nada[,2]),1)

# table(EN18[,2]<30)["TRUE"]/length(EN18[,2])
# 0.125
# table(EN18[,2]<60)["TRUE"]/length(EN18[,2])
# 0.25
# table(EN18[,2]<85)["TRUE"]/length(EN18[,2])
# 0.5

EN18.30.0<-replace(EN18[,2], EN18[,2]<30, 0)
EN18.30.5<-replace(EN18[,2], EN18[,2]<30, 15)
EN18.30.ld<-replace(EN18[,2], EN18[,2]<30, 30)
EN18.60.0<-replace(EN18[,2], EN18[,2]<60, 0)
EN18.60.5<-replace(EN18[,2], EN18[,2]<60, 30)
EN18.60.ld<-replace(EN18[,2], EN18[,2]<60, 60)
EN18.85.0<-replace(EN18[,2], EN18[,2]<85, 0)
EN18.85.5<-replace(EN18[,2], EN18[,2]<85, 42.5)
EN18.85.ld<-replace(EN18[,2], EN18[,2]<85, 85)

length(EN18[,2])
length(EN18[,2][EN18[,2]>=30])
length(EN18[,2][EN18[,2]>=60])
length(EN18[,2][EN18[,2]>=85])

sustmediasEN18<-round(rbind(c(mean(EN18[,2]), mean(EN18[,2]), mean(EN18[,2])),
c(mean( EN18.30.0), mean(EN18.30.5), mean(EN18.30.ld)), c(mean(EN18.60.0),
```

```

mean(EN18.60.5), mean(EN18.60.ld)), c(mean(EN18.85.0), mean(EN18.85.5),
mean(EN18.85.ld))),2)
nadamediasEN18<-round(rbind(censtats(EN18[,2], (EN18[,2]<=0))[,2],
censtats(EN18[,2], (EN18[,2]<=30))[,2], censtats(EN18[,2], (EN18[,2]<=60))[,2],
censtats(EN18[,2], (EN18[,2]<=85))[,2]),2)

sustsdEN18<-round(rbind(c(sd(EN18[,2]), sd(EN18[,2])), c(sd
(EN18.30.0), sd(EN18.30.5), sd(EN18.30.ld)), c(sd(EN18.60.0), sd(EN18.60.5),
sd(EN18.60.ld)), c(sd(EN18.85.0), sd(EN18.85.5), sd(EN18.85.ld))),2)
nadasdEN18<-round(rbind(censtats(EN18[,2], (EN18[,2]<=0))[,3], censtats(EN18[,2],
(EN18[,2]<=30))[,3], censtats(EN18[,2], (EN18[,2]<=60))[,3], censtats(EN18[,2],
(EN18[,2]<=85))[,3]),2)

valEN18<-cbind(sustmediasEN18, nadamediasEN18)
valEN18sd<-cbind(sustsdEN18, nadasdEN18)

valEN18<mean(EN18[,2])+0.125*mean(EN18[,2]) & valEN18>mean(EN18[,2])-
0.125*mean(EN18[,2])
valEN18sd<sd(EN18[,2])+0.125*sd(EN18[,2]) & valEN18sd>sd(EN18[,2])-
0.125*sd(EN18[,2])

write.xls(cbind(serie=EN18[,2], D_serie=EN18[,2]>0, seriep=EN18[,2],
D_seriep=EN18[,2]>30, seriem=EN18[,2], D_seriem=EN18[,2]>60, series=EN18[,2],
D_series=EN18[,2]>85, EN18.30.0, D_EN18.30.0=EN18[,2]>30, EN18.30.5,
D_EN18.30.5=EN18[,2]>30, EN18.30.ld, D_EN18.30.ld=EN18[,2]>30, EN18.60.0,
D_EN18.60.0=EN18[,2]>60, EN18.60.5, D_EN18.60.5=EN18[,2]>60, EN18.60.ld,
D_EN18.60.ld=EN18[,2]>60, EN18.85.0, D_EN18.85.0=EN18[,2]>85, EN18.85.5,
D_EN18.85.5=EN18[,2]>85, EN18.85.ld, D_EN18.85.ld=EN18[,2]>85), "EN18.xls")

# x<- EN18[,2]
# x<- EN18.30.0
# x<- EN18.30.5
# x<- EN18.30.ld
# x<- EN18.60.0
# x<- EN18.60.5
# x<- EN18.60.ld
# x<- EN18.85.0
# x<- EN18.85.5
# x<- EN18.85.ld

# x<- EN18[,2]
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias
bootstrap", xlab = "concentración de nitratos", ylab = "densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))

```

```
legend((quantile(boot,0.5)-6), 0.0012, pch=181, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")
```

```
bEN18<-data.frame(obs=EN18[,2],cen=EN18[,2]<85)
mean.sav <- function(bEN18){
mean(cenfit(bEN18$obs,bEN18$cen))[1]
}
resEN18 = censboot(bEN18,mean.sav,R=2000)
cimeanEN18<-c(quantile(resEN18$t,.025), quantile(resEN18$t,.975))
paste(round(cimeanEN18[1],2), round(cimeanEN18[2],2), sep="-")
```

```
bEN18<-data.frame(obs=EN18[,2],cen=EN18[,2]<85)
mean.sav <- function(bEN18){
mean(ros(bEN18$obs,bEN18$cen))
}
resEN18 = censboot(bEN18,mean.sav,R=2000)
cimeanEN18<-c(quantile(resEN18$t,.025), quantile(resEN18$t,.975))
paste(round(cimeanEN18[1],2), round(cimeanEN18[2],2), sep="-")
```

```
paste(round(mean(cenmle(EN18[,2],EN18[,2]<85)),2)[3],
round(mean(cenmle(EN18[,2],EN18[,2]<85)),2)[4], sep="-")
```

EN19

censura leve

```
JUIG002950nada<- read.table("JUIG002950nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG002950nada[,2], as.logical( JUIG002950nada[,3]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG002950nada[,2], as.logical(
JUIG002950nada[,3]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG002950nada[,2], as.logical( JUIG002950nada[,3]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG002950nada[,2], as.logical(
JUIG002950nada[,3]), conf.int=0.95)),1)[[4]], sep="-")
```

censura media

```
JUIG002950nada<- read.table("JUIG002950nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG002950nada[,2], as.logical( JUIG002950nada[,4]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG002950nada[,2], as.logical(
JUIG002950nada[,4]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG002950nada[,2], as.logical( JUIG002950nada[,4]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG002950nada[,2], as.logical(
JUIG002950nada[,4]), conf.int=0.95)),1)[[4]], sep="-")
```

censura severa

```
JUIG002950nada<- read.table("JUIG002950nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG002950nada[,2], as.logical( JUIG002950nada[,5]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG002950nada[,2], as.logical(
JUIG002950nada[,5]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG002950nada[,2], as.logical( JUIG002950nada[,5]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG002950nada[,2], as.logical(
JUIG002950nada[,5]), conf.int=0.95)),1)[[4]], sep="-")
```



```

round(mean(JUIG002950nada[,2]),1)
round(sd(JUIG002950nada[,2]),1)

# table(EN19[,2]<70)["TRUE"]/length(EN19[,2])
# 0.117
# table(EN19[,2]<170)["TRUE"]/length(EN19[,2])
# 0.235
# table(EN19[,2]<190)["TRUE"]/length(EN19[,2])
# 0.471

EN19.70.0<-replace(EN19[,2], EN19[,2]<70, 0)
EN19.70.5<-replace(EN19[,2], EN19[,2]<70, 35)
EN19.70.Id<-replace(EN19[,2], EN19[,2]<70, 70)
EN19.170.0<-replace(EN19[,2], EN19[,2]<170, 0)
EN19.170.5<-replace(EN19[,2], EN19[,2]<170, 85)
EN19.170.Id<-replace(EN19[,2], EN19[,2]<170, 170)
EN19.190.0<-replace(EN19[,2], EN19[,2]<190, 0)
EN19.190.5<-replace(EN19[,2], EN19[,2]<190, 95)
EN19.190.Id<-replace(EN19[,2], EN19[,2]<190, 190)

length(EN19[,2])
length(EN19[,2][EN19[,2]>=70])
length(EN19[,2][EN19[,2]>=170])
length(EN19[,2][EN19[,2]>=190])

sustmediasEN19<-round(rbind(c(mean(EN19[,2]), mean(EN19[,2]), mean(EN19[,2])),
c(mean(EN19.70.0), mean(EN19.70.5), mean(EN19.70.Id)), c(mean(EN19.170.0),
mean(EN19.170.5), mean(EN19.170.Id)), c(mean(EN19.190.0), mean(EN19.190.5),
mean(EN19.190.Id))),2)
nadamediasEN19<-round(rbind(censtats(EN19[,2], (EN19[,2]<=0))[,2],
censtats(EN19[,2], (EN19[,2]<=70))[,2], censtats(EN19[,2], (EN19[,2]<=170))[,2],
censtats(EN19[,2], (EN19[,2]<=190))[,2]),2)

sustsdEN19<-round(rbind(c(sd(EN19[,2]), sd(EN19[,2]), sd(EN19[,2])), c(sd
(EN19.70.0), sd(EN19.70.5), sd(EN19.70.Id)), c(sd(EN19.170.0), sd(EN19.170.5),
sd(EN19.170.Id)), c(sd(EN19.190.0), sd(EN19.190.5), sd(EN19.190.Id))),2)
nadasdEN19<-round(rbind(censtats(EN19[,2], (EN19[,2]<=0))[,3], censtats(EN19[,2],
(EN19[,2]<=70))[,3], censtats(EN19[,2], (EN19[,2]<=170))[,3], censtats(EN19[,2],
(EN19[,2]<=190))[,3]),2)

valEN19<-cbind(sustmediasEN19, nadamediasEN19)
valEN19sd<-cbind(sustsdEN19, nadasdEN19)

valEN19<mean(EN19[,2])+0.125*mean(EN19[,2]) & valEN19>mean(EN19[,2])-
0.125*mean(EN19[,2])
valEN19sd<sd(EN19[,2])+0.125*sd(EN19[,2]) & valEN19sd>sd(EN19[,2])-
0.125*sd(EN19[,2])

write.xls(cbind(serie=EN19[,2], D_serie=EN19[,2]>0, seriep=EN19[,2],
D_seriep=EN19[,2]>70, seriem=EN19[,2], D_seriem=EN19[,2]>170, series=EN19[,2],
D_series=EN19[,2]>190, EN19.70.0, D_EN19.70.0=EN19[,2]>70, EN19.70.5,
D_EN19.70.5=EN19[,2]>70, EN19.70.Id, D_EN19.70.Id=EN19[,2]>70, EN19.170.0,
D_EN19.170.0=EN19[,2]>170, EN19.170.5, D_EN19.170.5=EN19[,2]>170,
EN19.170.Id, D_EN19.170.Id=EN19[,2]>170, EN19.190.0,

```

```
D_EN19.190.0=EN19[,2]>190, EN19.190.5, D_EN19.190.5=EN19[,2]>190,
EN19.190.ld, D_EN19.190.ld=EN19[,2]>190), "EN19.xls")
```

```
# x<- EN19[,2]
# x<- EN19.70.0
# x<- EN19.70.5
# x<- EN19.70.ld
# x<- EN19.170.0
# x<- EN19.170.5
# x<- EN19.170.ld
# x<- EN19.190.0
# x<- EN19.190.5
# x<- EN19.190.ld
```

```
x<- EN19.190.0
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias
bootstrap", xlab = "concentración de nitratos", ylab = "densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))
legend((quantile(boot,0.5)-6), 0.0012, pch=181, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")
```

```
bEN19<-data.frame(obs=EN19[,2],cen=EN19[,2]<190)
mean.sav <- function(bEN19){
mean(cenfit(bEN19$obs,bEN19$cen))[1]
}
resEN19 = censboot(bEN19,mean.sav,R=2000)
cimeanEN19<-c(quantile(resEN19$t,.025), quantile(resEN19$t,.975))
paste(round(cimeanEN19[1],2), round(cimeanEN19[2],2), sep="-")
bEN19<-data.frame(obs=EN19[,2],cen=EN19[,2]<70)
mean.sav <- function(bEN19){
mean(ros(bEN19$obs,bEN19$cen))
}
resEN19 = censboot(bEN19,mean.sav,R=2000)
cimeanEN19<-c(quantile(resEN19$t,.025), quantile(resEN19$t,.975))
paste(round(cimeanEN19[1],1), round(cimeanEN19[2],1), sep="-")
```

```
paste(round(mean(cenmle(EN19[,2],EN19[,2]<190)),2)[3],
round(mean(cenmle(EN19[,2],EN19[,2]<190)),2)[4], sep="-")
```

```
# EN20
```

```
# censura leve
```



```
JUIG003594nada<- read.table("JUIG003594nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG003594nada[,2], as.logical( JUIG003594nada[,3]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG003594nada[,2], as.logical(
JUIG003594nada[,3]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG003594nada[,2], as.logical( JUIG003594nada[,3]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG003594nada[,2], as.logical(
JUIG003594nada[,3]), conf.int=0.95)),1)[[4]], sep="-")

# censura media
JUIG003594nada<- read.table("JUIG003594nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG003594nada[,2], as.logical( JUIG003594nada[,4]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG003594nada[,2], as.logical(
JUIG003594nada[,4]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG003594nada[,2], as.logical( JUIG003594nada[,4]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG003594nada[,2], as.logical(
JUIG003594nada[,4]), conf.int=0.95)),1)[[4]], sep="-")

# censura severa
JUIG003594nada<- read.table("JUIG003594nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG003594nada[,2], as.logical( JUIG003594nada[,5]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG003594nada[,2], as.logical(
JUIG003594nada[,5]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG003594nada[,2], as.logical( JUIG003594nada[,5]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG003594nada[,2], as.logical(
JUIG003594nada[,5]), conf.int=0.95)),1)[[4]], sep="-")

round(mean(JUIG003594nada[,2]),1)
round(sd(JUIG003594nada[,2]),1)

# table(EN20[,2]<66)["TRUE"]/length(EN20[,2])
# 0.153
# table(EN20[,2]<70)["TRUE"]/length(EN20[,2])
# 0.231
# table(EN20[,2]<86)["TRUE"]/length(EN20[,2])
# 0.461

EN20.66.0<-replace(EN20[,2], EN20[,2]<66, 0)
EN20.66.5<-replace(EN20[,2], EN20[,2]<66, 33)
EN20.66.ld<-replace(EN20[,2], EN20[,2]<66, 66)
EN20.70.0<-replace(EN20[,2], EN20[,2]<70, 0)
EN20.70.5<-replace(EN20[,2], EN20[,2]<70, 35)
EN20.70.ld<-replace(EN20[,2], EN20[,2]<70, 70)
EN20.86.0<-replace(EN20[,2], EN20[,2]<86, 0)
EN20.86.5<-replace(EN20[,2], EN20[,2]<86, 43)
EN20.86.ld<-replace(EN20[,2], EN20[,2]<86, 86)

length(EN20[,2])
length(EN20[,2][EN20[,2]>=66])
length(EN20[,2][EN20[,2]>=70])
length(EN20[,2][EN20[,2]>=86])

sustmediasEN20<-round(rbind(c(mean(EN20[,2]), mean(EN20[,2]), mean(EN20[,2])),
c(mean( EN20.66.0), mean(EN20.66.5), mean(EN20.66.ld)), c(mean(EN20.70.0),
```

```

mean(EN20.70.5), mean(EN20.70.ld)), c(mean(EN20.86.0), mean(EN20.86.5),
mean(EN20.86.ld))),2)
nadamediasEN20<-round(rbind(censtats(EN20[,2], (EN20[,2]<=0))[,2],
censtats(EN20[,2], (EN20[,2]<=66))[,2], censtats(EN20[,2], (EN20[,2]<=70))[,2],
censtats(EN20[,2], (EN20[,2]<=86))[,2]),2)

sustsdEN20<-round(rbind(c(sd(EN20[,2]), sd(EN20[,2])), c(sd
(EN20.66.0), sd(EN20.66.5), sd(EN20.66.ld)), c(sd(EN20.70.0), sd(EN20.70.5),
sd(EN20.70.ld)), c(sd(EN20.86.0), sd(EN20.86.5), sd(EN20.86.ld))),2)
nadasdEN20<-round(rbind(censtats(EN20[,2], (EN20[,2]<=0))[,3], censtats(EN20[,2],
(EN20[,2]<=66))[,3], censtats(EN20[,2], (EN20[,2]<=70))[,3], censtats(EN20[,2],
(EN20[,2]<=86))[,3]),2)

valEN20<-cbind(sustmediasEN20, nadamediasEN20)
valEN20sd<-cbind(sustsdEN20, nadasdEN20)

valEN20<mean(EN20[,2])+0.125*mean(EN20[,2]) & valEN20>mean(EN20[,2])-
0.125*mean(EN20[,2])
valEN20sd<sd(EN20[,2])+0.125*sd(EN20[,2]) & valEN20sd>sd(EN20[,2])-
0.125*sd(EN20[,2])

write.xls(cbind(serie=EN20[,2], D_serie=EN20[,2]>0, seriep=EN20[,2],
D_seriep=EN20[,2]>66, seriem=EN20[,2], D_seriem=EN20[,2]>70, series=EN20[,2],
D_series=EN20[,2]>86, EN20.66.0, D_EN20.66.0=EN20[,2]>66, EN20.66.5,
D_EN20.66.5=EN20[,2]>66, EN20.66.ld, D_EN20.66.ld=EN20[,2]>66, EN20.70.0,
D_EN20.70.0=EN20[,2]>70, EN20.70.5, D_EN20.70.5=EN20[,2]>70, EN20.70.ld,
D_EN20.70.ld=EN20[,2]>70, EN20.86.0, D_EN20.86.0=EN20[,2]>86, EN20.86.5,
D_EN20.86.5=EN20[,2]>86, EN20.86.ld, D_EN20.86.ld=EN20[,2]>86), "EN20.xls")

# x<- EN20[,2]
# x<- EN20.66.0
# x<- EN20.66.5
# x<- EN20.66.ld
# x<- EN20.70.0
# x<- EN20.70.5
# x<- EN20.70.ld
# x<- EN20.86.0
# x<- EN20.86.5
# x<- EN20.86.ld

x<- EN20.86.ld
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histograma de medias
bootstrap", xlab = "concentración de nitratos", ylab = "densidad")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))

```

```
legend((quantile(boot,0.5)-6), 0.0012, pch=181, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")
```

```
bEN20<-data.frame(obs=EN20[,2],cen=EN20[,2]<86)
mean.sav <- function(bEN20){
mean(cenfit(bEN20$obs,bEN20$cen))[1]
}
resEN20 = censboot(bEN20,mean.sav,R=2000)
cimeanEN20<-c(quantile(resEN20$t,.025), quantile(resEN20$t,.975))
paste(round(cimeanEN20[1],2), round(cimeanEN20[2],2), sep="-")
```

```
bEN20<-data.frame(obs=EN20[,2],cen=EN20[,2]<70)
mean.sav <- function(bEN20){
mean(ros(bEN20$obs,bEN20$cen))
}
resEN20 = censboot(bEN20,mean.sav,R=2000)
cimeanEN20<-c(quantile(resEN20$t,.025), quantile(resEN20$t,.975))
paste(round(cimeanEN20[1],1), round(cimeanEN20[2],1), sep="-")
```

```
paste(round(mean(cenmle(EN20[,2],EN20[,2]<86)),2)[3],
round(mean(cenmle(EN20[,2],EN20[,2]<86)),2)[4], sep="-")
```

EN21

censura leve

```
JUIG003866nada<- read.table("JUIG003866nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG003866nada[,2], as.logical( JUIG003866nada[,3]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG003866nada[,2], as.logical(
JUIG003866nada[,3]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG003866nada[,2], as.logical( JUIG003866nada[,3]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG003866nada[,2], as.logical(
JUIG003866nada[,3]), conf.int=0.95)),1)[[4]], sep="-")
```

censura media

```
JUIG003866nada<- read.table("JUIG003866nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG003866nada[,2], as.logical( JUIG003866nada[,4]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG003866nada[,2], as.logical(
JUIG003866nada[,4]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG003866nada[,2], as.logical( JUIG003866nada[,4]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG003866nada[,2], as.logical(
JUIG003866nada[,4]), conf.int=0.95)),1)[[4]], sep="-")
```

censura severa

```
JUIG003866nada<- read.table("JUIG003866nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG003866nada[,2], as.logical( JUIG003866nada[,5]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG003866nada[,2], as.logical(
JUIG003866nada[,5]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG003866nada[,2], as.logical( JUIG003866nada[,5]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG003866nada[,2], as.logical(
JUIG003866nada[,5]), conf.int=0.95)),1)[[4]], sep="-")
```

```
round(mean(JUIG003866nada[,2]),1)
```

```

round(sd(JUIG003866nada[,2]),1)

# table(EN21[,2]<52)["TRUE"]/length(EN21[,2])
# 0.128
# table(EN21[,2]<73)["TRUE"]/length(EN21[,2])
# 0.256
# table(EN21[,2]<114)["TRUE"]/length(EN21[,2])
# 0.487

EN21.52.0<-replace(EN21[,2], EN21[,2]<52, 0)
EN21.52.5<-replace(EN21[,2], EN21[,2]<52, 21)
EN21.52.ld<-replace(EN21[,2], EN21[,2]<52, 42)
EN21.73.0<-replace(EN21[,2], EN21[,2]<73, 0)
EN21.73.5<-replace(EN21[,2], EN21[,2]<73, 36.5)
EN21.73.ld<-replace(EN21[,2], EN21[,2]<73, 73)
EN21.114.0<-replace(EN21[,2], EN21[,2]<86, 0)
EN21.114.5<-replace(EN21[,2], EN21[,2]<86, 43)
EN21.114.ld<-replace(EN21[,2], EN21[,2]<86, 86)

length(EN21[,2])
length(EN21[,2][EN21[,2]>=52])
length(EN21[,2][EN21[,2]>=73])
length(EN21[,2][EN21[,2]>=114])

sustmediasEN21<-round(rbind(c(mean(EN21[,2]), mean(EN21[,2]), mean(EN21[,2])),
c(mean(EN21.52.0), mean(EN21.52.5), mean(EN21.52.ld)), c(mean(EN21.73.0),
mean(EN21.73.5), mean(EN21.73.ld)), c(mean(EN21.114.0), mean(EN21.114.5),
mean(EN21.114.ld))),2)
nadamediasEN21<-round(rbind(censtats(EN21[,2], (EN21[,2]<=0))[,2],
censtats(EN21[,2], (EN21[,2]<=52))[,2], censtats(EN21[,2], (EN21[,2]<=73))[,2],
censtats(EN21[,2], (EN21[,2]<=114))[,2]),2)

sustsdEN21<-round(rbind(c(sd(EN21[,2]), sd(EN21[,2]), sd(EN21[,2])), c(sd
(EN21.52.0), sd(EN21.52.5), sd(EN21.52.ld)), c(sd(EN21.73.0), sd(EN21.73.5),
sd(EN21.73.ld)), c(sd(EN21.114.0), sd(EN21.114.5), sd(EN21.114.ld))),2)
nadasdEN21<-round(rbind(censtats(EN21[,2], (EN21[,2]<=0))[,3], censtats(EN21[,2],
(EN21[,2]<=52))[,3], censtats(EN21[,2], (EN21[,2]<=73))[,3], censtats(EN21[,2],
(EN21[,2]<=114))[,3]),2)

valEN21<-cbind(sustmediasEN21, nadamediasEN21)
valEN21sd<-cbind(sustsdEN21, nadasdEN21)

valEN21<mean(EN21[,2])+0.125*mean(EN21[,2]) & valEN21>mean(EN21[,2])-
0.125*mean(EN21[,2])
valEN21sd<sd(EN21[,2])+0.125*sd(EN21[,2]) & valEN21sd>sd(EN21[,2])-
0.125*sd(EN21[,2])

write.xls(cbind(serie=EN21[,2], D_serie=EN21[,2]>0, seriep=EN21[,2],
D_seriep=EN21[,2]>52, seriem=EN21[,2], D_seriem=EN21[,2]>73, series=EN21[,2],
D_series=EN21[,2]>114, EN21.52.0, D_EN21.52.0=EN21[,2]>52, EN21.52.5,
D_EN21.52.5=EN21[,2]>52, EN21.52.ld, D_EN21.52.ld=EN21[,2]>52, EN21.73.0,
D_EN21.73.0=EN21[,2]>73, EN21.73.5, D_EN21.73.5=EN21[,2]>73, EN21.73.ld,
D_EN21.73.ld=EN21[,2]>73, EN21.114.0, D_EN21.114.0=EN21[,2]>114, EN21.114.5,

```

```
D_EN21.114.5=EN21[,2]>114, EN21.114.Id, D_EN21.114.Id=EN21[,2]>114),
"EN21.xls")
```

```
# x<- EN21[,2]
# x<- EN21.52.0
# x<- EN21.52.5
# x<- EN21.52.Id
# x<- EN21.73.0
# x<- EN21.73.5
# x<- EN21.73.Id
# x<- EN21.114.0
# x<- EN21.114.5
# x<- EN21.114.Id
```

```
x<- EN21.114.Id
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histogram bootstrap
means", xlab = "Nitrate concentration", ylab ="densitu")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))
legend((quantile(boot,0.5)-6), 0.0012, pch=181, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")
```

```
bEN21<-data.frame(obs=EN21[,2],cen=EN21[,2]<114)
mean.sav <- function(bEN21){
mean(cenfit(bEN21$obs,bEN21$cen))[1]
}
resEN21 = censboot(bEN21,mean.sav,R=2000)
cimeanEN21<-c(quantile(resEN21$t,.025), quantile(resEN21$t,.975))
paste(round(cimeanEN21[1],2), round(cimeanEN21[2],2), sep="-")
```

```
bEN21<-data.frame(obs=EN21[,2],cen=EN21[,2]<52)
mean.sav <- function(bEN21){
mean(ros(bEN21$obs,bEN21$cen))
}
resEN21 = censboot(bEN21,mean.sav,R=2000)
cimeanEN21<-c(quantile(resEN21$t,.025), quantile(resEN21$t,.975))
paste(round(cimeanEN21[1],1), round(cimeanEN21[2],1), sep="-")
```

```
paste(round(mean(cenmle(EN21[,2],EN21[,2]<114)),2)[3],
round(mean(cenmle(EN21[,2],EN21[,2]<114)),2)[4], sep="-")
```

```
# EN22
```

```
# censura leve
JUIG003940nada<- read.table("JUIG003940nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG003940nada[,2], as.logical( JUIG003940nada[,3]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG003940nada[,2], as.logical(
JUIG003940nada[,3]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG003940nada[,2], as.logical( JUIG003940nada[,3]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG003940nada[,2], as.logical(
JUIG003940nada[,3]), conf.int=0.95)),1)[[4]], sep="-")

# censura media
JUIG003940nada<- read.table("JUIG003940nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG003940nada[,2], as.logical( JUIG003940nada[,4]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG003940nada[,2], as.logical(
JUIG003940nada[,4]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG003940nada[,2], as.logical( JUIG003940nada[,4]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG003940nada[,2], as.logical(
JUIG003940nada[,4]), conf.int=0.95)),1)[[4]], sep="-")

# censura severa
JUIG003940nada<- read.table("JUIG003940nada.txt",header=T, dec=",")
paste(round(mean(cenmle(JUIG003940nada[,2], as.logical( JUIG003940nada[,5]),
conf.int=0.95)),1)[[3]], round(mean(cenmle(JUIG003940nada[,2], as.logical(
JUIG003940nada[,5]), conf.int=0.95)),1)[[4]], sep="-")
paste(round(mean(cenfit(JUIG003940nada[,2], as.logical( JUIG003940nada[,5]),
conf.int=0.95)),1)[[3]], round(mean(cenfit(JUIG003940nada[,2], as.logical(
JUIG003940nada[,5]), conf.int=0.95)),1)[[4]], sep="-")

round(mean(JUIG003940nada[,2]),1)
round(sd(JUIG003940nada[,2]),1)

# table(EN22[,2]<20)["TRUE"]/length(EN22[,2])
# 0.125
# table(EN22[,2]<24)["TRUE"]/length(EN22[,2])
# 0.375
# table(EN22[,2]<70)["TRUE"]/length(EN22[,2])
# 0.5

EN22.20.0<-replace(EN22[,2], EN22[,2]<20, 0)
EN22.20.5<-replace(EN22[,2], EN22[,2]<20, 10)
EN22.20.Id<-replace(EN22[,2], EN22[,2]<20, 20)
EN22.24.0<-replace(EN22[,2], EN22[,2]<24, 0)
EN22.24.5<-replace(EN22[,2], EN22[,2]<24, 12)
EN22.24.Id<-replace(EN22[,2], EN22[,2]<24, 24)
EN22.70.0<-replace(EN22[,2], EN22[,2]<70, 0)
EN22.70.5<-replace(EN22[,2], EN22[,2]<70, 35)
EN22.70.Id<-replace(EN22[,2], EN22[,2]<70, 70)

length(EN22[,2])
length(EN22[,2][EN22[,2]>=20])
length(EN22[,2][EN22[,2]>=24])
length(EN22[,2][EN22[,2]>=70])

sustmediasEN22<-round(rbind(c(mean(EN22[,2]), mean(EN22[,2]), mean(EN22[,2])),
c(mean( EN22.20.0), mean(EN22.20.5), mean(EN22.20.Id)), c(mean(EN22.24.0),
```



```

mean(EN22.24.5), mean(EN22.24.Id)), c(mean(EN22.70.0), mean(EN22.70.5),
mean(EN22.70.Id))),2)
nadamediasEN22<-round(rbind(censtats(EN22[,2], (EN22[,2]<=0))[,2],
censtats(EN22[,2], (EN22[,2]<=20))[,2], censtats(EN22[,2], (EN22[,2]<=24))[,2],
censtats(EN22[,2], (EN22[,2]<=70))[,2]),2)

sustsdEN22<-round(rbind(c(sd(EN22[,2]), sd(EN22[,2]), sd(EN22[,2])), c(sd
(EN22.20.0), sd(EN22.20.5), sd(EN22.20.Id)), c(sd(EN22.24.0), sd(EN22.24.5),
sd(EN22.24.Id)), c(sd(EN22.70.0), sd(EN22.70.5), sd(EN22.70.Id))),2)
nadasdEN22<-round(rbind(censtats(EN22[,2], (EN22[,2]<=0))[,3], censtats(EN22[,2],
(EN22[,2]<=20))[,3], censtats(EN22[,2], (EN22[,2]<=24))[,3], censtats(EN22[,2],
(EN22[,2]<=70))[,3]),2)

valEN22<-cbind(sustmediasEN22, nadamediasEN22)
valEN22sd<-cbind(sustsdEN22, nadasdEN22)

valEN22<mean(EN22[,2])+0.125*mean(EN22[,2]) & valEN22>mean(EN22[,2])-
0.125*mean(EN22[,2])
valEN22sd<sd(EN22[,2])+0.125*sd(EN22[,2]) & valEN22sd>sd(EN22[,2])-
0.125*sd(EN22[,2])

write.xls(cbind(serie=EN22[,2], D_serie=EN22[,2]>0, seriep=EN22[,2],
D_seriep=EN22[,2]>20, seriem=EN22[,2], D_seriem=EN22[,2]>724, series=EN22[,2],
D_series=EN22[,2]>70, EN22.20.0, D_EN22.20.0=EN22[,2]>20, EN22.20.5,
D_EN22.20.5=EN22[,2]>20, EN22.20.Id, D_EN22.20.Id=EN22[,2]>20, EN22.24.0,
D_EN22.24.0=EN22[,2]>24, EN22.24.5, D_EN22.24.5=EN22[,2]>24, EN22.24.Id,
D_EN22.24.Id=EN22[,2]>24, EN22.70.0, D_EN22.70.0=EN22[,2]>70, EN22.70.5,
D_EN22.70.5=EN22[,2]>70, EN22.70.Id, D_EN22.70.Id=EN22[,2]>70), "EN22.xls")

# x<- EN22[,2]
# x<- EN22.20.0
# x<- EN22.20.5
# x<- EN22.20.Id
# x<- EN22.24.0
# x<- EN22.24.5
# x<- EN22.24.Id
# x<- EN22.70.0
# x<- EN22.70.5
# x<- EN22.70.Id

x<- EN22.70.Id
meann<- function(x) mean(x)
meann(x)
boot <-numeric(10000)
for (i in 1:10000) boot[i] <- meann(sample(x,replace=T))
x<-boot
hist(boot,freq=FALSE, col = 'lightgreen',breaks=25, main= "Histogram bootstrapped
means", xlab = "nitrates concentration", ylab ="density")
curve(dnorm(x,mean(boot),sd(boot)),add=T, col = "blue" ,lw = 2)
abline(v= quantile(boot,0.975),col="red",lty=4,lw=3)
abline(v= quantile(boot,0.025) ,col="red", lty=4,lw=3)
abline(v= quantile(boot,0.5) ,col="red", lty=4,lw=1)
legend((quantile(boot,0.025)-6), 0.0012, pch=62, cex=0.75,
round(quantile(boot,0.025),1))

```

```
legend((quantile(boot,0.5)-6), 0.0012, pch=181, cex=0.75, round(quantile(boot,0.5),1))
legend((quantile(boot,0.975)-6), 0.0012, pch=60, cex=0.75,
round(quantile(boot,0.975),1))
paste(round(quantile(boot,0.025),1), round(quantile(boot,0.975),1), sep="-")
```

```
bEN22<-data.frame(obs=EN22[,2],cen=EN22[,2]<70)
mean.sav <- function(bEN22){
mean(cenfit(bEN22$obs,bEN22$cen))[1]
}
resEN22 = censboot(bEN22,mean.sav,R=2000)
cimeanEN22<-c(quantile(resEN22$t,.025), quantile(resEN22$t,.975))
paste(round(cimeanEN22[1],2), round(cimeanEN22[2],2), sep="-")
```

```
bEN22<-data.frame(obs=EN22[,2],cen=EN22[,2]<70)
mean.sav <- function(bEN22){
mean(ros(bEN22$obs,bEN22$cen))
}
resEN22 = censboot(bEN22,mean.sav,R=2000)
cimeanEN22<-c(quantile(resEN22$t,.025), quantile(resEN22$t,.975))
paste(round(cimeanEN22[1],1), round(cimeanEN22[2],1), sep="-")
```

```
paste(round(mean(cenmle(EN22[,2],EN22[,2]<70)),2)[3],
round(mean(cenmle(EN22[,2],EN22[,2]<70)),2)[4], sep="-")
```

Gráfico boxplot

```
EN21.52.0<-replace(EN21[,2], EN21[,2]<52, 0)
EN21.52.5<-replace(EN21[,2], EN21[,2]<52, 21)
EN21.52.ld<-replace(EN21[,2], EN21[,2]<52, 42)
EN21.73.0<-replace(EN21[,2], EN21[,2]<73, 0)
EN21.73.5<-replace(EN21[,2], EN21[,2]<73, 36.5)
EN21.73.ld<-replace(EN21[,2], EN21[,2]<73, 73)
EN21.114.0<-replace(EN21[,2], EN21[,2]<86, 0)
EN21.114.5<-replace(EN21[,2], EN21[,2]<86, 43)
EN21.114.ld<-replace(EN21[,2], EN21[,2]<86, 86)
```

```
par(mfrow=c(1,3))
par(bg="darkgreen",col.axis="white", lwd=2)
cenboxplot(EN21[,2], EN21[,2]<52, border = "white", col.lab="white", col.main="white")
cenboxplot(EN21[,2], EN21[,2]<73, border = "white", col.lab="white", col.main="white")
cenboxplot(EN21[,2], EN21[,2]<114, border = "white", col.lab="white",
col.main="white")
par(mfrow=c(1,1))
```

Gráfico boxplots por grupos

```
par(bg="darkgreen",col.axis="white", lwd=2)
bEN21<-data.frame(obs=EN21[,2],cen=EN21[,2]<49)
grupo<-c(rep(1,10), rep(2, 10))
cenboxplot(bEN21[,1], bEN21[,2], as.factor(grupo), border = "white", col.lab="white",
col.main="white")
```


Gráfico dispersion

```
par(bg="darkgreen",col.axis="white", lwd=2)
cenxyplot(1:length(EN21[,2]), rep("FALSE", length(EN21[,2])), JUIG003866[,2],
as.logical(JUIG003866[,3]), lty="dashed", xlab="número muestra", ylab="mg/L Nitrato",
main="Gráfico censurado dispersión JUIG003866", lwd=2, pch=16, xaxt="n",
font.axis=2, col="white", col.lab="white", col.main="white", col.ticks="white")
box(col="white")
```

Gráfico curva Kaplan-Meier

```
plot(cenfit(bEN21[,1], bEN21[,2]), col="white", col.lab="white")
```

Gráfico curva Kaplan-Meier dos muestras

```
par(bg="darkgreen",col.axis="white", lwd=2)
bEN9<-data.frame(obs=EN9[,2],cen=EN9[,2]<49)
grupo<-c(rep(1,10), rep(2, 10))
plot(cenfit(bEN9[,1], bEN9[,2], as.factor(grupo)) , col="white", col.lab="white")
cendiff(bEN9[,1], bEN9[,2], as.factor(grupo))
```

Gráfico ROS

```
par(bg="darkgreen",col.axis="white", lwd=2)
bEN21<-data.frame(obs=EN21[,2],cen=EN21[,2]<52)
plot(cenros(bEN21[,1], bEN21[,2]) , col="white", col.lab="white")
```

Gráfico MLE

```
plot(cenmle(bEN21[,1], bEN21[,2], dist="gaussian"),col="white", col.lab="white")
bEN21[,1]<-replace(bEN21[,1], which(bEN21[,1]==0), 1)
plot(cenmle(bEN21[,1], bEN21[,2], dist="lognormal"), col="white", col.lab="white")
```

Gráfico conjunto series con suavizado

```
JUIG002683<-read.zoo(EN17)
JUIG002924<-read.zoo(EN18)
JUIG002950<-read.zoo(EN19)
JUIG003594<-read.zoo(EN20)
JUIG003866<-read.zoo(EN21)
JUIG003940<-read.zoo(EN22)
z <- na.approx(merge(JUIG002683, JUIG002924, JUIG002950, JUIG003594,
JUIG003866, JUIG003940))
autoplot(z) +geom_line(size=0.75) + facet_free() + labs(x="date", y="mg/L Nitrates") +
geom_smooth(method = "loess" , se = FALSE)
an(cenmle(EN7[,2],EN7[,2]<54)),2)[4], sep="-")
```