

**Revista Internacional de  
Contaminación Ambiental**

Revista Internacional de Contaminación Ambiental  
ISSN: 0188-4999  
[claudio.amescua@atmosfera.unam.mx](mailto:claudio.amescua@atmosfera.unam.mx)  
Universidad Nacional Autónoma de México  
México

Grima-Olmedo, Carlos; Grima-Olmedo, Juan; Luque-Espinar,  
Juan Antonio; Pardo-Iguzquiza, Eulogio; Ramírez-Gómez, Álvaro  
**ANÁLISIS DE SUPERVIVENCIA A LA DETECCIÓN Y CORRECCIÓN DE LA  
CONTAMINACIÓN EN SERIES CON VALORES POR DEBAJO DEL LÍMITE DE DETECCIÓN**  
Revista Internacional de Contaminación Ambiental, vol. 35, núm. 1, 2019, Febrero-, pp. 165-178  
Universidad Nacional Autónoma de México  
México

DOI: <https://doi.org/10.20937/RICA.2019.35.01.12>

Disponible en: <https://www.redalyc.org/articulo.oa?id=37060163014>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

## ANÁLISIS DE SUPERVIVENCIA A LA DETECCIÓN Y CORRECCIÓN DE LA CONTAMINACIÓN EN SERIES CON VALORES POR DEBAJO DEL LÍMITE DE DETECCIÓN

Carlos GRIMA-OLMEDO<sup>1\*</sup>, Juan GRIMA-OLMEDO<sup>2</sup>, Juan Antonio LUQUE-ESPINAR<sup>3</sup>,  
Eulogio PARDO-IGUZQUIZA<sup>4</sup> y Álvaro RAMÍREZ-GÓMEZ<sup>5</sup>

<sup>1</sup> Departamento de Ingeniería Geológica y Minera, Escuela Técnica Superior de Ingenieros de Minas y Energía de Madrid, Universidad Politécnica de Madrid, Calle Ríos Rosas 21, 28003 Madrid, España

<sup>2</sup> Department of Research on Geological Resources, Instituto Geológico y Minero de España, Calle Cirilo Amorós, 42 Entrepalanta, 46004 Valencia, España

<sup>3</sup> Department of Research on Geological Resources, Instituto Geológico y Minero de España, Urb. Alcázar del Genil, 4-Edif. Zulema, Bajo, 18006 IGME Granada, España

<sup>4</sup> Department of Research on Geological Resources, Instituto Geológico y Minero de España, Calle Ríos Rosas 23, 28003 Madrid, España

<sup>5</sup> Departamento de Ingeniería Mecánica, Química y Diseño Industrial, Escuela Técnica Superior de Ingeniería y Diseño Industrial, Ronda de Valencia 3, 28012 Madrid, España

\*Autor para correspondencia; carlos.grima@upm.es

(Recibido junio 2017; aceptado abril 2018)

Palabras clave: datos censurados, directiva marco del agua, Kaplan-Meier, regresión robusta

### RESUMEN

El artículo 17 de la Directiva Marco del Agua insta a la Comisión y al Parlamento europeos a proteger el agua subterránea, y establece los criterios para determinar su estado químico y definir estándares de calidad ambiental. Si consideramos las características específicas de los contaminantes en agua subterránea, es preciso utilizar ensayos estadísticos adecuados. En el anexo IV de la Directiva de Aguas Subterráneas se establece que las medidas por debajo del límite de cuantificación (LC) deben sustituirse por la mitad del LC más alto, con la excepción de plaguicidas totales. Aunque es posible estimar valores medios en datos censurados con un único LC, su sustitución distorsiona los intervalos de confianza usados para verificar la tolerancia con umbrales. Las redes oficiales de monitoreo presentan escasos registros con valores por debajo del LC. Con objeto de establecer un marco metodológico para su evaluación, se han tomado valores de concentración de nitratos para realizar posteriormente una censura artificial. Se han seleccionado dos cuerpos de agua subterránea con diferentes características y frecuencias de monitoreo: la Plana de Sagunto, en la zona oriental de España, y un sector del Valle Central (Chile), situado en torno a los sistemas fluviales de los ríos Tinguirica y Antivero. Mediante la simulación de diferentes porcentajes de censura a los datos se han ensayado varios métodos. Se ha demostrado que las técnicas específicas de estimación de valores censurados, como Kaplan-Meier o regresión robusta ser las más eficientes para verificar el cumplimiento de los estándares ambientales.

Key words: censored data, water framework directive, Kaplan-Meier, robust regression

## ABSTRACT

Under article 17 of the Water Framework Directive the European Union was required to establish a framework to protect groundwater and set criteria for determining its chemical status and define environmental quality standards. Considering the specific characteristics of contaminants in groundwater, it is necessary to use adequate statistical tests. Annex IV of the Groundwater Directive establishes that all measurements below the limit of quantification (LOQ) must be replaced by half of the highest LOQ, with the exception of total pesticides. Although it is possible to estimate mean values in data censored with a single LOQ, its substitution distorts the confidence intervals used to verify tolerance with thresholds. The official monitoring networks lack series with values below the LOQ. In order to set up a methodological framework for its evaluation, nitrate concentration values have been taken in order to perform a subsequent artificial censorship. Two groundwater bodies with different monitoring characteristics and frequencies have been selected, Plana de Sagunto, in the eastern part of Spain, and a sector of the Central Valley (Chile), located around the systems of Tinguirica and Antivero rivers. Through the simulation of different percentages of censorship to data, several methods have been tested. Specific techniques for estimating censored values, such as Kaplan-Meier or robust regression have proven to be the most efficient to verify compliance with environmental standards.

## INTRODUCCIÓN

Los países de la Unión Europea afrontan el problema de la contaminación del suelo y los recursos hídricos (Grima et al. 2002), por lo que la Comisión Europea ha establecido programas de monitoreo de la calidad de los recursos naturales. Las redes de control de agua subterránea ponen de manifiesto la presencia en los acuíferos de muchos contaminantes sintéticos (Balderacchi et al. 2013) o de origen natural que no se muestrean habitualmente.

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 por debajo del límite correspondiente, ya que cualquier procedimiento analítico tiene una capacidad de detección finita. El análisis de series con este tipo de valores es complejo, aunque en los últimos años se han producido avances significativos para su interpretación.

Los problemas típicos que deben solucionarse en el caso de que existan valores por debajo del límite de detección (LD) (USEPA 2009) son el cálculo de los valores medios de concentración en una zona. Además, se tienen que establecer los intervalos de confianza de los parámetros considerados para su comparación con estándares, así como la probabilidad de que estos valores medios superen un umbral de concentración.

La Directiva Derivada de Aguas Subterráneas del Parlamento Europeo y el Consejo (PE-CUE 2006) complementa las provisiones de la Directiva Marco del Agua (PE-CUE 2000), y requiere la elaboración de normas de calidad para la protección de los acuíferos. Como consecuencia de la variabilidad natural en la composición química, así como de la inexistencia de redes de observación y control adecuadas de calidad del agua subterránea, se desestimó la elaboración de una lista de estándares de calidad que pudieran aplicarse en todas las masas de agua subterránea en Europa. La Comisión Europea desarrolla, a través del Proyecto BRIDGE, una metodología para establecer los criterios de evaluación del estado químico de los acuíferos para fijar los valores límite de contaminantes en el agua subterránea (Müller 2006). Estos últimos se definen como estándares de calidad establecidos por los Estados miembros, y representan la concentración de un contaminante específico que no debe sobrepasarse con el objeto de proteger la salud humana y el medio ambiente. Las masas de agua subterránea son sistemas complejos que presentan una alta variabilidad no sólo en su geometría y las características hidrodinámicas del medio geológico que los conforma, sino en sus propiedades fisicoquímicas (Ballesteros et al. 2001). Por lo tanto, no tiene sentido definir una concentración única para caracterizar la calidad de un acuífero en su conjunto (Blum et al. 2009).

Por ello, el establecimiento de valores umbrales implica considerar las características hidrogeológicas particulares de las redes de control de aguas

subterráneas y sus condicionantes específicos. Para el establecimiento de los valores umbrales con una base científica sólida, se aplicarán metodologías estadísticas basadas en pruebas de hipótesis, mediante las cuales se definirán incrementos estadísticamente significativos. 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 (Preziosi y del Bon 2014) calculados a partir de los datos, como con estándares de calidad fijados a partir de un subconjunto de datos o establecidos por las autoridades ambientales.

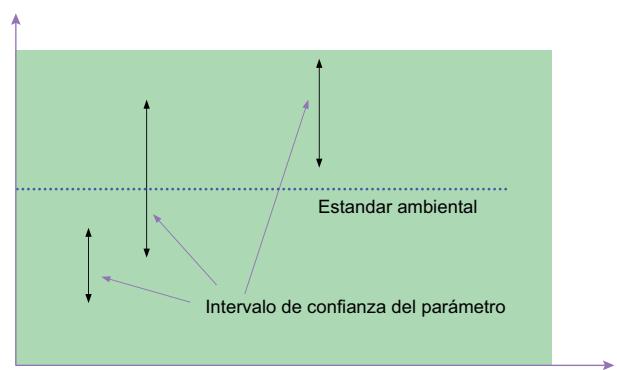
El resultado de un contraste estadístico de hipótesis puede implicar la clasificación de una masa de agua subterránea como en mal (o deficiente) estado químico desde el punto de vista de la Directiva de Aguas Subterráneas. Los principales objetivos del análisis del estado químico del cuerpo de agua subterránea son: *a)* comprobar el cumplimiento de los estándares de protección establecidos; *b)* verificar el empeoramiento significativo de su calidad, o *c)* establecer la existencia de tendencias significativas de aumento o descenso, en el caso de acciones de recuperación.

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 un acuífero (cuerpo de agua subterránea) a partir de valores de concentración de un contaminante por debajo del límite de detección (concentración mínima de sustancia que puede detectarse con fiabilidad), para contrastar la existencia de incrementos significativos en la concentración de un parámetro determinado (**Fig. 1**). El límite de

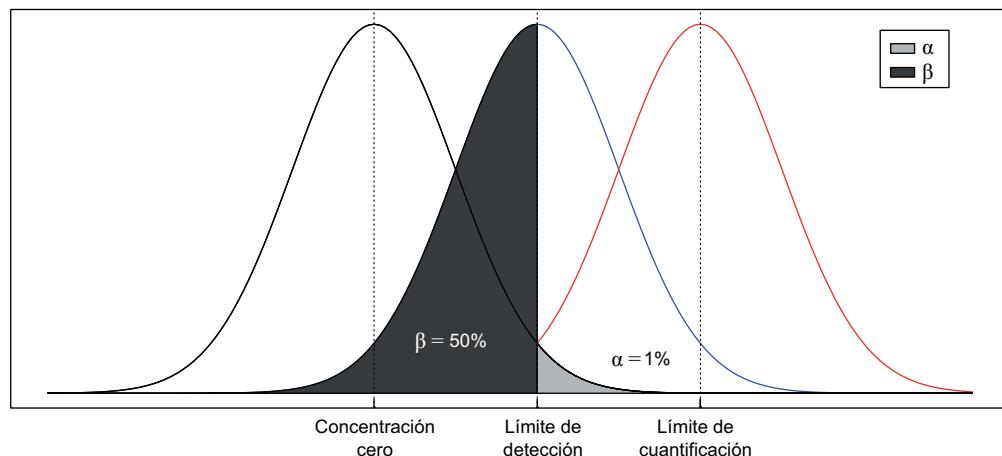
cuantificación (LC) suele definirse como la desviación estándar del blanco multiplicada por 10 (Saadati et al. 2013). De esta forma, el error  $\alpha$  (probabilidad de un falso positivo) es pequeño, del orden del 1 %.

## TÉCNICAS DE ANÁLISIS Y OBTENCIÓN DE 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 (**Fig. 2**). 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 químico mediante un valor único.



**Fig. 2.** Posición relativa del intervalo de confianza de la media o mediana en relación con el valor del estándar ambiental



**Fig. 1.** Límite de detección y límite de cuantificación (con base en la definición de Saadati et al. 2013)

Una de las ventajas de los intervalos de confianza frente a las pruebas 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 límites. Por otra parte, la anchura del intervalo también aporta información valiosa. Un intervalo de confianza estrecho indica 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 tamaño pequeño. El límite inferior de confianza es utilizado en el control operativo, 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. En este último caso se supone la masa de agua subterránea está afectada y el objetivo consiste en evaluar la efectividad de las medidas de recuperación implementadas para la inversión de las tendencias.

Cuando se analizan datos de concentraciones 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 para la determinación de estadísticos básicos en datos ambientales censurados son los siguientes: sustitución, máxima verosimilitud, regresión sobre los estadísticos de orden (Antweiler y Taylor 2008) y técnicas no paramétricas, en concreto el método de Kaplan-Meier (Singh y Totawattage 2013). Únicamente los dos últimos permiten el análisis de conjuntos de datos en los que hay 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.

Los métodos basados en sustitución son rápidos y muy atractivos por su sencillez. De hecho, en la Directiva 2006/118/CE (Aguas Subterráneas) se dice que “con el fin de evitar sesgos en la determinación de las tendencias, todas las mediciones por debajo del LC se cifrarán en la mitad del LC más alto registrado durante el periodo, con la excepción de los plaguicidas totales” (suma de todos los plaguicidas concretos detectados y cuantificados).

Sin embargo, la sustitución de los valores censurados por LC/2 proporciona resultados sesgados

(USEPA 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, la utilización de este tipo de métodos no es recomendable salvo en los siguientes casos: cuando la muestra sea demasiado pequeña (< 5), la proporción de valores censurados sea inferior al 10-15 % o existan procesos que generen 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 (USEPA 2009).

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 será generalmente deficiente. Esto 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.

La estimación por máxima verosimilitud 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 que se muestra en la ecuación 1 (Chacón y Míguez 1980):

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

Esta función tomará valores diferentes en función del valor de  $\theta$ . Considerando la expresión anterior como una función de  $\theta$ ,  $L(\theta/x_1, x_2, \dots, x_n)$  se denomina como función de verosimilitud. Entre los valores posibles de  $\theta$ , uno es tal que hace máxima la función de verosimilitud. En la práctica se suele tomar el logaritmo de esta función (ecuación 2):

$$\hat{\theta}_{mle} = \arg \max \left( \frac{\partial L(\theta/x_1, x_2, \dots, x_n)}{\partial \theta} \right) \quad (2)$$

Por tanto, suponiendo que los datos sigan una distribución determinada (generalmente normal o lognormal) se escogen los estadísticos que mejor ajusten los valores por encima del LC y el porcentaje de valores por debajo del mismo. Los estimadores de máxima verosimilitud son más precisos que las gráficas probabilísticas, ya que la raíz cuadrada del error cuadrático medio (RMSE, por sus siglas en inglés) es menor. Ambos métodos consiguen estimaciones sin sesgo cuando los datos presentan un ajuste perfecto a la distribución asumida y el número de datos es grande. En el caso de muestras de tamaño pequeño

( $n = 5$ ) la estimación es imprecisa y el sesgo grande (Gleit 1985). Igualmente, la presencia de valores anómalos puede distorsionar los estimadores de la media y la desviación estándar, provocando que los valores censurados extrapolados sean incluso negativos. Cuando el porcentaje de datos censurados es alto (superior al 40 %) no es recomendable en general el uso de métodos paramétricos.

El método de regresión robusta para el cálculo de estadísticos 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 LC 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.

Las ventajas obtenidas frente a los métodos distribucionales es que no se ven afectados por el ajuste de las distribuciones asumidas al utilizar datos reales. Las estimaciones de los valores extrapolados pueden reconvertirse a su escala original, evitando sesgos en muchas ocasiones inaceptables debido a la transformación (e.g., distribuciones lognormales en series muy asimétricas), resultando en una toma de decisiones incorrecta (USEPA 2002, Singh y Totawattage 2013). Al no haber un límite en el porcentaje de datos censurados —que pueden llegar al 80 %— los resultados deben interpretarse con precaución, especialmente en el caso de muestras de tamaño pequeño. Además, puede utilizarse 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.

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 sólo se sabe que ciertas concentraciones son menores que una cierta cantidad (generalmente el límite de detección). El tercer tipo 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 (ecuación 3):

$$S(t) = \text{Prob } (T > t) = 1 - F(t) \quad (3)$$

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

Los métodos no paramétricos se basan únicamente en los rangos de los datos (Antweiler y Taylor 2008) y no asumen que los datos se ajusten a una distribución conocida. 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. La aplicación del análisis de supervivencia (tiempo transcurrido hasta un suceso, como puede ser muerte del paciente) al análisis de conjuntos de datos con censura por la izquierda (concentración inferior al límite de detección) se realiza dando la vuelta a la serie y restando sus valores de una constante (nueva serie = constante - serie inicial). De esta forma pueden estimarse la función de supervivencia y los percentiles de los datos originales (Helsel 2005). Para cada  $t_i$  se definen:  $d_i$ : número de muertes en el momento  $i$ ; y  $n_i$ : 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  $S(t)$  (ecuación 4):

$$\hat{S}(t) = \prod_{i=1}^k \frac{b_i - d_i}{b_i} \quad (4)$$

donde  $b$  es el número total de observaciones (deteccadas y censuradas) y  $d$  el número de observaciones no censuradas.

En general se asume que los valores de concentración en agua subterránea siguen una distribución normal o lognormal. En el primer caso el intervalo de confianza viene definido por la Ecuación 5.

$$LIC_{1-\alpha} = \bar{x} - t_{(1-\alpha, n-1)} \frac{s}{\sqrt{n}} \quad LSC_{1-\alpha} = \bar{x} + t_{(1-\alpha, n-1)} \frac{s}{\sqrt{n}} \quad (5)$$

Donde  $LIC$  es el límite inferior del intervalo de confianza,  $LSC$  el límite superior del intervalo de confianza,  $\alpha$  el nivel de confianza,  $n$  el tamaño de la muestra,  $s$  la desviación estándar muestral y  $t_{(1-\alpha, n-1)}$  el percentil  $(1 - \alpha/2)$  de una distribución  $t$  con  $n - 1$  grados de libertad.

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 estimar el intervalo en escala logarítmica y después simplemente exponentiarlos a fin de obtener el intervalo de confianza en su escala original. La ecuación de Cox se utiliza para el cálculo de los límites de confianza (ecuación 6):

$$\begin{aligned} LIC &= \exp[\bar{y} + \frac{s_y^2}{2} - z \sqrt{\frac{s_y^2}{n} + \frac{s_y^4}{2(n-1)}}] \\ LSC &= \exp[\bar{y} + \frac{s_y^2}{2} + z \sqrt{\frac{s_y^2}{n} + \frac{s_y^4}{2(n-1)}}] \end{aligned} \quad (6)$$

donde  $LIC$  es el límite inferior del intervalo de confianza,  $LSC$  el límite superior del intervalo de confianza,  $\alpha$  el nivel de confianza,  $n$  el tamaño de la muestra,  $s$  la desviación estándar muestral e  $y$  el logaritmo natural de los datos

## ÁREA DE ESTUDIO Y DATOS

En el marco de la Estrategia Común de Implementación (ECI) de la Directiva Marco del Agua, la totalidad de los estados miembros de la Unión Europea y Noruega acordaron designar una serie de cuencas fluviales piloto para comprobar las diversas cuestiones desarrolladas en los documentos y guías de aplicación de la directiva. En España, se evaluaron estos documentos en la cuenca del Júcar, con el objetivo de proporcionar herramientas para la gestión de sus recursos hidráulicos.

Uno de los requerimientos básicos de las Directivas de Aguas consistió en identificar algunas presiones como la extracción de agua, la contaminación puntual y difusa, y la intrusión salina, y evaluar los impactos en cada masa de agua subterránea con objeto de conocer las causas que condicionan su mal estado químico. Para ello se trasladó la misma metodología desarrollada para las masas de agua superficial, definiendo el riesgo como una combinación de presiones definidas e impactos. La intrusión salina se debe al bombeo de agua subterránea de los pozos cercanos a la línea de costa y provoca la salinización

del agua en términos de su composición química. Su importancia es alta, debido a que los pozos se encuentran en zonas densamente pobladas, con actividad económica intensa y, por otra parte, es frecuente la existencia de ecosistemas terrestres dependientes del agua subterránea de alto valor ecológico.

En España, los documentos elaborados fueron evaluados en el ámbito territorial de la cuenca del Júcar (DHJ 2014), que se convirtió de esta manera en el centro de numerosos estudios (Estrela et al. 2004).

En el caso de Chile, la mayoría de la producción agrícola de este país se desarrolla en el Valle Central (Arumí y Oyarzún 2006), donde más de un 83 % del agua potable abastecida en las zonas rurales procede de la explotación de los acuíferos (Pizarro et al. 1999). En el **cuadro I** se muestran las estaciones seleccionadas en las cuencas piloto del Júcar (España) y de Valle Central (Chile). Debe destacarse el hecho de que no se ha identificado contaminación debida a fuentes puntuales. Por esa razón no hay series de datos con valores por debajo del LC. Como consecuencia, para evaluar el desempeño de métodos específicos de análisis, se han seleccionado series de datos disponibles, en este caso los nitratos, y a posteriori se han censurado artificialmente. Con ello, dado que se conocen los estadísticos de la muestra (media

**CUADRO I.** ESTACIONES SELECCIONADAS Y NÚMERO DE REGISTROS EN LA PLANA DE SAGUNTO (ESPAÑA) Y EL VALLE CENTRAL (CHILE)

| 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        |                 |          |
| Masa de agua subterránea, Valle Central    |          |                 |          |
| Código estación                            | Longitud | Código estación | Longitud |
| A  | 13       | J               | 13       |
| B  | 13       | K               | 13       |
| C  | 13       | L               | 13       |
| D  | 13       | M               | 13       |
| E  | 13       | N               | 13       |
| F  | 13       | O               | 13       |
| G  | 13       | P               | 13       |
| H  | 13       | Q               | 13       |
| I  | 13       |                 |          |

y desviación estándar), es posible analizar el sesgo que introduce cada una de las técnicas empleadas.

Con este fin se ha elegido el acuífero de la Plana de Sagunto, ya que se dispone de un conjunto de estaciones con registros históricos. Los materiales acuíferos que conforman la Plana de Sagunto se encuentran representados por niveles detriticos 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 (**Fig. 3**) hay 13 estaciones con valores por encima del valor umbral para los nitratos (50 mg/L). Las series son de tamaño variable, con datos comprendidos entre 1980 y 2010.

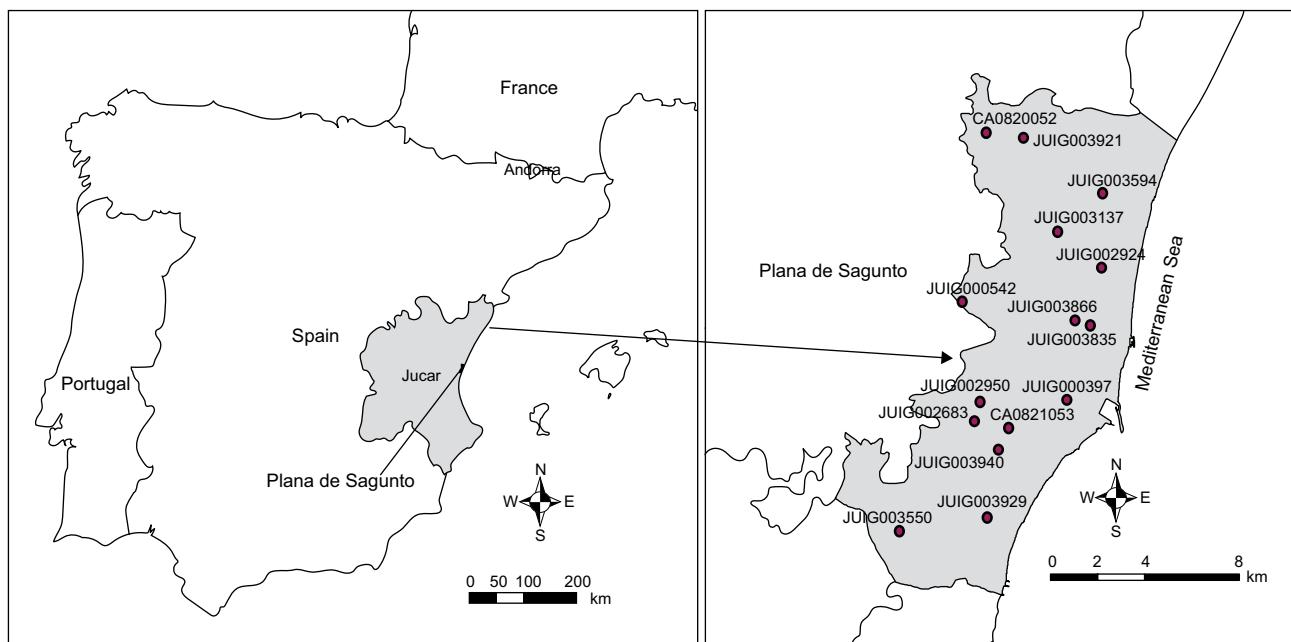
En Chile, el Valle Central se localiza entre los Andes y las cordilleras de la costa. De clima templado, esta zona produce el 83 % de las exportaciones agrícolas de Chile y el 96 % de la producción agrícola nacional (Golembeski 2004). En ella se ubican varias de las regiones vinícolas más conocidas de Chile y otros desarrollos agrícolas entre los que cabe citar el tabaco y el maíz, entre otras.

La conjunción entre un clima con bajas precipitaciones y una elevada disponibilidad de recursos hídricos (procedentes en gran parte del deshielo), en conjunto con la aplicación de manera generalizada de fertilizantes con alto contenido de nitrógeno, provocan la contaminación de las aguas subterráneas.

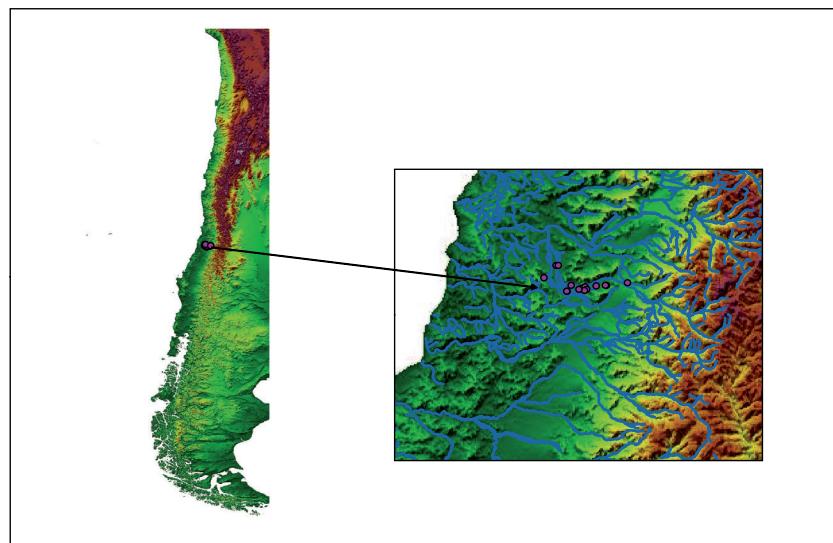
Debido a la utilización de pozos de poca profundidad para el abastecimiento de agua potable por parte de la población, especialmente trabajadores que desempeñan actividades agrícolas (Golembeski 2004), es particularmente relevante el control de las concentraciones de elementos potencialmente contaminantes del agua subterránea. La presencia potencial de otro tipo de compuestos en el agua subterránea, especialmente elementos sintéticos en concentraciones cercanas al límite de detección, requeriría la utilización de métodos específicos para su detección y corrección.

Se han seleccionado los datos de análisis de concentración de nitratos en agua subterránea de 17 pozos, muestreados en un área cercana a San Fernando, población situada a unos 140 km al sur de Santiago. Las mediciones se iniciaron en febrero de 2001 con frecuencia mensual y continuaron hasta febrero de 2002, lo que supone un total de 13 medidas en cada uno de los emplazamientos (**Fig. 4**).

El primer paso 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



**Fig. 3.** Zona de estudio en la Cuenca del Júcar (España)



**Fig. 4.** Zona de estudio en el Valle Central (Chile)

aplicación de modelos lineales a datos de concentración de contaminantes en agua subterránea suele ser inadecuada, debido a sus características intrínsecas (sesgo positivo, existencia de valores anómalos, etc.). De hecho, muchas variables de concentración de contaminantes en agua subterránea no están distribuidas normalmente y ni siquiera presentan una distribución simétrica (Montgomery et al. 1987). Para superar esta dificultad se utiliza la regresión local (LOESS) (Helsel y Hirsch 2002) o la regresión local ponderada (LOWESS) (Cleveland 1979). En la **figura 5** y el **cuadro I** se muestran las estaciones seleccionadas de Plana de Sagunto (España) y Valle Central (Chile), el número de registros existentes y las series temporales de las concentraciones de nitratos.

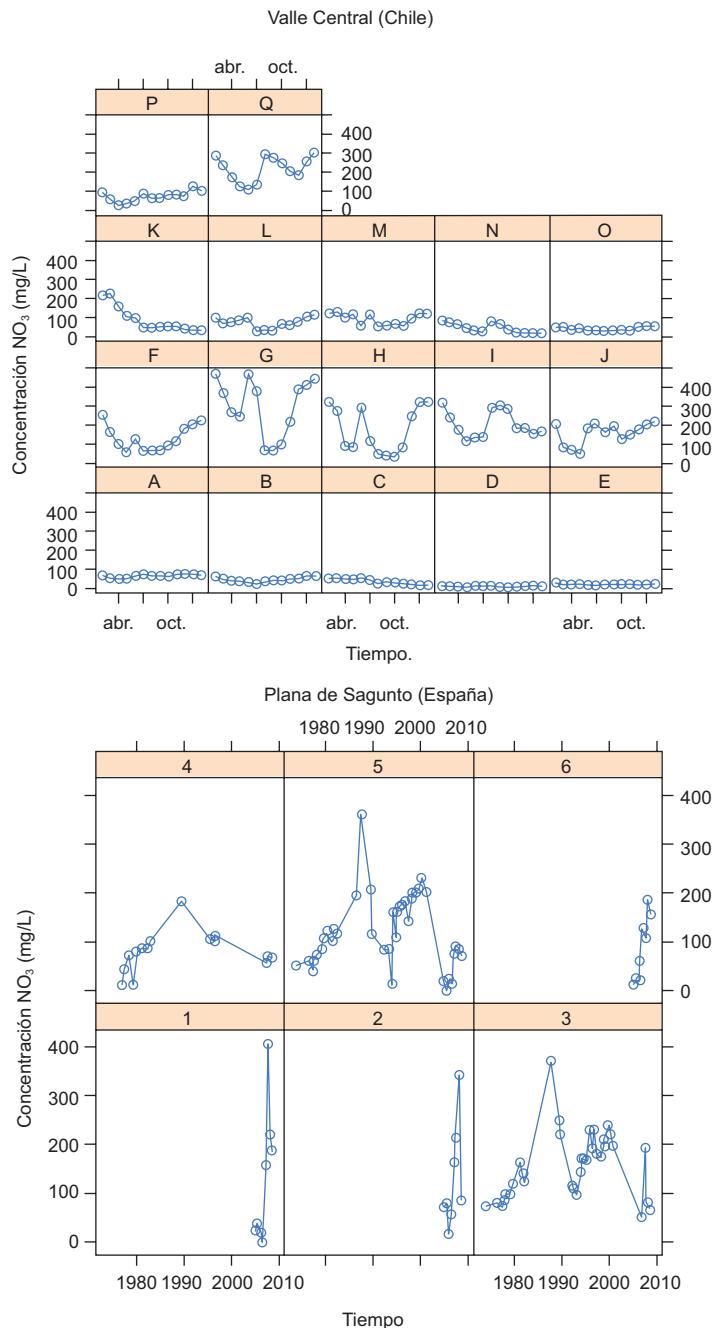
## DISCUSIÓN DE RESULTADOS

En el **cuadro II** se muestran los resultados obtenidos, indicándose de izquierda a derecha, los datos originales y los intervalos de confianza del valor medio calculado por los diferentes métodos. En la primera columna se han incluido las seis estaciones identificadas con su código, de las cuales se cuenta al menos con cuatro valores. En la segunda columna se indican la longitud de las series temporales originales y los límites de confianza de la media. En tercer lugar se indica el porcentaje de censura obtenido de forma artificial con objeto de realizar la simulación. Para ello se han seleccionado los valores correspondientes a los percentiles 12, 25 y

50 % de cada serie y se han hecho corresponder de forma arbitraria con los límites de cuantificación, con objeto de simular de forma aproximada dichos porcentajes de censura.

Los resultados obtenidos, que se presentan en el **cuadro II**, indican la necesidad de aplicar métodos específicos de análisis estadístico para el estudio de conjuntos de registros en los cuales no se han podido cuantificar algunas observaciones debido a las limitaciones de los métodos analíticos. Dado que las redes de monitoreo oficiales se han utilizado tradicionalmente para el análisis de componentes mayoritarios, no existen valores por debajo del límite de detección, ya que las concentraciones medidas son muy superiores a la precisión de los instrumentos. El desarrollo progresivo de la normatividad referente a la protección y conservación del medio ambiente por parte de las administraciones públicas, requiere que se midan concentraciones cada vez menores de sustancias potencialmente contaminantes para compararlas con estándares de calidad.

Para ello se han seleccionado series temporales con valores de concentración de nitratos cuyos estadísticos (media y desviación estándar) se han calculado con métodos tradicionales. A continuación, se han supuesto porcentajes de censura del 15 al 45 %, aproximadamente, y se han aplicado métodos de sustitución y de imputación para el cálculo de los intervalos de confianza de la media, con el objeto de comparar el límite superior de dichos intervalos con los estándares de calidad. Entre los primeros se ha simulado la sustitución de los valores censurados



**Fig. 5.** Series temporales en las zonas piloto seleccionadas

por 0, por la mitad del límite de detección y por este último valor. Entre los métodos de imputación se han considerado la estimación de máxima verosimilitud, la regresión robusta y el método de Kaplan-Meier.

Como puede observarse en el cuadro II, tanto en las series analizadas en la Plana de Sagunto como en el Valle Central, la sustitución de valores censurados

por 0 introduce sesgos a la izquierda en el cálculo de los intervalos de confianza de la media, que se incrementan en el mismo sentido que el porcentaje de valores por debajo del límite de detección. Del mismo modo, la sustitución de los valores censurados por el límite de detección introduce un sesgo a la derecha. En el caso de sustitución por la mitad del límite de

**CUADRO II. INTERVALOS DE CONFIANZA DE LA MEDIA ESTIMADOS EN LAS ESTACIONES SELECCIONADAS**

Plana de Sagunto (España). Intervalo de confianza del valor central

| Código de estación                    | Datos originales              | % datos censurados   | Valores sobre el límite de detección |  |   | Sustitución                             |  |   | Estimación máxima verosimilitud           | Regresión robusta | Kaplan-Meier |
|---------------------------------------|-------------------------------|----------------------|--------------------------------------|--|---|---|--|---|---|-------------------|--------------|
|                                       |                               |                      | 0                                    | 0.5 límite detección                   | Límite detección                        | 0                                       | 0.5 límite detección                       | Límite detección                        |   |                   |              |
| JUIG002683<br>media=133<br>sd=139.3   | N=8<br>LIC=47.5<br>LSC=230.3  | 12.5<br>25<br>50     | 7<br>6<br>4                          | 50.1-230.8<br>46.2-228.1<br>27.6-223.5 | 49.1-232.2<br>50.6-231.7<br>62.1-234.3  | 53.2-229.7<br>52.9-231.2<br>80.1-238.4  | 39.8-1455.2<br>33.7-2629.5<br>17.6-29791.1 | 49.5-231.2<br>47.7-236.2<br>--          | 39.6-231.2<br>40.7-233.1<br>136.5-265.4   |                   |              |
| JUIG002924<br>media=129.4<br>sd=106.1 | N=8<br>LIC=67.7<br>LSC=203.9  | 12.5<br>25<br>50     | 7<br>6<br>4                          | 63.3-202.5<br>50.9-199.7<br>26.6-192   | 67.3-202.3<br>65.4-200.2<br>59-198.4    | 71.6-203.8<br>77.9-204.9<br>94.7-213.9  | 64.3-317.4<br>57.3-366.7<br>37.3-604.1     | 73.1-202.0<br>71.7-207.3<br>--          | 64.2-205.1<br>69.1-207.7<br>73.8-214      |                   |              |
| JUIG002950<br>media=157<br>sd=68.9    | N=34<br>LIC=135.1<br>LSC=180  | 11.7<br>23.5<br>47.1 | 32<br>16<br>12                       | 129-178.2<br>64.3-139.1<br>43.5-119.8  | 132.7-180.4<br>123-171.5<br>120-166.5   | 135.5-180.4<br>179.9-206<br>194.6-216.7 | 133.2-187.3<br>101.9-191.1<br>84.4-181.7   | 135.7-180.3<br>143-193.1<br>132.3-198.1 | 135.2-180.6<br>177.7-204.9<br>192.5-215.7 |                   |              |
| JUIG003594<br>media=79.5<br>sd=40.2   | N=17<br>LIC=61.6<br>LSC=98.7  | 15.3<br>23.1<br>46.1 | 12<br>11<br>7                        | 44.1-92.8<br>39.4-89.8<br>18.6-75.5    | 60.8-97.6<br>59.7-96<br>54.7-91.6       | 76.1-102.9<br>78.1-104<br>88.8-110.4    | 50.5-147.5<br>47.7-149.8<br>32.8-163.2     | 69-100.9<br>67.4-101.4<br>49.8-101.8    | 74.4-102.6<br>75.7-103.3<br>85.3-109.1    |                   |              |
| JUIG003866<br>media=122.4<br>sd=75.1  | N=40<br>LIC=99.8<br>LSC=145.5 | 12.8<br>25.6<br>48.7 | 33<br>29<br>20                       | 94-143.1<br>85.6-139.1<br>81.2-136.3   | 99.2-145.9<br>99.5-145.7<br>100.1-145.8 | 105.1-148.1<br>113.1-152.1<br>118.3-155 | 148-1012.7<br>143.4-1276.6<br>140.1-1456   | 107-148.9<br>106.9-152.6<br>110.7-162.5 | 108.3-149.8<br>111.9-151.8<br>116.9-154.6 |                   |              |
| JUIG003940<br>media=85.9<br>sd=66.6   | N=8<br>LIC=42.9<br>LSC=130    | 12.5<br>37.5<br>50   | 7<br>5<br>4                          | 39.9-128.1<br>30.5-127.9<br>19.5-124   | 43.9-130.1<br>38-129.2<br>50.1-130.5    | 46.3-130.5<br>48.5-130.3<br>78.9-136.9  | 38.9-256.3<br>29.1-508.1<br>21.6-802.7     | 42.4-130.6<br>50.9-145.9<br>--          | 41.6-132.8<br>64.5-137.6<br>101.7-147.1   |                   |              |

LIC - LSC = Límite de Confianza Inferior-Superior

sd = desviación estándar

CUADRO III. INTERVALOS DE CONFIANZA DE LA MEDIA ESTIMADOS EN LAS ESTACIONES SELECCIONADAS

| Código de estación                 | Datos originales                       | % datos censurados      | Valle central (chile). Intervalo de confianza del valor central |  |   |   | Estimación máxima verosimilitud                 | Regresión robusta                      | Kaplan-Meier                                    |  |  |  |
|------------------------------------|--|-------------------------|---|--|---|---|---|--|---|--|--|--|
|                                    |  |                         | Sustitución   |  | Límite de detección                             |   |   |  |   |  |  |  |
|                                    |  |                         | Valores sobre el límite de detección                            | 0  |   |   |   |  |   |  |  |  |
| Datos censurados                   |  |                         |   |  |   |   |   |  |   |  |  |  |
| A<br>media = 65.05<br>sd = 9.86    | N = 13<br>LJC = 60.91<br>LSC = 69.90   | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 46.64-70.48<br>41.41-67.46<br>22.34-55.41      | 54.33-69.81<br>52.6-69.03<br>43.92-62.81        | 61.01-69.5<br>64.54-70.26<br>66.17-70.72        | 58.55-71.18<br>57.62-71.33<br>53.07-69.82       | 60-70.51<br>62-71.68<br>59-73.31       | 60.53-70.49<br>64.24-70.99<br>66.03-71.59       |  |  |  |
| B<br>media = 46.85<br>sd = 12.60   | N = 13<br>LJC = 41.47<br>LSC = 52.23   | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 33.62-51.62<br>29.55-49.78<br>16.3-43.4        | 38.53-52.29<br>36.89-51.47<br>31.79-48.03       | 43.13-52.79<br>43.92-52.84<br>45.17-53.09       | 38.96-55.37<br>37.95-55.53<br>33.23-55.17       | 40-53.66<br>40-54.37<br>36-58.41       | 42.04-54<br>43.3-54.38<br>44.49-54.95           |  |  |  |
| C<br>media = 36.03<br>sd = 14.51   | N = 13<br>LJC = 29.41<br>LSC = 42.35   | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 24.49-41.95<br>22.37-41.77<br>15.14-36.41      | 27.94-41.58<br>26.41-41.61<br>24.72-40.89       | 29.33-42.76<br>30.49-42.82<br>34.75-44.5        | 27.38-46.64<br>26.5-47.5<br>22.24-50.8          | 27-44.25<br>27-44.86<br>29-49.49       | 28.19-44<br>29.17-44.27<br>35.5-46.29           |  |  |  |
| D<br>media = 8.85<br>sd = 3.36     | N = 13<br>LJC = 7.43<br>LSC = 10.42    | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 6.53-10.29<br>5.22-10.32<br>3.68-8.71          | 7.12-10.39<br>6.72-10.56<br>6.83-9.95           | 8.16-10.61<br>7.94-10.69<br>10.1-11.14          | 6.6-12.4<br>6.35-12.67<br>5.21-12.81            | 7-10.72<br>7-10.76<br>8-11.3           | 7.64-10.73<br>8.08-10.8<br>9.97-11.32           |  |  |  |
| E<br>media = 21.07<br>sd = 3.69    | N = 13<br>LJC = 19.42<br>LSC = 22.77   | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 14.99-22.81<br>12.74-21.54<br>7.38-18.27       | 17.79-22.53<br>16.89-22.33<br>14.6-20.62        | 20.09-22.8<br>20.36-22.88<br>21.33-23.33        | 18.69-23.32<br>18.35-23.3<br>16.75-22.89        | 19-22.98<br>19-23.15<br>18-23.19       | 19.81-23.15<br>20.02-23.21<br>21.27-23.72       |  |  |  |
| F<br>media = 127.31<br>sd = 66.27  | N = 13<br>LJC = 99.03<br>LSC = 158.11  | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 84.99-151.88<br>76.96-150.86<br>49.81-136.37   | 91.71-154.43<br>89.3-154.58<br>87.38-149.58     | 97.61-154.76<br>99.35-155.88<br>119.6-164.08    | 89.87-179.94<br>86.53-185.62<br>70.59-208.77    | 87-166.45<br>82-163.22<br>75-179.74    | 92.14-163.87<br>92.52-164.19<br>118.17-173.59   |  |  |  |
| G<br>media = 296.12<br>sd = 150.19 | N = 13<br>LJC = 232.84<br>LSC = 363.23 | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 218.37-364.23<br>202.84-355.68<br>136-321.09   | 225.05-360.25<br>218.29-359.43<br>227.63-351.3  | 235.66-363.36<br>243.7-366.2<br>318.71-384.5    | 193.52-549.24<br>185.78-594.9<br>142.27-687.76  | 228-375.42<br>248-399.37<br>310-421.35 | 222.95-378.89<br>271.98-385.27<br>370.33-414.56 |  |  |  |
| H<br>media = 171.06<br>sd = 121.43 | N = 13<br>LJC = 116.93<br>LSC = 226.06 | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 107.67-221.57<br>101.96-222.12<br>78.57-207.64 | 116.71-224.14<br>112.93-224.34<br>108.16-221.31 | 114.81-222.83<br>124.98-224.36<br>140.78-233.19 | 98.93-362.67<br>94.63-404.42<br>72.66-640.89    | 113-234.46<br>106-248.67<br>107-291.39 | 107.37-238.17<br>119.29-241.58<br>138.91-249.31 |  |  |  |
| I<br>media = 203.48<br>sd = 71.33  | N = 13<br>LJC = 170.7<br>LSC = 234.53  | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 138.4-228.77<br>125.88-224.31<br>78.78-201.15  | 158.78-232.33<br>149.68-230.5<br>135.04-218.62  | 176.72-235.27<br>173.38-232.96<br>193.1-243.8   | 161.69-250.31<br>156.96-252.84<br>134.32-257.32 | 157-248.26<br>158-247.7<br>135-272.11  | 167.67-243.15<br>175.07-245.38<br>188.93-251.38 |  |  |  |

**CUADRO II. INTERVALOS DE CONFIANZA DE LA MEDIA ESTIMADOS EN LAS ESTACIONES SELECCIONADAS**

| Código de estación                | Datos originales                       | % datos censurados      | Valle central (chile). Intervalo de confianza del valor central |  |   |   | Estimación máxima verosimilitud                | Regresión robusta                      | Kaplan-Meier                                    |  |  |  |
|-----------------------------------|--|-------------------------|---|--|---|---|--|--|---|--|--|--|
|                                   |  |                         | Substitución  |  | Límite de detección                             | Límite de detección                             |  |  |   |  |  |  |
|                                   |  |                         | 0   | 0.5 límite de detección                        |   |   |  |  |   |  |  |  |
| J<br>media = 150.88<br>sd = 58.07 | N = 13<br>LIC = 126.19<br>LSC = 177.03 | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 112.52-175.47<br>101.99-174.51<br>60.2-150.75  | 120.53-178.47<br>120.28-173.91<br>116.99-168.22 | 130.45-177.33<br>136.03-177.33<br>172.14-188.24 | 110.53-218.81<br>106.52-226.08<br>86.18-227.4  | 121-185.25<br>140-188.31<br>158-196.79 | 126.32-182.75<br>146.08-184.6<br>177.28-192.85  |  |  |  |
| K<br>media = 87.30<br>sd = 68.65  | N = 13<br>LIC = 55.26<br>LSC = 115.81  | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 48.8-114.62<br>45.67-110.8<br>30.47-101.86     | 51.84-114.99<br>49.11-115.83<br>48.21-111.36    | 56.42-114.79<br>58.52-115.79<br>62.27-118.29    | 53.59-138.82<br>51.23-146.16<br>40.45-185.28   | 53-129.75<br>52-122.5<br>39-133.75     | 50.54-125.21<br>52.83-126.37<br>56.42-129.16    |  |  |  |
| L<br>media = 70.22<br>sd = 28.11  | N = 13<br>LIC = 58.40<br>LSC = 81.94   | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 50.54-82<br>46.69-81.09<br>26.94-69.58         | 54.33-81.72<br>53.67-81.18<br>51.66-78.36       | 58.73-82.83<br>59.37-82.41<br>74.38-87.3        | 52.16-96.92<br>50.44-100.08<br>41.35-101.41    | 59-86.79<br>60-87.17<br>58-91.45       | 55.49-95.67<br>65.52-87.19<br>73.62-90.1        |  |  |  |
| M<br>media = 90.81<br>sd = 30.88  | N = 13<br>LIC = 78.00<br>LSC = 104.06  | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 63.81-104.19<br>56.96-102.97<br>36.12-89.44    | 72.19-104.25<br>66.57-103.02<br>68.82-100.32    | 78.11-105.0<br>78.46-104.2<br>99.74-111.65      | 71.39-113.92<br>69.25-115.52<br>58.92-118.82   | 70-109.18<br>73-111.44<br>86-117.48    | 74.59-107.84<br>74.62-107.94<br>99.82-113.96    |  |  |  |
| N<br>media = 42.48<br>sd = 25.62  | N = 13<br>LIC = 31.51<br>LSC = 54.06   | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 27.67-52.8<br>25.56-51.8<br>18.73-48.36        | 29.57-52.41<br>28.44-53.19<br>27.21-54.79       | 31.15-53.84<br>31.25-53.13<br>38.28-59.02       | 27.76-67.6<br>26.65-71.38<br>21.22-89.06       | 29-54.37<br>27-57.81<br>25-66.37       | 28.53-56.54<br>28.73-56.67<br>39.17-60.26       |  |  |  |
| O<br>media = 38.86<br>sd = 10.31  | N = 13<br>LIC = 34.43<br>LSC = 43.29   | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 27.92-43.05<br>23.56-40.66<br>15.34-35.88      | 30.8-42.32<br>29.59-42.06<br>25.72-39.88        | 34.17-43.24<br>34.83-43.28<br>37.33-44.42       | 32.62-45.02<br>31.84-45.21<br>28.09-45.52      | 32-45.05<br>31-44.45<br>29-49.18       | 33.36-44.54<br>33.72-44.67<br>36.3-45.61        |  |  |  |
| P<br>media = 71.11<br>sd = 28.26  | N = 13<br>LIC = 58.59<br>LSC = 84.00   | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 52.62-83.48<br>44.57-80.59<br>28.35-70.75      | 56.1-82.67<br>57.33-82.78<br>50.8-77.92         | 60.46-83.38<br>66.04-85.11<br>70.93-87.31       | 53.38-97.13<br>51.41-98.62<br>42.45-100.98     | 59-83.6<br>59-88.54<br>60-89.26        | 61.04-87.18<br>64.44-88.09<br>74.23-92.16       |  |  |  |
| Q<br>media = 216.72<br>sd = 68.16 | N = 13<br>LIC = 187.59<br>LSC = 245.52 | 15.38<br>23.07<br>46.15 | 11<br>10<br>7   | 157.77-246.64<br>138.63-237.57<br>82.74-208.34 | 173.54-246.03<br>168.11-246.11<br>154.46-231.14 | 60.46-83.38<br>66.04-85.11<br>70.93-87.31       | 172.85-270.17<br>167.9-274.03<br>143.25-275.32 | 175-257.19<br>189-258.87<br>206-279.24 | 183.92-254.34<br>201.08-257.03<br>240.41-269.02 |  |  |  |

detección, el sesgo, de menor magnitud, depende igualmente del porcentaje de censura, pero se produce de forma errática, por lo que no es posible predecir su dirección. Cuando el número de registros es inferior a 10, como es el caso de algunas de las estaciones de la Plana de Sagunto, este efecto es más acusado, aunque para porcentajes elevados de censura ninguno de los métodos proporciona estimaciones precisas a medida que disminuye el número de datos.

Los métodos de regresión robusta y de Kaplan-Meier, por otra parte, son los que proporcionan resultados más aproximados a los datos reales y con longitudes menores del intervalo de confianza correspondiente. A medida que aumenta el porcentaje de censura, sin embargo, la longitud de los intervalos obtenidos también aumenta.

## CONCLUSIONES

Los métodos de sustitución introducen sesgos en las estimaciones de la media o la mediana, por lo que deben evitarse, ya que distorsionan los intervalos de confianza generados y por tanto disminuyen la potencia de los ensayos de conformidad con los estándares ambientales. En otras palabras, aumenta la probabilidad de falsos negativos, con las implicaciones que ello supone en la toma de decisiones con relación a la calidad de los recursos hídricos. Lo mismo sucede con la eliminación directa de los valores por debajo del LC, ya que ello supone introducir (o eliminar) señales en el conjunto de los registros.

Los métodos basados en regresión robusta y el método de Kaplan-Meier son los procedimientos recomendados, ya que proporcionan una mayor precisión. Cuando hay varios límites de cuantificación como resultado de la mejora en la precisión de los métodos analíticos o debido al cambio en las técnicas de análisis o laboratorio, el segundo es el de referencia, ya que puede manejar varios procedimientos.

Por tanto, la utilización de dichos métodos implica un mayor grado de protección al medio ambiente, ya que indican que existe falta de verificación del estándar ambiental o de detección de fugas de compuestos peligrosos, antes que los otros métodos contemplados. Además, se recomienda un mínimo de 10-15 muestras por encima del límite de detección para obtener resultados precisos.

La frecuencia de muestreo es un factor limitante al momento de obtener conclusiones en series temporales, ya que frecuencias de muestreo inferiores a un cuatrimestre pueden producir efectos de autocorrelación en las medidas, especialmente en terrenos

de permeabilidad media-alta. A este respecto, cabe citar que los datos analizados en el Valle Central de Chile presentan *p*-valores inferiores a 0.05 en los ensayos de correlación de residuos mediante la prueba tau de Kendall.

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 (cinco valores o menos) no es posible aplicar de forma precisa alguna técnica de análisis de datos censurados. En este caso la sustitución por la mitad del límite de detección es la alternativa viable, al menos hasta que existan suficientes valores para realizar un análisis apropiado.

En definitiva, la aplicación de técnicas estadísticas específicas para el análisis de valores de concentración de contaminantes por debajo del LC, obtenidos en las redes de monitoreo de la calidad de los recursos hídricos, es una parte esencial del proceso de toma de decisiones en cuanto a la clasificación de un cuerpo de agua subterránea como en buen o mal estado. Los intervalos de confianza del valor central obtenidos de esta forma también pueden aplicarse cuando es necesario comprobar la efectividad de las medidas de recuperación de un cuerpo de agua subterránea, y no se tienen más elementos para comparar con el límite superior del intervalo de confianza.

## REFERENCIAS

- Antweiler R.C. y Taylor H.E. (2008). Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets. I. Summary statistics. *Environ. Sci. Technol.* 42 (10), 3732-3738.  
DOI: 10.1021/es071301c
- Arumí J.L. y Oyarzún R.A. (2006). Las aguas subterráneas en Chile. *Boletín Geológico y Minero* 117 (1), 37-45.
- Balderacchi M., Benoit P., Cambierb P., Ekloc O.M., Garinid A., Gemitzie A., Gurelf M., Kloveg B., Nakich Z., Predaai E., Ruzicich S., Wachniewj P. y Trevisana M. (2013). Groundwater pollution and quality monitoring approaches at the European level. *Environ. Sci. Technol.* 43 (4), 323-408.  
DOI: 10.1080/10643389.2011.604259
- Ballesteros B.J., Rodríguez L., López J., Grima J., García O., Cobos J.R., Gómez J.D. y 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 (IGME)-Diputación Provincial de Alicante. Fondo documental IGME, Alicante, España, 179 pp.
- Ballesteros B.J., Barba-Romero J. y Navarro J.O. (2005). Caracterización adicional Plana de Sagunto. Memoria. Trabajos técnicos para la aplicación de la Directiva Marco del Agua en materia de aguas subterráneas. Base documental. Dirección General del Agua-Instituto Geológico y Minero de España, Valencia, España, 21 pp.
- Blum A., Pauwels H., Wenland F. y Griffioen J. (2009). Background levels under the Water Framework Directive. En: Groundwater monitoring (Quevauviller P., Fouillac A.-M., Grath J. y Ward R., Eds.). John Wiley and Sons, Chippenham, Reino Unido, pp. 145-153. DOI: 10.1002/9780470749685.ch9
- Müller D. (2006). D18: Final proposal for a methodology to set up groundwater threshold values in Europe. Background criteria for the identification of groundwater thresholds. Proyecto 006538 (SSPI), Viena, Austria, 63 pp.
- Chacón E. y Míguez F. (1980). Estadística aplicada. Rúgarte, Madrid, España, 787 pp.
- Cleveland WS. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.* 74 (368), 829-836. DOI: 10.1080/01621459.1979.10481038
- DHJ (2014). Plan hidrológico de cuenca 2009-2015. Demarcación Hidrográfica del Júcar. Anejo 12: Evaluación del estado de las masas de agua superficial y subterránea. Informe. Consejo del Agua de la Demarcación, Valencia, España, 293 pp.
- Estrela T., Maestu J., Fidalgo A., Pérez M.A., Fullana J. y Pujante A.M. (2004). Júcar pilot river basin. Provisional article 5 report. Pursuant to the Water Framework Directive. Confederación Hidrográfica del Júcar. Valencia, España. 189 pp.
- Gleit A. (1985). Estimation for small normal data sets with detection limits. *Environ. Sci. Technol.* 19 (12), 1201-1206. DOI: 10.1021/es00142a011
- Golembeski R.C. (2004). Agricultural practices and nitrate pollution in ground water in the Central Valley of Chile. Tesis de Maestría. Universidad de Carolina del Norte, Wilmington, EUA, 79 pp.
- Grima J., López J. y Ballesteros B.J. (2002). Sustainable management of groundwater resources with regard to contaminated land. RTD needs. *Wit. Trans. Ecol. Envir.* 55, 279-289. DOI: 10.2495/BF020271
- Helsel D.R. y Hirsch R.M. (2002). Statistical methods in water resources. En: Techniques of water-resources investigations of the United States geological survey (Groat C.G., Ed.). United States Geological Survey, Virginia, EUA, pp. 1-510.
- Helsel D.R. (2005). Non-detects and data analysis: Statistics for censored environmental data. Wiley-Interscience, Hoboken, EUA, 250 pp.
- Montgomery R.H., Loftis J.C. y Harris J. (1987). Statistical characteristics of groundwater quality variables. *Ground Water* 25 (2), 176-184. DOI: 10.1111/j.1745-6584.1987.tb02874.x
- PE-CUE (2000). DOUE-L-2000-82524. Marco comunitario de actuación en el ámbito de la política de aguas. Directiva. Parlamento Europeo y Consejo de la Unión Europea, Luxemburgo, 73 pp.
- PE-CUE (2006). DOUE-L-2006-82677. Relativa a la protección de las aguas subterráneas contra la contaminación y el deterioro. Directiva. Parlamento Europeo y Consejo de la Unión Europea, Estrasburgo, 13 pp.
- Pizarro R., Salazar C., Bravo C., Caro J., Bravo P., Carrasco P., Sunkel O., Brzovic F. y Romaggi M. (1999). Informe país estado del medio ambiente en Chile. LOM Ediciones, Santiago, Chile, 410 pp.
- Preziosi E. y del Bon A. (2014). Natural background level assessment in groundwaters: probability plot versus pre-selection method. *J. Geochem. Explor.* 143, 43-53. DOI: 10.1016/j.gexplo.2014.03.015
- Saadati N., Abdullah M.P., Zakaria Z., Sany S.B., Rezayi M. y Hassonizadeh H. (2013). Limit of detection and limit of quantification development procedures for organochlorine pesticides analysis in water and sediment matrices. *Chem. Cent. J.* 7 (1), 63-72. DOI: 10.1186/1752-153X-7-63
- Singh R.S. y Totawattage D.P. (2013). The statistical analysis of interval-censored failure time data with applications. *Open Journal of Statistics* 3 (2), 155-166. DOI: 10.4236/ojs.2013.32017
- USEPA (2002). Estimation of the exposure point concentration term using a gamma distribution. Informe NTIS PB2003-106618. United States Environmental Protection Agency, Las Vegas, NV, EUA, 25 pp.
- USEPA (2006). On the computation of a 95 % upper confidence limit of the unknown population mean based upon data sets with below detection limit observations. Informe 68-W-04-005. United States Environmental Protection Agency, Las Vegas, N.V., EUA, 123 pp.
- USEPA (2009). Statistical analysis of groundwater monitoring data at RCRA facilities. Unified guidance. Manual. Office of Resource Conservation and Recovery, United States Environmental Protection Agency, Washington, D.C., EUA, 884 pp.