

Implementación de un modelo de predicción de fallos orientado a la gestión y estrategias de mantenimiento en redes de distribución de agua potable

Implementation of a pipe failure prediction model as a support tool for water networks management strategies

Ramírez, R.^{a1}, Cobacho, R.^{b2}, Torres, D.^b, López-Jiménez, P.A.^c

^{a1}ITA-Dpto. Ingeniería Hidráulica y Medio Ambiente. Universitat Politècnica de València
E-mail: ^{a1} roraag@posgrado.upv.es, ^{a2} rcobacho@ita.upv.es

^bGlobal Omnium
E-mail: datorres@globalomnium.com

^cDpto. Ingeniería Hidráulica y Medio Ambiente. Universitat Politècnica de València
E-mail: palopez@upv.es

Recibido: 30/07/2019

Aceptado: 07/10/2019

Publicado: 31/10/2019

Citar como: Ramírez, R., Cobacho, R., Torres, D., López-Jiménez, P.A. (2019). Implementation of a pipe failure prediction model as a support tool for water networks management strategies. *Ingeniería del agua*, 23(4), 247-258. <https://doi.org/10.4995/la.2019.12151>

RESUMEN

En el presente trabajo se plantea el doble objetivo de, por una parte, alcanzar unas estimaciones fiables para la vida útil de las tuberías que conforman la red de distribución de un abastecimiento y, por otra, hacerlo mediante un modelo de, relativamente, baja complejidad y con poca necesidad de datos históricos. Dicho modelo se ha basado en la utilización de la distribución Weibull sobre registros de dos años de reparaciones. Los resultados alcanzados se han contrastado con los proporcionados por el modelo Weibull de riesgos proporcionales, de amplia difusión en la actualidad, aplicado sobre los mismos conjuntos de datos. La comparativa entre los resultados de ambos modelos ha resultado positiva tanto en la ubicación sobre la curva de bañera obtenida para cada grupo de tuberías, como en el error relativo existente entre ambas estimaciones de fallos futuros.

Palabras clave | vida útil; red hidráulica a presión; distribución Weibull; modelo de riesgos proporcionales; curva de bañera.

ABSTRACT

The work presented in this paper has a dual objective: on the one hand, reaching reliable estimates for the pipes lifespan in a water distribution network; on the other, doing so by means of a relatively low complex model with relatively low need for historical data. The model is based on using the Weibull distribution over two years of failure records. The results achieved are compared with those provided by the well-known Weibull proportional risk model, that have been applied on the same data sets. The comparison is positive in both the location on the Davis curve obtained for each group of pipes, and the relative error existing between both estimates of future failures.

Key words | pipe lifespan; failure prediction; water networks maintenance; bath curve.

INTRODUCCIÓN

La gestión de activos en un sistema de redes de distribución de agua potable resulta clave para una correcta toma de decisiones y permite, además, aplicar las acciones más adecuadas en cuanto a la operación, mantenimiento y manejo de inversiones. En estos sistemas, las tuberías son elementos fundamentales que se pueden ver afectados por diversos factores que condicionan su vida útil de servicio. El estudio de los fallos y predicción de supervivencia de estos elementos ha sido abordado en diversos trabajos, desarrollándose diferentes tipos de modelos. A modo de ejemplo, Sægrov et al. (1999) y Kleiner y Rajani (2001) hacen un resumen de diferentes modelos de tipo estadístico y determinístico.

Frente a la complejidad explícita de los modelos actuales, en este trabajo se presenta un modelo (el denominado *modelo propuesto*) que busca exactamente ser lo contrario, y poder servir, de este modo, como primer paso a los abastecimientos que se estén iniciando en este campo o a la formación especializada de postgrado. Dicho modelo propuesto es una adaptación de conceptos propios de la ingeniería del mantenimiento industrial y tiene tres ventajas principales: en primer lugar, un requerimiento limitado y básico de datos (en el caso de estudio se ha trabajado con dos años de roturas); en segundo lugar, una complejidad de cálculo baja (un único tipo de ajuste estadístico); y en tercer lugar, la obtención de unos resultados de interpretación clara y directa: un parámetro representativo de la vida útil de las tuberías de la red, una predicción del número de fallos futuros (a corto plazo) y una valoración del estado general de las tuberías (curva de bañera) según quede caracterizada la ocurrencia de sus fallos.

Como se explica en la sección correspondiente del artículo, debido a la escasez de los datos utilizados, no ha resultado posible realizar una validación del modelo estadísticamente rigurosa, ni así tampoco un análisis de sensibilidad con agrupaciones diversificadas de tuberías. Desde una perspectiva meramente práctica y a modo de validación alternativa, se ha planteado la comparación de los resultados ofrecidos por el modelo propuesto con los proporcionados, sobre los mismos conjuntos de datos, por uno de los modelos de mayor difusión en la actualidad, cual es el de riesgos proporcionales de Weibull (WPHM, por sus siglas en inglés, Eisenbeis, 1994; Debón et al. 2010; Kimutai et al., 2015).

DESCRIPCIÓN DE MODELOS

Modelo propuesto

El denominado modelo propuesto es de tipo paramétrico y univariable, además de sencillo y fácil de implementar. Este modelo sigue de cerca la metodología de Síntesis Universal de Medición CMD propuesta por Mora (2009), y parte de las siguientes hipótesis:

- Las tuberías deben analizarse por grupos cuyas propiedades sean lo más homogéneas posible. Los datos básicos necesarios, para cada tubería, son la fecha de instalación, el material y el diámetro.
- En cada grupo, el número de tuberías con fallos registrados debe ser, al menos, de 30. Este valor, como mínimo significativo, viene exigido por el procedimiento de análisis estadístico utilizado (máxima verosimilitud, Levin y Rubin, 1996).
- Por la estructura de datos utilizada, se considera solo el primer evento de fallo que registra cada tubería.
- Los datos de fallos disponibles para el análisis, lógicamente, se ubican dentro de un denominado periodo, o ventana, de observación temporal. Así, este conjunto de datos debe considerarse truncado por su izquierda (ya que se desconoce si ha habido o no fallos entre el momento de la instalación de cada tubería y el inicio de la ventana de observación). Por su derecha, el conjunto de datos puede estar censurado (en el caso de no haber tenido fallos durante la ventana, esto expresa la potencialidad todavía de un fallo posterior), o no censurado (en el caso de haber tenido al menos un fallo durante la ventana, esto expresa la certidumbre de la ocurrencia del fallo registrado). Una información detallada sobre la censura estadística puede encontrarse en Solano (2008).
- Para cada tubería se contempla el análisis de dos variables: su Edad (E, tiempo transcurrido desde su instalación hasta su primer fallo) y su Tiempo Hasta el Fallo (THF, tiempo transcurrido desde el inicio de la ventana de observación hasta su primer fallo).

- Las distribuciones de las variables E y THF, para cada grupo de tuberías a analizar, siguen un patrón que puede ser representado mediante una distribución de tipo paramétrico.

La distribución Weibull es la predilecta a ser empleada en este tipo de análisis debido a su versatilidad para adaptarse a distintas configuraciones (Ebeling, 2019; Solano, 2008; Mora, 2009). Además, su función de riesgo permite representar las distintas fases de la curva de bañera. Otras distribuciones como la lognormal, gamma y normal, pueden ser consideradas en caso de que la distribución Weibull no se ajuste apropiadamente.

La distribución Weibull en su forma más general consta de tres parámetros: forma, escala y posición. El tercero de ellos representa el tiempo en que la probabilidad de fallo es nula y suele omitirse debido a la complejidad de su estimación, pasando en estos casos a definirse la distribución solo con los parámetros de forma y escala. A continuación, se muestran las funciones básicas que describen esta distribución.

$$f(t) = \frac{\beta}{\eta} \left(\frac{t-\gamma}{\eta} \right)^{\beta-1} e^{-\left[\frac{t-\gamma}{\eta} \right]^\beta} \tag{1}$$

$$F(t) = 1 - S(t) = 1 - e^{-\left[\frac{t-\gamma}{\eta} \right]^\beta} \tag{2}$$

$$\lambda(t) = \frac{\beta}{\eta} \left(\frac{t-\gamma}{\eta} \right)^{\beta-1} \tag{3}$$

Donde $f(t)$ es la función de densidad, $F(t)$ es la función de fallo (distribución acumulada de la densidad), $S(t)$ es la función de supervivencia, y $\lambda(t)$ es la función de riesgo o tasa de fallos.

El parámetro de escala η_E que se obtiene tras el ajuste de E, conocido también como vida característica, está relacionado con la vida útil. Representa el tiempo en el que se espera que falle el 63.2% de los elementos. Esto se deduce al igualar $(t-\gamma_E)=\eta_E$ en la función de distribución acumulada, y se cumple al margen del valor que tome β_E . Aunque la vida característica no es la media de la distribución Weibull, ésta puede interpretarse como un indicador aproximado de posición (Solano, 2008). Por lo tanto, un alto valor de este parámetro representará un mayor tiempo de duración o vida útil.

El parámetro de forma β_{THF} que se obtiene tras el ajuste de THF determina la forma de la función de riesgo de la distribución. Así, ésta es decreciente para $\beta_{THF} < 0.95$, constante para $\beta_{THF} \approx 1$, cóncava o ligeramente creciente para $1.05 < \beta_{THF} < 2$ y creciente para $\beta_{THF} > 2$. De esta manera es como, a través del análisis de THF, es posible cubrir la curva de la bañera en sus diferentes fases.

La metodología para el desarrollo del modelo propuesto consta de diferentes procesos, los cuales han sido resumidos en tres etapas (Figura 1). La Etapa 1 consiste en la obtención y preparación de datos para el análisis. A partir de esto, se conforma una nueva base de datos donde cada una de las tuberías de la red se encuentra identificada y acompañada de diferentes variables explicativas. A continuación, se seleccionan los grupos de análisis que se componen de tuberías de características homogéneas agrupadas bajo ciertos criterios de interés. Para ello, resulta útil realizar previamente un análisis estadístico descriptivo de la red.

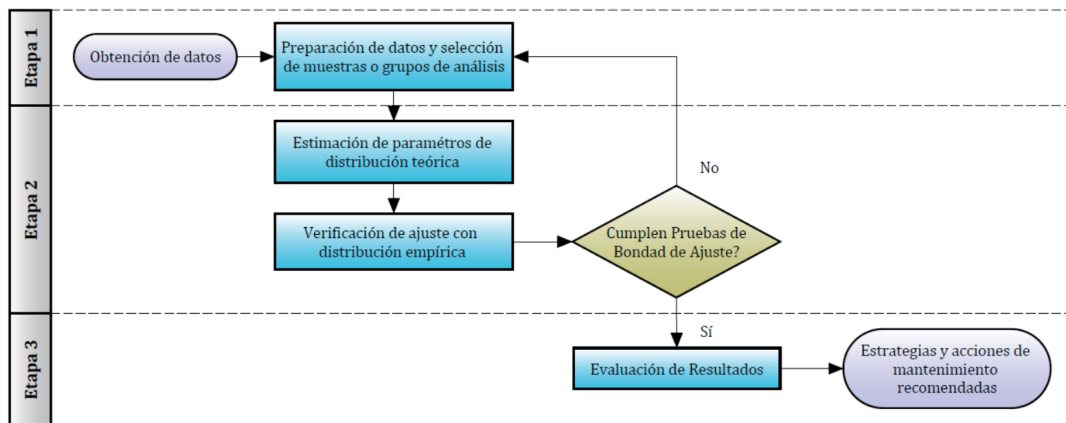


Figura 1 | Esquema de la metodología del modelo propuesto.

La Etapa 2 comprende el análisis estadístico de las variables a analizar (E y THF) para los grupos de tuberías seleccionados. Se calculan los parámetros de las distribuciones paramétricas por métodos de estimación de mínimos cuadrados o de máxima verosimilitud y se verifica el ajuste con las distribuciones empíricas. En caso de no ajustarse, se deberán reconformar los grupos y reiniciar el proceso de análisis, o bien, recolectar una mayor cantidad de datos hasta que estos puedan ser representados con una función paramétrica. La obtención estadística de los parámetros se ha realizado con ayuda del software de análisis estadístico R (Therneau y Lumley, 2015).

En la Etapa 3 se evalúan los resultados obtenidos para los grupos de tuberías analizados. De los resultados del análisis de la variable E se estima la vida útil. A partir de los resultados obtenidos del análisis de la variable THF se valora el estado de las tuberías y se predice el número de fallos esperados. Estos resultados deben ser finalmente abordados con el suficiente criterio técnico como para, a partir de ellos, proponer las estrategias o acciones de mantenimiento más idóneas a aplicar.

Modelo de riesgos proporcionales de Weibull (WPHM)

El WPHM es un modelo de tipo semiparamétrico y multivariable, también conocido como modelo de vida acelerada de Weibull. Una diferencia esencial con respecto al modelo propuesto es la variable de análisis. En el caso del WPHM dicha variable es τ . Tal como muestra la Figura 2, cada τ_i representa el tiempo transcurrido entre el inicio de la ventana de observación y el primer fallo (dato truncado), o entre un fallo y el siguiente (dato no censurado), o entre el último fallo y el final de la ventana de observación (dato censurado). Y cada valor τ_i se complementa con la información adicional (covariables) de la tubería en la que ha ocurrido. Dichas covariables pueden ser constantes (mismo valor para cada τ_i de una misma tubería, como el valor del diámetro) o dinámicas (distinto valor para cada τ_i de una misma tubería, como la denominada Edad Izquierda, EI_i).

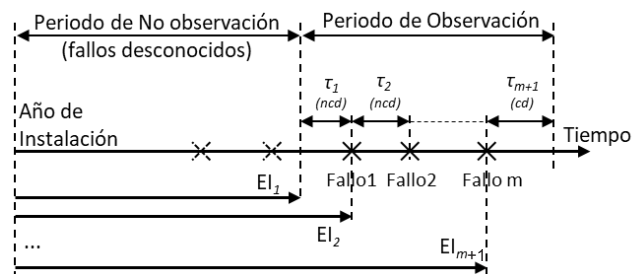


Figura 2 | Esquema de variables usadas para la estimar el logaritmo de la función de verosimilitud para el WPHM (Alvisi y Franchini, 2010).

A partir de este planteamiento (Le Gat y Eisenbeis, 2000; Solano, 2008), se asume que el vector de logaritmos naturales de τ se relaciona con un vector de covariables x a través de un modelo lineal válido y común para todas las tuberías (Ecuación 4 y Ecuación 5).

$$Y = \ln(\tau) = X\beta + \sigma W \quad (4)$$

$$\tau = e^{X\beta} e^{\sigma W} \quad (5)$$

Donde Y es el vector de los logaritmos naturales de τ , X es la matriz de covariables (de tantas filas como covariables y tantas columnas como valores τ_i haya en total), β y σ son los parámetros de ajuste de la regresión (siendo β el vector con los coeficientes de las covariables) y W es el vector de errores aleatorios, independientes e idénticamente distribuidos según una distribución de valores extremos, con una función de supervivencia denotada por S :

$$S(W) = \exp[-\exp(W)] \quad (6)$$

Como se ha indicado, los resultados de la regresión son los coeficientes que forman el vector β (siendo β_0 el primero de dichos coeficientes) y el parámetro σ . En particular, los parámetros β_0 y σ suelen denominarse *Intercept* y *Scale*, respectivamente. Se mantiene para los mismos esta notación por coherencia con las referencias citadas, pero es importante destacar que no son directamente iguales a los parámetros de forma ni escala ya presentados en el modelo propuesto. La equivalencia entre ellos se tratará más adelante. Finalmente, sustituyendo en la ecuación 6 la expresión de W derivada desde la Ecuación 4, se obtiene la función de supervivencia del WPHM:

$$S(\tau, \beta, x) = \exp\left[-\exp\left(\frac{\ln \tau - x^T \beta}{\sigma}\right)\right] = \exp\left[-\tau^{1/\sigma} \exp\left(-\frac{x^T \beta}{\sigma}\right)\right] \quad (7)$$

El nombre de modelo de vida acelerada se debe al efecto de aceleración o desaceleración que generan las variables explicativas sobre la variable τ . Así, las variables explicativas actúan multiplicativamente sobre τ o aditivamente sobre $\ln(\tau)$, (Solano, 2008).

Los parámetros principales de la regresión (*Intercept* y *Scale*) se estiman por el método de máxima verosimilitud. En síntesis (Le Gat y Eisenbeis, 2000), se aplica el logaritmo natural a la función de verosimilitud, expresada según la variable de tiempo entre fallos τ_i ($i=1,2,\dots,m$), e incluyendo los coeficientes de regresión β y las variables explicativas asociadas x (Ecuación 8). Para ello se ha utilizado el software R.

$$\ln[L(\tau_1, \tau_2, \dots, \tau_m, \beta, x)] = \sum_{i,cd} \ln(f(\tau_i, \beta, x)) + \sum_{i,cd} \epsilon(F(\tau_i, \beta, x)) \quad (8)$$

Finalmente, el proceso analítico para estimar el número futuro de fallos puede ser resuelto teóricamente mediante una distribución de convolución de sucesivos tiempos de fallo, lo cual resulta un proceso complicado. Por este motivo, se ha generalizado la utilización de un procedimiento aproximado mediante una simulación Monte Carlo. Para ello, a partir de la función de supervivencia (Ecuación 7), para cada tubería se generan tiempos de fallo aleatorios y luego se procede a ubicar cuántos de ellos están dentro del tiempo horizonte de estimación.

$$\tau = \left(\ln\left(\frac{1}{S}\right) \exp\left(\frac{x\beta}{\sigma}\right)\right)^\sigma \quad (9)$$

Durante el proceso de simulación Monte Carlo, se genera un número aleatorio entre cero y uno, que representa la supervivencia S (Ecuación 9). Así se genera el tiempo de fallo correspondiente, que, mientras no rebese el horizonte de la predicción, dará lugar a la generación sucesiva de nuevos tiempos. Esto continúa hasta que el tiempo obtenido para un fallo estimado sobrepase el horizonte de predicción. Durante este proceso, el número de fallos acumulados en cada tubería debe ser registrado, y en caso de existir variables dinámicas (como, por ejemplo, EI), su valor debe actualizarse continuamente. Este proceso se repite hasta 1000 veces para cada tubería, donde el valor promedio resultante se corresponde con el número de fallos esperados dentro del tiempo de estimación. Finalmente, para dicho periodo de estimación, desde el año T_0 hasta el año T_i , se suma el total de fallos de todas las tuberías (Ecuación 10).

$$M(T_0, T_i) = \sum_{p=1}^{n_p} m_p(T_0, T_i) \quad (10)$$

Donde M es el número de predicción de fallos para la red entera, T_0 es el año inicial del periodo de simulación, T_i es el año final de dicho periodo, m_p es el número de fallos para la tubería p , y n_p es número de tuberías de la red. Para conocer el número concreto de fallos previstos en un año determinado (por ejemplo, el último, T_i) basta con restar al total de fallos para todo el periodo (de T_0 a T_i), el total de estimados hasta desde el año inicial (T_0) hasta el anterior (T_{i-1}).

$$M(T_{i-1}, T_i) = \sum_{p=1}^{n_p} m_p(T_0, T_i) - \sum_{p=1}^{n_p} m_p(T_0, T_{i-1}) \quad (11)$$

CASO DE ESTUDIO

El presente estudio cuenta con datos del abastecimiento de una población mediterránea española, con más de 300 000 habitantes y 1000 km de red de tuberías. El sistema de gestión de información proporciona diferentes tipos de datos de la red, siendo necesario identificar aquellos datos completos que pueden afectar a la vida útil de los elementos. Para el caso de estudio se ha contado con una limitada cantidad de variables explicativas definidas para cada elemento de la red de distribución. Las principales han sido las siguientes: código del elemento, longitud, diámetro y fecha de instalación. Así mismo, se cuenta con un histórico de reparaciones correspondiente a los años 2017 y 2018. Este periodo es algo más corto que el considerado en otros trabajos ya citados, pero, como se verá, bien permite plantear y exponer el modelo propuesto y compararlo con el WPHM.

A efectos prácticos, la fecha de registro de reparaciones, se asume como fecha de ocurrencia de los fallos, en la cual se identifican los códigos de las tuberías afectadas. Esto permite cruzar información con la base de datos de tuberías existentes y definir otras variables o factores de influencia de fallos como la censura, tiempo entre fallos, etc.

La figura 3 describe someramente las características de la red de distribución y del histórico de fallos. La gráfica izquierda muestra los materiales de tuberías más representativos de la red, mientras que la gráfica derecha representa la frecuencia de fallo de los diferentes materiales para distintos rangos de edades.

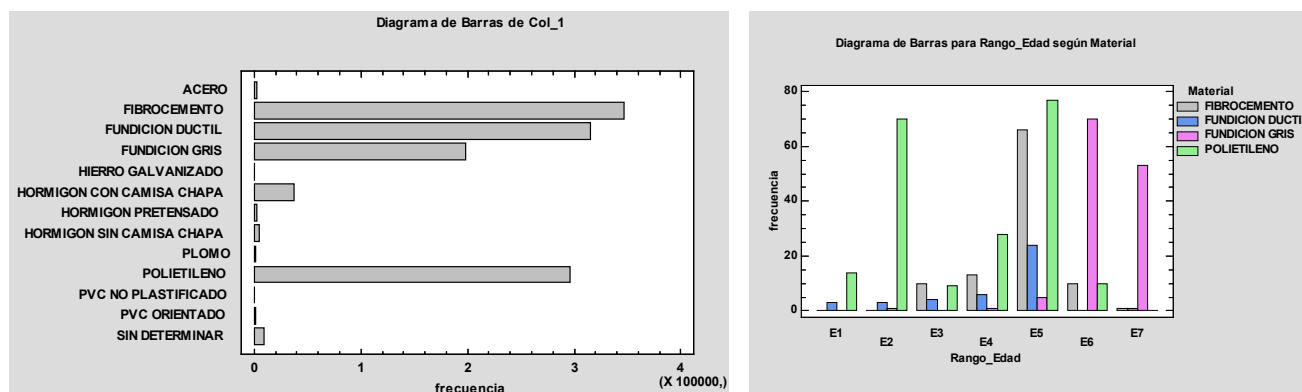


Figura 3 | (Izda.) Frecuencia de aparición de materiales tipo que componen la red de estudio en función de la longitud. (Dcha.) Elementos que presentan fallos por rango de edades y tipo de material.

Selección de grupos de análisis

Si bien la red está compuesta por una cierta diversidad de materiales, para este trabajo se han seleccionado los más significativos que cumplen con la cantidad mínima de ocurrencia de fallos para su análisis. Así, se parte de cuatro categorías de tuberías por material: fibrocemento (FC), fundición dúctil (FD), fundición gris (FG) y polietileno (PE). Esta estratificación responde además a la necesidad de diferenciar los procesos de rotura que ocurren en cada tipo de material, ya que se da por entendido que las variables explicativas no actúan de la misma manera en cada uno de ellos.

Para cumplir con pruebas de bondad de ajuste en el modelo propuesto, ha sido necesario, además, estratificar las tuberías por categorías de diámetro (Clase 2: diámetros entre 63 y 140 milímetros). Y en el caso de dos grupos concretos, también se ha necesitado acotar la edad de las tuberías analizadas. Todo ello se muestra en la tabla 1. Cabe subrayar que, para el método propuesto, es necesario contar con grupos o muestras de tuberías de características homogéneas. Por el contrario, esto no debe ser necesariamente así con el WPHM porque no requiere una estratificación ni agrupación de datos a tal nivel. Con todo, para poder contrastar los resultados de los dos modelos en este trabajo, ambos se han aplicado sobre los mismos grupos de tuberías.

Tabla 1 | Definición y características de los diferentes grupos de análisis

Grupo	Categ. Material	Categ. Diámetro	Edad (años)	Longitud Total (km)	Nº de 1 ^{os} fallos	Nº total de fallos
1	FC	2	≤ 50	156	52	64
2	FD	2	-	121	56	58
3	FG	2	≤ 60	84	36	51
4	PE	2	-	138	166	232

Para el modelo propuesto sólo se trabajará con los primeros fallos de cada tubería (6^a columna de la tabla 1), lo cual supone una cierta restricción, mientras que para el WPHM sí se tomará en consideración la totalidad de los fallos registrados (7^a columna de la tabla 1). Por otra parte, el cálculo de los parámetros de regresión de la distribución Weibull tanto para el modelo propuesto como para el modelo WPHM se ha realizado mediante el método de estimador de máxima verosimilitud MLE y la aplicación del software de análisis estadístico R.

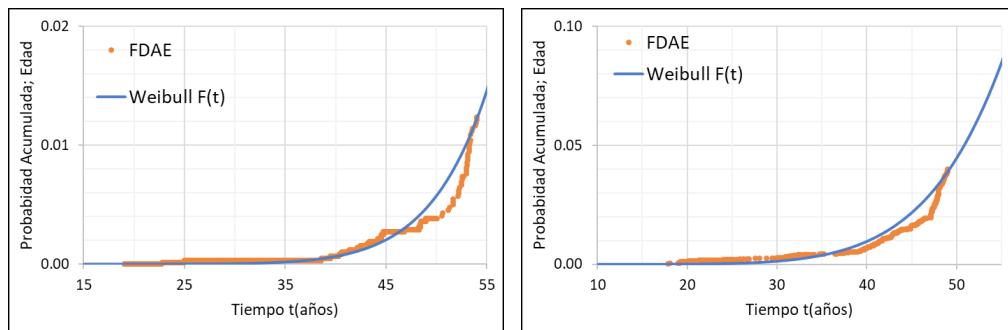
Resultados del modelo propuesto

Este modelo analiza las variables E y THF para llegar a una estimación del tiempo de vida útil, una valoración del estado de las tuberías y una predicción del número de fallos futuros. Para ello, se asume que ambas variables pueden ser representadas mediante una distribución Weibull, cuyo ajuste será verificado, mediante pruebas de bondad de ajuste, con respecto a la función de distribución empírica de los datos analizados.

La tabla 2 muestra los parámetros obtenidos para la variable E, a partir de los cuales es posible llegar a un estimador de la vida útil del conjunto de tuberías de los diferentes grupos. Dicho estimador es el parámetro de escala (η_E), que resulta en la misma unidad de tiempo que la variable analizada (años). Este valor está estrechamente relacionado con el valor medio de la distribución y, además, no se ve afectado por parámetro de forma (algo que sí ocurre con el citado valor medio). Específicamente, representa el tiempo en que se espera que falle el 63.2% de los elementos.

Tabla 2 | Parámetros y prueba de bondad de ajuste (Kolmogorov-Smirnov) para variable E (edad).

Grupo	Categoría Mat. / Diám. / edad	Parámetros		Prueba de bondad de ajuste K-S		
		β_E	η_E	Dmax	Dcr	Evaluación (Dcr > Dmax)
1	FC / 2 / ≤50	6.49	91.84	0.007	0.017	Cumple
2	FD / 2	9.82	84.48	0.003	0.014	Cumple
3	FG / 2 / ≤50	17.94	72.91	0.006	0.020	Cumple
4	PE / 2	6.96	77.90	0.010	0.018	Cumple

**Figura 4** | Ajuste de distribuciones paramétricas sobre la función acumulada empírica de datos observados para la variable E para los grupos 2 (izda.) y 4 (dcha.).

Por otro lado, los parámetros del ajuste de la variable THF permiten valorar el estado de la red y estimar el número de fallos futuros a corto plazo. La valoración del estado de las tuberías se realiza a partir del parámetro de forma (β_{THF}), del cual se deduce el valor de la tasa de fallos (Ecuación 3) y su ubicación entre las diferentes fases de la curva de bañera. A partir de ello, se puede definir entonces el tipo de acción de mantenimiento más adecuada. La tabla 3 muestra los valores de los parámetros estimados para la variable THF.

Tabla 3 | Parámetros y prueba de bondad de ajuste (Kolmogorov-Smirnov) para variable THF (tiempo hasta el fallo).

Grupo	Categoría Mat. / Diám. / edad	Parámetros		Prueba de bondad de ajuste K-S		
		β_{THF}	η_{THF}	Dmax	Dcr	Evaluación (Dcr > Dmax)
1	FC / 2 / ≤50	0.86	455.08	0.0010	0.017	Cumple
2	FD / 2	1.33	83.47	0.0008	0.014	Cumple
3	FG / 2 / ≤60	0.93	297.38	0.0003	0.020	Cumple
4	PE / 2	1.12	44.41	0.0030	0.018	Cumple

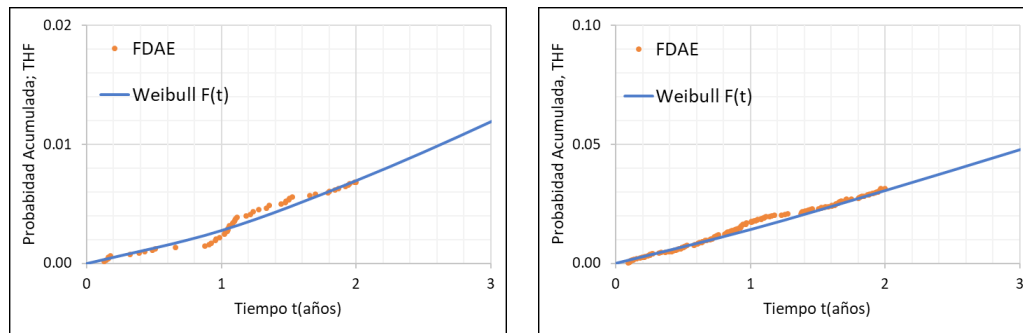


Figura 5 | Ajuste de distribuciones paramétricas sobre la función acumulada empírica de datos observados para la variable THF para los grupos 2 (izda.) y 4 (dcha.).

La estimación de fallos a futuro se realiza asumiendo que para el tiempo horizonte de predicción (5 años) se mantiene el comportamiento que describe la función paramétrica, es decir, el grupo de tuberías no va a variar su ubicación sobre la curva de bañera. Para ello, se calcula el valor de la función de riesgo $\lambda(t)$ para cada año y se aplica sobre el número total de tuberías del grupo (Tabla 4).

Tabla 4 | Predicción de fallos para distintos grupos de tuberías según el método propuesto.

Grupo	$\lambda(t) (\cdot 10^{-3})$					Nº fallos en cada año futuro					Fallos/km en cada año futuro				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	4.45	4.04	3.82	3.67	3.55	26	24	22	21	20	0.17	0.15	0.14	0.13	0.13
2	3.70	4.65	5.32	5.85	6.29	31	39	44	48	52	0.26	0.32	0.36	0.40	0.43
3	4.66	4.44	4.31	4.23	4.16	18	17	16	16	16	0.21	0.20	0.19	0.19	0.19
4	16.47	17.77	18.58	19.18	19.66	86	94	98	102	104	0.62	0.68	0.71	0.74	0.75

Resultados del WPHM

Este modelo sólo analiza la variable de tiempo entre fallos, τ , y la influencia sobre la misma de las covariables explicativas. Para ello se parte de una tabla, para cada grupo, que combina datos de tuberías y fallos, y en la que cada línea no corresponde a cada tubería del grupo, sino a cada tiempo de fallo de todas las tuberías, con todas las covariables correspondientes. Las tuberías que no han sufrido fallos también son incluidas, con tiempo de fallo igual a la duración total de la ventana de observación. La obtención de los coeficientes del ajuste, para cada grupo, se ha realizado mediante el software R y los resultados son los mostrados en la tabla 5.

Tabla 5 | Variables y parámetros estimados con el WPHM

	Grupo 1 (FC)		Grupo 2 (FD)		Grupo 3 (FG)		Grupo 4 (PE)	
	β_i	$\exp(\beta_i)$	β_i	$\exp(\beta_i)$	β_i	$\exp(\beta_i)$	β_i	$\exp(\beta_i)$
Intercept (β_0)	5.38	-	8.15	-	20.07	-	3.81	-
Scale	1.18	-	0.58	-	1.3	-	0.88	-
Ln Long. (β_1)	-0.206	0.813	-0.287	0.750	-0.174	0.839	-0.236	0.789
Diámetro (β_2)	0.006	1.006	-0.022	0.977	-0.065	0.937	0.016	1.016
Nº fallos (β_3)	-3.319	0.036	-2.353	0.095	-2.497	0.082	-0.728	0.482
EI (β_4)	0.041	1.042	-0.025	0.974	-0.109	0.896	-0.024	0.975
β_{THF} equivalente	0.85		1.72		0.77		1.14	

Siguiendo los trabajos descritos en Kimutai et al. (2015), los coeficientes β_i que más se alejan de cero, tanto positiva como negativamente, indican un mayor riesgo de fallo; mientras que el exponencial de cada coeficiente β_i indica la contribución de esa covariable al riesgo de fallo. Valores cercanos a 1 implican una contribución marginal.

La variable longitud ha sido introducida en el modelo aplicando el logaritmo natural sobre su valor, asumiendo así que la tasa de fallos es proporcional a la longitud de la tubería, tal como menciona da Costa (2011). Para los casos de tuberías con dimensiones menores a 1 metro, donde el logaritmo natural toma valores negativos, se ha restringido el valor mínimo que puede tomar esta covariable a 10^{-6} .

El signo negativo en el coeficiente del logaritmo natural de las longitudes se presenta en todos los grupos. Su interpretación es sencilla, ya que muestra la relación inversa entre la longitud de una tubería y la duración de su tiempo entre fallos. El signo positivo en los coeficientes del diámetro, grupos 1 y 4, indica que su tiempo entre fallos es mayor para tuberías con mayor diámetro, mientras que ocurre lo contrario para el caso de coeficientes negativos (grupos 2 y 3).

De los cuatro coeficientes analizados, el correspondiente a la covariable Nº fallos es el que muestra una mayor contribución al riesgo de fallo en las tuberías. Lo mismo se puede comprobar al ver el cálculo del exponencial del mismo, ya que se obtienen los resultados más lejanos a la unidad.

Le Gat y Eisenbeis (2000) recomiendan la incorporación de variables dinámicas relacionadas con la edad (como EI, con su coeficiente β_4) y el número de fallos previos (NOPF, según sus siglas en inglés), por considerarse de alta significancia. Sin embargo, el caso de la covariable NOPF puede derivar en problemas de cómputo al momento de realizar la simulación para la predicción del número de fallos. Así, puede ocurrir que los tiempos entre fallos que se generan sean cada vez más cortos, tendiendo a no alcanzar el tiempo horizonte de predicción y produciendo un número de fallos excesivamente alto, o bien, provocando que la simulación entre en un ciclo infinito. Esto se debe a la naturaleza exponencial de la ecuación utilizada para la estimación. Para evitar este problema Alvisi y Franchini (2010) omiten, directamente, el uso de esta covariable, mientras que da Costa (2011) propone reemplazar la covariable NOPF por $\ln(1+X_{NOPF})$, siendo X los valores que puede tomar la covariable NOPF. De manera alternativa, se puede considerar una covariable de tipo binario para forzar a la simulación a detenerse. En este trabajo se ha incluido la covariable Nº fallos (con coeficiente β_3), que representa el número de fallos durante el periodo de observación. Esto es una manera práctica de tomar en consideración la influencia del número previo de fallos, aunque, a diferencia del NOPF, se trate de una covariable constante y no dinámica.

Cabe hacer un apunte más sobre las diversas formas de parametrizar la distribución Weibull mediante al software *R*. Como se ha adelantado más arriba, la similar nomenclatura entre todas las variables utilizadas podría conducir a una cierta confusión, ya que la equivalencia entre los mencionados *Intercept* y *Scale* (y menos aún los β_i) para el WPHM no es directa respecto de los parámetros de forma (β_{THF}) y escala (η_{THF}) vistos para la variable THF del modelo propuesto, sino que se relacionan entre sí según las expresiones (Therneau y Lumley, 2015): $Scale = \frac{1}{\beta_{THF}}$; $Intercept = \ln(\eta_{THF})$.

En los resultados obtenidos (Tabla 5) para los grupos 2 y 4, correspondientes a FD y PE, el factor *Scale* resulta ser menor que la unidad. Esto supone valores del parámetro β_{THF} equivalente iguales a 1.72 y 1.14, respectivamente. Con ello se puede ubicar el estado de las tuberías de estos grupos en la etapa 1 de la fase 3 de la curva de bañera, es decir, la fase inicial de fallos por desgaste o envejecimiento. Por su parte, para las tuberías de FC y FG, se obtienen valores de β_{THF} equivalente iguales a 0.85 y 0.77, respectivamente, lo que ubica a estos grupos en la fase 1 de fallos tempranos.

Resulta paradójico el resultado para estos últimos materiales, FC y FG, sabiendo que son los más antiguos de los cuatro grupos. Para ello, pueden encontrarse, a priori, varias interpretaciones. Una de ellas, tal y como explican Le Gat y Eisenbeis (2000), puede deberse a que las tuberías más sensibles a los fallos de estos materiales ya han ido siendo sustituidas y desapareciendo de la base de datos, quedando solo las tuberías restantes, más antiguas y menos sensibles. En la misma línea ha podido influir también la restricción a la entrada en el grupo de las tuberías más viejas por cuestiones de bondad de ajuste estadístico de la distribución resultante. Todo esto puede verse reflejado para la FG en la tabla 1, donde se verifica que su longitud total es la menor respecto al resto de materiales. En el caso del FC, por su parte, presenta una incertidumbre mayor. Conviene tener presente, en todo caso, que en esta primera aproximación al problema se ha contado con una ventana de datos relativamente estrecha en comparación con la cantidad de datos truncados a la izquierda de la misma. Como conclusión, es razonable pensar, debido a la antigüedad total de estos materiales, que las tuberías de ambos grupos habrían realmente de ubicarse en una etapa más avanzada de desgaste de su vida útil, y en la medida en que mayores cantidades de datos vayan estando disponibles, esto se podrá confirmar con nuevos análisis.

Por último, la predicción del número de fallos se realiza siguiendo el procedimiento resumido más arriba. El resultado del número de fallos estimados para cada grupo (Tabla 6) y cada uno de los próximos 5 años se muestra en la tabla 6.

Tabla 6 | Predicción de número total de fallos utilizando WPHM.

Grupo	Categoría Mat. / Diám. / Edad	Nº fallos en cada año futuro					Fallos/km en cada año futuro				
		1	2	3	4	5	1	2	3	4	5
1	FC / 2 / <50 años	38	33	31	29	28	0.24	0.21	0.20	0.19	0.18
2	FD / 2	17	31	42	48	55	0.14	0.26	0.35	0.40	0.45
3	FG / 2 / < 60 años	29	25	25	26	27	0.35	0.30	0.30	0.31	0.32
4	PE / 2	113	129	138	143	150	0.82	0.93	1.00	1.04	1.09

DISCUSIÓN Y COMPARATIVA ENTRE LOS MODELOS

En general, un ajuste de modelos estadísticos debe complementarse con dos análisis adicionales. Por una parte, un análisis de sensibilidad, que cuantifique en qué medida una variación en los datos utilizados, o en la categorización hecha sobre los mismos, puede afectar al resultado del modelo; y por otra, una validación de los resultados (especialmente de las estimaciones a futuro), que confirme la mayor o menor exactitud de las mismas. La forma de plantear ambos análisis para el caso del modelo propuesto no es particularmente compleja. La sensibilidad del modelo puede valorarse planteando variaciones en el agrupamiento de tuberías y comprobando la alteración que ello supondría sobre los parámetros resultantes de forma y escala. Por su parte, la validación de las estimaciones se puede comprobar comparando dichas estimaciones con los registros reales de fallos en años sucesivos.

Como se ha adelantado más arriba, el modelo propuesto se ha aplicado en el abastecimiento del caso de estudio tan pronto como se ha dispuesto de una mínima cantidad de datos para ello (requiriendo, aun así, dos años de registros). Esta disponibilidad, rayana con la escasez, cuestiona la fiabilidad de un posible análisis de sensibilidad y, todavía, impide la validación a futuro. En cuanto al análisis de sensibilidad, la única manera de variar la agrupación de tuberías es ampliando los rangos de los grupos, especialmente con el diámetro (ya que de cada grupo actual no se pueden probar distintos subgrupos porque se han utilizado todos los datos disponibles). Y puede adelantarse con toda probabilidad que, el incluir en un mismo grupo tuberías de mayor variación en diámetro, va a aumentar su heterogeneidad interna, y a disminuir la fiabilidad de los resultados que se obtengan. En cuanto a las estimaciones a futuro, de nuevo, al haber utilizado la totalidad de los datos disponibles, no resulta posible, ajustar el modelo a partir de una selección de una parte de los mismos y comprobarlo con los datos no utilizados. En este caso, la validación podrá realizarse, como mínimo, cuando se conozcan los fallos registrados durante el presente 2019 o, con más fiabilidad, en un par de años más, ya que la estimación hecha abarca un horizonte de cinco años, y sobre todo, para los grupos 2 y 4, presenta una particular consistencia.

Por estas razones, la evaluación de los resultados proporcionados por el modelo propuesto se ha realizado mediante su comparación con los resultados del WPHM (del que ya se ha subrayado su amplia difusión actual) sobre los mismos datos. Dichos resultados son tanto la predicción del número de fallos para un tiempo horizonte, como la caracterización de la posición del grupo de tuberías sobre la curva de bañera. Para el primero de ellos, las respectivas predicciones se han mostrado en la tabla 4 y la tabla 6.

Para los grupos 1, 3 y 4, se aprecia que el modelo propuesto subestima el número de fallos respecto al WPHM. Para el grupo 2, por el contrario, hay una sobreestimación del número de fallos durante los primeros 3 años, mientras que para los siguientes dos años las predicciones prácticamente coinciden. Una posible explicación a esto es que el WPHM toma en consideración, mediante la covariable N° fallos, todos los fallos presentados en una tubería durante el periodo de observación, mientras que el modelo propuesto toma en consideración solo el primer fallo. La tabla 7 muestra el error relativo en la predicción del número de fallos para ambos modelos. En promedio (los cinco años para los cuatro grupos), las predicciones entre ambos grupos difieren en torno a un 20%. Si bien, desde un punto de vista estadístico, puede considerarse una diferencia estadísticamente significativa, desde un punto de vista práctico no es desproporcionada, sobre todo, tomando en cuenta que las alternativas al modelo propuesto son un modelo complejo como el WPHM o la ausencia completa de modelo. Tal variabilidad también podría estar relacionada con los datos en sí (problema ya apuntado), por lo que será preciso verificar la eficacia de dichas predicciones con nuevos datos observados tan pronto como sea posible.

Tabla 7 | Error relativo del modelo propuesto, respecto al modelo WPHM.

Grupo	Categoría Mat. / Diám. / Edad	Error relativo				
		T = 1	T = 2	T = 3	T = 4	T = 5
1	FC / 2 / <50 años	0.31	0.27	0.28	0.27	0.28
2	FD / 2	-0.80	-0.24	-0.06	0.00	0.05
3	FG / 2 / < 60 años	0.37	0.32	0.36	0.38	0.41
4	PE / 2	0.24	0.27	0.29	0.29	0.31

El segundo punto a destacar es que el factor de forma β resulta muy similar en ambos modelos (desviación promedio en torno al 12%), identificando de forma unívoca las mismas fases dentro de la curva de bañera para cada grupo. De aquí debe deducirse que, para este caso de estudio, ambos modelos describen de manera equivalente la función de supervivencia y riesgo de fallo.

Tabla 8 | Comparación del parámetro de forma beta para ambos modelos.

Grupo	Categoría Mat. / Diám. / edad	Parámetro de forma β		Fase (curva de bañera)
		Modelo Propuesto	WPHM	
1	FC / 2 / ≤ 50	0.86	0.85	Fase 1. Fallos tempranos
2	FD / 2	1.33	1.72	Fase 3. Fallos por desgaste
3	FG / 2 / ≤ 60	0.93	0.77	Fase 1. Fallos tempranos
4	PE / 2	1.12	1.14	Fase 3. Fallos por desgaste

CONCLUSIÓN

Un conocimiento suficientemente acotado de la vida útil de los distintos tipos de tuberías puede ser fundamental tanto para la gestión meramente técnica de la red, como, sobre todo, para su gestión económica. Las principales dificultades para ello, sin embargo, son dos: el disponer de un registro histórico de datos suficientemente amplio, y el contar con un modelo estadístico fiable que proporcione tales resultados.

Frente a modelos ya publicados, que presentan una complejidad significativa y, por tanto, una difícil barrera de entrada para abastecimientos que se acerquen por vez primera a esta problemática, en este trabajo se ha presentado un modelo simplificado (denominado modelo propuesto), resultado de adaptar técnicas de fiabilidad CMD (Mora, 2008) para componentes industriales al terreno de las tuberías de distribución de agua potable. Dicho modelo propuesto se ha aplicado a un caso de estudio con un histórico de datos reducido y se ha comparado con el modelo WPHM, como ejemplo de los actuales modelos complejos mencionados.

Los resultados obtenidos por ambos modelos y su comparación han arrojado conclusiones de interés. Es cierto que el modelo WPHM presenta la ventaja de no requerir una estratificación tan pormenorizada de las tuberías y de tomar en consideración variables explicativas que pueden afectar al proceso de fallo. Sin embargo, ello es a costa de un cálculo más complejo en dos etapas bien diferentes, ajuste estadístico multivariable y simulación, para las cuales se hace necesario software especializado, como

el R, y una cierta programación sobre bases de datos. Además, el tomar en consideración más factores, también implica mayores incertidumbres en datos y resultados, que se sitúan más allá de las necesidades iniciales de un abastecimiento en esta área.

Por su parte, el método propuesto destaca por la sencillez de su aplicación. Sólo necesita un único tipo de ajuste estadístico univariable que, aunque en este trabajo se ha hecho con R, bien podría acometerse con una hoja de cálculo. Cabe aclarar que, a pesar de su apuntada simplicidad, este ajuste implica una cierta exigencia en el campo de la estadística aplicada, y así están explicados cálculos y variables en este artículo. Con todo, no debe ser inasequible a la formación y capacidades ingenieriles de los profesionales en la gestión de un abastecimiento de agua. Además, las discrepancias encontradas entre los resultados de entre ambos modelos, están dentro de un margen asumible dadas las diferencias ya explicadas y el carácter inicial del trabajo hecho. Todo ello apunta, en definitiva, hacia nuevos desarrollos que contribuyan a mejorar estos primeros resultados y que, estando ya en marcha, serán objeto de futuras publicaciones.

AGRADECIMIENTOS

Los autores de este artículo desean agradecer el apoyo prestado, tanto de forma directa como a través de la Cátedra Aguas de Valencia de la UPV, por Global Omnium para el desarrollo de los trabajos aquí presentados.

REFERENCIAS

- Alvisi, S., Franchini, M. 2010. Comparative analysis of two probabilistic pipe breakage models applied to a real water distribution system. *Civil Engineering and Environmental Systems*, 27(1), 1-22. <https://doi.org/10.1080/10286600802224064>
- da Costa Martins, A. D. 2011. *Stochastic models for prediction of pipe failures in water supply systems*. Instituto Superior Tecnico, Lisboa, Portugal.
- Debón, A., Carrión, A., Cabrera, E., Solano, H. 2010. Comparing risk of failure models in water supply networks using ROC curves. *Reliability Engineering & System Safety*, 95(1), 43-48. <https://doi.org/10.1016/j.ress.2009.07.004>
- Ebeling, C. E. 2019. *An Introduction to Reliability and Maintainability Engineering*. Waveland Press.
- Eisenbeis, P. 1994. *Modelisation statistique de la prevision des defaillances des conduites d'eau potable*. Strasbourg 1.
- Kimutai, E., Betrie, G., Brander, R., Sadiq, R., Tesfamariam, S. 2015. Comparison of Statistical Models for Predicting Pipe Failures: Illustrative Example with the City of Calgary Water Main Failure. *Journal of Pipeline Systems Engineering and Practice*, 6(4), 4015005. [https://doi.org/10.1061/\(ASCE\)PS.1949-1204.0000196](https://doi.org/10.1061/(ASCE)PS.1949-1204.0000196)
- Kleiner, Y., Rajani, B. 2001. Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water*, 3(3), 131-150. [https://doi.org/10.1016/S1462-0758\(01\)00033-4](https://doi.org/10.1016/S1462-0758(01)00033-4)
- Le Gat, Y., Eisenbeis, P. 2000. Using maintenance records to forecast failures in water networks. *Urban Water*, 2(3), 173-181. [https://doi.org/10.1016/S1462-0758\(00\)00057-1](https://doi.org/10.1016/S1462-0758(00)00057-1)
- Levin, R. I., Rubin, D. S. (1996) *Estadística para administradores*; Prentice-Hall Hispanoamericana.
- Mora, L.A. 2009. *Mantenimiento planeación, ejecución y control*. Alfaomega Grupo Editor.
- Sægrov, S., Baptista, J. F. M., Conroy, P., Herz, R. K., LeGauffre, P., Moss, G., Schiatti, M. 1999. Rehabilitation of water networks: Survey of research needs and on-going efforts. *Urban Water*, 1(1), 15-22. [https://doi.org/10.1016/S1462-0758\(99\)00002-3](https://doi.org/10.1016/S1462-0758(99)00002-3)
- Solano, H. 2008. *Análisis de supervivencia en fiabilidad. Predicción en condiciones de alta censura y truncamiento: el caso de las redes de suministro de agua potable*. Riunet. Universitat Politècnica de València, Valencia (Spain). <https://doi.org/10.4995/thesis/10251/3792>
- Therneau, T. M., Lumley, T. 2015. *Package survival*. R Top Doc, 128.