UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE CAMINOS, CANALES Y PUERTOS

DEPARTAMENTO DE INGENIERÍA CIVIL:

HIDRÁULICA, ENERGÍA Y MEDIO AMBIENTE

TESIS DOCTORAL

**ANÁLISIS DE VARIABLES EXPLICATIVAS EN MODELOS DE
PREDICCIÓN DE ROTURAS EN REDES DE TUBERÍAS**

Autora

Patricia Gómez Martínez

Ingeniero de Caminos, Canales y Puertos (UPM)

Directores

Francisco J. Martín Carrasco

Dr. Ingeniero de Caminos, Canales y Puertos

Alfredo Granados García

Dr. Ingeniero de Caminos, Canales y Puertos

Mayo 2017

DEPARTAMENTO DE INGENIERÍA CIVIL: HIDRÁULICA, ENERGÍA Y MEDIO AMBIENTE
ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE CAMINOS, CANALES Y PUERTOS
UNIVERSIDAD POLITÉCNICA DE MADRID

TESIS DOCTORAL

**ANÁLISIS DE VARIABLES EXPLICATIVAS EN MODELOS DE
PREDICCIÓN DE ROTURAS EN REDES DE TUBERÍAS**

Autora

Patricia Gómez Martínez

Ingeniero de Caminos, Canales y Puertos (UPM)

Directores de Tesis

Francisco J. Martín Carrasco

Dr. Ingeniero de Caminos, Canales y Puertos

Alfredo Granados García

Dr. Ingeniero de Caminos, Canales y Puertos

Mayo 2017

Tribunal nombrado por el Sr. Rector Magnífico de la Universidad Politécnica de Madrid, el
día de de 2017.

Presidente:

Vocal:

Vocal:

Vocal:

Secretario:

Suplente:

Suplente:

Realizado el acto de defensa y lectura de la Tesis el día de de 2017, en
la Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos, habiendo
obtenido la calificación de

EL PRESIDENTE

LOS VOCALES

EL SECRETARIO

A mi padre, te echaré siempre de menos.

*A mi madre, por ayudarme, darme fuerzas y
enseñarme a ser mejor cada día.*

A mi hermano, no sé qué habría hecho sin ti.

*A Catalina, por devolverme la vida,
y a Arturo, por ser una nueva ilusión.*

A José Luis, por estar SIEMPRE.

AGRADECIMIENTOS

A través de estas líneas quiero expresar mi más profundo agradecimiento a todos aquellos que con su ayuda han hecho posible que desarrolle este trabajo y me han acompañado a lo largo del camino. En primer lugar a mis directores de tesis, Paco Martín Carrasco y Alfredo Granados, que desde el principio depositaron su confianza en mí y durante estos años me han orientado y apoyado.

En el ámbito laboral quiero dar las gracias a todos lo que han ido conduciendo mis pasos por el buen camino para terminar realizando esta investigación. En especial quiero agradecer el apoyo de Paco Cubillo, que fue el primero con el que empecé a hablar sobre mi idea de empezar la tesis, por su amistad y cercanía así como por su tiempo y asesoramiento en todo momento. También quiero agradecer a Luis Garrote su ayuda, disponibilidad y dedicación para orientar el trabajo. Y a Roberto Reglero, con quien me encontré un día por casualidad hace 10 años y despejó mis dudas sobre reconducir mi vida profesional hacia el mundo de la hidráulica.

Entre el ámbito profesional y personal, quiero dar las gracias a mi padre, Ricardo Gómez, que siempre fue y será mi ejemplo de Ingeniero de caminos, por la calidad humana y profesional que demostró y que aún hoy los que le conocieron me siguen transmitiendo. Ojalá estuvieras aquí, papá. Han pasado cuatro años y todos los días hay algo o alguien que me recuerda a ti.

En el terreno personal, mis padres María José y Ricardo han hecho posible que haya llegado el día en que termine la tesis. Ellos, siempre inseparables, me transmitieron los valores de esfuerzo, responsabilidad y sacrificio que han sido imprescindibles en este largo camino.

Y en estos últimos años mi agradecimiento más sentido a mi madre y a mi hermano Eduardo, que lo han sido todo. Mi hermano pendiente día a día, desde el país en el que estuviera, y mi madre, con su fuerza e intuición, nos ha motivado y orientado convirtiéndose en el pilar que nos sostiene frente a los imprevistos de la vida. También gracias a todos mis cuñados, y a mis suegros, por los momentos de descanso compartidos con ellos, que siempre han sido relajados y divertidos.

Le dedico también una línea a mi pequeña Catalina y a mi esperado Arturo, a los que no puedo querer más, gracias por dibujarme siempre una sonrisa.

Y a ti José Luis, que cambiaste mi vida desde que nos conocimos en aquella clase de Álgebra, gracias por todo. Tú sabes bien todo lo que esta Tesis ha supuesto, y sólo me queda decirte aquello que decidimos una vez, juntos hasta el punto de fuga.

RESUMEN

Esta Tesis plantea una metodología para la construcción de un modelo de predicción de roturas, basado en el análisis estadístico de los fallos de una infraestructura en servicio, que en este trabajo se particulariza para la red de distribución, y en su relación con las diferentes variables explicativas. Su aplicación permite identificar las variables con influencia significativa en la ocurrencia de roturas, cuantificar su repercusión en la capacidad predictiva de los modelos que se generen y seleccionar aquél modelo con mayor capacidad de predicción de las roturas de entre las diferentes opciones que se planteen según las múltiples combinaciones de variables consideradas.

Con el tiempo y el uso, las infraestructuras de abastecimiento se van deteriorando y van requiriendo cada vez mayores inversiones en reparación. En concreto, la red de tuberías presenta un número creciente de fallos anuales que hace que los gestores del abastecimiento busquen en todo momento reducir la frecuencia de las averías y roturas, siendo éste uno de los mayores retos a los que se enfrentan. La estimación de la vida útil de los elementos de la red requiere un criterio fiable para poder plantear adecuadamente las estrategias de renovación y mantenimiento.

Ante los diferentes planteamientos de modelos de predicción de roturas, ya sean físicos o estadísticos, determinísticos o probabilísticos, con diferentes formulaciones matemáticas, la presente Tesis evidencia la dependencia estadística de las roturas de tuberías respecto de las variables explicativas, y determina la combinación más adecuada para la construcción de modelos predictivos a través de la cuantificación de su influencia en la capacidad para predecir los fallos. La característica del planteamiento radica en que para la construcción de los modelos se seleccionan las variables predictivas apoyándose únicamente en un estudio estadístico objetivo de las roturas, basado en la evidencia científica sin adoptar ideas preconcebidas sobre la influencia de las diferentes variables en la ocurrencia de roturas.

La metodología planteada se divide en tres partes; en la primera de ellas se identifican las variables de las que dependen las roturas de la red de distribución. Para ello se toman todas las variables que a priori podrían tener significancia en la ocurrencia de roturas a partir de la literatura de referencia, y se realiza un test de significancia Kolmogorov-Smirnov que permite confirmar si existe relación de dependencia.

En la segunda parte se plantea la construcción de múltiples modelos predictivos basados en un análisis Bayesiano cuyo objetivo es identificar la probabilidad de ocurrencia de roturas como función de las diferentes variables explicativas seleccionadas en la etapa anterior. Para ello se analiza la función de distribución de la probabilidad de ocurrencia de las variables explicativas en una base de datos genérica y en una asociada a las roturas en la red de distribución. El proceso de generación de modelos va incrementando secuencialmente el orden de los mismos, partiendo del modelo más simple de orden uno, con una única variable, e incorporando paso a paso variables de manera adicional. De entre todos los modelos generados en un mismo orden, al incorporar una variable adicional se seleccionan los diez mejores; sobre esta selección se plantean todos los modelos de orden inmediatamente superior posibles al introducir una nueva variable explicativa independiente; este proceso se realiza con todas las variables explicativas disponibles. Se evalúan los modelos y se vuelven a elegir los diez con mejor comportamiento. El proceso prosigue incorporando variables de manera secuencial hasta que no se observa una mejoría significativa en el comportamiento entre un modelo y el de orden superior de acuerdo con el proceso de validación que se describe a continuación.

La combinación de variables en la generación de los diferentes modelos se realiza desde dos perspectivas, por un lado se plantean modelos con todas las variables independientes, y por otro se plantean modelos en los que dos variables se analizan de manera conjunta y el resto independientes.

La tercera parte de la metodología consiste en la evaluación de los modelos generados de forma objetiva y sistemática mediante la definición de un procedimiento de validación que cuantifica la calidad de los mismos en base a su capacidad de predicción de las roturas. Dicho planteamiento exige un análisis de sensibilidad previo que permita conocer la incertidumbre de los resultados de la validación para garantizar la bondad de la evaluación, y la minimización de la influencia de los parámetros que definen el procedimiento como son el tamaño de la muestra, el número de muestras y el periodo de ajuste.

El resultado esperado de la aplicación de la metodología es la definición de la mejor combinación de variables en el modelo predictivo de mayor calidad, es decir, aquél en el que se observa la mejor capacidad de predicción de las roturas.

Esta metodología se aplica a la red de distribución de la Comunidad de Madrid (España) gestionada por Canal de Isabel II. Aprovechando los datos disponibles en dicha red se analiza la influencia de las variables explicativas y se identifican los mejores modelos

predictivos. En este sentido, las variables con mejor comportamiento por sí solas son diámetro, material y año de instalación.

El análisis de los modelos desarrollados revela que si bien a medida que se incorporan variables explicativas de manera secuencial se va mejorando la capacidad predictiva, no todas las combinaciones de variables lo consiguen.

También se observa que la mejora se va suavizando por lo que no siempre compensará establecer un modelo predictivo con el mayor número de variables posibles pues la mejora en la predicción resulta insignificante de los modelos de mayor orden frente a la complejidad que implica el tratamiento de un mayor número de variables. Esta idea se traduce en que determinados modelos con un menor número de variables explicativas suponen mejores soluciones, es decir que no se observan ventajas relevantes al establecer modelos contruidos con un gran número de variables explicativas.

Los resultados obtenidos con los modelos de orden cuatro y orden cinco son muy similares lo que no justifica la aplicación de modelos más complejos simplificando así el tratamiento y la predicción de las roturas con menos variables, cuatro para el caso de estudio.

ABSTRACT

This Thesis proposes a methodology for the construction of a break prediction model based on the statistical analysis of the failures of an asset in service, in this work it is focused on the distribution network, and its relation with different explanatory variables. Its application allows the identification of variables with a significant influence on the occurrence of breaks, quantifying their impact on the predictive capacity of the generated models. It also allows selecting the model with the greatest break prediction capacity from the different options that arise from the multiple combinations of variables considered.

As aging water supply assets deteriorate, they require increasing investments for its repair and maintenance. Particularly the pipeline network has a growing number of annual failures that make water providers to seek for reducing the frequency of breaks and bursts, which is one of the biggest challenges they have to face. The estimation of the useful life of the network's elements requires a reliable criterion for proposing suitable renewal and maintenance strategies.

Given the different approaches of breaks prediction models, whether physical or statistical, deterministic or probabilistic, with different mathematical formulations, the present Thesis shows the statistical dependence of the pipe breaks on the explanatory variables. It also determines the variables' optimal combination for the construction of predictive models through the quantification of their influence on the models' ability to predict failures. The characteristic of the proposed methodology is that the predictive variables for constructing the models are selected according to an objective statistical study of the registered breaks. Hence, the models are built from scientific evidence without adopting preconceived ideas about the influence of the different variables in the occurrence of breaks.

The methodology is divided into three parts; in the first one, the variables on which the network breaks depend are identified. To do so, all the variables that might have significance in the occurrence of breaks from the reference literature are taken, and a Kolmogorov-Smirnov significance test is performed to confirm the dependence relation.

The second part proposes the construction of multiple predictive models based on a Bayesian analysis, with the aim of identifying the probability of occurrence of breaks as a

function of the different explanatory variables selected in the previous stage. For this purpose, the distribution functions of the likelihood of occurrence of the explanatory variables in a generic database and in one linked to the breaks in the distribution network are analyzed. The model generation process sequentially increases the order of the models, starting from the simplest model of order one, with a single variable, and incorporating step by step additional variables. From all the models generated in the same order, the ten best ones are selected. For each one of them, every additional variable is incorporated as independent variable to the model forming a higher order model. The models are then evaluated and the ten best performers are selected again. The process continues to incorporate variables sequentially until no significant improvement in behavior between a model and the higher order model is observed in accordance with the validation process described below.

The combination of variables in the generation of the different models is done from two perspectives, on the one hand models are built with all the independent variables, and on the other, models are built with two joint variables and the rest are treated as independent ones.

The third part of the methodology consists on the evaluation of the generated models from an objective and systematic approach. Therefore, a validation procedure that quantifies their breaks prediction capacity is proposed. This procedure firstly requires a sensitivity analysis that allows to know the uncertainty of the validation results in order to guarantee the suitability of the evaluation and to minimize the influence of the validation parameters that define the procedure. Such parameters are sample's size, number of samples and the adjustment period.

The expected result of the methodology is the definition of the best combination of variables for the most reliable predictive model. That model will be the one with the best pipe break predictive capacity observed within the multiple analyzed models.

This methodology is applied to the distribution network of the Madrid's region (Spain) managed by Canal de Isabel II. Taking advantage of the available networks data, the influence of the explanatory variables is analyzed and the best predictive models are identified. Hence, the variables with the best behavior acting independently are diameter, material and year of installation.

The analysis of the obtained models reveals that as explanatory variables are sequentially incorporated the predictive capacity is improved, but not all the variables combinations improve that capacity.

It is also observed that the improvement is reduced with each iteration. This means that it will not be worthy to apply a predictive model with the greatest number of possible variables since the improvement in the prediction is insignificant compared to the complexity that the treatment of a greater number of variables implies. Therefore, certain models with a smaller number of explanatory variables represent better solutions because no relevant advantages are observed when considering models built with a large number of explanatory variables.

The obtained results with the order four and order five models in the case study are very similar, because of that, the application of more complex models is not justified. This all results on a simplification of the treatment and prediction of the breaks with less variables, four for the case study.

ÍNDICE

1	INTRODUCCIÓN	1
1.1	Motivación	1
1.2	Objetivos	5
1.3	Organización del documento	6
2	ESTADO DEL ARTE	7
2.1	Eficiencia en la gestión de redes de distribución	7
2.2	Deterioro de las redes de distribución	11
2.3	Estudio de las roturas	12
2.4	Tipos de modelos predictivos	14
2.5	Modelos Bayesianos	22
2.6	Aplicación: modelos de vida útil	23
2.7	Resumen	25
3	METODOLOGÍA	27
3.1	Selección de variables	27
3.1.1	Variables explicativas principales	28
3.1.2	Variables explicativas secundarias	28
3.2	Dependencia estadística de las roturas respecto de variables	31
3.2.1	Fundamento teórico del test Kolmogorov-Smirnov	32
3.2.2	Metodología del test	32
3.2.3	Proceso de cálculo	33
3.3	Construcción de modelos	34
3.3.1	Probabilidad condicionada – Modelos bayesianos	34
3.3.2	Influencia de la agregación de variables	37
3.3.3	Construcción de modelos	39
3.4	Procedimiento de validación	41
3.4.1	Metodología	41
3.4.2	Valor medio de los radios de calidad – R1	44

3.4.3	Radio correspondiente a los valores medios de los parámetros – R2	45
3.5	Análisis de sensibilidad	47
3.5.1	Comparación cualitativa	48
3.5.2	Comparación cuantitativa	50
4	PARTICULARIZACIÓN DE LA METODOLOGÍA A LA RED DE CANAL DE ISABEL II	53
4.1	Red de distribución del Canal de Isabel II	53
4.1.1	Descripción de la red	53
4.1.2	Datos	53
4.2	Registros de roturas	55
4.3	Selección de variables	56
4.3.1	Variables explicativas principales	56
4.3.2	Variables explicativas secundarias	57
4.4	Dependencia estadística	58
4.5	Variables consideradas	58
5	ANÁLISIS DE SENSIBILIDAD PARA LA RED DE CANAL DE ISABEL II	61
5.1	Análisis del tamaño de las zonas de validación	61
5.1.1	Tamaño de las zonas de validación - Análisis cualitativo	61
5.1.2	Tamaño de las zonas de validación - Análisis cuantitativo	69
5.2	Análisis del tamaño de la muestra	72
5.2.1	Tamaño de la muestra – Análisis cualitativo	73
5.2.2	Tamaño de la muestra – Análisis cuantitativo	78
5.3	Análisis del periodo de validación	82
5.3.1	Comparación en dos periodos	82
5.3.2	Comparación cuantitativa	90
5.4	Resultado del análisis de sensibilidad	98
6	DEFINICIÓN DE LOS MODELOS PARA LA RED DE CANAL DE ISABEL II	99
6.1	Modelos de orden cero	99
6.2	Modelos de una variable	99
6.3	Modelos de orden dos	103
6.3.1	Modelos de dos variables conjuntas	103
6.3.2	Modelos de dos variables independientes	111
6.3.3	Resumen modelos de dos variables	119
6.4	Modelos de orden tres	120

6.4.1	Modelos de dos variables conjuntas y una independiente	121
6.4.2	Modelos de tres variables independientes	129
6.4.3	Resumen modelos de tres variables	137
6.5	Modelos de orden cuatro	139
6.5.1	Modelos de dos variables conjuntas y dos independientes	139
6.5.2	Modelos de cuatro variables independientes	147
6.5.3	Resumen modelos de cuatro variables	154
6.6	Modelos de orden cinco	155
6.6.1	Modelos de dos variables conjuntas y tres independientes	156
6.6.2	Modelos de cinco variables independientes	164
6.6.3	Resumen modelos de cinco variables	171
6.7	Mejores modelos	173
6.8	Nivel de desagregación	176
7	CONCLUSIONES Y APORTACIONES	179
7.1	Conclusiones	179
7.2	Aportaciones originales de la Tesis	185
7.3	Futuras líneas de investigación	186
	BIBLIOGRAFÍA	189
	ÍNDICE DE FIGURAS	197
	ÍNDICE DE TABLAS	205
	ANEXO 1: RANGOS DE LAS VARIABLES EXPLICATIVAS EN EL CASO DE ESTUDIO	207
	ANEXO 2: ANÁLISIS DE DEPENDENCIA DE VARIABLES EXPLICATIVAS	215
	ANEXO 3: RESULTADOS COMPLETOS DE TODOS LOS MODELOS PROPUESTOS	227

1 INTRODUCCIÓN

1.1 Motivación

El aumento de las necesidades de agua potable como consecuencia del crecimiento y desarrollo socioeconómico de la población, unido a la naturaleza finita de los recursos hídricos y a la obsolescencia de las infraestructuras, hace necesaria una mejor administración de los recursos, realizando una gestión y un uso eficiente de los mismos. Y para alcanzar una gestión eficiente, deben desarrollarse sistemas e instrumentos que optimicen el uso sostenible de los recursos disponibles.

La Comisión Europea considera desde hace varios años (EC, 2007) que la eficiencia y el ahorro es la primera estrategia a considerar para hacer frente a los problemas cada vez más frecuentes de escasez, y en 2010 (EC, 2010) estimó que en Europa se podría reducir el consumo total de agua en torno al 40%. Se necesitan medidas e instrumentos para mejorar la eficiencia en el uso del agua que apoyen el desarrollo económico a la vez que conserven los sistemas naturales (EEA, 2012). Tanto los recursos disponibles como la demanda de agua son variables en el tiempo, y ambos son susceptibles de sufrir cambios que no siempre pueden ser controlados por el gestor de la red. La disponibilidad de los recursos depende de las aportaciones hídricas y de las políticas adoptadas. En su gestión influyen ciertos factores no modificables que están relacionados con la distribución de las precipitaciones y las temperaturas, el comportamiento y uso de la cubierta vegetal o la recarga de acuíferos. Por otra parte, la demanda de agua varía en función de los hábitos y actividades de la población. Sin embargo, una de las medidas propuestas por la AWWA (2002) para gestionar la demanda y mejorar la garantía de suministro es disponer de redes estancas que minimicen las pérdidas. Parte de este consumo se encuentra asociado a las pérdidas de agua vinculadas a las roturas en las redes de distribución cuya ocurrencia tiende a aumentar con la edad de las tuberías si no se toman medidas. Este hecho pone de manifiesto la preocupación real existente entre las diferentes organizaciones nacionales e internacionales para la mejorar la operación y gestión de las redes de abastecimiento como estructuras que mejoran el uso del recurso. Resulta necesario investigar sobre prácticas y técnicas que permitan aumentar la eficiencia en la operación y gestión de los recursos hídricos, materiales y económicos respecto de los planteamientos actuales.

Los gestores de las redes de distribución centran sus esfuerzos en el diseño de estrategias que introduzcan mejoras, tanto en la conservación del recurso hídrico como en la conducta empresarial. Por un lado, buscan mejorar la garantía de suministro, entendida ésta como la capacidad para satisfacer la demanda manteniendo un determinado nivel de servicio. Y por otra parte, se persigue una mejor planificación de las inversiones en infraestructuras para tomar decisiones sobre las tareas de inspección, mantenimiento o renovación de la red que garanticen el suministro en todo momento de acuerdo a las condiciones de calidad preestablecidas.

Sin embargo, los sistemas de abastecimiento se componen de un conjunto complejo de infraestructuras que conectan las fuentes con el usuario final, y presentan importantes riesgos para asegurar el suministro. El agua, según sea su origen superficial o subterráneo, es captada mediante embalses o pozos, y es conducida hasta las estaciones de tratamiento de agua, de donde es conducida a través de una red de tuberías con estructura ramificada, generalmente conocida como red de aducción (Canal de Isabel II Gestión, 2012) hasta los depósitos o válvulas de control, para después ser transportada hasta los puntos de consumo y distribuida mediante las acometidas domiciliarias. Este último recorrido se realiza mediante una extensa red de tuberías que tienen dos funciones claramente diferenciadas: la red de transporte, constituida por tuberías troncales o arterias principales que generalmente tienen mayores diámetros y no presentan conexiones para satisfacer la demanda en su trayecto; y la red de distribución que suele discurrir por los núcleos urbanos con un trazado mallado de la que se deriva la red de acometidas para dar servicio a los clientes en cada punto de consumo.

Con el tiempo y el uso las infraestructuras de abastecimiento se van deteriorando y van requiriendo cada vez mayores inversiones en reparación. En concreto, la red de tuberías presenta un número creciente de fallos anuales que hace que los gestores del abastecimiento busquen en todo momento reducir la frecuencia de las averías y roturas. Garantizar el suministro implica cumplir con unos requisitos de calidad tanto en tiempo como en espacio, es decir de manera continua en cualquier punto de la red. Y para ello, minimizar el número de roturas es esencial por sus consecuencias en términos de continuidad, ya que los clientes perciben considerables interrupciones del servicio en función de los tiempos de identificación, localización y reparación de averías, que genera numerosas reclamaciones durante las operaciones y maniobras que se realizan hasta que se restablece de nuevo el servicio. También existe un riesgo potencial de contaminación del agua durante las tareas de reparación de las averías ante la posible intrusión de contaminantes y además se incurre en elevados costes sociales, medioambientales y económicos para los usuarios y las compañías de agua. Se presentan costes sociales

asociados a los daños a terceros, inundaciones, posibles cortes de tráfico, a la generación de ruido excesivo en la zona, costes de imagen de las gestoras del servicio y costes medioambientales asociados al volumen de agua perdida en roturas y averías y costes económicos vinculados a los gastos que conlleva la propia reparación por las tareas de localización y reparación que varían según las características de la rotura y el área de afección.

La red de tuberías representa un valor importante sobre el coste total de los sistemas de distribución por lo que su adecuada gestión es vital para las compañías de abastecimiento (Folkman 2012). Renaud et al. (2012) estiman que su valor representa más del 80% del valor total de las infraestructuras de distribución. Los beneficios más inmediatos de la predicción de las roturas, y en consecuencia de la anticipación a los fallos en la red, son la reducción de los costes de reparación, la reducción de los costes de agua perdida en las roturas, y en consecuencia la disminución del agua a suministrar con el consiguiente ahorro en la captación, tratamiento, transporte e impulsión cuando fuera necesario de dichos volúmenes de agua, la reducción de los costes de inspección para la localización de pérdidas por el mayor conocimiento sobre el deterioro de la red, la reducción de coste de avisos y quejas de clientes por suministrar un servicio más constante y fiable, y por último la mayor eficiencia de las inversiones en red y de los programas de renovación.

Las redes de distribución requieren continuas inversiones para la reposición de ciertos tramos de red, para el mantenimiento de la red y los equipos de control, o para disponer de nueva instrumentación como válvulas, caudalímetros, manómetros y estaciones de muestreo, para el desarrollo, calibración y mantenimiento de modelos hidráulicos de gestión de la red que simulen parámetros de presión, caudal, velocidad, tiempo de permanencia, arrastres de sedimentos y cloro residual, pero un porcentaje importante de dicha inversión se destina a la reparación de las roturas (AWWA, 2013).

Según sean el estado de conservación de la red, los condicionantes técnicos y operacionales, y la frecuencia de las roturas, los gestores deben tomar decisiones sobre si proceder a la reparación o a la renovación del tramo correspondiente. Para garantizar la adecuada prestación del servicio, anticiparse a los fallos resulta esencial. Sin embargo, conocer el estado en que verdaderamente se encuentra la red y predecir su comportamiento en el tiempo son tareas complicadas dado que las tuberías se encuentran enterradas generalmente.

Existen multitud de variables explicativas de las roturas en la red de distribución, y su estado puede ser muy distinto según la zona en la que se encuentren ya que se han visto afectadas por diversos aspectos que han ido condicionando su diseño en el tiempo. Por

una parte, existen condicionantes físicos y sociales, derivados de la propia ubicación y localización de los asentamientos, así como del crecimiento y desarrollo de los núcleos urbanos; y por otra parte, condicionantes técnicos y económicos derivados de las necesidades, las circunstancias y las prácticas de instalación y mantenimiento del pasado.

Esta tesis se centra en la búsqueda y desarrollo de instrumentos de mejora del conocimiento del sistema de distribución de agua mediante el desarrollo de modelos predictivos de roturas que posibiliten la adopción de las medidas adecuadas antes de que se desencadenen los fallos. En concreto se analiza la red de tuberías de distribución que satisface la demanda urbana, entendiendo esta última como la demanda del conjunto de usos residenciales, industriales, comerciales y públicos.

El objetivo de la investigación es evidenciar y cuantificar la dependencia estadística de las variables explicativas seleccionadas en las roturas a partir de un registro histórico de roturas para poder definir su combinación más adecuada en la construcción de un modelo de predicción de roturas en la red de distribución. Se busca desarrollar una guía que contemplando dichas variables explicativas permita desarrollar un estudio objetivo en el que se propongan modelos de predicción de roturas sin adoptar ninguna conclusión preconcebida, cuyas conclusiones surjan del mero análisis de los datos.

De esta forma se aporta una metodología que permite construir un modelo de predicción de roturas que apoye a los gestores del abastecimiento en sus decisiones sobre inversión para la gestión y el mantenimiento de la red. El procedimiento propuesto se inicia con la selección e identificación de las variables explicativas que influyen en el fenómeno de las roturas, en un segundo paso se plantean múltiples modelos predictivos combinando de diferentes formas las variables identificadas y por último se evalúan los resultados de cada uno de los modelos con el objetivo de seleccionar los que presenten una mayor capacidad predictiva de las roturas en las redes de distribución.

El desarrollo de nuevas técnicas de predicción de fallos contribuye tanto la reducción de las pérdidas de agua, a la mejora en la prestación del servicio por la minimización de los riesgos de discontinuidad y de contaminación potencial del agua y a la optimización de las inversiones permitiendo un mayor rendimiento hidráulico y económico de la red.

En la medida en que se dispone de datos más precisos sobre el estado de conservación de las tuberías se podrán establecer estrategias de gestión, mantenimiento y rehabilitación más adaptadas a la realidad de la red. Esto se traduce en una mejora en la eficiencia de las medidas tomadas, mejorando los resultados buscados con cada una de ellas.

1.2 Objetivos

El objetivo principal de esta Investigación es aportar una metodología para la construcción de un modelo de predicción de roturas, basado en el análisis estadístico de roturas en la red de distribución y su relación con las diferentes variables explicativas. También es objeto de este trabajo su aplicación a un caso de estudio real, con el fin último de definir la combinación de variables explicativas más adecuada que aporte un modelo con la mejor capacidad predictiva. El procedimiento metodológico propuesto consta de dos partes, una primera fase en la que se identifican las variables explicativas de las roturas y se construyen los modelos, y una segunda en la que se validan los modelos mediante dos métodos diferentes.

La construcción de modelos predictivos de roturas ayuda a los gestores del abastecimiento en su toma de decisiones sobre la planificación de inversiones en la red de distribución. Dada la importancia de esta herramienta de gestión de las redes, las metas concretas a alcanzar en la presente Tesis para la consecución del objetivo principal son las siguientes:

- a) Identificar las variables de las que dependen las roturas de la red de distribución, analizando su influencia en la ocurrencia de roturas. Para ello se requiere manejar diferentes fuentes de información con el fin de tener datos suficientes de todas las variables explicativas contempladas en este estudio.
- b) Definir un procedimiento para la construcción de modelos que pronostiquen las roturas. Se propone una formulación de modelos a partir de las variables previamente identificadas.
- c) Definir un procedimiento para evaluar los modelos generados de forma objetiva y sistemática. Para ello se realiza un análisis de sensibilidad que garantice la bondad de la evaluación resultante del procedimiento planteado.
- d) Evaluar los resultados que se obtienen con modelos contruidos con diferentes variables mediante aplicación directa en un caso real; determinando el modelo con un mejor comportamiento.

1.3 Organización del documento

La presente Tesis doctoral consta de siete capítulos. En primer lugar se hace una Introducción en la que se describe la motivación y se enumeran los objetivos que se quieren alcanzar con este trabajo.

En el segundo capítulo se aborda el Estado del Arte, que presenta la importancia de realizar una gestión eficiente de las redes de distribución, se expone la problemática asociada al deterioro de las mismas y se presentan los distintos estudios sobre la ocurrencia de las roturas y los diferentes modelos de predicción de fallos que se encuentran en la bibliografía.

El tercer apartado contempla la metodología desarrollada, que consta de tres partes diferenciadas. La metodología comienza por la selección de las variables explicativas y el estudio de la dependencia estadística de las roturas respecto de dichas variables, en segundo lugar se propone un método para proceder a la construcción de los modelos y por último se define el procedimiento de validación de los modelos, que incluye un estudio de sensibilidad a partir del cual se definen los requisitos a considerar en este último procedimiento para asegurar su fiabilidad.

Los capítulos cuarto, quinto y sexto plantean la aplicación de la metodología al caso de estudio de la red de distribución del Canal de Isabel II. En el capítulo cuarto se exponen las características de la red contemplada y se seleccionan las variables a considerar en los modelos que se generen para esta red. En el capítulo quinto se define el procedimiento de validación que será de aplicación para esta red en particular; y en el capítulo sexto se generan los modelos, se analizan los parámetros de calidad de cada uno de ellos tras aplicar el procedimiento de validación y se concluye definiendo los mejores modelos.

Por último, en el capítulo séptimo se incluyen las conclusiones del trabajo desarrollado, se presentan las aportaciones originales de la Tesis y se proponen futuras líneas de investigación.

2 ESTADO DEL ARTE

2.1 Eficiencia en la gestión de redes de distribución

Los servicios urbanos de abastecimiento deben procurar la máxima eficiencia con el fin de asegurar en todo momento un nivel de servicio adecuado (Cabrera-Béjar y Gueorguiev-Tzatchkov, 2012) que cumpla los requisitos establecidos por la legislación vigente.

Las redes de distribución de agua son, por lo general, grandes sistemas que demandan continuas mejoras para optimizar su diseño, y para alcanzar o mantener adecuados niveles de eficiencia (Geem, 2015). En la actualidad, la eficiencia es un parámetro clave para evaluar la gestión y operación de una red, y mejorar la eficiencia en la prestación y gestión del servicio de abastecimiento se ha convertido en una de las prioridades de los organismos, tanto públicos como privados, que gestionan las redes de distribución de agua (Akbarkhiavi y Imteaz, 2016). Jones and Sowby (2014) establecen que las prácticas para la optimización de los sistemas de abastecimiento deben estar orientadas a la eficiencia energética, al aseguramiento de la calidad del agua y al cumplimiento de un determinado nivel de servicio. Sin embargo, con el paso del tiempo las infraestructuras se van deteriorando y se incrementa el riesgo de que se produzcan fallos del sistema que condicionan la eficiencia de una red, ocasionando pérdidas de agua considerables e importantes costes económicos y sociales, que a su vez condicionan la calidad del servicio prestado (Kleiner et al., 1998). Estos factores pueden estar asociados a las características intrínsecas del sistema de distribución, por particularidades físicas como la edad de las infraestructuras o el diseño y construcción del mismo, o con condiciones operacionales como las altas presiones o la deficiente realización de las tareas de operación y mantenimiento. Todo ello hace que se requieran nuevas e importantes inversiones para garantizar la prestación del servicio en condiciones adecuadas y que sea necesario disponer de herramientas que ayuden a los gestores en la toma de decisiones gestionando la incertidumbre asociada a la planificación a medio o largo plazo (Campbell, et al., 2016). Todas ellas pasan por una mejor comprensión del funcionamiento del sistema y de los objetivos de gestión de los operadores del servicio.

Para optimizar el funcionamiento del sistema y maximizar la eficiencia global de la red los gestores del servicio buscan técnicas para la mejora de la certidumbre de la red, la adecuada gestión de presiones, la mejora en la calidad del agua y la mejora de balances

hídricos, y la garantía de abastecimiento frente a contingencias, operaciones y maniobras, y todo ello de acuerdo con un adecuado compromiso coste beneficio.

En la actualidad existen diversas herramientas que posibilitan el conocimiento más detallado de las redes distribución y su funcionamiento en el tiempo. Por una parte, el empleo de cartografía actualizada refleja los aspectos técnicos de la red y los sistemas de información recogen su estado hidráulico. Con el seguimiento y actualización periódica de estos datos se puede identificar el funcionamiento de la red y establecer las medidas correctoras necesarias para cumplir con las prestaciones requeridas.

Por otra parte, uno de los aspectos que caracterizan las redes de abastecimiento actuales es la falta de un patrón de diseño común, ya que las redes existentes son el resultado de diversos procesos de diseño a lo largo de su historia. Esta realidad dificulta un mantenimiento eficiente de las mismas y plantea la necesidad de establecer su reordenación. Así se han ido desarrollado nuevos planteamientos y actuaciones basadas en la división y monitorización de dichas redes en zonas o sectores de menor tamaño controlados, que han sido denominados en algunos contextos como District Metric Area (DMA).

Esta sectorización de las redes de distribución proporciona un mayor conocimiento de las mismas ya que al quedar subdividida en zonas más reducidas se simplifican las labores de diagnóstico y gestión de su funcionamiento (Gómez et al., 2014). Además, la combinación de esta solución con las diferentes técnicas de automatización y control de la red permite realizar un seguimiento en tiempo real sobre su funcionamiento hidráulico. Así, los datos recopilados pueden ser empleados por el gestor de la explotación para identificar anomalías y discrepancias con el estado óptimo de funcionamiento, y se facilita la toma de decisiones por parte de los operadores (Ulanicki, et al., 2003) ya que se pueden programar de forma más sencilla las tareas de mantenimiento, inspección o renovación de la red, minimizando las consecuencias de estas operaciones sobre la continuidad del servicio. De forma análoga, se puede lograr la reducción de pérdidas de agua no controlada y la localización de fugas (UKWIR, 1999), (IWA Loss-Group, 2007) así como una disminución en los tiempos de respuesta ante roturas, puesto que el menor tamaño de las zonas telecontroladas facilita su identificación, localización y cuantificación. Ciertos autores (Di Nardo, et al., 2013) mencionan que la sectorización permite una gestión adecuada las presiones en la red, con la consiguiente reducción de la presión máxima. Por un lado posibilita la regularización de las presiones estáticas; por otra parte, permite el control de la presión dinámica en el interior de cada sector hidráulico en función del caudal; y por último facilita la reducción de la presión máxima en la red. De ello se derivan

importantes mejoras como la reducción de las pérdidas de agua, la disminución de las roturas y averías en la red y en consecuencia el incremento de la vida útil de las tuberías, aunque también presenta posibles inconvenientes asociados a la disminución de presiones, como los cambios en la presión que puede percibir el consumidor final o la capacidad de la red para garantizar los caudales necesarios para la protección contra incendios. No obstante, a pesar de la mejora en la gestión, operación y control del abastecimiento, esta reorganización de las redes no aporta ventajas en cuando a la tarea de diagnóstico del deterioro y predicción de roturas. Por ello se requieren nuevas técnicas para la anticipación a los fallos en el sistema que se fundamenten en el conocimiento detallado de los procedimientos de envejecimiento y de las causas de fallo de las tuberías, y permitan valorar la vida útil de las infraestructuras para tomar las decisiones oportunas en cada momento.

A continuación se describe cómo la mejora en la predicción de las roturas contribuye a una mayor eficiencia de la red, y se alinea con los objetivos de gestión de los operadores.

Mejora de la certidumbre

La mejora en la capacidad predictiva de los fallos en la red contribuye a mejorar el nivel de certidumbre de la misma en cuanto al conocimiento de su estado de deterioro esperado y a las herramientas que se desarrollan para su planificación y gestión (MacGillivray, et al., 2006). Esto debería traducirse en una mayor eficiencia en la toma de decisiones sobre las inversiones en la red (Gomes et al., 2012). No obstante se requiere contrastar las herramientas de forma periódica a partir de información real y actualizada de campo.

Gestión de presiones

La gestión de presiones es un objetivo clave desde el punto de vista de la mejora de la eficiencia en las redes de distribución por las múltiples ventajas asociadas (Akbarkhiavi y Imteaz, 2016), y en este sentido la disminución de las roturas posibilita un régimen de presiones más homogéneo en el que se produzcan menos oscilaciones. Además, en la medida en que se reducen en el sistema los valores máximos de presión, se consigue una reducción tanto del volumen de pérdidas por fugas como de nuevas averías o roturas (Gomes et al., 2013) (Perez, et al., 2009), (Martínez-Codina et al., 2015), que en muchas ocasiones se encuentran ligadas a las maniobras en la red durante la reparación. Así se alarga la vida útil de las tuberías y se logra una mayor eficiencia de la red.

Mejora de la calidad del agua

La garantía de la calidad del agua es un objetivo destacado en la prestación del servicio, y en este sentido, la disminución de las roturas contribuye a reducir el riesgo potencial de contaminación del agua durante las tareas de reparación o renovación. Diversos autores han estudiado la variabilidad en la concentración de cloro residual para diferenciar las condiciones normales de las anomalías en la calidad del agua (Nejjari, et al., 2012).

Mejora de balances hídricos

La anticipación a las roturas aporta ventajas en la auditoría de las redes de distribución al minimizar el volumen de agua perdido en las fases de detección y localización de cada fallo. A partir de los datos obtenidos del balance hídrico de las redes de distribución de agua, se pueden establecer rendimientos que permiten evaluar la eficiencia del sistema y con ello la definición de un plan de acción que proponga, en base a criterios preestablecidos, las actuaciones a llevar a cabo para mejorarlo (Cabrera et al, 1999). El mayor conocimiento del funcionamiento de la red y la mejora en el balance hídrico como consecuencia de la disminución de las averías permitirá desarrollar auditorias con mayor precisión para conocer la eficiencia hidráulica, caracterizar mejor la curva de demanda, detectar fraudes, fugas (Li, et al., 2011) o diversos errores de medida y reducir los costes de mantenimiento (Herrera, et al., 2010). Las pérdidas de agua por fugas y roturas suponen costes económicos, energéticos y de tiempo que deben ser considerados para procurar una gestión eficiente del sistema ya que pueden afectar a su viabilidad financiera por las pérdidas de ingresos y el incremento de los costes operacionales (Adlan, et al., 2013).

Garantía de abastecimiento frente a contingencias, operaciones y maniobras

Los gestores de abastecimiento necesitan conocer el riesgo de no cumplir con un determinado nivel de servicio (Kandulu, J., 2013). El avance en el conocimiento de la red a través de la anticipación a los fallos permite una mejor planificación de las inversiones en infraestructuras para garantizar el suministro en todo momento de acuerdo a las condiciones de calidad preestablecidas. De esta forma se mejorará el nivel de servicio en términos de continuidad del servicio prestado.

Optimización de las inversiones

Una mejor predicción de las roturas permite una mayor eficiencia a través de la optimización de las inversiones de mantenimiento y renovación lo que permite maximizar los beneficios para un mismo nivel de inversión pues esta se podrá adaptar mejor a las

eventuales roturas. Se consigue por tanto un mayor conocimiento a pesar de las incertidumbres que redunda en una mejor aproximación a la vida útil de los elementos de la red (Savic, 2005), (Babayan, et al., 2005). Esto permite adaptar mejor el momento de renovación al final de la vida útil de los diferentes componentes maximizando así el periodo útil de las infraestructuras, y minimizando la inversión resultando por tanto en un beneficio para todos los involucrados ya sea el operador de la red o el usuario final.

2.2 Deterioro de las redes de distribución

El deterioro progresivo de los sistemas de abastecimiento se traduce en un aumento de la frecuencia de las roturas en las redes de distribución que tiene un impacto negativo en la calidad del servicio prestado, con consecuencias sobre la continuidad del suministro y la calidad del agua, además de los costes sociales, medioambientales y económicos asociados a su localización y reparación (Kunkel et al., 2008). El envejecimiento de las infraestructuras es una preocupación actual de los países desarrollados pero probablemente se convierta en un tema de estudio a nivel mundial en el futuro próximo (Hukka & Katko, 2015).

Anticipándose al fallo de las redes de distribución existentes se pueden obtener importantes mejoras en cuanto a su gestión, mantenimiento, operación y control. Mediante la construcción de modelos capaces de predecir las roturas, los gestores del abastecimiento lograrán una mayor eficiencia en la prestación del servicio, que se traducirá en un doble beneficio tanto para la propia empresa como para el usuario final. Por una parte, la empresa obtendrá mayores beneficios procedentes de una mejor planificación de las operaciones de mantenimiento e inspección de la red, la reducción de costes asociada a las roturas y del uso adecuado de recursos hídricos, materiales y humanos, cumpliendo e incluso excediendo sus expectativas. Y por otra parte el cliente se beneficiará de niveles de calidad superiores.

Generalmente la ocurrencia de roturas aumenta con la edad de las tuberías y los elementos de la red (Heywood & Starr, 2007). Sin embargo, mientras que algunos componentes pueden estar en servicio periodos superiores a los de diseño, otros elementos de edades inferiores presentan unas elevadas tasas de fallo y deben ser renovados pronto (Large et al., 2015). Esto puede explicarse por el hecho de que los fallos en las tuberías dependen de múltiples factores difíciles de caracterizar cuantitativamente (Babovic et al., 2002). El envejecimiento de las infraestructuras lleva asociado un deterioro natural pero hay algunos otros parámetros como la corrosión, la carga que

deben soportar las tuberías ya sea por presión exterior o asientos del terreno, la longitud y diámetros de la tubería, el material, la calidad de la instalación o incluso el propio histórico de roturas pueden influir los procesos de roturas (Boxall et al., 2007).

Por todo ello el conocimiento del grado de deterioro de las redes de distribución resulta de gran importancia para la gestión de las redes de abastecimiento. De hecho, algunas investigaciones (Xu et al., 2013), (Malm et al., 2012) se centran en la determinación de una estrategia de reposición basada en modelos de predicción de roturas. Según Pelletier et al. (2003) la identificación y comprensión de los factores que intervienen en el deterioro de las tuberías constituye un paso esencial para poder establecer una adecuada política de gestión de las infraestructuras de abastecimiento minimizando los impactos que las posibles incidencias puedan producir sobre el nivel de servicio prestado rentabilizando al mismo tiempo las inversiones destinadas a la reposición de las mismas. Según sus estudios, los síntomas de deterioro pueden ser muy variados identificando el aumento de las fugas de agua, la disminución de la capacidad hidráulica de las tuberías o el empeoramiento de la calidad del agua pero, de entre todos, el más que resulta más fácil de identificar y analizar es la frecuencia de las roturas en las redes de distribución.

En esta línea, algunas investigaciones definen la vida útil como el periodo de tiempo al final del cual el elemento deja de prestar la función para la que fue diseñado (Li & Mahmoodian, 2013). Por encima de este periodo los gestores deberán hacer frente a mayores costes de reparación (Selvakumar et al., 2014).

Así diferentes autores han desarrollado modelos para predecir las roturas con el objetivo de determinar no sólo las operaciones de mantenimiento preventivo sino también la estrategia de reposición (walski y Pellicia, 1982), (Klein y Moeschberger, 2005), (Watson, et al., 2004), (Economou, et al., 2007), (Park, et al., 2008), (Wang, et al., 2009), (Dridi, et al., 2009), (Bubtiena, et al., 2011), (Christodoulou, 2011), (Xu, et al., 2011), (Kimutai, et al. 2015), . Dado el gran número de variables que potencialmente pueden influir en el grado de deterioro estos estudios recalcan la importancia de disponer de suficientes datos y con la calidad suficiente como para poder establecer modelos predictivos suficientemente precisos.

2.3 Estudio de las roturas

En la actualidad se han desarrollado diversos estudios que tratan de identificar y cuantificar la influencia de los diferentes factores que pueden condicionar las roturas en la red (Babovic, et al., 2002), (Boxal, et al., 2007). Algunos de ellos (Martínez-Codina,

2015) se centran en la relación entre las presiones en la red y la probabilidad de roturas. Otros en cambio analizan la influencia de la edad de las tuberías como primer factor de deterioro natural pero existen muchos otros factores relacionados con las condiciones medioambientales y las características intrínsecas de las tuberías que tienen influencia en los fallos (Ahn, et al., 2005). Algunos de ellos son la corrosión externa, el nivel de solicitaciones a las que se ha visto sometida cada tubería, ya sea por presiones internas externas o incluso por asentamientos o desplazamientos del terreno, la longitud, el diámetro y el material de la tubería, las condiciones de instalación o incluso la propia historia de roturas de las tuberías.

Además se han desarrollado técnicas de evaluación de la condición de estado de las tuberías para descubrir las principales variables asociadas al proceso de deterioro (Yoo et al., 2014). La recopilación, análisis y transformación de la información relativa a la condición de estado en conocimiento permite tomar ciertas decisiones sobre la probabilidad de fallo y la planificación de la renovación (Casillas et al., 2015a) (Liu & Kleiner, 2013).

En este sentido, los estudios para determinar el estado de la infraestructura pueden clasificarse en directos o indirectos (Hunaidi, 2006). Los métodos directos incluyen inspecciones visuales automáticas o manuales, ensayos no destructivos y toma de muestras de las tuberías. Los métodos indirectos en cambio se apoyan en una auditoría del agua, control de caudales y medida de la resistividad del terreno evaluar el riesgo de corrosión. En cualquier caso, todas estas técnicas son relativamente costosas y pese a ello, implican nivel substancial de incertidumbre asociado al resultado de dichos estudios (Martínez-Codina et al., 2015).

Todos estos estudios confirman el amplio número de variables que influyen en la ocurrencia de roturas. Shamir & Howard (1979) clasifican las razones de las roturas en cuatro categorías: calidad y edad de la tubería, medioambiente que rodea a la tubería, calidad del equipo de instalación y condiciones de servicio, como la presión y los transitorios. Davies et al. (2001) analizan los factores de deterioro estructurales en tuberías, clasificándolos en factores de construcción, factores externos locales y otros factores. Otros autores (Røstum, et al., 1997) (Røstum, 2000) han clasificado las variables explicativas como variables estructurales o físicas, variables externas o medioambientales, variables internas o hidráulicas y variables relacionadas con el mantenimiento. Los factores de deterioro también pueden clasificarse en físicos, ambientales y de mantenimiento. En cualquier caso existe un debate abierto sobre la

cuantificación de la influencia de estos factores sobre las roturas de tuberías y el análisis de su capacidad para predecir las roturas.

A la hora de analizar los procesos de roturas resulta determinante el disponer de unos datos adecuados, el proceso de recolección de datos no siempre es adecuado y según apuntan varios autores, (Economu, et al., 2012), (Casillas, et al., 2015b), (Kabir, et al., 2016) y (Díaz, et al., 2016), desarrollar un modelo predictivo suficientemente fiable requiere de unos datos de partida depurados y amplios para eliminar las incertidumbres asociadas al proceso de toma de datos. Por este motivo algunos estudios (Dridi et al., 2009), (Economu, et al., 2007), (Watson et al., 2004) y (Arias-Nicolás, et al., 2015) se apoyan en un análisis bayesiano considerando variables aleatorias e incorporando información externa para construir una función de distribución y para considerar las incertidumbres asociadas a los parámetros del modelo.

Basado en esta idea, y para mejorar el conocimiento de los fenómenos de roturas, la construcción de modelos predictivos de roturas debe partir de datos fiables y la cuantificación de las incertidumbres asociadas a las predicción de roturas resulta un aspecto esencial (Kabir, et al., 2015b) (Xu et al., 2011).

2.4 Tipos de modelos predictivos

Con el objetivo de disponer información sobre las posibles roturas de los elementos de la red de distribución se han desarrollado múltiples modelos y aproximaciones para la determinación de la probabilidad de fallo o la duración estimada de la vida útil de los componentes de la red.

En este sentido, los modelos de predicción de roturas se pueden agrupar como modelos físicos y estadísticos (Kleiner & Rajani, 2001). Mientras los primeros buscan descubrir los mecanismos físicos asociados a las roturas, los modelos estadísticos se basan en la historia de datos de roturas para identificar los patrones de rotura en las tuberías (Wang, et al., 2009).

Los modelos físicos predicen por tanto, la ocurrencia de roturas mediante una simulación de los mecanismos físicos de roturas (Rajani y Tesfamariam, 2004). Estos modelos pueden ofrecer una buena aproximación del estado de la red si se dispone de información suficiente. El principal problema de este tipo de modelos es que requieren de un gran número de datos y ensayos, lo que encarece mucho su aplicación y no resulta viable para redes de distribución, limitándose en todo caso su aplicación a redes de transporte y

elementos específicos (Le Gat y Eisenbeis, 2000). Por esto los modelos estadísticos son una mejor opción para poder estudiar los fenómenos de roturas en una red de distribución.

Los modelos estadísticos se basan en el análisis del registro histórico de datos con el objetivo de hacer una predicción de su comportamiento repitiendo los patrones de comportamiento observados. Estos modelos también pueden ser clasificados como determinísticos o estocásticos (Da Costa Martins, 2011). En los modelos determinísticos el número de fallos se calcula a partir de funciones matemáticas basadas en variables explicativas pero requieren largas series de datos conocidos de las variables, mientras que en el caso de los modelos estocásticos se tiene en cuenta la naturaleza aleatoria de los fallos.

Los modelos estadísticos pueden ser clasificados en modelos univariados y modelos multivariados (Wang, 2006). Los modelos multivariados permiten un mejor entendimiento de la influencia de cada parámetro en la ocurrencia de los fallos aunque requieren un ajuste de cómo las variables actúan en la distribución de los fallos. En cuanto a los modelos univariados, requieren una división de los datos de entrada en grupos homogéneos asumiendo un ratio de fallo constante para cada grupo.

Otra clasificación de estos modelos divide los modelos predictivos en modelos de regresión, modelos de análisis de supervivencia y modelos heurísticos (Kabir, 2016). Los modelos de regresión se basan en aproximar la relación de dependencia entre una variable dependiente y una o varias variables independientes. Por otra parte, los modelos de supervivencia estudian el proceso de rotura desde un punto de vista de la fiabilidad del sistema analizando el deterioro de la red y la distribución de fallos en el tiempo modelizando el proceso desde un instante inicial y un evento final o rotura (Klein y Moeschberger, 2005). Por último los modelos heurísticos buscan una modelización válida para diferentes redes de distribución en aquellos casos en los que no se dispone de un registro de datos suficiente para plantear los procedimientos anteriores (Mailhot et al. 2000); otra de las características que incorporan es el hecho de que mientras unas tuberías presentan roturas, otras no, lo que tiene un fuerte impacto en el análisis de las roturas (Rostum, 2000).

Atendiendo a los primeros modelos de regresión, según la formulación matemática empleada se puede hablar de modelos determinísticos lineales o exponenciales.

Uno de los primeros modelos de predicción de roturas fue desarrollado por Shamir y Howard (1979). Este modelo emplea un análisis de regresión y una función exponencial

del tiempo y la edad de las tuberías, el modelo determina el número de roturas por unidad de longitud y año ($\text{km}^{-1}\text{año}^{-1}$) mediante la siguiente expresión:

$$N(t) = N(t_0) \cdot e^{A \cdot (t+g)}$$

Donde t es el momento de análisis en años, g es la edad de la tubería en el instante t , A es un coeficiente de crecimiento y $N(t_0)$ no es cero, lo que implica que se asume que toda tubería tiene siempre una frecuencia de rotura aunque puede ser menor en el comienzo de su vida útil.

Este modelo fue posteriormente evolucionado por Walski y Pellicia (1982) incorporando dos factores adicionales en el análisis: un parámetro relacionado con las roturas anteriormente registradas y el segundo parámetro estaba relacionado con el diámetro.

Dentro de este conjunto de modelos exponenciales cabe mencionar también la propuesta de Clark et al. (1982) que incorporaba múltiples variables: diámetro, presión, edad de la tubería, uso del suelo diferenciando entre uso residencial e industrial, número de reparaciones registradas, tipo de suelo en cuanto a su capacidad de producir corrosión, entre otros parámetros.

Dentro de los modelos determinísticos lineales cabe mencionar los desarrollados por Kettler y Goutler (1985), McMullen (1982) o Jacobs y Karney (1994). Este último ligaba la probabilidad de rotura a la longitud de la tubería y a la edad de la misma ligadas con el registro histórico de roturas.

References	Variables	Model
Shamir and Howard 1979	age	ERM ¹
Walski and Pelliccia 1982	age, diameter, and number of previous breaks	ERM
Clark et al. 1982	age, number of previous breaks	MLRM ² , MERM ³
Kettler and Goulter 1985	age, diameter	LRM ⁴
Goulter and Kazemi 1988	time, space from previous break	ERM
Jacobs and Karney 1994	length, age	LRM
Kleiner and Rajani 1999	age	ERM
Rajani and Kleiner 2001	age, freezing index, cumulative rain deficit, snapshot rain deficit, cumulative length of replaced water mains, cumulative length of cathodic protection retrofit	MERM
Kleiner and Rajani 2000, 2002	freezing index, rainfall deficit, cumulative length of replaced mains, cumulative length of cathodic protection	MERM
Boxall et al. 2007	diameter, age, length, material, soil corrosivity	PRM ⁵
Asnaashari et al. 2009	age, length, diameter, wall thickness, maximum pressure, pipe location, cover depth, failure history	MLRM, PRM
Wang et al. 2009	age, size, length	MLRM
Wood and Lence 2009	age, material, diameter, soil type	MERM
Yamijala et al. 2009	diameter, material, length, year of installation, time since last break, pressure, land use, soil type, temperature, rainfall, maximum soil moisture, max-min soil moisture, soil corrosivity	MLRM, MERM, GLM ⁶ , logistic GLM
Bubtienė et al. 2011	age, material, length, diameter, material, depth, type of soil, water quality	GLM
Christodoulou 2011	material, diameter, pipe type, traffic load in the vicinity, number of previous breaks, incident type	PRM

¹ERM: Exponential Regression Model

²MLRM: Multiple Linear Regression Model

³MERM: Multivariate Exponential Regression Model

⁴LRM: Linear Regression Model

⁵PRM: Poisson Regression Model

⁶GLM: Generalised Linear Model

Tabla 1. Modelos de regresión lineal y exponencial para la predicción de roturas (Kabir, 2016)

Dentro de los modelos clasificados como de supervivencia se emplean diferentes procedimientos de análisis como el proceso de regresión de Poisson (Asnaashari, et al., 2009), (Boxall, et al., 2007), el proceso de Poisson no homogéneo (Kleiner & Rajani, 2010), (Rogers, 2011), el modelo de Weibull (Dridi, et al., 2009), (Park, et al., 2008), los modelos multivariados exponenciales (Kleiner & Rajani, 2002), (Mailhot et al., 2000), la regresión de Cox (Park, et al., 2011), o la inferencia bayesiana (Dridi, et al., 2005), (Economou, et al., 2007), (Watson, et al., 2004) entre otros. La Tabla 2 recoge múltiples ejemplos de estos modelos.

References	Covariates	Cohorts/Groups	Methods
Kimutai et al. 2015	length, diameter, material, soil resistivity, freezing index, rain deficit	material, number of previous breaks	Cox-PHM ¹ , W-PHM ² , Poisson model
Economou et al. 2012	age, number of previous failures	material	Zero inflated NHPP
Fuchs-Hanusch et al. 2012	diameter, length, vintage	material	Cox-PHM
Rajani et al. 2012	water and air temperature	material	Zero inflated NHPP ³
Toumbou et al. 2012	number of pipe breaks, diameter, length, and installation dates	material, number of previous breaks	MEWM ⁴
Christodoulou 2011	material, diameter, pipe type, traffic load in the vicinity, number of previous breaks, incident type	material, number of previous breaks	ANN ⁵ , Kaplan-Meier estimator, PRM ⁶
Clark and Thurnau 2011	material, diameter, frailty factor	Material, diameter	Cox-PHM / shared frailty model
Osman and Bainbridge 2011	age, length, freezing index	material	MEM ⁷
Park et al. 2011	length, pipe internal pressure, land development, number of customer in a grid	material, number of previous breaks	Cox-PHM
Renaud et al. 2011	material, length, installation year, break history	material, number of previous breaks	Linear extension of the Yule process
Rogers 2011	age, material, diameter, leak type, number of previous breaks	material, number of previous breaks	NHPP
Alvisia and Franchina 2010	diameter, material, year of installation, number of breakages	installation year, material, number of previous breaks	MEWM, W-PHM
Clark et al. 2010	material, diameter, frailty factor	Material, diameter	Cox-PHM
Debón et al. 2010	age, length, diameter, pressure traffic, under a sidewalk, material	not mentioned	Cox-PHM, W-PHM, MRM ⁸
Fadaee and Tabatabaei 2010	type of failure, cumulative number of failures	type of failure	Power law model
Kleiner and Rajani 2010	age, diameter, material, length, service connection, vintage, soil type, freezing index, cumulative rain deficit, snapshot rain deficit, cathodic protection, retrofit cathodic protection	diameter	NHPP
Asnaashari et al. 2009	age, length, diameter, wall thickness, maximum pressure, pipe location, cover depth	material	MRM, PRM
Dridi et al. 2009	age	number of previous breaks	Bayesian MEWM

References	Covariates	Cohorts/Groups	Methods
Rogers and Grigg 2009	age, material, diameter, number of previous breaks	material, number of previous breaks	NHPP, MCDA ⁹
Gorji-Bandpy and Shateri 2008	material, diameter, length, year of installation, type of soil, land use	number of previous breaks	MEWM
Park et al. 2008b	material, length, diameter, installation time, pipe internal pressure, land development	material, number of previous breaks	Cox-PHM
Park et al. 2008a	age, previous failure times	material, diameter	Rates occurrence failure model
Boxall et al. 2007	diameter, age, length, material, Soil corrosivity	material	PRM
Park et al. 2007	material, diameter	survival time	Cox-PHM
Economou et al. 2007	age, pipe length, pipe diameter, pressure, absolute pressure	not mentioned	Bayesian NHPP
Dridi et al. 2005	age	number of previous breaks	Bayesian MEWM
Vanrenterghem-Raven 2004	age, length, diameter, material, traffic, soil, subway, location in the street, presence of ancient water zones	material, number of previous breaks	Cox-PHM, W-PHM
Watson et al. 2004	age	not mentioned	Bayesian NHPP
Pelletier et al. 2003	diameter, length, material, year of installation, type of soil, land use	number of previous breaks	MEWM
Kleiner and Rajani 2002	cumulative length of replaced mains, cumulative length of cathodic protection, freezing index, rainfall deficit	material	MEM
Le Gat and Eisenbeis 2000	age, length, diameter, types of pipe assembling, pressure, soil type, level of traffic, kind of supply	material, number of previous breaks	W-PHM
Mailhot et al. 2000	material, length, diameter, installation year, soil type, land use above the pipe	installation year	MEWM

¹Cox-PHM: Cox Proportional Hazard Models

²W-PHM: Weibull Proportional Hazard Models

³NHPP: Nonhomogeneous Poisson Process

⁴MEWM: Multivariate Exponential/Weibull Models

⁵ANN: Artificial Neural Network

⁶PRM: Poisson Regression Model

⁷MEM: Multivariate exponential model

⁸MRM: Multiple Regression Model

⁹MCDA: Multicriteria Decision Analysis

Tabla 2. Modelos de supervivencia (Kabir, 2016)

Los modelos heurísticos son aplicados ante la falta de datos disponibles en determinadas redes de estudio pues no todos los operadores disponen de una base de datos adecuada para poder plantear alguno de los procedimientos de regresión y análisis de supervivencia (Economou, et al., 2008), o bien los registros disponibles no tienen la calidad y fiabilidad necesaria. Como señalan Wood y Lence (2006) esto sucede por lo general en redes de

transporte en las que no se tiene un registro de roturas suficientemente amplio por tratarse de redes de menor tamaño en comparación con las redes de distribución donde la densidad de las tuberías es considerablemente mayor y permite disponer de un numeroso registro de datos.

Dentro de esta clasificación algunos de modelos cabe mencionar algunos como la propuesta de Yan y Vairavamoorthy (2003) que plantean un total de seis factores agrupados entre parámetros físicos y medioambientales o Christodoulou et al. (2003) que emplean un total de ocho variables. La Tabla 3 describe algunos de estos modelos:

References	Failure Index				Model/ Framework
	Structural Integrity	Hydraulic Capacity	Water Quality	Consequence	
Shi et al. 2013	diameter, age, material, temperature				Spatial analysis
Singh 2011	diameter, age, material, soil resistivity				Bayesian Analysis
Jafar et al. 2010	diameter, age, material, length, thickness, soil resistivity	pressure			ANN ¹
Fares and Zayed 2010	diameter, age, material, type of traffic, soil resistivity	velocity	turbidity, water age	population, land use	Fuzzy rule-based
Christodoulou et al. 2009	age, material, diameter, length, type of traffic				ANN & fuzzy logic
Francisque et al. 2009	diameter, age, material, water pH		turbidity, Free residual Cl, age	land use	AHP ² & fuzzy logic
Al-Barqawi and Zayed 2008	age, material, diameter, length, type of traffic	pressure			AHP & ANN
Amaitik and Amaitik 2008	age, diameter, soil resistivity	pressure			ANN
Al-Barqawi and Zayed 2006a	age, material, diameter, soil resistivity, type of traffic				ANN
Al-Barqawi and Zayed 2006b	age, material, type of traffic diameter, soil resistivity,	pressure			AHP
Kleiner et al. 2006	age				Fuzzy rule-based
Rogers 2006	age, material, diameter, soil resistivity				NHPP ³ & MCDA ⁴
Rogers and Grigg 2006	age, material, diameter, soil resistivity	pressure			NHPP & MCDA
Najafi and Kulandaivel 2005	age, material, diameter, length				ANN
Kleiner et al. 2004	age				Fuzzy MDP ⁵
Christodoulou et al. 2003	age, material, diameter, length, type of traffic				ANN
Jafar et al. 2003	age, material, diameter, thickness length, soil resistivity, type of traffic	pressure			ANN
Yan and Vairavamoorthy 2003	age, diameter, material, soil condition, road loading, surroundings				Fuzzy composite programming
Kettler and Goulter 1985	age, material, diameter				Regression analysis
O'Day 1982	diameter, material, soil resistivity				Function analysis

¹ANN: Artificial Neural Network; ²AHP: Analytic Hierarchy Process; ³NHPP: Non-homogeneous Homogenous Poisson Process; ⁴MCDA: Multicriteria Decision Analysis; ⁵MDP: Markov Deterioration Process

Tabla 3. Modelos de regresión lineal y exponencial para la predicción de roturas (Kabir, 2016)

En general, la mayoría de los modelos se apoyan en un planteamiento estadístico estimando el número de roturas en función de una serie de variables explicativas (Rostum, 2000), (Yamijala, 2007) a partir de los registros históricos registrados en la red. Como se recoge en esta revisión de los modelos existentes, hay un amplio espectro de técnicas para desarrollar este análisis. Kliner & Rajani (2011) señalan los análisis de

Poisson (Economou, 2010), vida acelerada (Lei, 1997), diagnóstico Bayesiano (Dridi et al. 2009, Watson et al. 2004). También existen planteamientos alternativos como las técnicas de aprendizaje (Berardi et al. 2008, Xu et al. 2011) o las redes neuronales (Achim et al. 2007).

2.5 Modelos Bayesianos

Dentro del espectro de modelos estadísticos analizados los modelos Bayesianos constituyen una rama importante por las ventajas que incorporan. En este sentido el planteamiento de modelos de roturas basados en un análisis Bayesiano ha sido objeto de varios estudios (Kulkarni et al., 1986), (Watson et al., 2004), (Babovic et al., 2002), (Poulakis et al., 2003).

Uno de los primeros planteamientos (Kulkarni et al., 1986) busca identificar los tramos de tuberías con mayor probabilidad de rotura para establecer una adecuada estrategia de renovación. El planteamiento agrupa las variables consideradas en dos grupos, por un lado las características estadísticas de la tubería tales como diámetro, longitud o tipo de suelo, entre otras; y por otro lado las variables dinámicas como la acumulación de roturas o la edad de las mismas. El planteamiento parte del teorema de Bayes con la siguiente expresión:

$$P[fallo|caraterísticas determinadas] = \frac{P_{c/f} \cdot P_f}{P_{c/f} \cdot P_f + P_{c/nf} \cdot (1 - P_f)}$$

Donde $P_{c/f}$ es la probabilidad de detectar ciertas características en la tubería que ha fallado y P_f la probabilidad de fallo. Dicha formulación puede reescribirse como:

$$P[fallo|caraterísticas determinadas] = \frac{P_{c/f} \cdot P_f}{P_c}$$

Siendo P_c la probabilidad de observar unas determinadas características en un tramo de tubería, independientemente de que se hayan registrados roturas en ella o no.

Sobre este planteamiento, Watson et al. (2004) incorporan el factor tiempo al referirse no sólo a la probabilidad de fallo sino también al número de roturas producidas en un determinado periodo de tiempo. Plantean un modelo bayesiano jerárquico en el que incorpora información de distintas fuentes a la vez que asumen similitud entre parámetros. En este sentido la información del teorema de Bayes actualiza los datos de otras tuberías de similares características y que no cuentan con series históricas tan

completas de roturas. Según los resultados obtenidos, se superan los inconvenientes de falta de datos de exceso de número de tuberías sin registro de roturas.

Posteriormente Watson (2005) y Watson et al. (2006) han desarrollado teóricamente el modelo anterior incorporando al modelo bayesiano un modelo de ocurrencia de las roturas basado en un proceso no homogéneo de Poisson. Mediante este planteamiento se asume que el tiempo entre roturas no es ni independiente ni distribuido idénticamente y que la probabilidad de fallo es la misma antes y después de que se repare una tubería.

El planteamiento de Economou et al. (2007) emplea de nuevo un modelo de Poisson no homogéneo agregado con el objetivo de obtener la probabilidad del número medio de roturas con el tiempo, condicionada a los datos disponibles. Posteriormente (Economou, et al., 2008) evolucionan dicho planteamiento con un modelo de Poisson no homogéneo inflado con ceros para explicar la ocurrencia de las roturas con el tiempo.

Otros modelos emplean la distribución de Weibull para caracterizar el tiempo entre la primera y la siguiente rotura, y el tiempo entre las subsecuentes roturas mediante una distribución exponencial (Dridi, et al., 2009).

En cualquier caso, una de las ventajas del planteamiento bayesiano es que se tienen en cuenta las incertidumbres asociadas a los datos disponibles como apuntan Kabir et al. (2015a).

2.6 Aplicación: modelos de vida útil

Una metodología generalmente aceptada para la representación del ciclo de vida de una tubería es la propuesta por Ascher y Feingold (1984). Según este planteamiento las roturas describen una curva en forma de “U”, “bathtub” curve en inglés, donde el eje horizontal es el tiempo y el vertical es la tasa de ocurrencia del fallo, Rate of Occurrence of Failure (ROCOF). Como se recoge en la Figura 1 (Kleiner & Rajani, 2001), se identifican tres tramos: fase inicial en la que tras la instalación hay un mayor riesgo de que ocurran roturas como resultado de una instalación deficiente, este tipo de roturas se manifiestan al comienzo del ciclo de vida y su frecuencia se reduce rápidamente. Una vez superada esa fase, cuando el sistema funciona correctamente se entra en la segunda fase; durante esta etapa la frecuencia de roturas es relativamente baja, la ocurrencia de las mismas durante esta etapa se justifica con la componente aleatoria asociada al fenómeno que puede estar originado en una carga externa excepcionalmente elevada, interferencias de agentes externos, etc. La tercera fase se corresponde con el periodo de desgaste, a

medida que avanza el tiempo aumenta la frecuencia de roturas debido al deterioro de la tubería.

Figura 1. Ciclo de vida de una tubería . "Bathtub curve" (Rajani y Klein, 2001).

Xu et al. (2013) afirman que hay dos aproximaciones que apoyan las decisiones sobre renovación de la red. Una de ellas se basa en la priorización de un conjunto de tuberías que se seleccionan bajo un determinado volumen de inversión y otro que se centra en la elección del momento óptimo para la sustitución de una tubería.

Con el objetivo de prever el comportamiento de la red y mejorar su gestión a través de una estrategia de reposición más adecuada se están desarrollando modelos de predicción de roturas (Xu, et al., 2013), (Malm, et al., 2012) (Renaud, et al, 2012). Se han planteado diferentes modelos partiendo del concepto de vida útil, entendido como el periodo de tiempo en el que una tubería deja de dar el servicio para el que fue diseñada incrementándose desde entonces los costes de reparación (Li, et al., 2013), (Selvakumar, et al., 2015).

Estos modelos son útiles para los gestores del abastecimiento ya que aportan una herramienta de ayuda a la programación de las campañas de mantenimiento preventivo y a la definición de los planes de renovación. No obstante, hay que hacer frente a la incertidumbre asociada a la calidad y la cantidad de los datos necesarios para construirlos

de forma robusta; así como establecer un modelo predictivo suficientemente preciso para basar la toma de decisiones sobre unos datos fiables.

2.7 Resumen

Los modelos de rotura de tuberías resultan imprescindibles en los planes de gestión de activos de las compañías que prestan el servicio de abastecimiento de agua (Engelhardt et al, 2000). Las decisiones relativas a la planificación de las operaciones de renovación de la red deben basarse en un conocimiento objetivo de la tasa esperada de roturas y su correspondiente influencia en la interrupción del servicio (Hong et al., 2006). Sin embargo, del estudio de la literatura disponible no puede deducirse con facilidad qué variables explicativas deben tenerse en cuenta los modelos de predicción de roturas. La mayor parte de los estudios determinan las variables a incluir en los modelos a priori, bien a partir de la experiencia del modelador o simplemente como consecuencia de la disponibilidad de datos.

Por todo lo expuesto en este apartado se deduce que si bien hay muchas alternativas y planteamientos para analizar la probabilidad de ocurrencia de roturas, no hay una metodología clara que permita seleccionar los factores explicativos que deban ser considerados en los modelos predictivos a emplear (Gómez-Martínez, et al., 2017). Igualmente la metodología desarrollada define la probabilidad de que se produzca la rotura en un periodo de tiempo determinado frente a otros planteamientos que en general se construyen en el tiempo y definen el número de veces que se van a producir las roturas en un periodo temporal determinado.

3 METODOLOGÍA

La metodología propuesta tiene como objetivo el generar la probabilidad esperada de rotura de cada uno de los tramos de la red de distribución en función de las características de los mismos que tengan una incidencia en dicha probabilidad.

Bajo esta premisa la metodología se estructura en las siguientes etapas:

- Identificación de las variables explicativas con incidencia en la probabilidad de fallo.
- Verificación de la dependencia estadística de las roturas respecto de las variables analizadas.
- Construcción de un amplio conjunto de posibles modelos estadísticos que establezcan la predicción de las roturas.
- Validación de los modelos empleados cuantificando su capacidad predictiva con el fin último de seleccionar el mejor modelo.
- Análisis de sensibilidad, como paso previo para la validación de los modelos limitando así la influencia de los parámetros empleados en los resultados obtenidos.

3.1 Selección de variables

La metodología que se presenta plantea un análisis estadístico sin influencias o ideas preconcebidas de cara al análisis de las roturas. En este sentido se realiza un amplio análisis de las posibles variables explicativas del fenómeno en cuestión.

Según este planteamiento, el análisis de las roturas se realiza de manera holística con la intención de incluir en él todas las posibles variables que puedan explicar su ocurrencia para obtener la mejor predicción posible.

En un primer paso se analizan las variables explicativas que describen los parámetros físicos de los diferentes tramos de tuberías. Estas variables son: la edad de las mismas, el material y el diámetro. Puesto que presentan una relación evidente con la ocurrencia de roturas de tuberías, se ha denominado este grupo como variables explicativas “principales”.

En un segundo paso se incluye cualquier otra posible variable explicativa del fenómeno de roturas en la red de distribución, identificando aquellas de las que se puede disponer información suficiente para poder ser estudiadas en los diferentes modelos predictivos. Las variables así identificadas se clasifican en variables hidráulicas o internas y ambientales dependiendo de su origen.

3.1.1 Variables explicativas principales

Las variables consideradas de manera habitual en los estudios de roturas de tuberías son tres:

- **Diámetro:** se toma como variable explicativa el diámetro nominal de la tubería.
- **Año de instalación:** mediante este parámetro se tiene en cuenta la edad de las tuberías.
- **Material:** puesto que existe una gran variedad de tipos de tuberías se analiza específicamente la influencia del material constituyente. Se realiza una clasificación entre los diferentes materiales identificados en la red a analizar.

3.1.2 Variables explicativas secundarias

Con el objetivo de englobar todas aquellas posibles variables explicativas dentro de los modelos predictivos se seleccionan varios grupos de variables que posteriormente son estudiadas para determinar la dependencia estadística de las roturas respecto de estas variables. Atendiendo a varios se plantean las posibles variables explicativas:

- **Factores internos:** se entienden por factores internos aquellos que van ligados a las variables que definen el comportamiento hidráulico dentro de cada una de las tuberías de la red.
- **Factores ambientales:** dentro de este grupo se incluyen aquellas variables que dependen tanto de las propiedades y características físicas de las tuberías como del entorno en el que se encuentran instaladas.

Con este primer planteamiento de posibles variables se valora la viabilidad de su incorporación a los modelos predictivos.

3.1.2.1 Factores internos

Velocidad del agua: dada la variabilidad de la velocidad del flujo de agua para cada tubería en función de las condiciones de explotación se toman tres velocidades de referencia: mínima, media y máxima. Estos valores se pueden obtener de los modelos hidráulicos del sistema analizado. Se establecen unos rangos de velocidades para establecer grupos homogéneos de tuberías.

Variaciones de presión: al igual que la velocidad del agua, la presión es variable en función de las condiciones de explotación. Por este motivo se adoptan igualmente tres valores de referencia: presión mínima, media y máxima. Para establecer unos grupos homogéneos de presiones se consideran escalones a la hora asignar una presión de referencia para cada una de las tuberías.

Presiones transitorias: puesto que no se disponen de valores de dichas presiones transitorias se descarta su incorporación en los modelos predictivos.

Calidad del agua: la calidad del agua puede tener influencia en las roturas por su incidencia en los fenómenos de corrosión interna de las tuberías, que podría desencadenar en rotura. Sin embargo, en general no se tienen datos suficientes para poder ser tenido en cuenta.

3.1.2.2 Factores ambientales

Tipo de terreno (características físicas y químicas): mediante esta variable se pretende incorporar la influencia de las características geológicas y geotécnicas en las que se localiza cada una de las tuberías.

Tipo de ubicación (transmisión de cargas al elemento): mediante esta variable se describe el modo en que las cargas superficiales, dependientes del tipo de uso del suelo, son transmitidas a las tuberías. En función del modo de instalación las cargas a resistir por la tubería varían pues el mecanismo de propagación de las presiones a nivel superficial hasta la tubería pueden ser diferentes según la ubicación de la misma. Aunque generalmente las tuberías se encuentran enterradas existen otras posibles ubicaciones como en superficie, en galería, en arqueta, aérea, hincada, en galería, etc.

Profundidad de enterramiento: se considera la profundidad de enterramiento de las diferentes tuberías de la red.

Tipo de junta: es un tipo de información difícil de obtener para las tuberías que generalmente se encuentran enterrados por lo que se descarta su utilización.

Periodo del año en que se produce (problemas de heladas o de calidad con la Tª): se descarta esta variable por la dificultad de disponer de registros climatológicos específicos en la localización y momento de la rotura.

Relación espesor/diámetro: esta característica depende de las normas y estándares que aplique cada fabricante. Puesto que dicha relación depende del tipo de material y timbraje no es resulta una variable significativa por estar implícitamente considerada en el diámetro y material de las tuberías.

Longitud del tubo: en general no se dispone información detallada de la longitud de cada uno de los tramos de tubería por lo que no es posible su incorporación al modelo.

Uso del suelo y exposición al tráfico (calzada, acera, etc.): mediante esta variable se reflejan los condicionantes mecánicos que solicitan a la tubería. En función del uso del suelo en superficie las cargas que deben resistir las tuberías varían, puesto que dichas cargas pueden ser uno de los parámetros determinantes en la ocurrencia de roturas su inclusión como variable explicativa en los modelos puede ser incidente.

Posición del nivel freático: dada la variabilidad de este parámetro con la localización de las tuberías resulta complicado disponer de información detallada para cada tubería sobre la posición del nivel freático. Por otra parte, dicho nivel es variable en función de la época del año. Por ambos motivos se descarta su utilización.

El resultado de este filtro se recoge en la Tabla 4 donde se indica su incorporación o no al estudio así como el motivo de su rechazo en su caso.

Tipo	Número	Variable secundaria	Código	Disponibilidad	Selección
Factor Interno	1	Velocidad del agua	V		Sí
	2	Variaciones de presión	P		Sí
	3	Presiones transitorias	Tran	Sin información	No
	4	Calidad del agua	Cal	Sin información	No
Factor Ambiental	5	Tipo de terreno	Terr		Sí
	6	Tipo de ubicación	Ubic		Sí
	7	Profundidad	Prof		Sí
	8	Tipo de junta	Junta	Sin información	No
	9	Periodo del año	Per	Sin información	No
	10	Espesor/diámetro	E-D	Sin información	No
	11	Longitud del tubo	Long		No
	12	Uso del suelo	Uso		Sí
	13	Nivel freático	NFreat	Sin información	No

Tabla 4. Variables secundarias.

3.2 Dependencia estadística de las roturas respecto de variables

La dependencia estadística de las roturas respecto de cada una de las variables seleccionadas se realiza mediante un estudio sistemático de las incidencias y cada una de las variables explicativas. Se plantea un análisis de dependencia de la función de distribución de las roturas según la variable y su comparación con la misma función de distribución, pero referida a todos los elementos.

Este estudio busca la evidencia empírica de la relación entre la variable analizada y las roturas, como metodología de estadística no paramétrica se plantea un test Kolmogorov-Smirnov (K-S). El test K-S plantea un procedimiento para aceptar o rechazar la hipótesis de que dos muestras pertenezcan o no a la misma distribución. Por tanto se establece una comparativa entre la distribución acumulada de los elementos donde se registran

roturas y la función de distribución genérica de todos los elementos. El test mide el grado de ajuste entre la distribución de la muestra (condicionada a roturas) y la distribución teórica que debe seguir la muestra (distribución genérica). En la medida en que se identifica una diferenciación entre ambas se concluye que existe una dependencia estadística de las roturas respecto de la variable analizada.

3.2.1 Fundamento teórico del test Kolmogorov-Smirnov

El test de Kolmogorov-Smirnov (K-S) es una prueba de significación estadística no paramétrica empleada para determinar la bondad de ajuste de dos distribuciones de probabilidad entre sí. Se busca por tanto determinar si el conjunto de datos condicionado a una determinada variable explicativa proviene de la distribución total de la red. Mide por tanto el grado de ajuste entre la distribución de la muestra y la distribución teórica que se supone debe seguir la muestra

3.2.2 Metodología del test

La metodología del test de K-S se basa en el contraste de la siguiente hipótesis:

- H_0 : Los datos analizados siguen la distribución teórica.
- H_1 : Los datos analizados no siguen la distribución teórica.

Según este test el estadístico de contraste viene definido como:

$$D = \max_{1 \leq i \leq n} |F_c(x_i) - F_g(x_i)|$$

Siendo: $F_g(x)$ la función de distribución acumulada teórica o genérica, $F_c(x)$ la distribución acumulada analizada o condicionada, x_i son los valores de la distribución que se han obtenido y D_α es el nivel de significancia del estadístico de contraste. Es decir que D es la mayor diferencia absoluta observada entre la función de distribución acumulada empírica de los valores condicionados, y la función de distribución genérica de todos los valores.

Para la toma de dediciones se establece un valor crítico del estadístico de contraste por encima del cual se rechaza la hipótesis H_0 :

- Si $D \leq D_\alpha$ se acepta la hipótesis H_0 , lo que indica que los datos obtenidos siguen la misma distribución. En estos casos, la variable explicativa analizada no afecta a la probabilidad de ocurrencia de roturas.

- Si $D > D_\alpha$ se rechaza la hipótesis H_0 , que significa que los datos obtenidos no siguen la misma distribución y por tanto la variable explicativa sí influye en la probabilidad de ocurrencia de roturas.

El valor del estadístico de contraste viene determinado por el nivel de significancia α que se fija en el contraste así como de los grados de libertad de la muestra.

3.2.3 Proceso de cálculo

La aplicación del test K-S incluye los siguientes pasos:

1º. Se ordenan las distribuciones de los datos condicionados y no condicionados en orden creciente de tal forma que; $x_1 \leq x_2 \leq \dots \leq x_n$.

2º. Se crea las Funciones de Distribuciones Acumuladas condicionada y genérica:

$$F_c(x) = \begin{cases} 0 & x < x_1 \\ \frac{1}{n} & x_1 \leq x < x_2 \\ \dots & \\ \frac{i}{n_i} & x_i \leq x < x_{i+1} \\ \dots & \\ 1 & x_n \leq x \end{cases} \quad F_g(x) = \begin{cases} 0 & x < x_1 \\ \frac{1}{n} & x_1 \leq x < x_2 \\ \dots & \\ \frac{i}{n_i} & x_i \leq x < x_{i+1} \\ \dots & \\ 1 & x_n \leq x \end{cases}$$

3º. Se considerarán las diferencias entre las funciones tanto las superiores como las inferiores, por lo que se obtendrá dos máximos:

$$D^+ = \max_{1 \leq i \leq n} [F_c(x_i) - F_g(x_i)]$$

$$D^- = \max_{1 \leq i \leq n} [F_g(x_i) - F_c(x_i)]$$

Se tomará por tanto un valor de D:

$$D = \max[D^+, D^-]$$

Una vez fijado el nivel de significancia α se establece la comparación entre el estadístico de contraste y el valor crítico asociado al nivel de significancia seleccionado:

Cuando se verifica que $D \leq D_\alpha$ se acepta la hipótesis H_0 , la variable explicativa analizada no afecta a la probabilidad de ocurrencia de roturas.

En cambio, si $D > D_\alpha$ se rechaza la hipótesis H_0 , y por tanto la variable explicativa sí influye en la probabilidad de ocurrencia de roturas.

3.3 Construcción de modelos

Identificadas las variables explicativas que tienen incidencia sobre la probabilidad de ocurrencia de roturas en las tuberías de la red de distribución se plantean múltiples modelos predictivos con el fin de seleccionar el que ofrezca una mayor capacidad de predicción.

Puesto que el estudio no parte de una idea o estudio preconcebido, sino que surge del análisis de los datos, se debe plantear un amplio conjunto de modelos predictivos combinando las diferentes variables explicativas.

3.3.1 Probabilidad condicionada – Modelos bayesianos

Los modelos predictivos empleados en la metodología se basan en el teorema de Bayes y la probabilidad condicionada. El objetivo de estos modelos es identificar la probabilidad de ocurrencia de una rotura como función de una determinada variable explicativa (o un conjunto de ellas).

Estos modelos se basan en las observaciones registradas y pretenden determinar la probabilidad de rotura de un elemento de la red en un periodo de tiempo determinado a partir del conocimiento de su comportamiento histórico. Para ello se parte de una hipótesis básica consistente en que la probabilidad de fallo de una tubería en un periodo de tiempo es expresable en términos de una variable (o un conjunto de ellas) e independientemente de su emplazamiento.

Este planteamiento permite realizar el estudio probabilístico función del tiempo como única variable. Con ello, se tiene que la función de probabilidad de fallo en función del tiempo, $F(t)$, es la función de probabilidad que gobierna el fallo en función del tiempo. Partiendo de esta definición se puede analizar el tiempo al fallo y deducir así la probabilidad de fallo en un intervalo:

La probabilidad de “fallo” hasta el tiempo t sería:

$$p_f(0, f) = Pr[Tiempo al fallo < t] = \int_0^t f(\tau) d\tau = F(t)$$

La probabilidad de “no fallo” (seguridad) hasta el tiempo t sería:

$$p_f(0, f) = Pr[Tiempo al fallo > t] = \int_t^{\infty} f(\tau) d\tau = 1 - F(t)$$

Por la propia definición excluyente de ambas funciones se tiene:

$$p_s(0, t) = 1 - p_f(0, t) \text{ y } p_f(0, t) = 1 - p_s(0, t)$$

Consecuentemente, la probabilidad de “fallo” en el intervalo $[t, t+\Delta t]$ sería:

$$\begin{aligned} p_f(t, t + \Delta t) &= Pr[(\text{Tiempo al fallo} > t) \text{ y } (\text{Tiempo al fallo} < t + \Delta t)] \\ &= \int_t^{t+\Delta t} f(\tau) d\tau = F(t + \Delta t) - F(t) \end{aligned}$$

La complejidad de este planteamiento reside en la caracterización de la función de distribución de probabilidad. En este sentido existen dos tratamientos diferentes, por un lado se puede plantear un desarrollo teórico basado en la descripción del fenómeno físico y como la variable en cuestión determina la variación de la probabilidad de fallo o a partir de observaciones.

En esta metodología se opta por la opción de estimar dicha probabilidad partiendo de observaciones. Para ello se emplean los registros de roturas como fuente de información fundamental para el estudio. El procedimiento se plantea suponiendo que durante el periodo de tiempo del que se disponen registros (si este es reducido) no varían sustancialmente las variables explicativas de los fallos y por tanto se plantea una inferencia bayesiana.

En este planteamiento se considera una muestra condicionada a las roturas y la genérica. Esta muestra genérica engloba todos los elementos que cumplen una característica determinada mientras que la condicionada a las roturas únicamente incluye la fracción de estos elementos en los que se registran roturas.

Para el caso de una única variable explicativa X , se definen dos sucesos:

Suceso A: Ocurre un fallo en el sistema

Suceso B: la variable explicativa toma un valor dentro del intervalo x y $x+\Delta x$

La probabilidad de que ocurra una rotura $Pr(A)$ en un determinado elemento de la red y en un periodo determinado viene definida por la expresión:

$$P(A) = \frac{N_f}{N_T}$$

Donde N_f es el número de fallos registrados en el periodo y N_T es el número de componentes del sistema.

La probabilidad genérica describe la información sobre las roturas (en ausencia de información sobre las variables explicativas). Si se dispone de información adicional, que

la variable explicativa tome un valor en el intervalo $[x, x+\Delta x]$, se puede estimar la probabilidad de ocurrencia de que la variable tome un valor en ese intervalo como:

$$\Pr(B) = \Pr(x \leq X < x + \Delta x) = F_X(x + \Delta x) - F_X(x)$$

Donde F_X es la función de distribución de la variable explicativa X .

A partir de las observaciones se puede determinar la probabilidad de que aquellos elementos en los que se produce la rotura la variable explicativa tome un valor del intervalo $[x, x+\Delta x]$ según la siguiente expresión:

$$\Pr(B|A) = F_{FX}(x + \Delta x) - F_{FX}(x)$$

Donde $\Pr(B|A)$ es la probabilidad de que la variable explicativa X de un elemento en el que se produce una rotura tome un valor en el intervalo $[x, x+\Delta x]$ y F_{FX} es la función de distribución de la variable explicativa X sobre los elementos en los que se produce una rotura.

Aplicando el teorema de Bayes se puede estimar la probabilidad de ocurrencia de una rotura con la variable explicativa dentro del intervalo $[x, x+\Delta x]$ según la siguiente expresión:

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)} = \frac{F_{FX}(x + \Delta x) - F_{FX}(x)}{F_X(x + \Delta x) - F_X(x)} \frac{N_f}{N_T}$$

Donde $\Pr(A|B)$ es la probabilidad de fallo de un componente mientras la variable explicativa X tome un valor en el intervalo $[x, x+\Delta x]$.

Por tanto, la probabilidad de fallo condicionada a una variable explicativa se puede estimar partiendo de la distribución de probabilidad de la variable genérica (sin estar condicionada a las roturas) y la distribución de probabilidad de la misma variable condicionada a los elementos en los que se registró una rotura. De esta forma para un intervalo determinado de la variable explicativa, el ratio de las probabilidades es un factor multiplicador aplicable a la distribución de probabilidad de fallo genérica determinada por la relación entre N_f y N_T (relación entre número de componentes del sistema y número de roturas registradas).

La formulación de este factor multiplicación es la siguiente:

$$P(A|B) = P_R \cdot P(A)$$

Donde, P_R es la “relación de probabilidad”, por la que se multiplica la probabilidad genérica $P(A)$. Si P_R es superior a uno, la probabilidad genérica de rotura se incrementa como consecuencia de la variable explicativa y en caso contrario se reduce.

Este planteamiento puede extenderse cuando se combinan múltiples variables explicativas. Para ello debe verificarse si las variables son independientes o no. Cuando las variables son independientes la función de densidad conjunta es el producto de las funciones de densidad marginales.

Con el objetivo de estimar la probabilidad de fallo conjunta en función de varias variables explicativas se debe determinar la función de distribución de probabilidad del conjunto de variables explicativas. Cuando las variables son independientes, la función de densidad conjunta se obtiene mediante el producto de las funciones de densidad marginales de cada una de las variables. En este escenario, la probabilidad de rotura se obtiene multiplicando la probabilidad genérica por el producto de las relaciones de probabilidad de cada una de las variables explicativas:

$$P(A|B) = P_{R1} \cdot P_{R2} \cdot \dots \cdot P_{Rk} \cdot P(A)$$

donde las P_{Ri} son las relaciones de probabilidad de cada una de las variables explicativas.

Por el contrario, si existe dependencia se debe estimar la función de distribución conjunta. Para limitar la complejidad del tratamiento matemático y para no aumentar en exceso el nivel de información disponible, el análisis se ha limitado a grupos de dos variables. Finalmente el resto de variables se incorpora.

Dado que en este caso el tratamiento matemático es más complejo y se incrementa la necesidad así la formulación a emplear es la siguiente:

$$P(A|B) = P_{R1-2} \cdot P_{R3} \cdot \dots \cdot P_{Rk} \cdot P(A)$$

siendo las P_{R1-2} la relación de probabilidad de las dos variables analizadas conjuntamente y las restantes P_{Ri} son las relaciones de probabilidad de las variables independientes.

3.3.2 Influencia de la agregación de variables

Las variables consideradas también se clasifican desde el punto de vista numérico, en este sentido se tienen variables cualitativas, numéricas discretas y numéricas continuas como se recoge en la Tabla 5:

Tipo de variable	Variable explicativa
Numérica discreta	Año de instalación
	Diámetro
Cualitativa	Tipo de ubicación
	Tipo de terreno
	Material
	Uso del suelo
Numérica continua	Presión (mínima, media máxima)
	Profundidad de enterramiento
	Velocidad (mínima, media, máxima)

Tabla 5. Clasificación numérica de las variables

Para la definición de las funciones de probabilidad y el estudio de las diferentes variables explicativas se pueden considerar diversos niveles de desagregación. De esta forma se amplía el espectro de modelos propuestos con el objetivo de aumentar la capacidad predictiva de los mismos. Se considera en todo caso que el nivel de desagregación es el mismo para todas las variables que intervienen en el modelo. En concreto se dividen las variables en un número de entre dos y diez grupos, más un caso adicional con el máximo grado de desagregación posible.

Las discretizaciones se definen intentando que el número de elementos en las diferentes categorías establecida sea similar. El procedimiento parte de la función de distribución genérica de la variable en la que se definen k intervalos; se toman subgrupos de tamaño $1/k$ y se determina el valor correspondiente de la variable. El planteamiento se recoge en la Figura 2:

Figura 2. Discretización de las variables explicativas

Para el caso de modelos de dos variables el procedimiento comienza con el análisis de todas las posibles parejas de desagregación para las dos variables. En el caso de múltiples variables, se repite el proceso con las dos primeras variables (las que se consideran conjuntas en los modelos que presentan este tipo de variables o bien las dos primeras para aquellos modelos que presentan todas las variables independientes). Seguidamente se evalúa la calidad del modelo de dos variables eligiendo la combinación de desagregación óptima. Sobre ese modelo de dos variables se incorpora la tercera variable; para ésta se elige el nivel óptimo de desagregación mediante el indicador de calidad. Secuencialmente se van incorporando variables adicionales con sus respectivos niveles de desagregación.

3.3.3 Construcción de modelos

Como resultado del análisis estadístico previo, se tienen una serie de variables explicativas con incidencias en la probabilidad de fallo de la red de distribución. El siguiente paso es establecer un modelo predictivo que las combine con el objetivo de lograr una predicción lo más precisa posible sobre la probabilidad de rotura de un determinado elemento de la red.

Para determinar el mejor modelo predictivo, continuando con el planteamiento de la metodología según el cual se analizan todas las variables desde un punto de vista puramente estadístico se analiza un amplio espectro de modelos predictivo generados con diferentes números de variables, formas de combinar y niveles de desagregación.

Mediante el procedimiento de evaluación de la capacidad predictiva (ver apartado 3.5) se determina qué modelo es el más adecuado de una manera objetiva.

Cada uno de estos modelos viene definido por tres parámetros:

- a) Número de variables que se analizan conjuntamente.
- b) Número de variables que se analizan independientemente.
- c) Nivel de desagregación: número de intervalos en que se agrupan las variables.

La notación empleada refleja estos tres parámetros (a, b, c), siguiendo el mismo orden anteriormente descrito.

El número total de variables explicativas consideradas se denomina orden del modelo; esto es la suma de las variables conjuntas e independientes.

El nivel de referencia inicial es el modelo de orden cero, es decir la predicción obtenida como relación entre el número de elementos en los que se registra rotura en el periodo y el número total de elementos de la red analizada.

En la medida que se incorporan variables al modelo se busca una mejora en la predicción. Dado el amplio número de variables explicativas consideradas en el estudio las posibles combinaciones de las mismas y órdenes de los modelos se pueden producir un volumen de análisis excesivo. Por ello el análisis se realiza de manera progresiva incorporando variables al modelo de orden cero analizando la respuesta de la predicción en cada paso.

Puesto que se considera que un modelo en el que se incorporen todas las variables no sería práctico, en esta metodología se analizan modelos de orden 1 a 5. En cualquier caso la metodología y los procedimientos empleados son válidos para modelos de orden superior.

Bajo estas premisas se plantean de forma progresiva los diferentes modelos en orden creciente:

- Modelo de orden 0: referencia que se obtiene simplemente con el número de roturas observadas.
- Modelos de orden 1: se analizan todas las variables explicativas de manera independiente con los diferentes niveles de desagregación.
- Modelos de orden 2: se analizan todas las posibles combinaciones de variables (tomadas de dos en dos) tanto de manera independiente (0, 2, k) como de manera conjunta (2, 0, k).

- Modelos de orden 3: para reducir el número de combinaciones posibles se construyen los modelos de orden 3 partiendo de los diez mejores modelos de orden 2 agregando una tercera variable independiente. Se tienen por tanto modelos con tres variables independientes (0, 3, k) y modelos con dos variables conjuntas más una independiente (2, 1, k).
- Modelos de orden n: se repite de manera análoga el procedimiento tomando los diez mejores modelos de orden n-1 a los que se incorpora una variable independiente adicional.

En todos los modelos se consideran diez niveles de desagregación más uno adicional que considera el mayor nivel de desagregación posible.

3.4 Procedimiento de validación

Dado el amplio espectro de modelos predictivos planteados se requiere de un procedimiento de validación que cuantifique la capacidad predictiva de cada uno de ellos de forma que se puedan elegir los más adecuados. En este sentido se plantea procedimiento basado en la división de los registros de roturas, empleándose un conjunto de datos para la calibración del modelo y el resto de datos para su validación.

3.4.1 Metodología

El procedimiento de validación se realiza por comparación del número de roturas previsto por el modelo (N_p) y el número de roturas observado (N_o) en el periodo y muestra de validación.

En este sentido, la extracción de muestras sigue un criterio geográfico mediante el cual se identifica un punto central aleatorio entorno al cual se define una superficie de muestra con un tamaño limitado (L_{min} , L_{max}). De esta forma tomando un punto de referencia dentro de la red estudiada se toma una porción de la misma como muestra sobre la que se realiza la validación. Esta se lleva a cabo a partir de los pares de valores (N_{pi} , N_{oi}) obtenidos para la muestra i. Para la validación de un modelo se toma un número amplio de muestras que permite tener datos suficientes sobre el comportamiento predictivo del mismo.

La Figura 3 muestra cómo se construye la muestra y se seleccionan los elementos de la red dentro de la misma. De igual forma se observan los puntos de validación a emplear.

Figura 3. (A) Ejemplo de área de muestra de validación: punto central en rojo y elementos de la red en azul.
(B) Puntos de validación procesados.

Para la validación del modelo se emplean parte de los registros de roturas de cada una de las muestras, parte de los datos se utilizan para la calibración del modelo y los datos restantes para la validación y comparación de la predicción con las roturas observadas. Se tiene por tanto para cada periodo de validación $n-1$ periodos de contraste, resultando un total de $n*(n-1)$ validaciones en función del periodo analizado (n años) según se muestra en la Tabla 6.

Validación	Ajuste del modelo			
	1	2	$n-1$	n
1	-	G	H	I
2	A	-	J	K
$n-1$	B	D	-	L
n	C	E	F	-

Tabla 6. Periodos de ajuste del modelo y validación

Con este planteamiento se realiza un ajuste por regresión lineal entre los valores obligando a su paso por el origen de coordenadas. En ese análisis se obtienen dos estadísticos de calidad:

- Pendiente de la recta de regresión (p).
- Coeficiente de correlación R^2 (c).

Ambos parámetros serán mejores cuanto más se aproximen a la unidad. En la Figura 4 siguiente se muestra la interpretación gráfica de ambos parámetros. Cada uno de los puntos de la figura siguiente es el resultado de la predicción realizada para cada muestra sobre la que se aplica el modelo en cuestión.

Figura 4. Interpretación de los estadísticos de ajuste

Como estadístico de calidad del modelo se toma la distancia al punto óptimo: pendiente de la recta de ajuste igual a 1 y coeficiente de correlación igual a 1. Cuya expresión es la siguiente:

$$r = \sqrt{(p - 1)^2 + (c - 1)^2}$$

Donde r es el estadístico de calidad, p es la pendiente de la recta de ajuste y c es la dispersión (coeficiente de correlación). Estos valores son calculados para cada una de las $n*(n-1)$ validaciones que se realizan con cada modelo se tienen los correspondientes estadísticos de calidad

Con este planteamiento la selección del mejor modelo puede hacerse considerando dos alternativas:

- Valor medio de los radios de calidad.
- Radio correspondiente a los valores medios de los radios.

Con el objetivo de tener la mayor fiabilidad en la validación de los modelos se clasifican atendiendo a ambos criterios y se seleccionan los que presenten mejores estadísticos globales de calidad. De esta forma se dispone de una metodología que proporciona un criterio cuantitativo para determinar la calidad de los modelos empleados permitiendo la comparación directa entre ellos y la elección de la mejor predicción.

El procedimiento así descrito es objeto de un análisis de sensibilidad (ver apartado siguiente) con el objetivo de calibrar la metodología, y poder eliminar la incertidumbre asociada al posible condicionamiento de los resultados por el procedimiento de validación empleado.

3.4.2 Valor medio de los radios de calidad – R1

Esta validación se realiza con el valor medio de los radios de calidad obtenidos para cada uno de los periodos de validación. La fórmula a considerar en este procedimiento de validación es la siguiente:

$$R1 = \frac{\sum_{i=1}^{n \cdot (n-1)} r_i}{n \cdot (n-1)}$$

Se obtiene también la desviación típica mediante la siguiente fórmula:

$$SD = \sqrt{\frac{1}{n \cdot (n-1)} \sum_{i=1}^{n \cdot (n-1)} (r_i - R1)^2}$$

La interpretación gráfica del procedimiento de validación se muestra en la Figura 5. El radio medio se representa en el eje vertical mientras que la desviación típica se indica en el eje horizontal. Se representan los 12 resultados obtenidos para un mismo modelo en un mismo color; resaltándose con un punto grueso el promedio de los radios. En marrón se indica el valor de referencia sin considerar variables explicativas (modelo de orden cero). En azul se representa el mejor modelo predictivo. El espacio vertical que separa ambas líneas de referencia es la mejora que se obtiene con la incorporación de las variables empleadas en el modelo.

Figura 5. Interpretación del Valor medio de los radios de calidad R1

3.4.3 Radio correspondiente a los valores medios de los parámetros – R2

El segundo procedimiento de validación emplea el radio correspondiente a los valores medios de cada uno de los estadísticos de calidad. Para ello es preciso obtener el valor promedio de la pendiente y el valor promedio del coeficiente de correlación según las expresiones siguientes:

$$P = \frac{\sum_i^{n \cdot (n-1)} p_i}{n \cdot (n-1)}$$

$$C = \frac{\sum_i^{n \cdot (n-1)} c_i}{n \cdot (n-1)}$$

Donde P es el valor medio de la pendiente de ajuste, p_i es el valor de la pendiente de ajuste para cada una de las $n \cdot (n-1)$ validaciones, C es el valor medio del coeficiente de correlación y c_i es el valor de dicho coeficiente para cada una de las $n \cdot (n-1)$ validaciones.

La interpretación gráfica de este nuevo vector (P, C) se muestra en la Figura 6 donde cada una de las letras corresponde a cada una de las $n \cdot (n-1)$ validaciones empleadas. Se señala en azul el punto medio de coordenadas (P, C) y el estadístico de calidad como distancia de este punto al punto óptimo de coordenadas (1, 1). De esta forma se muestra la representación de cada una de las posibles validaciones de un mismo modelo y el

resultado (valor medio de los estadísticos de calidad) que se empleará en más adelante para comparar el comportamiento de los diferentes modelos.

Figura 6. Interpretación gráfica de los valores medios de los estadísticos de calidad.

Este planteamiento por tanto sigue la formulación siguiente en la que se calcula el radio o distancia al punto óptimo (1, 1):

$$R2 = \sqrt{(P - 1)^2 + (C - 1)^2}$$

Donde R2 es el estadístico de calidad, P y C son los valores promedios de la pendiente de ajuste y la dispersión según se describe con anterioridad.

En la Figura 7 se muestra la comparativa de los diferentes modelos combinando el estadístico de calidad R2 y las diferentes validaciones analizadas para cada orden de modelo. Los resultados de P y C para cada modelo se presentan con puntos de un mismo color, siendo cada uno de los puntos el resultado de cada validación. El punto de partida (modelo de orden cero) se indica con una línea discontinua marrón, mientras que el mejor modelo crea un arco con trazo discontinuo verde pasando por un único punto correspondiente con la representación del mejor modelo.

Figura 7. Interpretación del Radio R^2 correspondiente a los valores medios de los parámetros

3.5 Análisis de sensibilidad

Para chequear el proceso de validación se realiza un análisis de sensibilidad. El objetivo es cuantificar la incertidumbre de los resultados en función de los estadísticos de calidad del procedimiento de validación aplicado, y por ello, de los parámetros empleados en cada procedimiento. Por ello, el análisis realizado se centra en los tres parámetros siguientes:

- Tamaño de las zonas que componen la muestra.
- Tamaño de la muestra (número de zonas).
- Periodo de validación.

Este análisis de sensibilidad se desarrolla a partir de un análisis cualitativo y otro cuantitativo de los resultados obtenidos para diferentes ajustes de cada uno de los parámetros anteriores.

El estudio del tamaño de las zonas se realiza modificando los límites superiores e inferiores del tamaño (L_{max} y L_{min}), considerando en todos los casos un número de elementos comprendidos dentro de la zona, independientemente de su tamaño.

En cuanto al tamaño de la muestra entendido como número de unidades que se toman para la validación de los modelos; es con este número de unidades con las que se realiza el ajuste de regresión que determina los estadísticos de calidad. En este sentido se realiza el análisis de sensibilidad con tamaños de muestra entre 100 y 500 unidades.

Para el periodo de validación se realiza una descomposición del periodo total de datos en n años completos realizándose un conjunto de $n*(n-1)$ validaciones diferentes tomando cada vez un periodo para calibrar el modelo y los restantes periodos (años) para su validación.

Una vez que se calibran estos parámetros se plantea un procedimiento de validación de los modelos que minimiza la dependencia de los resultados en el método de validación elegido.

3.5.1 Comparación cualitativa

Para realizar el análisis cualitativo se aplica la metodología propuesta en apartados anteriores para diferentes valores de un mismo parámetro del procedimiento de validación y se identifica el modelo con la mejor capacidad predictiva. Comparando los resultados obtenidos, cuando se verifica que los resultados se mantienen a pesar de que se adopten diferentes valores de cada uno de los parámetros se considera el procedimiento de validación suficientemente robusto.

Se analizan los resultados de los estadísticos de calidad (pendiente, dispersión y radio) para cada modelo estudiando el comportamiento con tres tipos de gráfico: resultados de dispersión y pendiente de ajuste, ajuste y dispersión de los resultados y gráficos de escalas de colores donde se comparan los valores del radio como parámetro de calidad para los modelos analizados.

Primero se comparan los gráficos de comportamiento del radio correspondiente a los valores medios de los estadísticos como en la Figura 8. A continuación se incluye un ejemplo de la sensibilidad de los modelos al tamaño de la muestra.

Figura 8. Análisis de sensibilidad – Comparación del Radio correspondiente a los valores medios de los parámetros. A la izquierda para tamaño de muestra 100 y a la derecha para tamaño de muestra 500.

En un segundo paso se comparan los gráficos de ajuste-dispersión, Figura 9:

Figura 9. Análisis de sensibilidad – Comparación del ajuste y dispersión. A la izquierda para tamaño de muestra 100 y a la derecha para tamaño de muestra 500.

En ambos casos se analiza la tendencia general de los resultados, en la medida que las gráficas son análogas el procedimiento de validación no es muy sensible a la variación del parámetro, por el contrario comportamientos diferentes en las gráficas denotan que los resultados del procedimiento de validación se ven afectados por la variación de este parámetro.

Por último se comparan los gráficos de resumen de resultados donde se emplea una escala de colores para representar el indicador de calidad, radio, Figura 10.

Figura 10. Análisis de sensibilidad – Comparación de los gráficos resumen de resultados por colores. A la izquierda para tamaño de muestra 100 y a la derecha para tamaño de muestra 500.

3.5.2 Comparación cuantitativa

La comparación cuantitativa se realiza comparando los resultados de los estadísticos de calidad (pendiente, dispersión y radio) en función del factor a analizar. Estos resultados se comprueban para todos los modelos de una variable (orden uno), y para las dos variables que presentan mejores resultados en los modelos de orden dos. La figura siguiente muestra el resultado obtenido para el factor del tamaño de la muestra como ejemplo del análisis de sensibilidad:

Figura 11. Análisis de sensibilidad – Comparación cuantitativa

Mediante las gráficas anteriores (Figura 11) se aprecia como al aumentar o disminuir el factor analizado varía el comportamiento del modelo, es decir de los estadísticos de calidad obtenidos, en mayor o menor medida. De su análisis se pueden extraer conclusiones importantes sobre la sensibilidad del método de validación a cada uno de los parámetros estudiados.

4 PARTICULARIZACIÓN DE LA METODOLOGÍA A LA RED DE CANAL DE ISABEL II

4.1 Red de distribución del Canal de Isabel II

4.1.1 Descripción de la red

La metodología propuesta se aplica a la red del Canal de Isabel II, compañía encargada de la gestión del ciclo integral del agua en la Comunidad de Madrid, España, con más de 6,5 millones de habitantes y 179 municipios. En esta región la empresa ha implementado más de 800 sectores hidráulicos con una inversión de más de 15 millones de euros. La red de distribución así configurada se compone de más de 370.000 tramos de tuberías de distribución con una longitud total de más de 14.000 km.

Esta red dispone de un sistema de información geográfica (GIS), un importante sistema de automatización y telecontrol y bases de datos de incidencias y actuaciones vinculadas al abastecimiento. Gracias a ello, se pueden identificar múltiples variables y parámetros de cada uno de los tramos de la red por lo que se tiene una caracterización detallada de la misma. Esta característica, unida a la configuración sectorizada mediante la cual la red se divide en sectores monitorizados (también denominados district metered areas o DMAs), permite tener un conocimiento detallado del funcionamiento de la red. Adicionalmente el equipo de I+D+i de la empresa dispone de modelos hidráulicos calibrados de toda la red que periódicamente son actualizados. Estos modelos aportan información específica sobre las condiciones de la red, particularizadas para cada tramo de tubería, que permiten analizar su comportamiento con detalle.

4.1.2 Datos

El planteamiento del estudio parte de la información disponible en los diferentes sistemas y registros del Canal de Isabel II. Con esta información se pueden caracterizar los diferentes tramos de la red de distribución de Canal de Isabel II para poder realizar el estudio estadístico asociado a la ocurrencia de roturas. En este sentido se han empleado las siguientes fuentes:

4.1.2.1 Sistema de información geográfica

En este sistema de información geográfica del Canal de Isabel II se encuentran caracterizados los diferentes elementos de la red para los que se incluyen características como: diámetro, año de instalación, material, ubicación geográfica, etc. Así el sistema recoge un gran volumen de datos de todos los elementos de la red de manera georeferenciada. En este sentido se recogen los siguientes tipos de elementos: tubos, caudalímetros, contadores, desagües, dispositivos de purga, filtros, puntos de acometida, puntos de medida, válvulas, ventosas, sectores, etc.

El sistema está diseñado para disponer de un inventario de las redes de abastecimiento y saneamiento gestionadas por Canal de Isabel II, también se emplea para el almacenamiento y la gestión de la información de toda la red de manera unificada, permite la difusión de la información a todos los profesionales de la compañía. El sistema también dispone de diferentes aplicaciones de análisis, edición y consulta de la información de forma que todos los datos disponibles puedan ser aprovechados para la gestión más adecuada de la red.

4.1.2.2 Sistema de gestión de incidencias

El sistema de gestión de incidencias del Canal de Isabel II recoge incidencias, avisos, trabajos y mantenimiento de la red gestionada. En esta base de datos se registran diferentes tipos de sucesos asociados con anomalías en el servicio prestado.

Los avisos son reclamaciones que pueden estar relacionadas con el servicio prestado en términos de calidad y de cantidad. Estos registros son comunicados tanto por departamentos internos de la empresa que identifican algún tipo de anomalía como por clientes que realizan comunicaciones al percibir algún tipo de deficiencia en el servicio prestado.

Las incidencias son anomalías efectivas que se crean una vez que los avisos han sido analizados y confirmados.

Los trabajos identifican actividades realizadas con una planificación y antelación que pueden alterar el normal funcionamiento de la red.

Las actuaciones reflejan todas aquellas operaciones que se realizan para resolver una incidencia o realizar un determinado trabajo.

4.1.2.3 Modelos hidráulicos

En el Canal de Isabel II se han desarrollado modelos numéricos de comportamiento hidráulico de la red de distribución que proporciona, entre otros valores, los de las variables hidráulicas básicas presión y velocidad para todos los componentes de la red bajo distintas hipótesis de funcionamiento. Estos modelos están calibrados con los datos de la red obtenidos en tiempo real y son revisados con una periodicidad semestral.

4.1.2.4 Otras fuentes

Para complementar la caracterización de la red se emplean otras fuentes de información externas a la propia gestión de Canal de Isabel II. De esta forma la caracterización del uso del suelo se extrae del Sistema de Información sobre Ocupación del Suelo de España SIOSE, coordinado por el Instituto Geográfico Nacional y el Centro Nacional de Información Geográfica. La caracterización geotécnica se realiza mediante el Sistema de Información Geográfica de la Comunidad de Madrid, para lo cual se consulta el Mapa Geológico Digital Continuo GEODE realizado por el Instituto Geológico y Minero de España.

4.2 Registros de roturas

Para realizar el estudio de roturas se han empleado la información recopilada por Canal de Isabel II al respecto. En el año 2010 se creó una base de datos en el que se recogen múltiples datos sobre las roturas registradas para poder identificar las causas de la roturas. La ventaja de esta base de datos es que incorpora un gran número de variables que permiten caracterizar las roturas:

- Identificación del elemento: diámetro espesor, material, longitud, antigüedad, etc.
- Localización: municipio y coordenadas X e Y.
- Características del terreno de asiento y de la zanja: tipo de terreno, presencia de yesos, nivel freático, tipo de relleno de zanja, tipo y espesor de la cama de asiento, etc.
- Datos hidráulicos: se incorporan datos de presión, ya sea estática o dinámica registrada en los sistema de información de Canal de Isabel II, oscilaciones de presión, etc.

- Datos recabados in situ durante la reparación de la rotura identificada: presencia de raíces o humedades, uso del suelo, tráfico, corrosión, fisuras, aplastamiento, localización de la rotura (en la pared del tubo, en una junta o pieza especial, etc.).
- Análisis técnico de la rotura: resultados del estudio realizado con la información recabada como por ejemplo la causa y el modo del fallo, observaciones o particularidades identificadas, etc.

Para el presente estudio se emplean los datos de esta base de datos en el periodo 2010 a 2015.

4.3 Selección de variables

Con la información disponible según las fuentes anteriormente descritas se selecciona una serie de variables que pueden incorporarse al estudio estadístico por disponerse de información suficiente y fiable. De entre todas las variables seleccionadas únicamente se estudiarán aquellas en las que se observe una incidencia estadística significativa en la ocurrencia de roturas siguiendo el procedimiento descrito en la metodología planteada. Los rangos de los valores de cada una de las variables explicativas seleccionadas se recogen en el Anexo 1.

4.3.1 Variables explicativas principales

Las variables consideradas en la aplicación de la metodología planteada en la red de Canal de Isabel II son las siguientes:

- Diámetro: se toma como variable explicativa el diámetro nominal de la tubería. Los diámetros observados en la serie estudiada oscilan entre 20 y 3000 mm.
- Año de instalación: se dispone de datos desde el 1 de noviembre de 2010 al 2 de septiembre de 2015. El tratamiento de esta variable se explica tras el análisis de sensibilidad.
- Material: los materiales identificados en la serie de tubería estudiadas son los siguientes: acero, caña, fibrocemento, fundición dúctil, fundición gris, hormigón, polietileno y PVC. Dentro de las dos últimas categorías se diferencian entre aquellas que cumplen con la norma del Canal de Isabel II y las que no.

4.3.2 Variables explicativas secundarias

Con la información disponible para el caso de la red de estudio las variables secundarias se agrupan en los siguientes factores:

- Factores internos.
- Factores físicos o medioambientales.

Factores internos:

Velocidad del agua: se establecen unos rangos de velocidades para establecer grupos homogéneos de tuberías, si bien durante el proceso de formulación de los modelos y su posterior validación se plantean diferentes niveles de desagregación según el rango de velocidades observadas que oscilan entre 0 y 1 m/s.

Variaciones de presión: los valores que se emplean son escalones de presión de 10 mca hasta los 140 mca, escalones de 20 mca hasta 200 y un límite superior absoluto de 320 mca.

Presiones transitorias: puesto que no se disponen de valores de dichas presiones transitorias se descarta su incorporación en los modelos predictivos.

Calidad del agua: la calidad del agua puede tener influencia por su incidencia en los fenómenos de corrosión interna de las tuberías, lo cual podría desencadenar en rotura. Sin embargo, en general no se tienen datos suficientes para poder ser tenido en cuenta.

Factores físicos o medioambientales:

Tipo de terreno (características físicas y químicas): la caracterización del tipo de terreno se realiza empleando el Mapa Geológico Digital Continuo (GEODE), resultado de la agrupación, correlación y digitalización de la Serie MAGNA (1:50.000) del Instituto Geológico y Minero de España (IGME). Dicha información se incorpora a la información del sistema GIS para la asignación del tipo de terreno correspondiente a cada una de las tuberías de la red. En función de la capacidad portante de los terrenos identificados se establece la siguiente clasificación: masas de agua, yesos, arcillas, limos, arenas, gravas y rocas.

Tipo de ubicación (transmisión de cargas al elemento): las ubicaciones detectadas son: tubería enterrada, en bypass, en galería, hincada, en superficie, tuberías de aspiración y tuberías de impulsión.

Profundidad de enterramiento: se considera la profundidad de enterramiento de las diferentes tuberías de la red obtenida mediante un modelo digital de elevaciones en el que se incorporan todos los datos disponibles de los diversos elementos de la misma (válvulas, desagües, dispositivos de purga y ventosas). La cota considerada para cada tubería es la asociada al punto medio de cada una ellas. Para la elaboración de dicho MDE se filtran ciertos valores de profundidad para evitar distorsiones en el modelo. Los puntos eliminados son: profundidades de 0 metros y mayores de 10 metros y elementos que se encuentran en ubicaciones tales como exterior depósito, galerías, superficie o almenara.

Tipo de junta: no se dispone de información suficiente.

Uso del suelo y exposición al tráfico (calzada, acera, etc.): los usos del suelo se toman del Sistema de Información sobre ocupación del Suelo de España (SIOSE). Adicionalmente, se incorpora a dicho sistema la información propia del sistema GIS del Canal de Isabel II que identifica con mayor grado de detalle las líneas de transporte (calles, carreteras, caminos, metro, cercanías y AVE). De cara al estudio de las roturas se identifican tres categorías de uso de suelo que podrían tener más relevancia por las cargas de tráfico que son zonas de transportes, zonas urbanas e industriales con algo menos de cargas de tráfico y zonas libres que incluyen zonas forestales, agrícolas y zonas con poca o ninguna vegetación.

4.4 Dependencia estadística

Según se plantea en la metodología desarrollada, resulta necesario verificar la dependencia estadística de las variables seleccionadas para poder ser incluidas en el procedimiento de creación y validación de modelos. Empleando los registros de roturas y la información de las mismas se realizan sendos análisis de dependencia estadística tipo K-S.

Tal y como se recoge en el Anexo 2, los resultados de los análisis de inferencia estadística concluyen que todas las variables seleccionadas presentan una dependencia estadística y por tanto son susceptibles de incluirse en el procedimiento de creación de modelos predictivos.

4.5 Variables consideradas

Resumiendo los apartados anteriores se emplean las siguientes variables explicativas para caracterizar las roturas (Tabla 7).

Número	Variable explicativa considerada	Código	Tipo
1	Año de instalación	Año	Principal
2	Material	Mat	Principal
3	Diámetro	Diam	Principal
4	Tipo de ubicación	Ubic	Secundaria ambiental
5	Tipo de terreno	Terr	Secundaria ambiental
6	Profundidad de enterramiento	Prof	Secundaria ambiental
7	Uso del suelo	Uso	Secundaria ambiental
8	Presión mínima	P min	Secundaria interna
9	Presión media	P med	Secundaria interna
10	Presión máxima	P max	Secundaria interna
11	Velocidad mínima	V min	Secundaria interna
12	Velocidad media	V med	Secundaria interna
13	Velocidad máxima	V max	Secundaria interna

Tabla 7. Variables consideradas.

5 ANÁLISIS DE SENSIBILIDAD PARA LA RED DE CANAL DE ISABEL II

En este apartado se recogen los resultados obtenidos de la aplicación del análisis de sensibilidad descrito en la metodología propuesta a los datos de la red de distribución de Canal de Isabel II. Se analiza cada uno de los tres factores del procedimiento de validación:

- Tamaño de las zonas que componen la muestra.
- Tamaño de la muestra (número de zonas).
- Periodo de validación.

Y se concluye con el resultado de todo el proceso y la definición de la validación obtenida tras el análisis de sensibilidad. Se minimiza así la influencia de dichos parámetros en los resultados obtenidos dotando al resultado obtenido de la aplicación de la metodología de una mayor objetividad y capacidad predictiva.

5.1 Análisis del tamaño de las zonas de validación

Como se ha explicado en la metodología, con el objetivo de contrastar la influencia del tamaño de las zonas de validación en los resultados obtenidos se realiza un doble análisis: cualitativo y cuantitativo. La conclusión tras este análisis es la definición de los límites inferior y superior ($L_{\min}$ y $L_{\max}$) con los que se definen las áreas de validación empleadas para el ajuste y la validación de los modelos empleados.

5.1.1 Tamaño de las zonas de validación - Análisis cualitativo

A través del análisis cualitativo se estudian los resultados de un conjunto de variables significativas comparando el resultado obtenido con diferentes tamaños. En este sentido se estudian los resultados para muestra de tamaños:

- a) $L_{\min} = 10.000$ m y $L_{\max} = 40.000$ m
- b) $L_{\min} = 40.000$ y $L_{\max} = 70.000$ m

En ambos casos se fija el periodo de ajuste entre 1/11/2010 y 02/04/2013 y el periodo de validación 03/04/2013 y 02/09/2015. El tamaño de las muestras en ambos casos es de 200 unidades.

En un primer paso se muestran los resultados obtenidos para los modelos de orden cero, Figura 12.

Figura 12. Análisis cualitativo tamaño de muestra – resultados modelos de orden cero. (a) Muestras 10.000-40.000 m. (b) Muestras 40.000-70.000 m.

Del análisis de dichos modelos de orden cero se observa que al aumentar el tamaño de la zona el comportamiento mejora. Se observa que en la figura (b) se tiene una pendiente de ajuste más próxima a uno y un radio también ligeramente más próximo a uno.

El siguiente paso es la comparación de los resultados obtenidos en los modelos de orden uno y un conjunto de modelos de orden dos como se recoge en las tablas siguientes (Tabla 8 y Tabla 9).

Variable	Zonas de 10-40 km				Zonas de 40-70 km			
	Desagreg	Pendiente	Coef. R^2	Radio	Desagreg	Pendiente	Coef. R^2	Radio
Diámetro	1	0.918	0.935	0.105	1	1.037	0.947	0.065
Año	5	0.994	0.974	0.026	2	1.014	0.967	0.036
Terreno	4	0.933	0.936	0.092	1	1.037	0.947	0.065
Material	0	0.965	0.990	0.037	0	0.957	0.983	0.046
Uso del suelo	1	0.918	0.935	0.105	0	1.026	0.949	0.057
Presión med	1	0.998	0.942	0.058	1	1.084	0.940	0.103
Presión máx	1	0.998	0.942	0.058	1	1.084	0.940	0.103
Presión min	1	0.998	0.942	0.058	1	1.084	0.940	0.103
Velocidad med	0	0.951	0.938	0.079	1	1.055	0.945	0.078
Velocidad máx	0	0.951	0.939	0.078	1	1.055	0.945	0.078
Velocidad min	0	0.942	0.937	0.086	2	1.055	0.945	0.078
Profundidad	7	0.928	0.947	0.089	0	1.018	0.952	0.052

Tabla 8. Comparación de los estadísticos de calidad del modelo de orden uno en función del tamaño de la zona de validación

Los resultados de los modelos de orden uno son significativamente diferentes, observándose diferencias entre los radios obtenidos para un mismo modelo. Además hay variación en los modelos que ofrecen mejores resultados según el tamaño de la zona. En este sentido no se identifica una correlación entre los mejores modelos en un planteamiento y en otro.

Variable	Zonas de 10-40 km					Zonas de 40-70 km				
	Variab. 2	Desag.	Pendiente	Coef. R ²	Radio	Variab. 2	Desag.	Pendiente	Coef. R ²	Radio
Diám	Año	0	0.991	0.990	0.013	Año	8	0.974	0.978	0.034
Año	Mat	0	0.993	0.991	0.011	P min	3	1.010	0.971	0.031
Terr.	Año	5	0.995	0.977	0.024	Año	2	1.014	0.967	0.036
Mat.	Año	0	0.993	0.991	0.011	P min	2	0.997	0.979	0.021
Uso	Año	5	0.996	0.976	0.024	Año	2	1.007	0.968	0.033
P med	Año	3	0.975	0.981	0.031	Mat.	2	1.007	0.978	0.023
P máx	Año	3	0.985	0.982	0.023	Mat.	2	1.009	0.977	0.024
P min	Mat	2	1.022	0.990	0.024	Mat.	2	0.997	0.979	0.021
V med	Mat	0	0.986	0.991	0.017	Año	4	0.998	0.966	0.034
V máx	Mat	0	0.989	0.990	0.015	Mat.	0	0.972	0.982	0.033
V min	Mat	0	0.984	0.991	0.019	Año	4	0.993	0.967	0.034
Prof	Año	5	1.009	0.980	0.022	Año	2	1.015	0.967	0.036

Tabla 9. Comparación de los estadísticos de calidad del modelo de orden dos en función del tamaño de la zona de validación

En cuanto a los resultados de los modelos de orden dos, de nuevo se encuentran diferencias, obteniéndose por lo general mejor capacidad predictiva con las zonas de menor tamaño. En cuanto a los mejores modelos, se observan diferencias entre las combinaciones que ofrecen mejores resultados en un caso y en otro, aunque cabe destacar que en los casos en los que la segunda variable es el año hay más similitudes.

El siguiente paso de la validación es la comparación de los gráficos del análisis de regresión así como los que recogen los valores de los radios medios (R2). Para el caso de los modelos de una variable se observa que el comportamiento es diferente en un caso y en otro.

Figura 13. Análisis cualitativo tamaño de la zona de validación – resultados R2 para modelos de una variable. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.

La comparación de los ajustes de regresión de la variable año o material (ver Figura 14 y Figura 15) presenta diferencias aunque los valores del ajuste son similares (pendiente y regresión).

Figura 14. Análisis cualitativo tamaño de la zona de validación – resultados regresión lineal para modelos año. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km..

Figura 15. Análisis cualitativo tamaño de la zona de validación – resultados regresión lineal para material. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.

Los modelos de orden dos que se forman con la variable año también ofrecen resultados diferentes como se puede ver en la figura siguiente.

Figura 16. Análisis cualitativo tamaño de la zona de validación – resultados R^2 para la variable año. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.

Según la Figura 17 el ajuste de la variable año con material (la combinación que ofrece el mejor resultado de orden dos en las zonas de validación pequeñas) presenta un comportamiento diferente para las zonas grandes teniendo una pendiente peor (0.993 para zonas pequeñas y 0.967 para las zonas grandes). De nuevo se aprecian diferencias de comportamiento entre un tamaño y otro.

Figura 17. Análisis cualitativo tamaño de la zona de validación – resultados regresión lineal para año con material. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.

Para los modelos de orden dos en los que interviene la variable diámetro también presentan diferencias de comportamiento como se aprecia en la Figura 17. Si bien es cierto que se muestran algunas coincidencias como el hecho de que el mejor modelo se forma combinando diámetro con año. Sin embargo, comparando el ajuste de regresión del modelo diámetro-año (ver Figura 19) se observa que en el caso de las zonas de menor tamaño la pendiente ajuste es mejor al igual que la dispersión.

Figura 18. Análisis cualitativo tamaño de la zona de validación – resultados R2 para diámetro. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.

Figura 19. Análisis cualitativo tamaño de la zona de validación – resultados regresión lineal para diámetro con año. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.

Por último se analiza el gráfico resumen de resumen de resultados en los que se presenta el valor del indicador de calidad para todas las combinaciones de variables (Figura 20).

Figura 20. Análisis cualitativo tamaño de la zona de validación – resultados estadísticos de calidad de todas las combinaciones de variables – escala de colores. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.

En este gráfico se observa como la tendencia de la variación del resultado con ambos tamaños de áreas es muy diferente y se aprecia que en el caso de las áreas de validación grandes (40.000-70.000 m) hay muchas más combinaciones que no mejoran los resultados del modelo de orden cero (recuadros en gris).

Queda claro tras este análisis cualitativo que el comportamiento de los modelos se ve condicionado por el tamaño de las zonas de validación.

5.1.2 Tamaño de las zonas de validación - Análisis cuantitativo

El análisis cuantitativo del tamaño de la zona de validación se basa en los estadísticos de calidad obtenidos para diferentes modelos. Partiendo de un total de 500 zonas con tamaños variables entre 10.000 y 70.000 m, se consideran grupos de muestras de 200 elementos con un tamaño medio creciente. Para realizar este análisis se ordenan las zonas por orden creciente y se van seleccionando muestras de 200 elementos. Esta selección se hace de manera progresiva incrementando 25 puestos en cada muestra y seleccionando los 200 elementos correlativos en cada paso de 25. De esta forma se disponen de 13 grupos sobre los que se realiza un análisis de los estadísticos de calidad, estos resultados (pendiente, dispersión y radio) se muestran en los gráficos siguientes tomando el tamaño medio para cada uno de los grupos.

Se observa que el tamaño de las zonas de validación presenta incidencia en los resultados. En el caso de los modelos de orden 1 se aprecia como existen variaciones importantes en función del tamaño (no hay tramos horizontales en las gráficas, Figura 21). Otra característica importante es el hecho de que las líneas no siempre permanecen paralelas lo que muestra la influencia del tamaño de la zona en los resultados.

Figura 21. Análisis cuantitativo tamaño de la zona de validación – resultados estadísticos de calidad. Modelos orden 1. (a) Pendiente. (b) Dispersión. (c) Radio.

Estas diferencias son más acusadas si se analiza el radio obtenido en los modelos de orden 2, tanto los que se construyen con la variable año (Figura 22) como los que se generan con la variable diámetro (Figura 23).

Figura 22. Análisis cuantitativo tamaño de la zona de validación – resultados estadísticos de calidad. Modelos orden 2, año. (a) Pendiente. (b) Dispersión. (c) Radio.

Figura 23. Análisis cuantitativo tamaño de la zona de validación – resultados estadísticos de calidad. Modelos orden 2, diámetro. (a) Pendiente. (b) Dispersión. (c) Radio.

5.2 Análisis del tamaño de la muestra

El tamaño de la muestra se define como el número de zonas que se analizan (de un tamaño determinado) sobre las que se realiza la predicción de roturas y se compara con el número de roturas registrado. El tamaño de la muestra define por tanto el volumen de datos a analizar mediante la regresión lineal.

Se parte de un análisis con 100 muestras y se incrementa hasta 500. El proceso se plantea con un análisis cualitativo y otro cuantitativo.

5.2.1 Tamaño de la muestra – Análisis cualitativo

El análisis cualitativo se realiza con todas las variables de manera independiente y con una selección de los modelos de 2 variables, tomando aquellos que ofrecen mejores resultados. Se realiza una comparación entre los resultados para un tamaño de muestra de 100 unidades y otro de 500. Se mantiene el periodo de ajuste constante (01/11/2010 a 02/04/2013) así como el de validación (03/04/2013 a 02/09/2015). El tamaño de las zonas se toma dentro de los límites: $L_{\min} = 10.000$ m y $L_{\max} = 40.000$ m.

Como primera aproximación se comparan los estadísticos de calidad obtenidos para los modelos analizados con los dos tamaños de muestra. Como se presenta en la Tabla 10, para los modelos de una única variable los resultados son similares, obteniéndose estadísticos de calidad cualitativamente muy parecidos con ligeras diferencias poco significativas.

Variable	Muestra de 100 unidades				Muestra de 500 unidades			
	Desagreg	Pendiente.	Coef. R^2	Radio	Desagreg	Pendiente.	Coef. R^2	Radio
Diámetro	1	0.918	0.935	0.105	1	0.938	0.948	0.081
Año	5	0.994	0.974	0.026	5	1.006	0.979	0.021
Terreno	4	0.933	0.936	0.092	4	0.953	0.949	0.069
Material	0	0.965	0.990	0.037	0	0.966	0.9990	0.035
Uso del suelo	1	0.918	0.935	0.105	1	0.938	0.948	0.081
Presión med	1	0.998	0.942	0.058	1	1.018	0.953	0.051
Presión máx	1	0.998	0.942	0.058	1	1.018	0.953	0.051
Presión min	1	0.998	0.942	0.058	1	1.018	0.953	0.051
Velocidad med	0	0.951	0.938	0.079	0	0.971	0.950	0.058
Velocidad máx	0	0.951	0.939	0.078	0	0.971	0.951	0.057
Velocidad min	0	0.942	0.937	0.086	0	0.961	0.945	0.064
Profundidad	7	0.928	0.947	0.089	7	0.947	0.957	0.068

Tabla 10. Comparación de los estadísticos de calidad del modelo de orden uno en función del tamaño de muestra

Analizando los modelos de orden dos (ver Figura 24), los resultados son muy similares tanto por las combinaciones que ofrecen los mejores resultados como los valores de los resultados de los modelos.

Variable	Muestra de 100 unidades					Muestra de 500 unidades				
	Variab. 2	Desag.	Pendiente	Coef. R^2	Radio	Variab. 2	Desag.	Pendiente	Coef. R^2	Radio
Diám	Año	0	0.996	0.991	0.010	Año	0	0.996	0.991	0.010
Año	Mat	0	0.997	0.991	0.009	Mat	0	0.998	0.991	0.009
Terr.	Año	5	0.995	0.975	0.025	Año	5	1.009	0.980	0.022
Mat.	Año	0	0.997	0.991	0.009	Año	0	0.998	0.991	0.009
Uso	Año	5	0.996	0.975	0.025	Año	5	1.009	0.980	0.022
P med	Año	3	0.981	0.981	0.027	Año	3	0.988	0.983	0.021
P máx	Año	3	0.991	0.981	0.020	Año	3	0.988	0.983	0.017
P mín	Mat	2	1.027	0.990	0.029	Mat	2	1.028	0.991	0.029
V med	Mat	0	0.991	0.991	0.013	Mat	0	0.991	0.991	0.013
V máx	Mat	0	0.994	0.991	0.011	Mat	0	0.994	0.991	0.011
V mín	Mat	0	0.989	0.991	0.014	Mat	0	0.989	0.991	0.014
Prof	Mat	0	0.981	0.991	0.021	Mat	0	0.983	0.991	0.020

Tabla 11. Comparación de los estadísticos de calidad del modelo de orden dos en función del tamaño de muestra

Analizando los gráficos del análisis de regresión así como los que recogen los valores de los radios medios (R^2), se observan las mismas tendencias de comportamiento para los dos tamaños de muestra comparados.

Figura 24. Análisis cualitativo tamaño de muestra – resultados R^2 para modelos de una variable. (izq) tamaño de muestra 100 unidades. (der) tamaño de muestra 500 unidades.

Figura 25. Análisis cualitativo tamaño de muestra – resultados regresión lineal para la variable año. (izq) tamaño de muestra 100 unidades. (der) tamaño de muestra 500 unidades.

Figura 26. Análisis cualitativo tamaño de muestra – resultados regresión lineal para la variable material. (izq) tamaño de muestra 100 unidades. (der) tamaño de muestra 500 unidades.

Figura 27. Análisis cualitativo tamaño de muestra – resultados R2 para año. (izq) tamaño de muestra 100 unidades. (der) tamaño de muestra 500 unidades.

Figura 28. Análisis cualitativo tamaño de muestra – resultados regresión lineal para la variables año y material. (izq) tamaño de muestra 100 unidades. (der) tamaño de muestra 500 unidades.

Figura 29. Análisis cualitativo tamaño de muestra – resultados R2 para la variable diámetro. (izq) tamaño de muestra 100 unidades. (der) tamaño de muestra 500 unidades.

Figura 30. Análisis cualitativo tamaño de muestra – resultados regresión lineal para la combinación de variables diámetro y año. (izq) tamaño de muestra 100 unidades. (der) tamaño de muestra 500 unidades.

El análisis cualitativo concluye con la comparación de los estadísticos de calidad de todas las combinaciones de variables estudiadas en el análisis de sensibilidad (modelos de orden 2). Este análisis se realiza mediante una escala de colores en la que el indicador de calidad del modelo de orden cero se toma como referencia, es decir que todas aquellas combinaciones que obtengan resultados peores que el modelo de orden cero se muestran en color gris.

Figura 31. Análisis cualitativo tamaño de muestra – resultados de los estadísticos de calidad de todas las combinaciones de variables – escala de colores.

En la Figura 31 se observa como la tendencia de la variación del resultado con ambos tamaños de muestra es similar.

5.2.2 Tamaño de la muestra – Análisis cuantitativo

Para el análisis cuantitativo se analizan los resultados de los estadísticos de calidad para tamaños de muestra entre 100 y 500 unidades. Como en el paso anterior, se analizan todos los modelos de una variable. Para los modelos de orden dos se analiza la combinación de variables que arrojan los mejores resultados: año y diámetro. Las gráficas siguientes muestran los resultados obtenidos representando en el eje horizontal el tamaño de la muestra y el eje vertical los estadísticos de calidad (de izquierda a derecha): pendiente de ajuste, dispersión y radio.

Figura 32. Análisis cuantitativo tamaño de muestra – resultados de los estadísticos de calidad. Modelos orden 1. (a) Pendiente. (b) Dispersión. (c) Radio.

Figura 33. Análisis cuantitativo tamaño de muestra – resultados de los estadísticos de calidad. Modelos orden 2, año. (a) Pendiente. (b) Dispersión. (c) Radio.

Figura 34. Análisis cuantitativo tamaño de muestra – resultados de los estadísticos de calidad. Modelos orden 2, diámetro. (a) Pendiente. (b) Dispersión. (c) Radio.

El análisis de los resultados mostrados en las gráficas anteriores (Figura 32, Figura 33 y Figura 34) muestra como los resultados obtenidos por encima de las 200 unidades no varían significativamente pues la posición relativa de cada una de las líneas (correspondiente a cada uno de los modelos analizados) no varía su posición relativa. También se observa como al aumentar el tamaño de la muestra la capacidad predictiva va mejorando ya que la pendiente de ajuste se aproxima a la unidad, la dispersión va estabilizándose lentamente y el radio obtenido se va reduciendo.

5.3 Análisis del periodo de validación

La influencia del periodo de validación en los resultados del proceso planteado se analiza en dos pasos.

En un primer paso se analizan los resultados dividiendo los datos en dos periodos con la misma duración, un periodo de ajuste y un periodo de validación. Dado que el periodo de estudio comprende disponibles (desde el 1 de noviembre de 2010 al 2 de septiembre de 2015), se divide en dos periodos idénticos de 2,42 años. Se realiza una comparación cualitativa ajustando el modelo en el primer periodo y evaluando en el segundo periodo los resultados obtenidos con los datos observados y posteriormente la situación inversa.

En un segundo paso se realiza una comparación cuantitativa similar a la realizada con los dos parámetros anteriores dividiendo los datos en cuatro periodos. En este caso se ajusta el modelo con uno de los periodos y se valida con los otros tres; este proceso se repite con todas las combinaciones posibles (12).

5.3.1 Comparación en dos periodos

El primer paso es realizar el modelo de orden cero obtenido con las dos combinaciones de ajuste y validación aplicada a los dos periodos de igual duración. Con esta división se tienen las siguientes series de datos para cada periodo como se recoge en la Tabla 12.

Periodo 1: 01/11/2010 a 02/04/2013			Periodo 2: 03/04/2013 a 02/09/2015		
Número de roturas	Número de roturas identificadas	% de roturas identificadas	Número de roturas	Número de roturas identificadas	% de roturas identificadas
6.347	4.355	68,61%	5.558	5.083	91,45%

Tabla 12. Análisis periodo de validación – comparación en dos periodos. Datos observados en cada periodo.

Como se muestra en la Figura 35 el comportamiento es análogo, si bien la pendiente de ajuste en un caso es inferior a uno y en el otro superior, la distancia al óptimo es similar así como la dispersión, siendo ésta 0,94 en ambos casos.

Figura 35. Análisis periodo de validación – comparación en dos periodos, modelo de orden cero. (izq) Ajuste en periodo 1 y validación en periodo 2. (der) Ajuste en periodo 2 y validación en periodo 1.

El siguiente paso es la comparación de los resultados de todos los modelos de una variable como se recoge en la Tabla 13. Se observa un comportamiento similar en términos de radio obtenido salvo en las variables de presión y velocidad donde se observa una mayor diferencia en los resultados obtenidos.

Variable	Periodo 1 (01/11/2010-02/04/2013)				Periodo 2 (03/04/2013-02/09/2015)			
	Desag.	Pendiente	Coef. R^2	Radio	Desag.	Pendiente	Coef. R^2	Radio
Diám	1	1.076	0.946	0.093	1	0.938	0.948	0.081
Año	4	0.995	0.962	0.038	5	1.006	0.979	0.021
Terr.	1	1.076	0.946	0.093	4	0.953	0.949	0.069
Mat.	4	0.974	0.987	0.029	0	0.966	0.990	0.035
Uso	0	1.041	0.945	0.069	1	0.938	0.948	0.081
P med	1	1.170	0.954	0.176	1	1.018	0.953	0.051
P máx	1	1.170	0.954	0.176	1	1.018	0.953	0.051
P min	1	1.170	0.954	0.176	1	1.018	0.953	0.051
V med	1	1.104	0.948	0.116	0	0.971	0.950	0.058
V máx	1	1.104	0.948	0.116	0	0.971	0.991	0.057
V min	1	1.104	0.948	0.116	0	0.961	0.949	0.064
Prof	1	1.076	0.946	0.093	1	0.947	0.957	0.068

Tabla 13. Análisis periodo de validación – comparación en dos periodos. Comparación de los estadísticos de calidad de los modelos de orden uno en función del ajuste del periodo de validación – dos periodos.

Analizando los resultados de los mejores modelos de orden dos (Tabla 14) se observan ciertas coincidencias en cuanto a los dos mejores modelos (combinaciones de variables) y los resultados de calidad obtenidos (radio) y en aquellos en los que el material es la segunda variable.

Variable	Periodo 1 (01/11/2010-02/04/2013)					Periodo 2 (03/04/2013-02/09/2015)				
	Variab. 2	Desag.	Pendiente	Coef. R^2	Radio	Variab. 2	Desag.	Pendiente	Coef. R^2	Radio
Diám	Año	9	1.001	0.988	0.012	Año	0	0.996	0.991	0.010
Año	Mat	0	0.999	0.989	0.011	Mat	0	0.998	0.991	0.009
Terr.	Mat	4	0.982	0.989	0.022	Año	5	1.009	0.980	0.022
Mat.	V min	6	1.001	0.989	0.011	Año	0	0.998	0.991	0.009
Uso	Mat	4	0.984	0.987	0.021	Año	5	1.009	0.980	0.022
P med	Diam	3	1.009	0.958	0.043	Año	3	0.988	0.983	0.021
P máx	Diam	2	1.003	0.958	0.042	Año	3	0.988	0.983	0.017
P mín	Diam	4	1.023	0.963	0.043	Mat	2	1.028	0.991	0.029
V med	Mat	7	1.003	0.989	0.012	Mat	0	0.991	0.991	0.013
V máx	Mat	6	1.004	0.989	0.012	Mat	0	0.994	0.991	0.011
V mín	Mat	6	1.001	0.989	0.011	Mat	0	0.989	0.991	0.014
Prof	Mat	4	0.992	0.988	0.014	Mat	0	0.983	0.991	0.020

Tabla 14. Análisis periodo de validación – comparación en dos periodos. Comparación de los estadísticos de calidad de los modelos de orden dos en función del ajuste del periodo de validación – dos periodos.

El siguiente paso es el análisis de los gráficos de resultados de calidad para los modelos de orden uno (Figura 36). Se aprecian ciertas similitudes y diferencias, por un lado los modelos de año y material presentan un comportamiento muy similar, en cambio otras variables presentan dispersiones diferentes.

Figura 36. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados R2 para modelos de una variable. (izq) Periodo 1 ajuste y 2 validación. (der) Periodo 2 ajuste y 1 validación.

Analizando el gráfico de ajuste de regresión para la variable año se aprecia una tendencia muy similar que confirma lo observado en el gráfico anterior (Figura 37).

Figura 37. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados regresión lineal para año. (izq) Periodo 1 ajuste y 2 validación. (der) Periodo 2 ajuste y 1 validación.

Como con la variable año, la variable material presentaba un comportamiento similar entre los dos periodos de validación empleados. Este aspecto se confirma con el análisis del gráfico de ajuste de regresión.

Figura 38. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados regresión lineal para material. (izq) Periodo 1 ajuste y 2 validación. (der) Periodo 2 ajuste y 1 validación.

Analizando los modelos de orden dos en los que interviene la variable año se observa un gráfico que en términos generales es similar para ambos modelos de validación (Figura 39). En ambos casos los mejores resultados se obtienen combinando año con diámetro y material.

Figura 39. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados R2 para año. (izq) Periodo 1 ajuste y 2 validación. (der) Periodo 2 ajuste y 1 validación.

Enlazando con el gráfico anterior, la curva de ajuste del modelo año con material muestra resultados similares con una dispersión muy reducida en ambos casos así como una pendiente próxima a la unidad (Figura 40).

Figura 40. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados regresión lineal para año con material. (izq) Periodo 1 ajuste y 2 validación. (der) Periodo 2 ajuste y 1 validación.

Analizando los modelos de orden dos en los que interviene la variable diámetro se obtienen mayores diferencias. Como es lógico, la combinación de las variables diámetro y año ofrece similares valores como se ha visto con anterioridad (Figura 41). Esto se confirma con el gráfico del ajuste de regresión del modelo diámetro con año Figura 42.

Figura 41. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados R2 para diámetro. (izq) Periodo 1 ajuste y 2 validación. (der) Periodo 2 ajuste y 1 validación.

Figura 42. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados regresión lineal para diámetro con año. (izq) Periodo 1 ajuste y 2 validación. (der) Periodo 2 ajuste y 1 validación.

El último paso de la comparación del periodo de validación en dos periodos es la representación de los resultados de calidad en un gráfico con escala de colores. Dicho gráfico se ajusta con el resultado de orden cero. De esta forma las combinaciones que tienen un radio mayor que el obtenido para orden cero se marcan en gris.

Como puede observarse hay claras diferencias (parte central de la Figura 43), pero algunas variables sí presentan resultados similares como los modelos de la variable año y material.

Figura 43. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados estadísticos de calidad de todas las combinaciones de variables – escala de colores.

Tras el análisis en dos periodos se concluye que los resultados sí se ven afectados por este factor. Dicho resultado es consecuente con los datos de partida pues para el primer periodo las roturas registradas son del orden de 4.000 mientras que para el segundo periodo son superiores a 5.000, existiendo un sesgo inicial. Por tanto queda claro que el periodo de validación condiciona los resultados de los modelos predictivos. Se hace por tanto necesario un análisis cuantitativo más detallado para establecer un procedimiento de validación a partir de un periodo que ofrezca mayores garantías.

5.3.2 Comparación cuantitativa

Una vez que se comprueba que el periodo de validación afecta a los resultados obtenidos se plantea un análisis cuantitativo dividiendo la información en periodos de un año completo cada uno. Se tienen por tanto cuatro series de datos: 2011, 2012, 2013 y 2014. Con esta división se tienen los siguientes datos para cada periodo (Tabla 15):

Año	Núm. roturas	Núm. roturas identificadas
2011	2761	1758
2012	2496	1773
2013	2298	1970
2014	2363	2237

Tabla 15. Análisis periodo de validación – comparación cuantitativa. Datos de roturas por periodo.

Con este planteamiento en el que se tienen cuatro periodos de validación se realiza una verificación múltiple de forma que se realiza el ajuste con cada uno de los periodos y se realiza la validación con los tres restantes. Se tiene por tanto un total de $n \cdot (n-1)$ validaciones, que para el caso de 4 periodos suponen 12 validaciones como se muestra en la Tabla 16.

Validación	Ajuste del modelo			
	2011	2012	2013	2014
2011	-	G	H	I
2012	A	-	J	K
2013	B	D	-	L
2014	C	E	F	-

Tabla 16. Periodos de ajuste del modelo y validación

Se analiza en un primer paso la evaluación de cada uno de los modelos de orden cero en los otros tres periodos restantes. Es decir que se evalúa cada modelo tres veces, obteniéndose 12 validaciones según se muestran en la Figura 44.

Figura 44. Análisis cualitativo periodo de validación – comparación cuantitativa. Ajuste modelo de orden cero. (a) Modelo ajuste 2011, validado 2012. (b) Modelo ajuste 2011, validado 2013. (c) Modelo ajuste 2011, validado 2014. (d) Modelo ajuste 2012, validado 2011. (e) Modelo ajuste 2012, validado 2013. (f) Modelo ajuste 2012, validado 2014. (g) Modelo ajuste 2013, validado 2011. (h) Modelo ajuste 2013, validado 2012. (i) Modelo ajuste 2013, validado 2014. (j) Modelo ajuste 2014, validado 2011. (k) Modelo ajuste 2014, validado 2012. (l) Modelo ajuste 2014, validado 2013.

Se observa en la figura anterior que la validación de un modelo con un mismo de ajuste en los restantes periodos es similar, mientras que las diferencias entre modelos con diferente periodo de ajuste son más significativas.

Como con los otros dos factores que condicionan la metodología de validación planteada se analizan los todos los modelos de orden uno y los modelos de orden dos generados con la variable año y diámetro, a partir de los estadísticos de calidad (pendiente, dispersión y radio). Se agrupan las 12 validaciones en función del periodo que se emplea para el ajuste de cada uno de los modelos, esta agrupación permite ver diferencias relativas entre modelos en función del periodo de ajuste empleado así como diferencias entre las validaciones resultantes con un mismo periodo de ajuste.

Se analizan primero los resultados de los modelos de orden 1 (Figura 45). Como primer resultado la variabilidad de los resultados en cuanto a estadísticos de calidad refleja que existe una significativa diferencia en los resultados entre unos periodos y otros. Adicionalmente se observa que el orden de los modelos no es constante, es decir que el periodo influye en la capacidad predictiva de los diferentes modelos. De hecho, dentro de un mismo periodo de ajuste, el orden de los modelos se ve alterado (los resultados de los modelos se cruzan). Todo ello permite concluir que los periodos de validación y ajuste son influyentes en los resultados del procedimiento de validación.

Figura 45. Análisis cuantitativo periodo de validación – resultados estadísticos de calidad. Modelos orden 1.
(a) Pendiente. (b) Dispersión. (c) Radio.

Al analizar los resultados de los estadísticos de calidad en los modelos de orden 2 se observan resultados similares. Por un lado las gráficas tienen una variabilidad significativa entre los diferentes periodos de ajuste de validación. Y por otra parte se observan trazados cruzados y con ordenación diferente, es decir que el comportamiento de una determinada combinación de variables es diferente según en qué periodo es validado o qué periodo se ajusta. El orden de los trazados es también relevante pues, de nuevo, según el periodo de ajuste o validación considerado, el mejor resultado se obtiene con una pareja de variables diferente (Figura 46 y Figura 47).

Figura 46. Análisis cuantitativo periodo de validación – resultados estadísticos de calidad. Modelos orden 2 - diámetro. (a) Pendiente. (b) Dispersión. (c) Radio.

Figura 47. Análisis cuantitativo periodo de validación – resultados estadísticos de calidad. Modelos orden 2 - año. (a) Pendiente. (b) Dispersión. (c) Radio.

El análisis concluye con la comparativa de los gráficos de resultados de los estadísticos de calidad. Según se aprecia en la Figura 48, existe una similitud en los resultados dentro de un mismo ajuste, se aprecia una misma tendencia para un modelo dado en los tres periodos en los que se valida. Sin embargo las diferencias entre unos modelos y otros son significativas dependiendo del periodo de ajuste empleado. Esto denota que el periodo con el que se ajusta el modelo influye en los resultados pues si se analiza un mismo periodo validado con diferentes periodos de ajuste los resultados son diferentes.

Modelo de 2011

Validado en 2012

Validado en 2013

Validado en 2014

Modelo de 2012

Validado en 2011

Validado en 2013

Validado en 2014

Modelo de 2013

Validado en 2011

Validado en 2012

Validado en 2014

Modelo de 2014

Validado en 2011

Validado en 2012

Validado en 2013

Figura 48. Análisis cualitativo periodo de validación. Resultados de los estadísticos de calidad de todas las combinaciones de variables – escala de colores.

La conclusión tras el análisis de sensibilidad del periodo de validación es que este parámetro influye claramente en los resultados de la validación. Los estadísticos de calidad obtenidos se ven afectados y por tanto para elegir el modelo que permita obtener los mejores resultados resulta necesario analizar diferentes periodos de ajuste y validación como se ha expuesto para poder seleccionar el que presente un mejor comportamiento global.

5.4 Resultado del análisis de sensibilidad

El resultado del análisis de sensibilidad es la definición del proceso de validación con los siguientes parámetros:

- Tamaño de muestra: 500 unidades.
- Tamaño de cada unidad de la muestra: entre 10 y 100 km.
- Periodo de ajuste: los modelos se ajustan con los datos de los cuatro periodos disponibles y se validan con el resto. De esta forma se tienen 12 combinaciones de validación posible en función del periodo.

Los resultados así obtenidos se tratan con el procedimiento de validación anteriormente descrito.

6 DEFINICIÓN DE LOS MODELOS PARA LA RED DE CANAL DE ISABEL II

Con los datos recopilados en la red de distribución del Canal de Isabel II y una vez filtrados, se aplica la metodología descrita en apartados anteriores. Se analizan modelos predictivos de manera secuencial, comenzando con los modelos de orden cero e incorporando variables explicativas de manera progresiva. El objetivo del análisis de los diferentes modelos es seleccionar aquél que permita pronosticar la tasa de roturas esperable con mayor fiabilidad.

6.1 Modelos de orden cero

Los modelos de orden cero se construyen únicamente con las roturas observadas sin tener en cuenta ninguna variable explicativa. Suponen por tanto el nivel de referencia sobre el que se incorporan las diferentes variables para mejorar la capacidad predictiva.

El resultado de los 12 casos de validación que se obtienen a partir de las distintas combinaciones de periodos de ajuste y validación realizadas, se recoge en la Tabla 17.

R2			R1	
Pen. Med.	Coef. R^2 Med.	Radio P. Med.	Med. de Radio	Des. Tip. de Radio
1,107	0,975	0,110	0,157	0,125

Tabla 17. Modelo de orden cero. Estadísticos de calidad.

6.2 Modelos de una variable

Tomando como referencia los resultados de los modelos de orden cero se realiza la validación de los modelos de una variable. En este análisis se estudian diferentes niveles de desagregación desde uno a diez intervalos más el nivel de máxima desagregación posible.

En la Tabla 18 se recogen los estadísticos de calidad obtenidos para las diferentes variables explicativas. En esta primera aproximación se identifican las variables más influyentes: diámetro, material y año. Estas tres variables son las que presentan unos mejores estadísticos al analizarse por separado. Se aprecia una diferencia clara entre estas tres variables y el resto en cuanto al radio obtenido y la pendiente de ajuste media. Las variables internas como presión y velocidad, así como el terreno o el uso del suelo, presentan resultados muy similares en cuanto a su capacidad predictiva.

Variable	R2			R1	
	Pen. Med.	Coef. R ² Med.	Radio P. Med.	Med. de Radio	Des. Tip. de Radio
Diámetro	1,021	0,971	0,036	0,075	0,053
Material	1,030	0,980	0,036	0,085	0,066
Año	1,049	0,980	0,053	0,097	0,084
Uso del suelo	1,099	0,975	0,102	0,149	0,119
Profundidad	1,100	0,976	0,103	0,151	0,122
Presión máx	1,112	0,975	0,115	0,153	0,127
Presión med	1,112	0,975	0,115	0,153	0,127
Presión mín	1,111	0,975	0,114	0,154	0,127
Terreno	1,108	0,975	0,111	0,156	0,125
Velocidad máx	1,108	0,975	0,111	0,156	0,125
Velocidad mín	1,108	0,975	0,111	0,156	0,125
Velocidad med	1,107	0,975	0,110	0,157	0,125

Tabla 18. Modelo de orden cero. Estadísticos de calidad.

Se analizan en detalle los resultados de cada una de las 12 validaciones realizadas, los resultados de los estadísticos de calidad (pendiente de la recta de ajuste, coeficiente de correlación y radio) se muestran en la Figura 49 junto con los valores de referencia (modelo de orden cero). Se aprecia un claro paralelismo entre los resultados de la pendiente de ajuste, sin embargo en cuanto a la dispersión ese paralelismo no es tan marcado. Es decir que en términos de pendiente de ajuste los resultados son más homogéneos. La combinación de ambos estadísticos en el radio arroja resultados

similares, lo que supone en términos globales que, partiendo de determinados periodos de ajuste se pierde capacidad predictiva en modelos de una variable.

Cabe destacar que las gráficas de resultados (Figura 49) son constantes en las últimas validaciones (H, J, F, I, K y L), y que éstas corresponden a las validaciones realizadas con modelos ajustados con el periodo de 2013 y 2014. Por tanto, como dichas validaciones cubren todos los periodos con igual resultado de nivel de predicción, se tienen periodos de validación similares en los que no hay una variabilidad muy significativa que no pueda ser reflejada por los modelos de una variable. Este hecho confirma lo observado en el análisis de sensibilidad del factor del periodo de validación donde se observa que los resultados se ven claramente influenciados por dicho factor.

Figura 49. Modelos de orden uno. Resultados de los estadísticos de calidad para cada variable en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

Se completa el análisis de los modelos de una variable con la Figura 50 donde se muestran los gráficos de resultados globales. En el primero de ellos (Figura 50.a) se muestran las 12 validaciones de cada una de las variables estudiadas representadas por su radio medio así como por la desviación estándar de estos radios dentro de una misma variable. En este gráfico se muestra la diferencia de capacidad predictiva anteriormente señalada, siendo diámetro, material y año las variables que ofrecen los mejores resultados. El resto de variables se concentran en el mismo rango de predicción con un radio medio próximo a 0,15 al igual que el modelo de orden cero.

En la figura Figura 50.b se representan las validaciones mediante su dispersión y pendiente de ajuste, de forma implícita el radio se representa como la distancia geométrica entre cada uno de los puntos y la coordenada (1, 1). En esta figura se aprecia con mayor claridad que ciertas validaciones de otras variables diferentes a las tres de mejor comportamiento (diámetro, material y año) ofrecen buenos resultados, quedando por dentro de la curva de radio 0,03 correspondiente al radio de la variable diámetro.

Figura 50. Modelos de orden uno. Resultados globales de los estadísticos de calidad para cada variable en cada validación. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

Cabe destacar en este análisis que las tres variables con mayor capacidad predictiva por sí solas son diámetro, material y año. La media del radio obtenido con dichas variables es respectivamente: 0,075, 0,085 y 0,097 frente al valor inicial del modelo de orden cero de 0,157. De igual manera la dispersión se ve reducida desde 0,125 a 0,053 para el caso del diámetro. Por tanto, la mejora en la capacidad predictiva al incorporar una variable explicativa es significativa.

Se detecta también que las variables hidráulicas de presión y velocidad carecen de capacidad predictiva por sí solas. En los modelos de orden superior se estudia si su combinación con otras variables mejora el comportamiento del modelo, pues por sí mismas no consiguen mejorar el resultado obtenido con el modelo de una única variables

6.3 Modelos de orden dos

Una vez analizados los modelos de orden uno, se incorpora una segunda variable explicativa a los modelos predictivos. En todos los casos se analizan de nuevo varios niveles de agregación de componente, tomando resultados con uno a diez intervalos más el modelo de máxima desagregación posible. Dicho nivel de desagregación considerado en los modelos de orden dos es el mismo en las dos variables consideradas.

Se estudian primero los modelos de dos variables conjuntas y a continuación los modelos de dos variables independientes.

6.3.1 Modelos de dos variables conjuntas

El planteamiento de este conjunto de modelos se realiza partiendo de los resultados de los modelos de una variable, en este sentido se analizan los modelos de dos variables conjuntas en los que intervienen las variables diámetro, material y año pues éstas eran las variables que destacaron por tener mayor capacidad predictiva por sí mismas. Es por tanto lógico esperar que la incorporación de variables explicativas adicionales puedan mejorar los resultados en el análisis de los modelos de orden uno.

En este análisis se tiene por un lado el nivel de referencia de los modelos de orden cero así como los mejores modelos de orden uno: material 0,075, diámetro 0,085 y año 0,097. Estos últimos son los valores a mejorar con la incorporación de una segunda variable al modelo.

Se analizan en primera instancia los resultados de todos los modelos posibles en los que intervienen las anteriores variables y posteriormente se analizan los resultados globales.

6.3.1.1 Diámetro

La variable diámetro es la que tiene una mejor capacidad predictiva de entre todas las variables analizadas en el paso anterior. Partiendo de ese nivel de referencia (0,075), en la Figura 51 se recogen los resultados de los tres estadísticos de calidad para los modelos de orden dos en los que interviene la variable diámetro, donde cada ordenada corresponde

a una de las 12 validaciones realizadas. En términos de pendiente de ajuste los resultados son similares con independencia de la validación (gráfica con tendencia horizontal). En cambio la dispersión sí presenta diferencias más significativas entre unas validaciones y otras aunque se mantiene un cierto paralelismo, es decir, la dispersión observada presenta desviaciones análogas para todos los modelos en determinadas validaciones.

Figura 51. Modelos de dos variables conjuntas. Resultados de los estadísticos de calidad para cada variable combinada con material en cada diámetro. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

Los resultados globales para estos modelos de orden dos se recogen en la Figura 52, el principal comentario es que la incorporación de variables adicionales no mejora la capacidad predictiva cuando interviene el diámetro pues no se mejora el valor conseguido con el modelo de orden uno (0,075). En este sentido los mejores resultados se obtienen combinando la variable diámetro con profundidad, terreno y material. Esas

combinaciones consiguen un radio medio de 0,076 sin mejorar el resultado del modelo de orden uno.

Se observa también que a medida que mejora el radio medio se reduce la desviación estándar, es decir que la predicción adquiere una mayor fiabilidad en términos globales.

Figura 52. Modelos de dos variables conjuntas. Resultados globales de los estadísticos de calidad en modelos de orden dos incluyendo la variable diámetro. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.3.1.2 Material

La variable material presenta un radio medio de 0,085 en el modelo de orden uno, al incorporarse una variable adicional de manera conjunta se obtienen los resultados recogidos en la Figura 53.

De la primera se observa un claro paralelismo en el comportamiento de la pendiente de ajuste entre los diferentes modelos y validaciones. En el comportamiento global, comparando el radio obtenido en cada simulación se observa que por lo general los resultados obtenidos con los modelos de orden dos son mejores que los resultados del modelo de orden cero con independencia de la validación y las variables consideradas.

Figura 53. Modelos de dos variables conjuntas. Resultados de los estadísticos de calidad para cada variable combinada con material en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

En la Figura 54 se recogen los resultados globales, cabe destacar en este caso que el comportamiento de todos los modelos es análogo, concentrándose la mayor parte de las validaciones próximas al mejor modelo con pocos puntos peores que el modelo de orden cero (por fuera de la curva marrón). En particular los mejores resultados se obtienen con la combinación de material con diámetro (0,075), mejorando ligeramente el resultado de orden uno.

Figura 54. Modelos de dos variables conjuntas. Resultados globales de los estadísticos de calidad en modelos de orden dos incluyendo la variable material. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.3.1.3 Año

Los modelos de orden dos planteados a partir de la variable año ofrecen un comportamiento mejor que los de orden uno, observándose claras mejoras en los resultados (Figura 55).

Figura 55. Modelos de dos variables conjuntas. Resultados de los estadísticos de calidad para cada variable combinada con año en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

En concreto los modelos formados con año y diámetro mejoran significativamente los resultados de dicho modelo, obteniéndose un radio medio de 0,078 frente a 0,097 que se obtiene en el modelo de orden uno para año (Figura 56). La combinación con material también mejora los resultados llegando a un radio medio de 0,092.

Figura 56. Modelos de dos variables conjuntas. Resultados globales de los estadísticos de calidad en modelos de orden dos incluyendo la variable año. (a) Valor medio de los radios en función de su desviación estándar.

(b) Pendiente y coeficiente correlación

6.3.1.4 Resultados globales

Cabe destacar que los mejores modelos siempre incorporan la variable diámetro lo que confirma el resultado del modelo de orden uno en el que el diámetro era la variable con mayor capacidad predictiva. En cuanto al coeficiente R^2 medio, la diferencia es más reducida y entre los 10 modelos seleccionados la variación es mínima.

En general, según la Tabla 19, los resultados obtenidos no mejoran significativamente los resultados obtenidos con los modelos de orden uno.

Variable 1	Variable 2	R2			R1	
		Pen. Med.	Coef. R ² Med.	Radio P. Med.	Med. de Radio	Des. Tip. de Radio
Diámetro	Profundidad	1.021	0.972	0.035	0.076	0.055
Diámetro	Material	1.025	0.977	0.034	0.076	0.059
Diámetro	Terreno	1.021	0.972	0.035	0.076	0.055
Diámetro	Uso del suelo	1.022	0.971	0.036	0.077	0.056
Diámetro	Año	1.022	0.974	0.034	0.078	0.062
Diámetro	Velocidad máx	1.024	0.972	0.037	0.084	0.063
Diámetro	Velocidad min	1.031	0.973	0.041	0.084	0.063
Terreno	Material	1.031	0.981	0.036	0.086	0.066
Diámetro	Velocidad med	1.027	0.973	0.039	0.086	0.066
Material	Uso del suelo	1.034	0.981	0.039	0.088	0.071

Tabla 19. Modelo de dos variables conjuntas. Estadísticos de calidad, mejores modelos.

En la Figura 57 se muestran los resultados globales de calidad para los modelos con dos variables conjuntas. Se marca en gris el mejor resultado de orden 1 y en marrón el de orden cero.

Figura 57. Modelos de dos variables conjuntas. Resultados globales de los estadísticos de calidad en los mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

Una vez analizados todos los modelos de dos variables conjuntas se observa que ciertos modelos mejoran la capacidad predictiva de los modelos de una variable aunque no se consiga una mejora respecto del mejor modelo de orden uno. De los modelos de dos variables conjuntas, el mejor es el formado por las variables diámetro y profundidad presenta una media de radio de 0,076 mientras que el mejor modelo de orden uno tenía un radio de 0,075. Es decir, que en términos absolutos no se mejora la capacidad predictiva.

En términos relativos sí se aprecia una mejora significativa pues los diez mejores modelos presentan una media de radio muy próxima al mejor modelo de orden dos (0,076) y de orden uno (0,075).

Cabe destacar que la variable diámetro está presente en la mayoría de los mejores modelos, lo que confirma la capacidad predictiva intrínseca a esta variable.

6.3.2 Modelos de dos variables independientes

Pese a que se ha observado que existe una dependencia estadística entre las variables, los modelos de dos variables independientes se analizan de forma análoga a los modelos de dos variables conjuntas. Sin embargo, resulta interesante analizar los modelos de orden dos con ambas variables de manera independiente puesto que la función de distribución conjunta presenta más incertidumbre que la estimación de las funciones de distribución marginales.

6.3.2.1 Diámetro

La variable diámetro presenta el mejor resultado en cuanto a capacidad predictiva dentro de los modelos de una variable con una media del radio de 0,075, y al incorporar variables adicionales de manera independiente conformando los diferentes modelos de orden dos se obtienen resultados que no mejoran esa predicción (Figura 58 y Figura 59). Así, el mejor modelo de orden dos es el formado por las variables diámetro y terreno, con el que se obtiene una media del radio de 0,076. Es decir, la capacidad predictiva no mejora y prácticamente iguala el resultado del modelo de orden uno.

Si se analiza el indicador R^2 , se aprecia una ligera mejoría pues se pasa de un radio medio de 0,034 a un valor de 0,035 para el modelo formado por diámetro y terreno.

Figura 58. Modelos de dos variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con material en cada diámetro. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

Figura 59. Modelos de dos variables independientes. Resultados globales de los estadísticos de calidad en modelos de orden dos incluyendo la variable diámetro. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.3.2.2 Material

Al analizar los resultados de los modelos de orden dos generados con dos variables independientes incluyendo la variable material (Figura 60), se observa un comportamiento en términos de pendiente de ajuste bastante uniforme. Para todas las validaciones los diferentes modelos ofrecen resultados similares siendo la pendiente de ajuste dependiente del modelo y no tanto de la validación.

El análisis de la dispersión ofrece una mayor variabilidad en función de la validación observada aunque en general el orden de los modelos es similar pues se mantiene un cierto paralelismo entre las líneas que definen el comportamiento de cada modelo.

La combinación de ambos resultados observada en el gráfico de la distribución del radio presenta una cierta homogeneidad en cuanto a la capacidad predictiva de los modelos, donde claramente los que incluyen las variables velocidad máxima, profundidad o velocidad mínima presentan siempre el peor comportamiento. Cabe destacar que si bien en los modelos de peor comportamiento hay una diferencia significativa en la capacidad predictiva, los que mejor comportamiento ofrecen resultados más homogéneos frente a las diferentes validaciones; estos modelos son los que incluyen las variables diámetro, año, uso o velocidad media.

Figura 60. Modelos de dos variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con material en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

Al analizar los resultados globales de estos modelos, Figura 61, se aprecia una concentración de resultados de la gran mayoría de modelos en torno a una media de radio de 0,10. Siendo el mejor modelo el que combina material con profundidad (0,084), mejorando mínimamente los resultados del modelo de orden uno que ofrece un resultado de media de radio de 0,085.

Figura 61. Modelos de dos variables independientes. Resultados globales de los estadísticos de calidad en modelos de orden dos incluyendo la variable material. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.3.2.3 Año

Los resultados obtenidos con la variable año presentan un comportamiento similar al estudiado previamente con las variables diámetro y material en cuanto a la pendiente de ajuste. Sin embargo, la dispersión obtenida es muy sensible a cada validación obteniéndose resultados dispares, no sólo en cuanto a la dispersión obtenida sino al orden de los modelos. En el segundo gráfico de la Figura 62 se observa que las líneas de cada modelo no siguen un paralelismo muy claro lo que indica la variabilidad de los resultados según la validación.

Si se analiza el gráfico del radio, se tiene un resultado similar al de la variable material, observándose una clara diferencia entre los mejores modelos en los que la diferencia de radio entre las diferentes validaciones es muy reducida frente a la variabilidad de los resultados de los modelos con peor comportamiento. Dentro de los primeros se incluyen los que combinan la variable año con las variables diámetro, velocidad media o uso del suelo.

Figura 62. Modelos de dos variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con año en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

El análisis de los resultados globales según la Figura 63, muestra como por lo general no se mejoran los resultados del modelo de orden uno (representado con trazo azul), únicamente la combinación de año con profundidad ofrece una cierta mejora pasando de 0,097 de media de radio en el modelo de orden uno a 0,083 en el modelo de orden dos. En ambos gráficos se observa una concentración de resultados en el entorno de radio 0,11.

Figura 63. Modelos de dos variables independientes. Resultados globales de los estadísticos de calidad en modelos de orden dos incluyendo la variable año. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.3.2.4 Resultados globales

Una vez analizados todos los modelos posibles que se obtienen de la combinación de las mejores variables de orden uno, es decir diámetro, material y año con el resto de variables, se puede concretar que si bien algunos modelos mejoran la capacidad predictiva de los modelos de orden uno, el resultado no ofrece una mejora significativa. Como se observa en la Tabla 20, el mejor modelo de dos variables en términos de media de radio es el que combina diámetro y terreno con un valor de 0,076. Otras combinaciones como las que se tienen en dicha tabla ofrecen resultados muy similares en el entorno de un radio de 0,080. Sin embargo, comparando estos resultados con los de los modelos de orden uno, no se tiene una mejora pues el mejor modelo de orden uno (diámetro) tiene una media de radio de 0,075.

Variable 1	Variable 2	R2			R1	
		Pen. Med.	Coef. R ² Med.	Radio P. Med.	Med. de Radio	Des. Tip. De Radio
Diámetro	Terreno	1.019	0.971	0.034	0.076	0.054
Diámetro	Velocidad máx	1.025	0.972	0.037	0.079	0.058
Diámetro	Velocidad med	1.026	0.972	0.038	0.080	0.059
Diámetro	Uso del suelo	1.019	0.968	0.037	0.081	0.059
Diámetro	Velocidad min	1.028	0.972	0.039	0.082	0.061
Año	Profundidad	1.062	0.982	0.064	0.083	0.087
MAterial	Profundidad	1.054	0.991	0.055	0.084	0.087
Diámetro	Profundidad	1.025	0.970	0.039	0.084	0.064
Material	Uso del suelo	1.032	0.979	0.038	0.087	0.068
Diámetro	Profundidad	1.030	0.975	0.039	0.087	0.066

Tabla 20. Modelo de dos variables independientes. Estadísticos de calidad, mejores modelos.

En la Figura 64 se muestran los resultados globales de los estadísticos de calidad; en general se observa que todos los modelos están muy próximos al mejor resultado (diámetro y terreno) que a su vez ofrece un resultado similar al mejor modelo de orden uno (diámetro, 0,075). Es decir, que el salto cualitativo que se produce entre el modelo de orden cero (representado con trazo marrón en la Figura 64) y el mejor modelo de orden uno, sin destacar mejora alguna con los modelos de orden dos.

Figura 64. Modelos de dos variables independiente. Resultados globales de los estadísticos de calidad en los mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

Una vez analizados todos los modelos de dos variables independientes se observa que ciertos modelos mejoran la capacidad predictiva de algunos de los modelos de una variable pero no se mejora el resultado obtenido con el mejor modelo de orden uno (0,075).

Como en el caso de los modelos de dos variables conjuntas, en general se mejora la capacidad predictiva de los modelos pero no en términos absolutos. Es decir, que hay más modelos con un comportamiento próximo al mejor modelo de orden uno, pero que no logran mejorar los resultados de dicho modelo (diámetro). Igualmente, la variable diámetro está presente en gran parte de los diez mejores modelos de orden dos.

6.3.3 Resumen modelos de dos variables

Del análisis realizado de los modelos de orden dos, tanto con dos variables conjuntas como dos variables independientes, se observa que no hay una mejora de la capacidad predictiva en términos absolutos. Es decir que la mejora que se tiene desde el modelo de orden cero (0.160) al modelo mejor modelo de orden uno (0,075) no se repite en los modelos de orden dos (0,076). Como se aprecia en la Figura 65 y Figura 66, si bien no hay una mejora en términos absolutos, en términos relativos se observa que los mejores modelos se concentran próximos al mejor modelo de orden dos. Como es lógico, al incorporar dos variables al modelo predictivo se obtienen por lo general mejores resultados, si bien no se logra superar la capacidad de predicción obtenida con el mejor modelo de orden uno.

Figura 65. Evolución del radio medio con el orden del modelo con dos variables conjuntas y el resto independientes.

Figura 66. Evolución del radio medio con el orden del modelo con todas las variables independientes.

6.4 Modelos de orden tres

Según se describe en la metodología y siguiendo la estructura de los modelos de orden dos, se estudian en un primer paso los modelos de dos variables conjuntas y una independiente. De nuevo en cada caso se parte de los mejores resultados obtenidos en la fase anterior, es decir se parte de los modelos de orden dos.

6.4.1 Modelos de dos variables conjuntas y una independiente

Como se ha explicado con anterioridad, estos modelos de orden tres se construyen partiendo de los diez mejores modelos de orden dos formados por dos variables conjuntas. Sobre estos se incorpora una variable independiente adicional para mejorar la capacidad predictiva.

6.4.1.1 Diámetro – profundidad

Los resultados de los diferentes modelos de orden tres generados con las variables conjuntas diámetro y profundidad a las que se les añade una tercera variable independiente se recogen en la Figura 67. Se aprecia que gran parte de los modelos ofrecen un comportamiento muy similar pues las líneas de cada uno de ellos son prácticamente coincidentes, únicamente los modelos que incorporan la variable año y material se alejan algo más presentando una peor pendiente. El análisis de la dispersión es semejante al anterior en cuanto a que estas dos variables presentan un comportamiento más errático, siendo el resto de modelos similares con un comportamiento análogo representado por líneas sensiblemente paralelas. En cualquier caso cabe destacar la variabilidad del comportamiento de los modelos entre las diferentes validaciones pues no hay una trayectoria o tendencia horizontal en ninguno de los gráficos. De nuevo se pone de manifiesto la influencia del periodo de ajuste en los resultados de calidad obtenidos.

Figura 67. Modelos de dos variables conjuntas y una independiente. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-profundidad en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

Analizando el comportamiento global de los modelos según se muestra en la Figura 68, se confirma la concentración de resultados de gran parte de los modelos y cómo los modelos que incorporan la variable año y material presentan un resultado diferente con peor capacidad predictiva. De entre todos los modelos los que mejor comportamiento presentan son los que incorporan las variables de presión y de velocidad, siendo el mejor resultado el de presión mínima (0,066). Se mejora por tanto el resultado del modelo diámetro profundidad que presentaba una media de radio de 0,076.

Por tanto la incorporación de una variable independiente adicional a los modelos que consideran las variables diámetro y profundidad de forma conjunta mejora la capacidad predictiva salvo en casos puntuales como la variable año y material.

Figura 68. Modelos de dos variables conjuntas y una independiente. Resultados globales de los estadísticos de calidad modelos de orden tres incluyendo las variable diámetro-profundidad. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.4.1.2 Diámetro – material

El comportamiento de los modelos de orden tres considerando las variables diámetro y material de manera conjunta es análogo al observado en los modelos de diámetro y profundidad. En este caso todos los modelos presentan un comportamiento muy similar a excepción del que incorpora la variable año. De igual forma, la dispersión es algo más errática sobre todo para las variables año y profundidad. En términos globales el radio obtenido es similar en los diferentes modelos presentando una variabilidad importante en función de la validación analizada.

Figura 69. Modelos de dos variables conjuntas y una independiente. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-material en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

Si se analizan los resultados globales, Figura 70, se confirma este comportamiento de los modelos concentrándose los resultados en el entorno de un radio de 0,070. De nuevo los mejores modelos son los que incorporan las variables de velocidad y presión siendo el óptimo el que incorpora la velocidad mínima (0,062). Este resultado mejora los resultados del modelo de orden dos correspondiente (0,076).

Figura 70. Modelos de dos variables conjuntas y una independiente. Resultados globales de los estadísticos de calidad modelos de orden tres incluyendo las variable diámetro-material. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.4.1.3 Diámetro – terreno

Este conjunto de modelos de orden tres en los que se incorporan las variables diámetro y terreno de manera conjunta presenta un resultado similar a los anteriores. Como se muestra en la Figura 71 se aprecia un paralelismo entre todos los modelos para las diferentes validaciones en términos de la pendiente de ajuste, observándose que los modelos que incorporan las variables año y material presentan un comportamiento claramente peor. Este comportamiento también se observa en los resultados de la dispersión y del radio medio. De nuevo se aprecia una notable variabilidad del comportamiento de los diferentes modelos en función de la validación analizada.

Figura 71. Modelos de dos variables conjuntas y una independiente. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-terreno en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

El análisis del comportamiento global de los diferentes modelos se recoge en la Figura 72. De nuevo se observa un comportamiento muy similar entre todos los modelos salvo los mencionados anteriormente. Los mejores resultados vuelven a obtenerse con las variables de presión y velocidad, siendo el mejor modelo el que introduce la variable presión mínima obteniendo un radio de 0,066 lo que reduce el resultado del modelo de orden dos correspondiente (0,076).

Figura 72. Modelos de dos variables conjuntas y una independiente. Resultados globales de los estadísticos de calidad en modelos de orden tres incluyendo las variable diámetro-terreno. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.4.1.4 Mejores modelos

Una vez analizados todos los modelos de orden 3 en los que intervienen dos variables de manera conjunta y una tercera de manera independiente se recogen los 10 mejores modelos en la Tabla 21. Los mejores resultados se obtienen para las combinaciones diámetro-material y diámetro-año combinándose con las variables de presión o velocidad y puntualmente con la variable profundidad.

Con estos modelos se mejoran los resultados de los modelos de una variable (0,075) y los de dos variables conjuntas (0,076) obteniéndose un radio medio de 0,062 en la combinación de diámetro-material con velocidad mínima.

Variables conjuntas	Variable independiente	R2			R1	
		Pen. Med.	Coef. R ² Med.	Radio P. Med.	Med. de Radio	Des. Tip. De Radio
Diam-Mate	Velocidad min	1.011	0.980	0.023	0.062	0.051
Diam-Año	Presión máx	1.010	0.979	0.023	0.063	0.055
Diam-Mate	Profundidad	1.034	0.988	0.036	0.063	0.065
Diam-Mate	Velocidad med	1.012	0.980	0.023	0.063	0.051
Diam-Año	Presión med	1.008	0.979	0.022	0.063	0.054
Diam-Mate	Velocidad máx	1.012	0.980	0.024	0.064	0.052
Diam-Año	Presión min	1.006	0.979	0.022	0.064	0.052
Diam-Año	Profundidad	1.039	0.982	0.043	0.065	0.065
Diam-Mate	Presión min	1.015	0.981	0.024	0.065	0.051
Diam-Terr	Presión min	1.013	0.975	0.028	0.066	0.049

Tabla 21. Modelo de dos variables conjuntas y una independiente. Estadísticos de calidad, mejores modelos.

Una vez descartados los modelos que presentan un peor comportamiento, se observan los resultados globales en la Figura 73. En todos ellos la capacidad predictiva es muy similar, resultado lógico pues partiendo de los resultados de los modelos de orden dos se seleccionan aquellos que tienen un mejor comportamiento y que por tanto mejoran los resultados del modelo de orden dos. No obstante, aunque la reducción es menos significativa en el paso de modelo de orden cero a modelo de una variable, se consigue obtener una mejor predicción en los resultados (reduciendo el radio) al mismo tiempo que se reduce la dispersión de los resultados.

En todos los modelos analizados se vuelve a poner de manifiesto la relevancia que las diferentes validaciones y periodos de ajuste tienen en el comportamiento de los mismos.

Figura 73. Modelos de dos variables conjuntas y una independiente. Resultados globales de los estadísticos de calidad en los mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.4.2 Modelos de tres variables independientes

Como se ha explicado con anterioridad, estos modelos de orden tres se construyen partiendo de los diez mejores modelos de orden dos formados por dos variables independientes: diámetro-terreno, diámetro velocidad máxima y diámetro-velocidad media. Sobre estos se incorpora una variable independiente adicional para mejorar la capacidad predictiva.

6.4.2.1 Diámetro – terreno

Si bien se observa una variabilidad de los resultados en función de la validación analizada, todos los modelos siguen una misma tendencia salvo los que incorporan la variable año y material. Sin embargo, éstas últimas son dos de las variables que mejor comportamiento presentan en los modelos de orden uno. Este comportamiento se muestra en la Figura 74, donde tanto para la pendiente de ajuste, como la dispersión o el radio obtenido la tendencia es similar.

Figura 74. Modelos de tres variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-terreno en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

Analizando el comportamiento global de los modelos según se muestra en la Figura 75, se observa que el comportamiento de todos los modelos es muy similar y en general se mejoran los resultados del modelo de orden dos. Las excepciones son los modelos que incorporan las variables año y material, claramente por encima de dicho resultado; así como los que incorporan el uso y la profundidad, que presentan resultados próximos al del modelo de orden dos (0,082 y 0,084 frente a 0,076). Los mejores modelos por tanto son los que incorporan las variables de presión y velocidad, alcanzándose el óptimo con la presión mínima, este modelo reduce el radio medio desde 0,076 a 0,065.

Figura 75. Modelos de tres variables independientes. Resultados globales de los estadísticos de calidad modelos de orden tres incluyendo las variable diámetro-terreno. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.4.2.2 Diámetro – velocidad máxima

Los modelos de orden tres que incorporan las variables diámetro y velocidad máxima ofrecen unos resultados muy similares a los de los modelos que incluyen diámetro y terreno. Como se muestra en la Figura 76, en términos de pendiente de ajuste el comportamiento entre los modelos es muy similar salvo para el caso de las variables año y material. De igual forma, los resultados de dispersión y radio son similares. En todos los modelos se observa una fuerte variabilidad predictiva de los modelos entre una validación y otra, sin embargo en términos relativos el comportamiento del conjunto de los modelos sí es el mismo.

Figura 76. Modelos de tres variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-velocidad máxima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

Los resultados antes enunciados se confirman al observar la Figura 77, donde se recogen los resultados globales para estos modelos de orden 3. A diferencia de los modelos que incorporan las variables diámetro y terreno, en este caso sólo los modelos que incorporan las variables de presión mejoran ligeramente los resultados del modelo de orden dos, representado en trazo gris en dicha figura; siendo el mejor resultado en cuanto a radio medio de 0,072 obtenido con la variable presión mínima. Los resultados analizados en términos de dispersión y pendiente de ajuste media (R^2) son similares.

Figura 77. Modelos de tres variables independientes. Resultados globales de los estadísticos de calidad en modelos de orden tres incluyendo las variable diámetro-velocidad máxima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.4.2.3 Diámetro – velocidad media

De forma similar a los modelos anteriormente descritos, los modelos de orden tres que incorporan las variables diámetro y velocidad media, como variables independientes, presentan un comportamiento homogéneo salvo los modelos que incorporan las variables año y material. En este caso el comportamiento entre los diferentes modelos en términos relativos, mientras que en términos absolutos se sigue apreciando una variabilidad de los estadísticos de calidad en función de la validación analizada.

Figura 78. Modelos de tres variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-velocidad media en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

El análisis de los resultados globales de estos modelos es similar al de los modelos de diámetro y velocidad máxima, Figura 78. Todos los modelos ofrecen resultados similares a los del modelo de orden dos (0,076), siendo los que incorporan las variables año o material los que tienen un peor comportamiento, por encima del modelo de orden cero. De nuevo la incorporación de las variables de presión mejora los resultados del modelo de orden dos aunque ligeramente, situándose el mejor modelo con una radio de 0,073 obtenido con la incorporación de la presión mínima. Los resultados para las variables de velocidad son también similares aunque ligeramente peores que la referencia del modelo de orden dos.

Figura 79. Modelos de tres variables independientes. Resultados globales de los estadísticos de calidad en modelos de orden tres incluyendo las variables diámetro-velocidad media. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.4.2.4 Mejores modelos

Una vez analizados todos los modelos de orden 3 formados por tres variables independientes se recogen los 10 mejores modelos en la Tabla 22. Los mejores resultados se obtienen para las combinaciones diámetro-terreno y diámetro-uso a las que se incorporan las variables de presión. Los resultados en todos estos modelos son muy similares con un radio medio de entre 0,065 y 0,069, obteniéndose el mejor resultado para la combinación diámetro-terreno-presión mínima. Estos resultados suponen una mejoría respecto del mejor modelo de dos variables independientes formado por las variables diámetro y terreno con un radio medio de 0,076. Se obtiene por tanto una ligera mejora en los resultados en términos de radio medio. En este sentido, tanto la dispersión como la pendiente media mejora así como la desviación típica del radio.

Cabe destacar que la combinación de las mejores variables de orden uno en un único modelo no configuran la mejor predicción. De hecho, partiendo de los resultados del mejor modelo de orden dos con variables independientes, formado por diámetro y terreno, la incorporación de variables que a priori tienen una buena capacidad predictiva como año o material no ofrece resultados mejores que otras variables como las de presión o velocidad.

Variables anteriores	Variable adicional	R2			R1	
		Pen. Med.	Coef. R ² Med.	Radio P. Med.	Med. de Radio	Des. Tip. De Radio
Diam-Terr	Presión min	1.011	0.975	0.027	0.065	0.048
Diam-Terr	Presión med	1.012	0.975	0.027	0.066	0.048
Diam-Terr	Presión máx	1.013	0.976	0.028	0.066	0.049
Diam-Uso	Presión min	1.013	0.974	0.030	0.067	0.048
Diam-Uso	Presión med	1.014	0.974	0.030	0.068	0.049
Diam-Uso	Presión máx	1.007	0.974	0.027	0.068	0.049
Diam-Terr	Velocidad máx	1.014	0.973	0.030	0.069	0.049
Mate-Uso	Velocidad mi	1.018	0.980	0.027	0.069	0.053
Diam-Terr	Velocidad med	1.016	0.973	0.031	0.070	0.051
Año-Prof	Presión min	1.046	0.982	0.049	0.070	0.072

Tabla 22. Modelo de tres variables independientes. Estadísticos de calidad, mejores modelos.

Puesto que los mejores modelos parten de las combinaciones diámetro-terreno y diámetro-uso, los resultados se concentran en una misma zona superando ligeramente los resultados de dichos modelos de orden dos, representado el mejor de ellos (diámetro-terreno) con trazo gris, Figura 80.

Cabe mencionar que aun cuando el radio medio obtenido para el modelo año-profundidad-presión mínima es similar al mejor resultado (0,070) la desviación estándar de los resultados es sensiblemente superior a la del resto de modelos seleccionados.

Figura 80. Modelos de tres variables independientes. Resultados globales de los estadísticos de calidad en los mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.4.3 Resumen modelos de tres variables

A diferencia de los resultados observados con los modelos de orden dos analizados, la incorporación de una tercera variable permite mejorar la capacidad predictiva ligeramente Figura 81. En este sentido se consigue una reducción en el mejor modelo de orden tres formado por diámetro y material como variables conjuntas y velocidad mínima, pasando de un radio medio de 0,076 a un radio en dicho modelo de orden tres de 0,062. Apenas se observan diferencias entre los mejores modelos de tres variables independientes (Figura 82) y los de dos variables conjuntas y una independiente (Figura 81). Ambos conjuntos de modelos se comparan en la Figura 83 sin observarse diferencias apreciables.

Figura 81. Evolución del radio medio con el orden del modelo con dos variables conjuntas y el resto independientes.

Figura 82. Evolución del radio medio con el orden del modelo con todas las variables independientes.

Figura 83. Evolución del radio medio con el orden del modelo comparativa entre modelos con dos variables conjuntas y el resto independientes (rojo) y modelos con todas las variables independientes (azul)

6.5 Modelos de orden cuatro

Siguiendo con la metodología propuesta se construyen los modelos de orden cuatro partiendo de los mejores resultados de los modelos de orden tres. Por un lado se toman los tres mejores modelos formados por dos variables conjuntas y una independiente y los tres mejores modelos de tres variables independientes, a todos ellos se les incorpora una cuarta variable independiente adicional.

6.5.1 Modelos de dos variables conjuntas y dos independientes

De entre los diferentes modelos de orden tres de dos variables conjuntas y una independiente se toman los diez mejores siéndolos tres mejores: diámetro – material y velocidad mínima (0,062), diámetro – año y presión máxima (0,063) y diámetro – material y profundidad (0,063). A partir de ellos se construyen diferentes modelos de orden cuatro incorporando una cuarta variable independiente. Se analizan todos los modelos así generados y se eligen los 10 mejores.

6.5.1.1 Diámetro – material – velocidad mínima

Los resultados de los diferentes modelos planteados se recogen en la Figura 84. Modelos de dos variables conjuntas y dos independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-material-velocidad mínima en cada

validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio. Se aprecia un comportamiento idéntico en términos de pendiente de ajuste para todos los modelos salvo el caso de la variable año. En cuanto a la dispersión de los resultados, se observa una mayor variabilidad entre los diferentes modelos destacando las desviaciones observadas con las variables profundidad y año. En cuanto al radio resultante en cada validación se tiene una misma tendencia en todos los modelos, salvo la variable año.

Figura 84. Modelos de dos variables conjuntas y dos independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-material-velocidad mínima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

Analizando los resultados globales, en términos de radio medio se observa que únicamente el modelo que incluye la variable profundidad (0,052) mejora los resultados del modelo de orden 3 de referencia (0,062, marcado con trazo gris). Esta apreciación no se mantiene con el criterio de validación de R^2 pues mediante este indicador según la

Figura 85, se observa que varios modelos mejoran los resultados del modelo de orden 3. Se observa por tanto que al incorporar una variable adicional se mejora ligeramente la capacidad predictiva y al mismo tiempo la desviación estándar de los resultados también mejora, reduciéndose en todos los casos.

Figura 85. Modelos de dos variables conjuntas y dos independientes. Resultados globales de los estadísticos de calidad en modelos de orden cuatro incluyendo las variable diámetro-material-velocidad mínima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.5.1.2 Diámetro – año – presión máxima

El modelo de orden cuatro formado por las variables conjuntas diámetro y año y la variable independiente presión máxima presenta un radio medio de 0,063. Partiendo de ese punto la incorporación de las diferentes variables ofrece los resultados recogidos en la Figura 86y Figura 87.

Atendiendo a la distribución de los estadísticos de calidad en función de la validación se observa de nuevo que todos los modelos presentan unos resultados muy similares en cuanto a pendiente de ajuste salvo la variable material que sigue un comportamiento claramente peor. Este mismo comportamiento se observa en la gráfica de dispersión y de radio. En todos los gráficos se observa una destacada variabilidad en los resultados en función de la validación como se observa en todos los modelos del estudio.

Figura 86. Modelos de dos variables conjuntas y dos independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-año-presión máxima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

Los resultados globales de estos modelos, Figura 87, muestran unos resultados muy próximos a los del modelo de orden 3, únicamente la variable profundidad con un radio medio de 0,052 mejora los resultados de manera apreciable. La incorporación de las variables de velocidad y el tipo de terreno mejoran mínimamente los resultados en términos de radio aunque la dispersión se mantiene o empeora ligeramente. Si se analiza el comportamiento mediante el indicador R2 dicha mejora es inexistente pues los valores de estos modelos igualan pero no mejoran los obtenidos con el modelo de orden tres.

Figura 87. Modelos de dos variables conjuntas y dos independientes. Resultados globales de los estadísticos de calidad modelos de orden cuatro incluyendo las variable diámetro-año-presión máxima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.5.1.3 Diámetro – material - profundidad

El modelo formado por las variables conjuntas diámetro y material más la variable independiente profundidad se combina con el resto de variables de manera análoga a los dos modelos anteriores. De nuevo se observa un comportamiento muy similar en términos de pendiente de ajuste, algo más errático en cuanto a la dispersión obtenida donde destaca el modelo que incorpora la variable uso del suelo al tener un comportamiento sensiblemente diferente al resto de modelos. El comportamiento del radio en cada validación sigue el mismo orden que la pendiente de ajuste, es decir que todos los modelos presentan resultados muy similares para una validación dada. De nuevo los resultados entre validaciones son muy diferentes.

Figura 88. Modelos de dos variables conjuntas y dos independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-material-profundidad en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

En términos globales se observa en la Figura 89 que los resultados mejoran para casi todas las variables salvo las variables año, uso del suelo, y terreno, si bien la mejora que se obtiene es contenida pasando de un radio medio de 0,063 en el modelo de orden 3, representado en trazo gris, a un radio de 0,052 para el modelo de orden cuatro que incorpora la variable velocidad mínima. En términos de R^2 el resultado también es de mejora en los modelos de las variables velocidad y presión, reduciéndose el radio promedio de 0,023 a 0,018 para la variable velocidad mínima.

Figura 89. Modelos de dos variables conjuntas y dos independientes. Resultados globales de los estadísticos de calidad modelos de orden cuatro incluyendo las variable diámetro-material-profundidad. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.5.1.4 Mejores modelos

Una vez analizados los diferentes modelos de orden cuatro con dos variables conjuntas y dos independientes, se seleccionan los diez mejores modelos como se muestra en la Tabla 23. Comparando los resultados con los modelos de orden 3 se observa una mejora del radio medio del mejor modelo, pasando de 0,062 en el modelo diámetro, material más velocidad mínima a 0,052 en el modelo formado por las variables conjuntas diámetro y año y las variables independientes presión media y profundidad.

Variables conjuntas	Variables independientes	R2			R1	
		Pen. Med.	Coef. R ² Med.	Radio P. Med.	Med. de Radio	Des. Tip. De Radio
Diam-Año	P med - Prof	1.033	0.981	0.038	0.052	0.053
Diam-Año	P min - Prof	1.029	0.983	0.034	0.052	0.051
Diam-Año	P máx - Prof	1.035	0.981	0.040	0.052	0.054
Diam-Mat	V min - Prof	1.014	0.989	0.018	0.052	0.054
Diam-Mat	Prof – V med	1.018	0.989	0.021	0.052	0.054
Diam-Mat	Prof – V máx	1.019	0.990	0.022	0.054	0.054
Diam-Mat	Prof – P min	1.024	0.989	0.026	0.054	0.055
Diam-Mat	Prof – P max	1.025	0.989	0.027	0.055	0.058
Diam-Mat	Prof – P med	1.029	0.989	0.031	0.055	0.057
Diam-Año	Prof – V max	1.032	0.983	0.036	0.058	0.057

Tabla 23. Modelo de dos variables conjuntas y dos independientes. Estadísticos de calidad, mejores modelos.

Los resultados globales de estos modelos se recogen en la Figura 90. Tomando la referencia del mejor modelo de orden tres, representado en trazo gris, se observa una mejoría generalizada en los resultados tanto en términos de radio medio como de R².

Figura 90. Modelos de dos variables conjuntas y dos independientes. Resultados globales de los estadísticos de calidad mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.5.2 Modelos de cuatro variables independientes

Los modelos de cuatro variables independientes se construyen incorporando una cuarta variable independiente a los tres mejores modelos de orden tres formado por variables independientes.

6.5.2.1 Diámetro – terreno – presión mínima

Como en otros casos, los resultados de los indicadores de calidad presentan una marcada dependencia de la validación pues todas las gráficas muestran perfiles variables (alejados de la horizontal), Figura 91. Como en otros modelos se observan diferencias muy acusadas entre la precisión obtenida en una validación respecto de otra.

Figura 91. Modelos de cuatro variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro – terreno - presión mínima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

En términos globales, ver Figura 92, el resultado de estos modelos es prácticamente el mismo que en los modelos de orden 3. Los resultados en términos de radio medio son prácticamente los mismos salvo las variables año y material que presentan un comportamiento claramente peor. De esta forma el mejor modelo es el que incorpora la variable velocidad máxima, obteniendo un radio medio de 0,064 mientras que el modelo de orden 3 arrojaba un radio medio de 0,065.

Figura 92. Modelos de cuatro variables independientes. Resultados globales de los estadísticos de calidad modelos de orden cuatro incluyendo las variables diámetro – terreno – presión mínima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.5.2.2 Diámetro – terreno – presión media

En este otro conjunto de modelos de orden cuatro los resultados son análogos. En este caso el conjunto de modelos tiene un comportamiento similar entre sí salvo los que incorporan las variables año y material como se ve en la Figura 93. Este comportamiento se concreta tanto para pendiente de ajuste donde estos dos últimos modelos tienen un peor ajuste, como en la dispersión donde su comportamiento es claramente peor que el resto de los modelos.

Figura 93. Modelos de cuatro variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro – terreno – presión media en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

Los resultados de los indicadores de calidad globales como se muestra en la Figura 94 confirman las tendencias antes señaladas. La capacidad predictiva como en el conjunto de modelos analizados en el apartado anterior apenas se ve mejorada. Los modelos de orden cuatro presentan un comportamiento ligeramente peor que el modelo de orden tres de referencia, marcado con una línea de trazo gris (0,066). Únicamente los modelos que incorporan las variables velocidad máxima (0,064) y velocidad media (0,065) y velocidad mínima (0,066) consiguen mejorar mínimamente o igualar el comportamiento del modelo de orden tres. En términos de R^2 el resultado es análogo pues únicamente el modelo que incorpora la variable velocidad máxima consigue mejorar ligeramente el resultado del modelo de orden tres pasando de 0,027 a 0,026. El resto de variables de velocidad igual al modelo de orden tres al igual que el modelo que incorpora la variable uso del suelo, este

último modelo en el indicador de media del radio se aproxima también (0,069) al valor del modelo de orden tres.

Figura 94. Modelos de cuatro variables independientes. Resultados globales de los estadísticos de calidad modelos de orden cuatro incluyendo las variable diámetro – terreno – presión media. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.5.2.3 Diámetro – terreno – presión máxima

Al igual que los dos conjuntos de modelos anteriores, todos los modelos formados por cuatro variables independientes incluyendo las variables: diámetro, terreno y presión máxima tienen un comportamiento similar. Únicamente los modelos que incorporan las variables año y material presentan diferencias con el resto de modelos siendo su comportamiento peor como se muestra en la Figura 95.

Figura 95. Modelos de cuatro variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro – terreno – presión máxima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

El resultado global de los modelos se muestra en la Figura 96, donde se observa que el comportamiento de prácticamente todos los modelos es similar al del modelo de orden tres de referencia. Únicamente el modelo que incorpora la variable velocidad máxima consigue mejorar mínimamente el radio obtenido en aquel modelo (0,066), marcado con trazo gris, obteniendo un radio medio de 0,065.

Figura 96. Modelos de cuatro variables independientes. Resultados globales de los estadísticos de calidad modelos de orden cuatro incluyendo las variables diámetro – terreno – presión máxima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.5.2.4 Mejores modelos

De entre todos los modelos de cuatro variables independientes analizados se seleccionan los diez mejores en la Tabla 24. Prácticamente todos los modelos incorporan las variables diámetro y terreno y se complementan con algunas de las variables de presión y otra de velocidad. De entre las múltiples combinaciones la que presenta una mejor capacidad predictiva es la formada por diámetro, terreno, presión mínima y velocidad máxima, con un radio medio de 0,064. Comparando con el resto de modelos seleccionados se observa que las diferencias entre todos ellos son mínimas.

De igual forma, si se comparan estos resultados con los de los mejores modelos de tres variables independientes las diferencias son prácticamente inexistentes. Es decir que la incorporación de una variable adicional respecto del modelo de orden tres no supone una ventaja pues el esfuerzo de cálculo asociado no consigue mejorar la predicción.

Variables anteriores	Variable adicional	R2			R1	
		Pen. Med.	Coef. R ² Med.	Radio P. Med.	Med. de Radio	Des. Tip. De Radio
Diam-Terr-P min	V max	1.010	0.976	0.026	0.064	0.047
Diam-Terr-P med	V max	1.011	0.976	0.026	0.064	0.048
Diam-Terr-P max	V max	1.011	0.976	0.026	0.065	0.048
Diam-Terr-V med	P min	1.011	0.976	0.027	0.065	0.048
Diam-Terr-P min	V min	1.011	0.975	0.027	0.065	0.048
Diam-Terr-V med	P med	1.012	0.976	0.027	0.065	0.049
Diam-Terr-P med	V min	1.012	0.976	0.027	0.066	0.048
Diam-Terr-P max	V med	1.006	0.976	0.025	0.066	0.049
Diam-Terr-P max	V min	1.013	0.976	0.028	0.066	0.049
Diam-Terr-P min	V max	1.013	0.974	0.029	0.067	0.048

Tabla 24. Modelo de cuatro variables independientes. Estadísticos de calidad, mejores modelos.

Analizando los resultados globales en la Figura 97, se confirma el comportamiento idéntico de todos los modelos con ligeras diferencias insignificantes entre ellos.

Figura 97. Modelos de cuatro variables independientes. Resultados globales de los estadísticos de calidad mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.5.3 Resumen modelos de cuatro variables

Una vez analizados los dos grandes grupos de modelos de orden cuatro, es decir los formados por dos variables conjuntas y dos variables independientes (Figura 98) y por otro lado los modelos de cuatro variables independientes (Figura 99) se aprecia una ligera mejoría en la capacidad predictiva, similar a la que se obtiene al pasar del modelo de orden dos al de orden tres. En concreto, el radio medio en el mejor modelo de orden tres formado por las variables conjuntas material y diámetro junto con la variable independiente velocidad mínima es de 0,062. En este caso, el mejor modelo de orden cuatro formado por dos variables conjuntas y dos variables independientes consigue reducir el radio medio a 0,052, dicho modelo es el formado por las variables conjuntas diámetro y año junto con las variables independientes presión media y profundidad. Se observa una ligera diferencia en los modelos de cuatro variables independientes (Figura 99) pues no consiguen reducir el radio medio de manera significativa siendo el mejor de ellos el formado por las variables diámetro, terreno, presión mínima y velocidad máxima con un radio de 0,064. La comparativa entre ambos grupos de modelos se muestran en la Figura 100.

Figura 98. Evolución del radio medio con el orden del modelo con dos variables conjuntas y el resto independientes.

Figura 99. Evolución del radio medio con el orden del modelo con todas las variables independientes.

Figura 100. Evolución del radio medio con el orden del modelo comparativa entre modelos con dos variables conjuntas y el resto independientes (rojo) y modelos con todas las variables independientes (azul)

6.6 Modelos de orden cinco

El último conjunto de modelos analizado es el de orden cinco. Aplicando la metodología descrita, se toman los mejores modelos de dos variables conjuntas y dos variables independientes así como los de cuatro variables independientes y se combinan con una quinta variable independiente. Los resultados obtenidos se describen en los apartados siguientes.

6.6.1 Modelos de dos variables conjuntas y tres independientes

El punto de partida para construir los modelos de dos variables conjuntas y tres independientes son los diez mejores modelos de orden cuatro, dos variables conjuntas y dos independientes. Sobre estos se incorpora una quinta variable independiente y se validan los diferentes modelos. Se plantean por tanto modelos con las variables conjuntas diámetro y año complementados con las variables de presión y profundidad de manera independientes.

6.6.1.1 Diámetro – año con presión media y profundidad

Como en otros modelos de orden 4, el comportamiento de los modelos dentro de un mismo conjunto es muy similar. En este caso únicamente el modelo que incorpora la variable material presenta un comportamiento algo diferente en cuanto a los estadísticos de calidad como se muestra en la Figura 101. Se observa que el comportamiento de los modelos es más homogéneo en cuanto a las diferentes validaciones, es decir que la capacidad predictiva de los modelos se ve menos influenciada que en otros casos dada la tendencia horizontal de las gráficas de los estadísticos de calidad, especialmente la pendiente de ajuste y el radio.

Figura 101. Modelos de dos variables conjuntas y tres independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-año con presión media y profundidad en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

En términos de calidad global, los resultados de todos los modelos son muy similares, como se anticipaba en la Figura 102, el único modelo que se aparta del resto es el que incorpora la variable material. En términos generales no se mejora la capacidad predictiva del modelo de orden cuatro de referencia (0,052). El mejor modelo del conjunto analizado, obtenido al incorporar la variable velocidad mínima, obtiene un radio medio de 0,053 por lo que no se consigue mejorar la precisión. Al analizar el comportamiento según el radio del punto medio se observa que el mejor resultado se obtiene con la variable velocidad mínima.

Figura 102. Modelos de dos variables conjuntas y tres independientes. Resultados globales de los estadísticos de calidad modelos de orden cinco incluyendo las variable diámetro-año con presión media y profundidad.

(a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.6.1.2 Diámetro – año con presión mínima y profundidad

Como en el conjunto anterior, según se observa en la Figura 103, todos los modelos presentan un comportamiento muy similar salvo el que incorpora la variable material que presenta un peor comportamiento.

Figura 103. Modelos de dos variables conjuntas y tres independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-año con presión mínima y profundidad en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

Los resultados de globales de calidad, Figura 104, muestran como los diferentes modelos presentan resultados muy similares sin conseguir mejorar los resultados obtenidos con el modelo de referencia de orden cuatro. En este caso el mejor modelo es el que incorpora la variable terreno como variable independiente complementaria obteniendo el mismo radio que en el modelo de orden cuatro, 0,052. En términos de radio del punto medio el mejor modelo en cambio es el que incorpora la variable velocidad mínima, así mismo se observan ligeras diferencias en el orden de los mejores modelos. En cualquier caso, según este criterio tampoco se consigue mejorar el resultado obtenido con el modelo de orden cuatro de referencia que obtenía un radio del punto medio de 0,034 frente al valor de 0,035 obtenido en el modelo de orden cinco.

Figura 104. Modelos de dos variables conjuntas y tres independientes. Resultados globales de los estadísticos de calidad modelos de orden cinco incluyendo las variable diámetro-año con presión mínima y profundidad.
(a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.6.1.3 Diámetro – año con presión máxima y profundidad

Los modelos formados por estas cuatro variables, ver Figura 105, presentan un comportamiento idéntico a los que se forman con las variables diámetro y año como variables conjuntas y presión media y profundidad como variables independientes.

Figura 105. Modelos de dos variables conjuntas y tres independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-año con presión máxima y profundidad en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

En términos globales, Figura 106, el mejor modelos se obtiene con la inclusión de la variable velocidad mínima, obteniéndose un radio medio de 0,053 lo que no mejora el resultado del modelo de orden 4 de referencia el cual obtiene un radio medio de 0,052. El resto de los modelos de orden cinco así formados presenta un comportamiento similar con resultados próximos entre sí salvo el caso de la variable material que presenta un comportamiento claramente peor.

Figura 106. Modelos de dos variables conjuntas y tres independientes. Resultados globales de los estadísticos de calidad modelos de orden cinco incluyendo las variable diámetro-año con presión máxima y profundidad.
(a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.6.1.4 Mejores modelos

Una vez analizados todos los modelos de orden cinco formados por dos variables conjuntas y tres independientes se obtienen unos resultados idénticos. De entre los diez mejores modelos de orden cinco, recogidos en la Tabla 25, únicamente el modelo formado por las variables conjuntas diámetro y material junto con las variables independientes profundidad, velocidad media y velocidad máxima consigue reducir el radio medio desde 0,052 a 0,051. El resto de los mejores modelos obtienen resultados similares.

Si se analiza el criterio del radio del punto medio, los modelos de orden cinco no consiguen mejorar ni igualar los resultados del mejor modelo de orden cuatro que obtenía un radio del punto medio de 0,021. En el caso de los modelos de orden cinco el mejor modelo según este criterio, formado por diámetro y material como variables conjuntas con las variables independientes velocidad mínima, profundidad y velocidad máxima obtiene un radio del punto medio de 0,023.

Variables conjuntas	Variables independientes	R2			R1	
		Pen. Med.	Coef. R ² Med.	Radio P. Med.	Med. de Radio	Des. Tip. De Radio
Diam-Mat	Prof - V med - V max	1.028	0.990	0.029	0.051	0.059
Diam-Año	P min – Prof - Terr	1.035	0.982	0.039	0.052	0.054
Diam-Año	P min – Prof – V max	1.032	0.983	0.036	0.052	0.053
Diam-Año	P min – Prof – V min	1.030	0.983	0.035	0.052	0.052
Diam-Mat	V min – Prof – V max	1.021	0.990	0.023	0.052	0.055
Diam-Año	P med – Prof - V min	1.034	0.983	0.038	0.053	0.053
Diam-Año	P max – Prof - V min	1.037	0.981	0.041	0.053	0.054
Diam-Año	Prof - V max - P med	1.037	0.982	0.041	0.053	0.054
Diam-Año	P med – Prof - Terr	1.038	0.982	0.042	0.053	0.054
Diam-Año	P min - Prof – V med	1.034	0.983	0.038	0.053	0.054

Tabla 25. Modelo de dos variables conjuntas y tres independientes. Estadísticos de calidad, mejores modelos.

Los resultados globales de estos modelos se recogen en la Figura 107, los resultados mostrados apenas ofrecen diferencias respecto a los obtenidos con los diez mejores modelos de orden cuatro.

Figura 107. Modelos de dos variables conjuntas y tres independientes. Resultados globales de los estadísticos de calidad mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.6.2 Modelos de cinco variables independientes

Siguiendo la metodología propuesta, se toman los diez mejores modelos de cuatro variables independientes y se estudia su comportamiento al introducir una quinta variable independiente adicional.

6.6.2.1 Diámetro – terreno – presión mínima – velocidad máxima

Como en otros casos, se aprecia una misma tendencia de comportamiento entre todos los modelos con la excepción de dos variables, en este caso las variables año y material. Sigue existiendo una notable variabilidad de los resultados en función de la validación tal y como se muestra en la Figura 108.

Figura 108. Modelos de cinco variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro – terreno - presión mínima – velocidad máxima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

En términos globales, como se observa en la Figura 109, no se consigue una mejora de la capacidad predictiva pues el mejor modelo, formado al incorporar la variable velocidad mínima, obtiene un radio medio igual al modelo de orden 4 pero no mejora dichos resultados (0,064). En cuanto al criterio R^2 , radio del punto medio, tampoco se mejoran los resultados del modelo de orden cuatro, el mejor modelo como con el criterio del radio medio es el que incorpora la velocidad mínima como quinta variable.

Figura 109. Modelos de cinco variables independientes. Resultados globales de los estadísticos de calidad modelos de orden cinco incluyendo las variables diámetro – terreno – presión mínima – velocidad máxima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.6.2.2 Diámetro – terreno – presión media – velocidad máxima

El comportamiento de este conjunto de modelos es muy similar al desarrollado en el apartado anterior, respecto del que únicamente se reemplaza la presión mínima por presión media. Parecen lógicos por tanto los resultados observados en la Figura 110.

Figura 110. Modelos de cinco variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro – terreno – presión media – velocidad máxima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

Los resultados en términos globales, Figura 111, son análogos a los descritos anteriormente, en este caso el mejor modelo en cuanto a radio medio vuelve a ser el que incorpora la velocidad mínima aunque en este caso se obtiene un resultado ligeramente peor (0,065), por lo que no se iguala el comportamiento del modelo de orden 4. Si se analiza el criterio del radio del punto medio se tiene que el mejor modelo es el que incorpora la variable uso del suelo que solo consigue igualar el resultado del modelo de orden cuatro (0,026).

Figura 111. Modelos de cinco variables independientes. Resultados globales de los estadísticos de calidad modelos de orden cinco incluyendo las variable diámetro – terreno – presión media – velocidad máxima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.6.2.3 Diámetro – terreno – presión máxima – velocidad máxima

De forma similar a los otros conjuntos de modelos de cinco variables independientes, en este caso el comportamiento de los modelos es parejo entre sí salvo de nuevo, los que incorporan las variables año y material como se muestra en la Figura 112.

Figura 112. Modelos de cinco variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro – terreno – presión máxima – velocidad máxima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.

Como se observa en la Figura 113, el mejor resultado en cuanto al radio medio vuelve a obtenerse con la velocidad mínima (0,065) seguido del uso del suelo (0,068), y de nuevo no se mejora el resultado del modelo de orden 4. Si se considera el criterio del radio del punto medio, otra vez se invierte la posición siendo el mejor modelo el que incorpora la variable uso del suelo obteniendo un resultado de 0,026 que iguala pero no mejora el resultado del modelo de orden cuatro.

Figura 113. Modelos de cinco variables independientes. Resultados globales de los estadísticos de calidad modelos de orden cinco incluyendo las variables diámetro – terreno – presión máxima – velocidad máxima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.6.2.4 Mejores modelos

De entre el amplio conjunto de modelos formados por cinco variables independientes se recogen en la Tabla 26 los diez mejores. La comparación de esta tabla con los diez mejores modelos de cuatro variables independientes pone de manifiesto que la incorporación de una variable adicional no mejora el comportamiento pues se tiene resultados muy similares y no se mejora en ningún caso la capacidad predictiva.

Los mejores resultados se obtienen para las combinaciones diámetro y terreno complementados con variables de velocidad y presión. El mejor resultado se obtiene con el modelo formado por diámetro, terreno, presión mínima, velocidad máxima y velocidad mínima. Aunque los resultados del resto de modelos seleccionados son muy similares con diferencias insignificantes.

Variables anteriores	Variable adicional	R2			R1	
		Pen. Med.	Coef. R ² Med.	Radio P. Med.	Med. de Radio	Des. Tip. De Radio
Diam-Terr-P min- V max	V min	1.011	0.976	0.026	0.064	0.047
Diam-Terr-P med- V min	V max	1.011	0.976	0.026	0.065	0.048
Diam-Terr-P max- V max	V min	1.012	0.976	0.027	0.065	0.048
Diam-Terr-P min- V min	V med	1.012	0.976	0.027	0.065	0.048
Diam-Terr-V med- P med	V min	1.011	0.976	0.026	0.066	0.049
Diam-Terr-P max- V med	V min	1.012	0.977	0.026	0.066	0.049
Diam-Uso-P min- V max	V min	1.013	0.975	0.029	0.067	0.048
Diam-Uso-P min- V max	Terr	1.014	0.975	0.029	0.067	0.049
Diam-Terr-P med- V max	Uso	1.008	0.975	0.026	0.068	0.050
Diam-Terr-P max- Vmax	Uso	1.008	0.975	0.026	0.068	0.050

Tabla 26. Modelo de cinco variables independientes. Estadísticos de calidad, mejores modelos.

Si se analizan los gráficos de calidad global, Figura 114, se observa que ningún modelo mejora el umbral del mejor modelo de orden cuatro, marcado en azul coincidente con el mejor modelo pues ambos presentan un radio medio de 0,064. Se aprecia que todos los modelos seleccionados tienen unos resultados muy similares tanto en términos de radio medio como de desviación de dichos valores.

En términos de radio del punto medio los resultados son idénticos, no se observa una mejora respecto de los modelos de orden cuatro. Es decir, que todos los modelos presentan unos resultados muy similares y en ningún caso se mejora la capacidad predictiva de aquellos modelos.

Figura 114. Modelos de cinco variables independientes. Resultados globales de los estadísticos de calidad mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación

6.6.3 Resumen modelos de cinco variables

Una vez analizados todos los modelos de cinco variables, se observa que tanto para los modelos de dos variables conjuntas y tres independientes (Figura 115) como los modelos de cinco variables independientes (Figura 116), no se tiene una mejora de la capacidad predictiva de los modelos por el hecho de incorporar una variable adicional. En este sentido, no queda justificado el uso de modelos de orden 5 pues no suponen una ventaja en el análisis de las roturas y por tanto su aplicación supone un esfuerzo injustificado pues el uso de una quinta variable no permite mejorar los resultados. Ambos comportamientos se comparan en la Figura 117.

Figura 115. Evolución del radio medio con el orden del modelo con dos variables conjuntas y el resto independientes.

Figura 116. Evolución del radio medio con el orden del modelo con todas las variables independientes.

Figura 117. Evolución del radio medio con el orden del modelo comparativa entre modelos con dos variables conjuntas y el resto independientes (rojo) y modelos con todas las variables independientes (azul)

La tendencia de reducción de la dispersión de los diferentes modelos observada en los primeros modelos se consolida en los modelos de orden cinco. Este dato reafirma la idea de que incorporar variables adicionales no supondrá una mejora de la capacidad predictiva. Los resultados convergentes en todas las combinaciones analizadas hacia un mismo radio medio hace pensar que se llega al valor asintótico que se puede obtener con un modelo predictivo; quedando la distancia entre el punto origen (pendiente de ajuste 1,000 y dispersión 1,000) y el mejor modelo observado, condicionada por la componente aleatoria correspondiente a un fenómeno como el de las roturas en tuberías de una red distribución.

6.7 Mejores modelos

Una vez analizado todo el espectro de modelos predictivos desde el modelo de orden cero a los modelos de orden cinco se identifican los mejores modelos para cada uno de los grupos de modelos tal y como se muestra en la Tabla 27.

Orden	Variables Conjuntas	Variables Independientes	Med. de Radio
0	-	-	0,157
1	-	Diámetro	0,075
2	Diámetro-Profundidad	-	0,076
	-	Diámetro-Terreno	0,076
3	Diámetro-Material	Velocidad min	0,062
	-	Diámetro-Terreno-Presión min	0,065
4	Diámetro-Año	Presión med-Profundidad	0,052
	-	Diámetro-Terreno-Presión min-Velocidad max	0,064
5	Diámetro-Material	Profundidad-Velocidad med-Velocidad max	0,051
	-	Diámetro-Terreno-Presión min-Vel. max-Vel. min	0,064

Tabla 27. Mejores modelos por cada orden y grupo.

Destaca la presencia de la variable diámetro en todos los modelos ya sea en los formulados a partir de variables independientes o los de dos variables conjuntas complementadas con variables independientes.

Cabe destacar que la variable terreno está presente en todas las combinaciones de modelos con todas sus variables independientes pero en ninguno de los mejores modelos que incorporan dos variables conjuntas.

De las tres mejores variables independientes según los resultados del modelo de orden uno (diámetro, material y año), destaca que la variable año sólo está presente en uno de los mejores modelos y la variable material sólo aparece en dos de los modelos. Puesto que estas tres variables presentan unos resultados claramente mejores que el resto de manera independiente, en el modelo de orden uno, era de esperar que estuvieran presentes en los mejores modelos. En cambio las variables de velocidad y presión que por sí solas no aportan una capacidad predictiva significativa, es decir que presentan malos resultados en el modelo de orden uno, complementan los modelos de mejor forma que las variables material y año que a priori presentan una mejor capacidad predictiva por sí mismos.

Analizando los modelos formados únicamente con variables independientes se observa una secuencia consistente de las variables seleccionadas. A pesar de que se plantean las

múltiples combinaciones en el modelo de orden superior, conforme se van incorporando las variables se mantienen las variables seleccionadas en el modelo de orden inferior. Es decir que la mejor variable de orden uno, diámetro, se le incorpora en el modelo de orden dos la variable terreno, posteriormente el mejor modelo de orden tres añade la variable Presión mínima. La secuencia se repite con los modelos de orden 4 y 5 que incorporan las variables velocidad máxima y mínima.

Esta secuencia lógica en la formación del mejor modelo no se aprecia en los modelos que incorporan dos variables conjuntas. En este caso hay un par de variables conjuntas que se repite (diámetro-material), pero las variables independientes que les acompañan son diferentes en cada orden de modelo. Esta variabilidad observada en el caso de variables conjuntas se justifica en parte por la mayor complejidad en el momento de estimar el modelo de las dos variables conjuntas pues se necesitan muchos más datos, y por tanto tiene menos calidad. Adicionalmente, puede tener incidencia en la diversidad de variables independientes que acompañan a las variables conjuntas el hecho de que la diferencia entre el mejor modelo y los restantes en cada orden es muy pequeña, tal y como se observa en los resultados de los modelos planteados. Así se justifica que exista poca diferencia en cuanto a la capacidad predictiva del modelo según la variable independiente que se incorpore.

Orden	Variables conjuntas	Variable Independiente	R2			R1	
			Pen. Med.	Coef. R ² Med.	Radio P. Med.	Med. de Radio	Des. Tip. de Radio
0	-	-	1,107	0,975	0,110	0,157	0,125
1	-	-	1,021	0,971	0,036	0,085	0,066
2	2	-	1,021	0,972	0,035	0,076	0,055
	-	2	1,019	0,971	0,034	0,076	0,054
3	2	1	1,011	0,980	0,023	0,062	0,051
	-	3	1,011	0,975	0,027	0,065	0,048
4	2	2	1,033	0,981	0,038	0,052	0,053
	-	4	1,010	0,976	0,026	0,064	0,047
5	2	3	1,028	0,990	0,029	0,051	0,059
	-	5	1,011	0,976	0,026	0,064	0,047

Tabla 28. Resultados de los mejores modelos

El análisis de los resultados de los mejores modelos de cada una de las categorías estudiadas se recoge en la Tabla 28. Respecto de los resultados obtenidos en el modelo de orden cero donde el radio medio es de 0,157 se obtiene una clara mejoría con el modelo de una variable, donde se reduce el radio a 0,085. Las sucesivas combinaciones de variables consiguen reducir el radio hasta 0,051. Si bien, dado que esa mejoría es apenas perceptible entre el modelo de orden cuatro y el modelo de orden cinco, ante la similitud de resultados resulta más adecuado emplear un modelo de orden cuatro por la menor complejidad en el tratamiento de los datos. Si se analiza en cambio el criterio del radio del punto medio, el mejor modelo es el de orden 3 con dos variables conjuntas.

También se aprecia que, dentro de un mismo orden, por lo general el comportamiento de los modelos que incorporan dos variables conjuntas es ligeramente mejor que el modelo con todas las variables independientes en términos de media de radio. Si se analiza el radio del punto medio este comportamiento no se detecta.

6.8 Nivel de desagregación

El nivel de desagregación es otro parámetro que repercute en los resultados obtenidos, en este sentido la metodología contempla un número de grupos entre uno y diez más un caso adicional con el mayor nivel de desagregación posible. Sin embargo, y con el objetivo de limitar el número de combinaciones posibles, la metodología desarrollada plantea un mismo nivel de desagregación para todas las variables incluidas en un mismo modelo. Es decir que no se contempla la posibilidad de que distintas variables dentro de un mismo modelo predictivo tengan niveles de desagregación diferentes. En este sentido se tomaba el nivel que mejor comportamiento ofrecía pero constante dentro del mismo modelo.

Para analizar la posible influencia de un nivel de desagregación variable entre las diferentes variables explicativas de un mismo modelo se han realizado varios análisis complementarios consistentes en el estudio de los modelos con niveles de desagregación variables. Este análisis se limita a las combinaciones óptimas de cada tipo de modelo.

Para realizar este análisis complementario se comprueban todos los niveles de desagregación posibles para las dos variables conjuntas o para las dos primeras variables independientes según el caso. De entre todos los modelos así generados se selecciona la combinación con mejor comportamiento, y sobre ésta se ha incorporado al modelo una tercera variable. Como con las dos variables iniciales, se realizan múltiples combinaciones para buscar el nivel de desagregación óptimo. Como con las construcciones de modelos se repite el proceso de manera secuencial.

Siguiendo este procedimiento la Tabla 29 recoge los resultados de desagregación de variables obtenidos para los modelos construidos con dos variables consideradas conjuntamente y el resto independientes. Se observa que los valores del radio medio obtenido no son los mismos que en apartados anteriores pues como se ha explicado antes, en aquellos análisis el nivel de desagregación era constante para todas las variables.

Orden modelo	Variables conjuntas	Variables independientes	Radio Med.	V1	V2.	V3	V4	V5
2	Diam-Prof	-	0.128	6	94	-	-	-
3	Diam-Mat	V min	0.116	2	7	153	-	-
4	Diam-Año	P med - Prof	0.115	8	4	58	3	-
5	Diam-Material	Prof - V med - Vmax	0.106	2	7	5	9	9

Tabla 29. Desagregación de variables para modelos con dos variables analizadas conjuntamente

De igual forma, la Tabla 30 recoge los resultados para los modelos con todas las variables independientes.

Orden modelo	Variables independientes	Radio Med.	V1	V2.	V3	V4	V5
2	Diam-Terr	0.129	8	4	-	-	-
3	Diam-Terr-P min	0.126	8	4	54	-	-
4	Diam-Terr-P min- V max	0.126	8	4	54	6	-
5	Diam-Terr-P min- V max- V min	0.126	8	4	54	6	6

Tabla 30. Desagregación de variables para modelos con todas las variables independientes

Se observa que las variables hidráulicas (presión y velocidad) junto con la profundidad toman niveles de desagregación entre cuatro y seis intervalos y puntualmente con el nivel de desagregación máxima. De entre los modelos así seleccionados, la mejor combinación, es decir la de menor radio medio, es la formada por diámetro y terreno como variables conjuntas y las variables profundidad, velocidad media y velocidad máxima.

Teniendo en cuenta los resultados de las dos familias de modelos, el pequeño incremento de rendimiento frente a la complejidad que introduce una variable adicional, se opta por el modelo de orden cuatro con dos variables conjuntas y dos independientes con la siguiente desagregación:

- Diámetro, con ocho intervalos: (0-80], (80-100], (100-140], (140-150], (150-200], (200-250], (250-300], (300-3000].
- Año, con cuatro intervalos: [1900-1940], [1941-1993], [1994-2004], [2005-2015].
- Presión media, con toda la desagregación posible (58 valores).
- Profundidad, con tres intervalos: (0-1,2], (1,2-1,4], (1,4-10].

7 CONCLUSIONES Y APORTACIONES

Las conclusiones generadas tras la elaboración de la Tesis se estructuran en tres bloques; por un lado las conclusiones generales obtenidas del trabajo realizado, por otra parte se concretan las aportaciones originales de la Tesis y por último se plantean las futuras líneas de investigación que pueden desarrollarse como continuación de los estudios realizados.

7.1 Conclusiones

Tras los estudios y análisis realizados para la elaboración de la Tesis se extraen las siguientes conclusiones:

1. Propuesta de una metodología para identificar y cuantificar la influencia de las diferentes variables explicativas en la ocurrencia de roturas de una red de distribución.

Se ha desarrollado una metodología basada en un análisis estadístico de las roturas de una red de distribución con el objetivo de identificar las variables explicativas de los fallos y poder seleccionar las más adecuadas para la construcción de modelos de predicción de roturas en base a su influencia sobre la capacidad predictiva de los modelos.

Mediante esta metodología se ha seleccionado un conjunto de variables explicativas clasificadas como variables físicas, ambientales e hidráulicas, con el fin de comprobar si éstas presentan influencia estadísticamente significativa en la ocurrencia de roturas en la red de distribución. El resultado es que prácticamente todas las variables analizadas tienen influencia significativa. Ante esta situación y con el objetivo de generar un modelo predictivo suficientemente preciso sin llegar a un modelo excesivamente complejo incorporando todas las variables, se establece un procedimiento para construir modelos que de manera secuencial vayan incrementando el número de variables explicativas que consideran y poder validarlos para cuantificar la capacidad predictiva de cada uno.

2. Estudio objetivo basado en el tratamiento y análisis estadístico de los datos.

El planteamiento de la Tesis es el tratamiento estadístico de los datos disponibles sin ideas preconcebidas sobre las variables explicativas que se apoye únicamente en la evidencia científica. En este sentido toda la metodología parte de los datos disponibles y

el tratamiento de todas las variables explicativas se efectúa con un mismo procedimiento que permite la combinación de las mismas en diferentes modelos predictivos así como la validación de los mismos de forma objetiva.

El estudio desarrollado en la Tesis busca generar el conocimiento necesario para desarrollar un modelo que valore la vida útil de los componentes de la red de abastecimiento. Para ello se plantea un análisis estadístico y conceptual del comportamiento de la infraestructura en servicio que se concreta en la red de tuberías de la distribución.

3. Metodología para la construcción de modelos de predicción de roturas en una red de distribución.

Partiendo de los datos y registros disponibles, y continuando con la metodología desarrollada para la identificación de las variables explicativas, se desarrolla un procedimiento para seleccionar el modelo más adecuado de entre las múltiples configuraciones posibles según su capacidad para predecir roturas. El planteamiento se basa en la elaboración de un conjunto de modelos de una manera secuencial a partir de un número creciente de variables, y la validación de los mismos, que determina los que mejor se adaptan a la red de estudio.

Para ello, el proceso de construcción de modelos se basa en una inferencia bayesiana mediante la cual se realiza una comparación que permite determinar cómo se modifica la probabilidad de rotura a partir de la información disponible sobre las variables explicativas. Se analiza así la función de distribución de la probabilidad de ocurrencia de las variables explicativas en una base de datos genérica y en una asociada a las roturas en la red de distribución.

El procedimiento va incrementando secuencialmente el orden de los modelos, partiendo del modelo más simple de orden uno, con una única variable explicativa, y se van incorporando paso a paso variables de manera adicional. Para generar los modelos de orden dos se analizan todas las posibles combinaciones de variables (tomadas de dos en dos) tanto de manera independiente como de manera conjunta. De entre los modelos así generados se seleccionan los diez que presentan un mejor comportamiento y sobre esta selección se plantean todos los modelos de orden inmediatamente superior posibles al introducir una variable explicativa independiente adicional; este proceso se realiza con todas las variables explicativas disponibles. Una vez que se evalúan todos los modelos se eligen los diez con mejor comportamiento y se repite el proceso para generar los modelos de un orden mayor.

Con este planteamiento se observa la evolución del comportamiento de los modelos pudiendo determinar la idoneidad o no de emplear un modelo con mayor número de variables en función de la mejora de los estadísticos de calidad obtenidos en cada orden de modelo.

4. Metodología para la validación de los modelos planteados.

Junto con el procedimiento de construcción de modelos se establece una metodología para la validación de los diferentes modelos generales.

Para definir el proceso de validación se ha propuesto un análisis de sensibilidad que permite confirmar los requisitos que debe cumplir en cuanto a los parámetros del procedimiento de validación como son el tamaño de la muestra, el tamaño de las zonas de validación y el periodo de validación. Mediante este análisis se busca limitar la alteración de los resultados por el proceso de validación.

Se plantea un procedimiento que permite valorar de forma cuantitativa la capacidad de predicción de cada modelo para poder seleccionar el más adecuado. Se basa en la división de los registros de roturas en un marco temporal y geográfico, empleándose un conjunto de datos para la calibración del modelo y el resto de datos para su validación. De esta forma la extracción de muestras sigue un criterio geográfico mediante el cual se identifica un punto central aleatorio entorno al cual se define una superficie de muestra con un tamaño limitado. Así, tomado un punto de referencia dentro de la red estudiada se toma una porción de la misma como muestra sobre la que se realiza la validación.

Dada la influencia del periodo de ajuste y validación en los resultados observados, en el procedimiento de validación se emplean parte de los datos para la calibración del modelo y los datos restantes para la validación y comparación de la predicción con las roturas observadas. Se tiene por tanto para cada periodo de validación $n-1$ periodos de contraste, resultando un total de $n*(n-1)$ validaciones en función del periodo analizado (n años).

La validación se fundamenta en un ajuste de regresión lineal de todas las muestras caracterizado por la pendiente de la recta de regresión y el coeficiente de correlación. Ambos parámetros se combinan de manera cuadrática para obtener el estadístico de calidad “radio”, definido como la distancia al punto óptimo, es decir a pendiente de ajuste 1 y coeficiente de correlación 1. Las diferentes validaciones en función de los periodos de ajuste y validación se efectúan de dos formas complementarias. Por un lado establece la media de los radios de cada modelo calculando el promedio de los diferentes radios. Por otro lado se calcula el radio del punto medio, es decir la distancia al punto óptimo de un

punto teórico resultante del promedio de las pendientes de ajuste y de la dispersión. Ambos estadísticos de calidad son empleados en el análisis de los diferentes modelos.

Con esta metodología se puede cuantificar la influencia que tienen las variables explicativas en la capacidad predictiva del modelo resultante lo que permitiría valorar su inclusión en los modelos de vida útil que dan soporte a la planificación de las operaciones de renovación de la red.

Tras la aplicación de la metodología planteada:

5. Se han identificado las variables explicativas más representativas para la red de Canal de Isabel II, así mismo se determina la combinación más adecuada para construir modelos predictivos de roturas en su red de distribución.

Como resultado del proceso aplicado a la red de Canal de Isabel II se identifican las variables explicativas más adecuadas para la construcción de los modelos. Para dicha red de distribución las variables explicativas con significancia estadística identificadas son: año de instalación, material, diámetro, tipo de ubicación, tipo de terreno, profundidad de enterramiento, uso del suelo, variables de presión (mínima, media y máxima) y de velocidad (mínima, media y máxima).

De entre todos los modelos planteados, los que ofrecen mejores resultados son los modelos de orden 4 y 5 con dos variables conjuntas y el resto independientes. De éstos, el mejor resultado se obtiene con la siguiente combinación de variables: diámetro y material como variables conjuntas y las variables independientes profundidad, velocidad media y velocidad máxima. Sin embargo, el modelo de orden cuatro presenta un comportamiento prácticamente igual con la ventaja de requerir una variable menos; la combinación de orden cuatro con el mejor resultado es la de diámetro y año como variables conjuntas y las variables independientes presión media y profundidad.

Se observa que la variable diámetro es la única común en todas las configuraciones óptimas de los diferentes órdenes de modelo. Destaca también el hecho que las tres variables independientes con mejor capacidad predictiva, diámetro, material y año, cuando están presentes en un mismo modelo no mejoran el comportamiento y únicamente se combinan de dos en dos en los mejores modelos. En cambio, las variables de velocidad y presión que por sí solas no aportan una capacidad predictiva significativa, es decir con malos resultados en el modelo de orden uno, complementan los modelos de orden superior de mejor forma que las variables material y año, que a priori presentan una mejor capacidad predictiva. Este resultado confirma la necesidad de realizar un

planteamiento estadístico sin ideas preconcebidas sobre la combinación de las variables explicativas que conforman los modelos.

Cabe destacar que la variable terreno está presente en todas las combinaciones de modelos con todas sus variables independientes pero en ninguno de los mejores modelos que incorporan dos variables conjuntas.

En los modelos formados únicamente con variables independientes se observa una secuencia lógica en la que para cada mejor modelo de orden superior se incluyen las variables del orden inmediatamente inferior más una variable complementaria.

En cambio, en los modelos que incorporan dos variables conjuntas no se aprecia una tendencia tan clara, observándose una mayor heterogeneidad en las variables que van superando el proceso de selección. Esto se debe fundamentalmente a la mayor complejidad para estimar el modelo de las dos variables conjuntas pues requiere un mayor volumen de datos. No obstante, se aprecia que ciertos pares de variables tienen una capacidad predictiva equivalente apareciendo frecuentemente. En esta misma línea, se ha observado una gran similitud en los resultados obtenidos con las diferentes variables independientes que acompañan a las dos variables conjuntas. Esto explica que, dado que la capacidad predictiva entre unas variables y otras es muy similar, el mejor modelo de cada orden puede presentar diversidad en las variables seleccionadas.

6. Número de variables óptimo.

La relación observada entre el número de variables incorporadas a los modelos y la capacidad predictiva de los mismos sugiere que no se tiene una ventaja clara al incorporar modelos con un número elevado de variables predictivas pues no se mejora significativamente dicha capacidad por encima de un cierto número de variables.

Se observa que la mayor mejora de precisión se produce del modelo orden cero al modelo con una única variable. Los pasos posteriores (analizados hasta orden 5) mejoran la precisión de una manera más suave. Para el caso de estudio los resultados entre los modelos de orden cuatro y cinco la capacidad predictiva obtenida es prácticamente la misma por lo que el mejor modelo de entre todos los analizados es el modelo de orden cuatro pues ante una misma capacidad predictiva el mejor modelo es el de menor orden por tener una mayor facilidad de implantación y tratamiento.

En general se observa que varias combinaciones de variables presentan una capacidad predictiva muy similar por lo que no resulta tan determinante elegir una u otra.

7. El análisis de sensibilidad previo a la validación de los diferentes modelos revela que es necesario calibrar los parámetros que gobiernan el proceso de validación de los modelos.

El análisis de sensibilidad ha sido necesario para ajustar los tres parámetros que gobiernan el proceso de validación de los modelos: tamaño de las zonas de validación, tamaño de la muestra y periodo de validación.

El análisis del tamaño de la zona de validación determina que las zonas definidas con un lado del cuadrado de selección de entre 10 y 40 km presentan un mejor comportamiento frente a zonas de validación de mayor tamaño.

Para el análisis del número de muestras de estudio, se observa que por encima de un umbral de 200 unidades los resultados no mejoran significativamente por lo que se establece el tamaño de la muestra en 500.

8. Pese al proceso de ajuste de los parámetros de validación, los resultados muestran como el periodo de ajuste y validación influye significativamente en los resultados obtenidos.

Por este motivo la metodología desarrollada considera un proceso de validación múltiple para minimizar la influencia de los diferentes periodos de ajuste y validación en los resultados. En el caso de la red de Canal de Isabel II los registros disponibles se han agrupado en cuatro periodos de un año completo por lo que se han realizado 12 validaciones; resultantes de validar el modelo de cada año en los tres años restantes.

9. Se cuantifica la mejora introducida en la capacidad predictiva por los diferentes órdenes de los modelos.

Como es lógico se observa que a medida que se incorporan variables explicativas al modelo predictivo se consigue un mejor comportamiento reduciendo el radio medio. En concreto para la red de Canal de Isabel II el radio medio pasa desde 0,157 en el modelo de orden 0, a un valor de 0,075 en el modelo de orden uno y a unos valores para los modelos de dos variables conjuntas: 0,076 (0,076), 0,062 (0,065), 0,052 (0,064) y 0,051 (0,064) para los modelos de orden 2 a 5 respectivamente, entre paréntesis los resultados para los modelos con todas las variables independientes.

De entre los múltiples modelos se observa un comportamiento ligeramente mejor en los modelos que incluyen dos variables conjuntas frente a los que se construyen con todas las variables independientes.

En este sentido, la aplicación a la red de Canal de Isabel II muestra como la mejora obtenida al incorporar un mayor número de variables al modelo predictivo es limitada lo que permite establecer modelos con un buen comportamiento sin necesidad de incorporar un número de variables muy elevado.

7.2 Aportaciones originales de la Tesis

Las aportaciones originales de la Tesis se agrupan desde un punto de vista metodológico y desde la observación de los resultados obtenidos al aplicar la metodología propuesta a la red de distribución de Canal de Isabel II.

Desde un punto de vista metodológico:

1. Propuesta de una metodología para identificar y cuantificar la influencia de las diferentes variables explicativas en la ocurrencia de roturas de una red de distribución mediante el análisis estadístico y conceptual del comportamiento de la red.
2. Estudio objetivo, basado en la evidencia científica, en el que no se adopta ninguna conclusión preconcebida sobre las variables explicativas del fenómeno de las roturas, los modelos surgen únicamente del tratamiento y análisis estadístico de los datos.
3. Metodología para la construcción y validación de modelos de predicción de las roturas en una red de distribución partiendo de los datos históricos disponibles, que permite seleccionar el que presenta un mejor comportamiento de acuerdo a un criterio cuantitativo.

Tras la aplicación de la metodología planteada:

4. Se identifican las variables explicativas más representativas para la red de Canal de Isabel II. Así mismo se determina la combinación más adecuada para construir modelos predictivos. Se observa que la variable diámetro es la única común en todas las configuraciones óptimas.
5. La relación observada entre el número de variables incorporadas a los modelos y la capacidad predictiva de los mismos sugiere que no se tiene una ventaja clara al incorporar modelos con un número elevado de variables predictivas pues no se mejora significativamente dicha capacidad por encima de un cierto número de variables.

6. El análisis de sensibilidad propuesto revela que es necesario calibrar los parámetros que gobiernan el proceso de validación de los modelos.
7. Pese al proceso de ajuste de los parámetros de validación, los resultados muestran como el periodo de ajuste y validación influye significativamente los resultados obtenidos. Por este motivo la metodología desarrollada considera un proceso de validación múltiple para minimizar la influencia de los diferentes periodos de ajuste y validación en los resultados.
8. Se cuantifica la mejora introducida en la capacidad predictiva por los diferentes órdenes de los modelos. El mayor incremento de precisión se produce del modelo orden cero al modelo con una única variable. Los pasos posteriores (analizados hasta orden 5) mejoran la precisión de una manera más suave.

7.3 Futuras líneas de investigación

Partiendo del estudio desarrollado en la Tesis, se plantean varias líneas de investigación que pueden dar continuidad a dicho estudio como son las siguientes:

- a) Extensión de la metodología a otros elementos de la red de abastecimiento como redes de transporte, acometidas y otros elementos específicos propios de un sistema de abastecimiento como válvulas, dispositivos de purga, desagües, ventosas, etc.
- b) Aplicación de la metodología a otras redes de distribución para comprobarla influencia de las variables explicativas y la selección del modelo óptimo.
- c) Incorporación de los modelos predictivos formulados a un modelo de vida útil que permita evaluar el deterioro de cada uno de los elementos de la red.
- d) Evaluar la cantidad y calidad de los datos empleados en la construcción de los modelos junto al análisis cuantitativo de validación de los modelos.
- e) Incorporación de la variable coste al estudio, evaluando el coste de sustitución, el coste asociado a la posible rotura y el ahorro que se puede conseguir con una adecuada reposición de elementos.
- f) Integración en las políticas de renovación de la red. De esta forma la capacidad de predecir las roturas de los elementos de la red y la evaluación económica de las opciones de renovación puede ser aprovechada para optimizar las inversiones en

infraestructuras minimizando el impacto de las posibles roturas de forma que se mejore el nivel de servicio prestado.

BIBLIOGRAFÍA

1. Achim, D., Ghotb, F. & McManus, K., 2007. Prediction of water pipe asset life using neuronal networks. *Journal of Infrastructure System* 13(1), pp. 26-30.
2. Adlan, M. N., Akasseh, J. M. A., Abustan, H. I., Hanif, A. B. M. 2013. Identifying the appropriate time band to determine the minimum night flow: a case study in Kinta Valley, Malaysia. *Water Science and Technology: Water Supply*, 13.2, 328-336.
3. Ahn, J.C., Lee, S.W., Lee, G.S., Koo J.Y., 2005. Predicting water pipe breaks using neural network. *Water Suppl*, 5(3-4), pp. 159-172.
4. Akbarkhiavi, S. P., and Imteaz, M. A., 2016. Using a new pressure index for water distribution systems upgradation improvement evaluation. *Water Science and Technology: Water Supply*, 17(1), 1339-1348. DOI: 10.2166/ws.2016.065
5. American Water Works Association, AWWA, 2013. Buried no longer: Confronting America's Water Infrastructure Challenge. AWWA, Denver, 37pp.
6. American Water Works Association, AWWA, 2002. Drought Management Handbook. AWWA Bookstore, Denver. Colorado. USA.
7. Ascher, H. and Feingold, H., 1984. Repairable systems: Modelling, inference, misconceptions and their causes, M. Dekker, New York.
8. Asnaashari, A., McBean, E. A., Shahrour, I., & Gharabaghi, B., 2009. Prediction of watermain failure frequencies using multiple and Poisson regression. *Water Science and Technology: Water Supply*, 9(1), 9-19.
9. Arias-Nicolás, J. P., Martín, J., Ruggeri, F. and Suárez-Llorens, A., 2015. A Robust Bayesian Approach to an Optimal Replacement Policy for Gas Pipelines. *Entropy*, 17, pp. 3656-3678, DOI 10.3390/e17063656.
10. Babaya, A. V., Kapelan, Z., Savic, D. A. and Walters, G. A. 2005. Least-Cost Design of Water Distribution Networks under Demand Uncertainty. *Journal of Water Resources Planning and Management*, ASCE 131(5), 375-382.
11. Babovic, V., Drécourt, J. P., Keijzer, M., and Hasen, P. F., 2002. A data mining approach to modeling of water supply assets. *Urban Water*, 4(4), pp. 401-414.
12. Berardi, L., Giustolisi, O., Kapelan, Z. & Savic, D.A., 2008. Development of pipe deterioration models for water distribution systems using EPR. *Journal of Hydroinformatics*, 10.2, 113-126.
13. Boxall, J., O'Hagan, A., Pooladsaz, S., & Saul, A.J., 2007. Estimation of burst rates in water distribution mains. *Water management*, 160(2): 73-82.
14. Cabrera, E., Almandoz, J., Arregui, F., García-Serra, J., 1999. Auditoría de redes de distribución de agua. *Ingeniería del Agua*, 6, 4, pp 291-303.
15. Cabrera-Béjar, J. A., Gueorguiev-Tzatchkov, V., 2012. Modelación de redes de distribución de agua con suministro intermitente. *Tecnología y Ciencias del Agua*, 3(2), 5-25.

16. Campbell, E.; Izquierdo, I.; Montalvo, I.; Pérez-García, R., 2016. A Novel Water Supply Network Sectorization Methodology Based on a Complete Economic Analysis, Including Uncertainties. *Water*, 8, 179.
17. Canal de Isabel II Gestión, 2012. Normas para redes de abastecimiento. Madrid: Canal de Isabel II Gestión.
18. Casillas, M. V., Garza-Castañón, L. E., Puig V., and Vargas-Martinez, A., 2015a. Leak Signature Space: An Original Representation for Robust Leak Location in Water Distribution Networks. *Water*; 7, pp. 1129-1148, DOI 10.3390/w7031129.
19. Casillas, M. V., Garza-Castañón, L. E. and Puig V., 2015b. Optimal Sensor Placement for Leak Location in Water Distribution Networks using Evolutionary Algorithms. *Water*, 7, 6496-6515, DOI 10.3390/w7116496
20. Christodoulou, S., Aslani, P., & Vanrenterghem, A., 2003. Risk analysis framework for evaluating structural degradation of water mains in urban settings using neurofuzzy systems and statistical modeling techniques. *Proceedings of World water and environmental resources congress, ASCE, Reston, Va., USA, June 23-26*, 1-9.
21. Clark, R., Stafford, C., & Goodrich, J., 1982. Water distribution systems: a spatial and cost evaluation. *Journal of the Water Resources Planning and Management Division*, 108(WR3), 243-256.
22. Da Costa Martins, A. D., 2011. Stochastic models for prediction of pipe failures in water supply systems, Master Thesis, Universidade Técnica de Lisboa, Lisboa.
23. Davies, J., Clarke, B., Whiter, J. & Cunningham, R., 2001. Factors influencing the structural deterioration and collapse of rigid sewer pipes. *Urban Water*, 3, 73-89.
24. Di Nardo, A., Di Natale, M., Santonastaso, G., Tzatchkov, V., and Alcocer-Yamanaka, V., 2013. Water Network Sectorization Based on Graph Theory and Energy Performance Indices *Journal of Water Resources Planning and Management*. DOI 10.1061/(ASCE)WR.;1; 1943-5452.0000364.
25. Díaz, S., Mínguez, R., González, J., 2016. Stochastic approach to observability analysis in water networks. *Ingeniería del Agua*, 20.3, pp. 139-152, DOI 10.4995/la.2016.4625.
26. Dridi, L., Mailhot, A., Parizeau, M. & Villeneuve, J.-P., 2009. Multiobjective approach for pipe replacement based on bayesian inference of break model parameters. *Journal of Water Resources Planning and Management*, pp. 344-354.
27. Dridi, L., Mailhot, A., Parizeau, M., & Villeneuve, J. P., 2005. A strategy for optimal replacement of water pipes integrating structural and hydraulic indicators based on a statistical water pipe break model. *Proceedings of the 8th International Conference on Computing and Control for the Water Industry*, 65-70.
28. EC, European Commision, 2007. Water scarcity and droughts. In depth assessment. Second Interim Report-june 2007. DG Environment – European Commision
29. EC, European Commision, 2010. Water scarcity and drought in the European Union, Factsheet August 2010. European Commission, Brussels.
30. Economou, T., Kapelan, Z. and Bailey, T. C., 2012. On the prediction of underground water pipe failures: zero inflation and pipe-specific effects. *Journal of Hydroinformatics*, 14 (4), 872-883, DOI 10.2166/hydro.2012.144.

31. Economou, T., 2010. Thesis: Bayesian modelling of recurrent pipe failures in urban water systems using non-homogeneous Poisson processes with latent structure. University of Exeter. Exeter, UK.
32. Economou, T., Kapelan, Z., & Bailey, T.C., 2008. A zero-inflated Bayesian model for the prediction of water pipe bursts. In Proceedings of the 10th Annual Water Distribution Systems Analysis Conference WDSA2008, Van Zyl, J.E., Ilemobade, A.A., Jacobs, H.E. (eds.), August 17-20, 2008, Kruger National Park, South Africa.
33. Economou, T., Kapelan, Z., and Bailey, T.C., 2007. An aggregated hierarchical Bayesian model for the prediction of pipe failures. Proceedings of 9th International Conference on Computing and Control for the Water Industry (CCWI), Leicester, UK.
34. European Environment Agency, EEA, 2012. Towards efficient use of water resources in Europe. EEA Report No 1. ISBN: 978-92-9213-275-0
35. Engelhardt, M., Skipworth, P.J., Savic, D.A., Saul, A.J. & Walters, G.A., 2000. Rehabilitation strategies for water distribution networks: a literature review with a UK perspective. *Urban Water* 2, 153-170.
36. Folkman, S., 2012. Water mains break rate in the USA and Canada: a comprehensive study. Utah State University, 28, Logan, Utah, United States.
37. Geem, Z.W., 2015. Multiobjective Optimization of Water Distribution Networks Using Fuzzy Theory and Harmony Search. *Water*, 7, 3613–3625.
38. Gomes, R., Sousa, J., and Marques, A. S., 2013. The influence of pressure/leakage relationships from existing leaks in the benefits yielded by pressure management. *Water Utility Journal*, 5, 25-32.
39. Gomes, R., Marques, A.S., Sousa, J., 2012. Decision support system to divide a large network into suitable District Metered Areas. *Water science and technology*, vol. 65, no. 9, 1667-1675.
40. Gómez, P., Cubillo, F., Martín F.J., 2014. Proceedings of the 12th International Conference on Computing and Control for the Water Industry, CCWI2013, September 2013, Perugia, Italy.
41. Gómez-Martínez, P., Cubillo, F., Martín-Carrasco F.J., Garrote, L., 2017. Statistical Dependence of Pipe Breaks on Explanatory Variables. *Water*, 9, 158; doi:10.3390/w9030158
42. Herrera, M., Canub, S., Karatzogloub, A., Perez-García, R., and Izquierdo, J., 2010. An approach to water supply clusters by semi-supervised learning. Conference: International Environmental Modelling and Software Society (iEMSs), At Ottawa, Canada
43. Heywood, G., and Starr, M., 2007. Development of national Deterioration Models. UKWIR London, UK, Report No 07/RG/05/20, 162.
44. Hong, H., Allounche, E. & Trivedi, M., 2006. Optimal scheduling of replacement and rehabilitation of water distribution systems. *Journal of Infrastructure Systems*, 12, 184-191.
45. Hukka, J.J.; Katko, T.S., 2015. Appropriate Pricing Policy Needed Worldwide for Improving Water Services Infrastructure. *Journal - American Water Works Association*, 107, E37–E46.

46. Hunaidi, O., 2006. Condition assessment of water pipes. Proceedings of EPA Workshop on Innovation and Research for Water Infrastructure in the 21st Century, 1–6.
47. IWA-Loss-Group, 2007. District Metered Areas: Guidance Notes. IWA Eds. 4, 16, 21, 22, 95.
48. Jacobs P., & Karney, B., 1994. GIS development with application to cast iron water main breakage rates. In Proceedings of 2nd International Conference on Water Pipeline Systems, Edinburgh. Mechanical Engineering Publication Ltd, London.
49. Jones, S. C., Sowby, R. B., 2014. Water system optimization: Aligning energy efficiency, system performance, and water quality. Journal - American Water Works Association, 106 -6, 66-71, dx.doi.org/10.5942/jawwa.2014.106.0087.
50. Kabir, G., 2016. Planning repair and replacement program for water mains: a Bayesian framework. University of British Columbia.
51. Kabir, G., Tresfamariam, S. and Sadiq, R., 2016. Bayesian model averaging for the prediction of water main failure for small to large Canadian municipalities. Canadian Journal of Civil Engineering, 43(3), pp. 233-240, DOI 10.1139/cjce-2015-0374.
52. Kabir, G., Tresfamariam, S., & Sadiq, R., 2015a. Prediction of water mains failure: a Bayesian approach. In the proceedings of the 12th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP), Vancouver, Canada, July, 2015.
53. Kabir, G.; Tresfamariam, S.; Loepky, J.; Sadiq, R., 2015b. Integrating Bayesian linear regression with ordered weighted averaging: Uncertainty analysis for predicting water main failures. ASCE-ASME J. Risk Uncertain. Eng. Syst. A Civ. Eng. 1, 3.
54. Kandulu, J., 2013. Are we risking too much? Estimating the risk-mitigating benefit of urban water supply augmentation options. Water Science & Technology, IWA, 531-539.
55. Kettler, A. J., & Goulter, I. C., 1985. An analysis of pipe breakage in urban water distribution networks. Canadian Journal of Civil Engineering, 12(2), 286-93.
56. Klein, J. P., & Moeschberger, M. L., 2005. Survival analysis: techniques for censored and truncated data. Springer Science & Business Media.
57. Kleiner, Y., & Rajani, B. B., 2010. I-WARP: Individual water main renewal planner. Drinking Water Engineering and Science, 3, 71-77.
58. Kleiner, Y., & Rajani, B., 2002. Forecasting variations and trends in water-main breaks. Journal of infrastructure systems, 8(4), 122-131.
59. Kleiner, Y., and Rajani, B., 2001. Comprehensive review of structural deterioration of water mains: Statistical models. Urban Water; 3(3), pp. 131–150.
60. Kleiner, Y., Adams, B.J. and Rogers J.S., 1998. Long-term planning methodology for water distribution system rehabilitation. Water Resources and Research 34 (8), 2039-2051.
61. Kulkarni, R. B., Golabi, K., and Chuang, J., 1986. Analytical techniques for selection of repair-or-replace options for cast iron gas piping systems - Phase I. Gas Research Institute, PB87-114112, Chicago, IL.
62. Kunkel, G., Laven, K. and Mergelas, B., 2008. Does Your City Have High Risk Pipes? Journal of American Water Works Association, 100(4), 70-74.

63. Large, A., Le Gat, Y., Elachachi, S. M., Renaud, E., Breyse, D., Tomasian, M., 2015. Improved modelling of 'long-term' future performance of drinking water pipes. *Journal of Water Supply: Research and Technology*, 64 (4), pp. 404-414, DOI:10.2166/aqua.2015.115.
64. Le Gat, Y., & Eisenbeis, P., 2000. Using maintenance records to forecast failures in water networks. *Urban Water*, 2(3), 173-181.
65. Lei, J., 1997. Statistical approach for describing lifetimes of water mains – Case Trondheim Municipality. SINTEF Civil and Environmental Engineering, Report No. 22F007.28, Trondheim, Norway.
66. Li, C.Q. and Mahmoodian, M., 2013. Risk based service life prediction of underground cast iron pipes subjected to corrosion. *Reliability Engineering and System Safety*, 119, 102-108.
67. Li, W., Ling, W., Liu, S., Zhao, J., Liu, J. Chen, Q., Qiang, Z., Qu J., 2011. Development of systems for detection, early warning, and control of pipeline leakage in drinking water distribution: A case study. *Journal of Environmental Sciences*, 23,11, 1816-1822.
68. Liu, Z., Kleiner, Y., 2013. State of the art review of inspection technologies for condition assessment of water pipes. *Measurement*, 46, pp.1-15.
69. MacGillivray, B. H., Hamilton, P. D., Strutt, J. E. , Pollard, S. J. T., 2006, Risk analysis strategies in the water utility sector: an inventory of applications for better and more credible decision-making. *Critical Reviews in Environmental Science and Technology*, 36:85-139, doi: 10.1080/10643380500531171.
70. Mailhot, A., Pelletier, G., Noel, & J.P.V., 2000. Modelling the evolution of the structural state of water pipe networks with brief recorded pipe break histories: Methodology and application. *Water Resources Research*, 36(10): 3053–3062.
71. Malm, A., Ljunggren, O., Bergstedt, O., Pettersson, T.J.R., Morrison, G.M., 2012. Replacement predictions for drinking water networks through historical data, *Water Research*, 46 (7), 2149–2158.
72. Martínez Codina, A., 2015. Tesis doctoral: Metodología de análisis estadístico de roturas en redes de distribución de agua. E.T.S.I. Caminos, Canales y Puertos (UPM).
73. Martínez-Codina A., Cueto-Felgueroso, L., Castillo, M. and Garrote, L., 2015. Use of Pressure Management to Reduce the Probability of Pipe Breaks: A Bayesian Approach. *Journal of Water Resources Planning and Management*, 141 (9).
74. McMullen, L. D., 1982. Advanced Concepts in Soil Evaluation for Exterior Pipeline Corrosion, *Proceeding AWWA Annual Conference*, Miami.
75. Nejari, F., Pérez, R., Puig, V., Quevedo, J., Sarrate, R., Cugueró, M. A., Sanz, G., Mirats, J. M. 2012. Abnormal quality detection and isolation in water distribution networks using simulation models. *Drinking Water Engineering and Science*, 5, 67-72.
76. Park, S., Jun, H., Agbenowosi, N., Kim, B.J., & Lim, K., 2011. The proportional hazards modelling of water main failure data incorporating the time-dependent effects of covariates. *Water Resources Management*, 25(1), 1–19.
77. Park, S., Jun, H., Kim, B., & Im, G., 2008. Modelling of water main failure rates using the log-linear ROCOF and the power law process. *Water Resources Management*, 22(9), 1311–1324.

78. Pelletier, G., Mailhot, A. & Villeneuve, J.P., 2003. Modelling Water Pipe Breaks-Three Case Studies. *Journal of Water Resources Planning and Management*, 129 (2), 115-123.
79. Pérez, R., Puig, V., Pascual, J., Peralta, A., Landeros, E. and Jordanas, L., 2009. Pressure sensor distribution for leak detection in Barcelona water distribution network, *Wa. Sci. Technol.*, 9, 715– 721.
80. Poulakis, Z., Valougeorgis, D. and Papadimitiou, C., 2003. Leakage detection in water pipe networks using a bayesian probabilistic framework. *Probabilistic Engineering Mechanics*, 315-27.
81. Rajani, B., & Tesfamariam, S., 2004. Uncoupled axial, flexural, and circumferential pipe soil interaction analyses of partially supported jointed water mains. *Canadian Geotechnical Journal*, 41(6), 997-1010.
82. Renaud, E., Le Gat, Y. and Poulton, M., 2012. Using a break prediction model for drinking water networks asset management: From research to practice. *Water Science and Technology: Water Supply*, 12 (5) 674-682, DOI: 10.2166/ws.2012.040.
83. Rogers, P. D., 2011. Prioritizing water main renewals: case study of the Denver water system. *Journal of Pipeline Systems Engineering and Practice*, 2(3), 73-81.
84. Røstum, J., 2000. Master Thesis: Statistical modeling of pipe failures in water networks.. Norwegian University of Science and Technology, Trondheim.
85. Røstum, J., Dören, L. and Schilling, W., 1997. The concept of business risk used for rehabilitation of water networks. Norwegian University of Science and Technology, IVB-report B2-1997-2, 67-75.
86. Savic, A. D. 2005. Coping with Risk and Uncertainty in Urban Water Infrastructure Rehabilitation Planning. *Proceedings Acqua e città - i convegno nazionale di idraulica urbana s'agnello, Italy*, September 2005.
87. Selvakumar, A., Matthews, J.C., Condit, W., Sterling R., 2015. Innovative research program on the renewal of aging water infrastructure systems. *Journal of Water Supply: Research and Technology*, 64 (2), 117-129, DOI 10.2166/aqua.2014.103.
88. Shamir, U. & Howard, C. D., 1979. An Analytic Approach to Scheduling Pipe Replacement. *Journal of American Water Works Association*, 71(5), 248-258.
89. Ulanicki, B., Prescott, S.L. and Shipley, N., 2003. Analysis of district metered areas (DMAs) performance. *Advances in water supply management proceedings of the CCWI2003*, 59-67.
90. UKWIR, 1999. *A Manual of DMA Practice*. UKWIR Eds., London, United Kingdom.
91. Walski, T.M., & Pelliccia, A., 1982. Economic analyses of watermain breaks. *Journal of American Water Works Association*, 74(3), 140-147.
92. Wang, Y., Moselhi, O., and Zayed, T., 2009. Study of the suitability of existing deterioration models for water mains, *Journal of Performance of Constructed Facilities*, 40, 40–46.
93. Wang, Y., 2006. Master Thesis: Deterioration and condition rating analysis of water mains. Quebec, Canada.
94. Watson, T.G. and Mason, A., 2006. A hierarchical bayesian model for the maintenance of water pipe networks. *7th International Conference on Hydroinformatics*. Nice, France.

95. Watson, T.G., 2005. Tesis doctoral: A hierarchical bayesian model and simulation software for the maintenance of pipe networks. New Zeland, Department of Civil and Resource Engineering, University of Auckland.
96. Watson, T.G., Christian, C.D., Mason, A.J., Smith, M.H., and Meyer, R., 2004. Bayesian-based pipe failure model. *Journal of Hydroinformatics*, 6(4), 259-264.
97. Wood, A., & Lence, B. J., 2006. Assessment of water main break data for asset management. *Journal of American Water Works Association*, 98(7), 76-86.
98. Xu, Q., Chen Q., Li, W., Ma, J. and Blanckaert, K., 2013. Optimal pipe replacement strategy based on break rate prediction through genetic programming for water distribution network. *Journal of Hydro-environment Research*, 1-7.
99. Xu, Q.; Chen, Q.; Li, W.; Ma, J., 2011. Pipe break prediction based on evolutionary data-driven methods with brief recorded data. *Reliability Engineering & System Safety*, 942–948.
100. Yamijala, S., 2007. Statistical Estimation of Water Distribution System Pipe Break Risk. Masters Thesis. Texas A&M University. Texas, USA.
101. Yan, J., & Vairavamoorthy, K., 2003. Fuzzy Approach for Pipeline Condition Assessment. In *Pipeline Technologies, Security, and Safety*, ASCE, Baltimore, MD, USA, 467-476.
102. Yoo, D.G.; Kang, D.; Jun, H.; Kim, J.H., 2014. Rehabilitation Priority Determination of Water Pipes Based on Hydraulic Importance. *Water*, 6, 3864-3887.

ÍNDICE DE FIGURAS

Figura 1. Ciclo de vida de una tubería . “Bathtub curve” (Rajani y Klein, 2001).	24
Figura 2. Discretización de las variables explicativas	39
Figura 3. (A) Ejemplo de área de muestra de validación: punto central en rojo y elementos de la red en azul. (B) Puntos de validación procesados.	42
Figura 4. Interpretación de los estadísticos de ajuste	43
Figura 5. Interpretación del Valor medio de los radios de calidad R1.....	45
Figura 6. Interpretación gráfica de los valores medios de los estadísticos de calidad.	46
Figura 7. Interpretación del Radio R2 correspondiente a los valores medios de los parámetros	47
Figura 8. Análisis de sensibilidad – Comparación del Radio correspondiente a los valores medios de los parámetros. A la izquierda para tamaño de muestra 100 y a la derecha para tamaño de muestra 500.	49
Figura 9. Análisis de sensibilidad – Comparación del ajuste y dispersión. A la izquierda para tamaño de muestra 100 y a la derecha para tamaño de muestra 500.	49
Figura 10. Análisis de sensibilidad – Comparación de los gráficos resumen de resultados por colores. A la izquierda para tamaño de muestra 100 y a la derecha para tamaño de muestra 500.	50
Figura 11. Análisis de sensibilidad – Comparación cuantitativa	51
Figura 12. Análisis cualitativo tamaño de muestra – resultados modelos de orden cero. (a) Muestras 10.000-40.000 m. (b) Muestras 40.000-70.000 m.....	62
Figura 13. Análisis cualitativo tamaño de la zona de validación – resultados R2 para modelos de una variable. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.	65
Figura 14. Análisis cualitativo tamaño de la zona de validación – resultados regresión lineal para modelos año. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.....	65
Figura 15. Análisis cualitativo tamaño de la zona de validación – resultados regresión lineal para material. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.....	66
Figura 16. Análisis cualitativo tamaño de la zona de validación – resultados R2 para la variable año. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.....	66
Figura 17. Análisis cualitativo tamaño de la zona de validación – resultados regresión lineal para año con material. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.....	67
Figura 18. Análisis cualitativo tamaño de la zona de validación – resultados R2 para diámetro. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.	67
Figura 19. Análisis cualitativo tamaño de la zona de validación – resultados regresión lineal para diámetro con año. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.....	68
Figura 20. Análisis cualitativo tamaño de la zona de validación – resultados estadísticos de calidad de todas las combinaciones de variables – escala de colores. (izq) Tamaño de zona de validación 10-40 km. (der) Tamaño de zona de validación 40-70 km.	68

Figura 21. Análisis cuantitativo tamaño de la zona de validación – resultados estadísticos de calidad. Modelos orden 1. (a) Pendiente. (b) Dispersión. (c) Radio.....	70
Figura 22. Análisis cuantitativo tamaño de la zona de validación – resultados estadísticos de calidad. Modelos orden 2, año. (a) Pendiente. (b) Dispersión. (c) Radio.....	71
Figura 23. Análisis cuantitativo tamaño de la zona de validación – resultados estadísticos de calidad. Modelos orden 2, diámetro. (a) Pendiente. (b) Dispersión. (c) Radio.	72
Figura 24. Análisis cualitativo tamaño de muestra – resultados R2 para modelos de una variable. (izq) tamaño de muestra 100 unidades. (der) tamaño de muestra 500 unidades.	74
Figura 25. Análisis cualitativo tamaño de muestra – resultados regresión lineal para la variable año. (izq) tamaño de muestra 100 unidades. (der) tamaño de muestra 500 unidades.	75
Figura 26. Análisis cualitativo tamaño de muestra – resultados regresión lineal para la variable material. (izq) tamaño de muestra 100 unidades. (der) tamaño de muestra 500 unidades.	75
Figura 27. Análisis cualitativo tamaño de muestra – resultados R2 para año. (izq) tamaño de muestra 100 unidades. (der) tamaño de muestra 500 unidades.	76
Figura 28. Análisis cualitativo tamaño de muestra – resultados regresión lineal para la variables año y material. (izq) tamaño de muestra 100 unidades. (der) tamaño de muestra 500 unidades.	76
Figura 29. Análisis cualitativo tamaño de muestra – resultados R2 para la variable diámetro. (izq) tamaño de muestra 100 unidades. (der) tamaño de muestra 500 unidades.	77
Figura 30. Análisis cualitativo tamaño de muestra – resultados regresión lineal para la combinación de variables diámetro y año. (izq) tamaño de muestra 100 unidades. (der) tamaño de muestra 500 unidades.	77
Figura 31. Análisis cualitativo tamaño de muestra – resultados de los estadísticos de calidad de todas las combinaciones de variables – escala de colores.	78
Figura 32. Análisis cuantitativo tamaño de muestra – resultados de los estadísticos de calidad. Modelos orden 1. (a) Pendiente. (b) Dispersión. (c) Radio.	79
Figura 33. Análisis cuantitativo tamaño de muestra – resultados de los estadísticos de calidad. Modelos orden 2, año. (a) Pendiente. (b) Dispersión. (c) Radio.	80
Figura 34. Análisis cuantitativo tamaño de muestra – resultados de los estadísticos de calidad. Modelos orden 2, diámetro. (a) Pendiente. (b) Dispersión. (c) Radio.	81
Figura 35. Análisis periodo de validación – comparación en dos periodos, modelo de orden cero. (izq) Ajuste en periodo 1 y validación en periodo 2. (der) Ajuste en periodo 2 y validación en periodo 1.	83
Figura 36. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados R2 para modelos de una variable. (izq) Periodo 1 ajuste y 2 validación. (der) Periodo 2 ajuste y 1 validación.	86
Figura 37. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados regresión lineal para año. (izq) Periodo 1 ajuste y 2 validación. (der) Periodo 2 ajuste y 1 validación.....	86
Figura 38. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados regresión lineal para material. (izq) Periodo 1 ajuste y 2 validación. (der) Periodo 2 ajuste y 1 validación.	87
Figura 39. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados R2 para año. (izq) Periodo 1 ajuste y 2 validación. (der) Periodo 2 ajuste y 1 validación.	87
Figura 40. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados regresión lineal para año con material. (izq) Periodo 1 ajuste y 2 validación. (der) Periodo 2 ajuste y 1 validación.....	88

Figura 41. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados R2 para diámetro. (izq) Periodo 1 ajuste y 2 validación. (der) Periodo 2 ajuste y 1 validación.	88
Figura 42. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados regresión lineal para diámetro con año. (izq) Periodo 1 ajuste y 2 validación. (der) Periodo 2 ajuste y 1 validación.	89
Figura 43. Análisis cualitativo periodo de validación – comparación en dos periodos. Resultados estadísticos de calidad de todas las combinaciones de variables – escala de colores.	89
Figura 44. Análisis cualitativo periodo de validación – comparación cuantitativa. Ajuste modelo de orden cero. (a) Modelo ajuste 2011, validado 2012. (b) Modelo ajuste 2011, validado 2013. (c) Modelo ajuste 2011, validado 2014. (d) Modelo ajuste 2012, validado 2011. (e) Modelo ajuste 2012, validado 2013. (f) Modelo ajuste 2012, validado 2014. (g) Modelo ajuste 2013, validado 2011. (h) Modelo ajuste 2013, validado 2012. (i) Modelo ajuste 2013, validado 2014. (j) Modelo ajuste 2014, validado 2011. (k) Modelo ajuste 2014, validado 2012. (l) Modelo ajuste 2014, validado 2013.	92
Figura 45. Análisis cuantitativo periodo de validación – resultados estadísticos de calidad. Modelos orden 1. (a) Pendiente. (b) Dispersión. (c) Radio.	94
Figura 46. Análisis cuantitativo periodo de validación – resultados estadísticos de calidad. Modelos orden 2 - diámetro. (a) Pendiente. (b) Dispersión. (c) Radio.	95
Figura 47. Análisis cuantitativo periodo de validación – resultados estadísticos de calidad. Modelos orden 2 - año. (a) Pendiente. (b) Dispersión. (c) Radio.	96
Figura 48. Análisis cualitativo periodo de validación. Resultados de los estadísticos de calidad de todas las combinaciones de variables – escala de colores.	97
Figura 49. Modelos de orden uno. Resultados de los estadísticos de calidad para cada variable en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	101
Figura 50. Modelos de orden uno. Resultados globales de los estadísticos de calidad para cada variable en cada validación. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación.	102
Figura 51. Modelos de dos variables conjuntas. Resultados de los estadísticos de calidad para cada variable combinada con material en cada diámetro. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	104
Figura 52. Modelos de dos variables conjuntas. Resultados globales de los estadísticos de calidad en modelos de orden dos incluyendo la variable diámetro. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación.	105
Figura 53. Modelos de dos variables conjuntas. Resultados de los estadísticos de calidad para cada variable combinada con material en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	106
Figura 54. Modelos de dos variables conjuntas. Resultados globales de los estadísticos de calidad en modelos de orden dos incluyendo la variable material. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación.	107
Figura 55. Modelos de dos variables conjuntas. Resultados de los estadísticos de calidad para cada variable combinada con año en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	108
Figura 56. Modelos de dos variables conjuntas. Resultados globales de los estadísticos de calidad en modelos de orden dos incluyendo la variable año. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación.	109
Figura 57. Modelos de dos variables conjuntas. Resultados globales de los estadísticos de calidad en los mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación.	110

Figura 58. Modelos de dos variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con material en cada diámetro. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	112
Figura 59. Modelos de dos variables independientes. Resultados globales de los estadísticos de calidad en modelos de orden dos incluyendo la variable diámetro. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación.....	113
Figura 60. Modelos de dos variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con material en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	114
Figura 61. Modelos de dos variables independientes. Resultados globales de los estadísticos de calidad en modelos de orden dos incluyendo la variable material. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación.....	115
Figura 62. Modelos de dos variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con año en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	116
Figura 63. Modelos de dos variables independientes. Resultados globales de los estadísticos de calidad en modelos de orden dos incluyendo la variable año. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	117
Figura 64. Modelos de dos variables independiente. Resultados globales de los estadísticos de calidad en los mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	119
Figura 65. Evolución del radio medio con el orden del modelo con dos variables conjuntas y el resto independientes.	120
Figura 66. Evolución del radio medio con el orden del modelo con todas las variables independientes.	120
Figura 67. Modelos de dos variables conjuntas y una independiente. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-profundidad en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	122
Figura 68. Modelos de dos variables conjuntas y una independiente. Resultados globales de los estadísticos de calidad modelos de orden tres incluyendo las variable diámetro-profundidad. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación.....	123
Figura 69. Modelos de dos variables conjuntas y una independiente. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-material en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	124
Figura 70. Modelos de dos variables conjuntas y una independiente. Resultados globales de los estadísticos de calidad modelos de orden tres incluyendo las variable diámetro-material. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación.....	125
Figura 71. Modelos de dos variables conjuntas y una independiente. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-terreno en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.....	126
Figura 72. Modelos de dos variables conjuntas y una independiente. Resultados globales de los estadísticos de calidad en modelos de orden tres incluyendo las variable diámetro-terreno. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación.....	127
Figura 73. Modelos de dos variables conjuntas y una independiente. Resultados globales de los estadísticos de calidad en los mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	129
Figura 74. Modelos de tres variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-terreno en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	130

Figura 75. Modelos de tres variables independientes. Resultados globales de los estadísticos de calidad modelos de orden tres incluyendo las variable diámetro-terreno. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	131
Figura 76. Modelos de tres variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-velocidad máxima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	132
Figura 77. Modelos de tres variables independientes. Resultados globales de los estadísticos de calidad en modelos de orden tres incluyendo las variable diámetro-velocidad máxima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	133
Figura 78. Modelos de tres variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-velocidad media en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	134
Figura 79. Modelos de tres variables independientes. Resultados globales de los estadísticos de calidad en modelos de orden tres incluyendo las variables diámetro-velocidad media. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	135
Figura 80. Modelos de tres variables independientes. Resultados globales de los estadísticos de calidad en los mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	137
Figura 81. Evolución del radio medio con el orden del modelo con dos variables conjuntas y el resto independientes.	138
Figura 82. Evolución del radio medio con el orden del modelo con todas las variables independientes.	138
Figura 83. Evolución del radio medio con el orden del modelo comparativa entre modelos con dos variables conjuntas y el resto independientes (rojo) y modelos con todas las variables independientes (azul).....	139
Figura 84. Modelos de dos variables conjuntas y dos independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-material-velocidad mínima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	140
Figura 85. Modelos de dos variables conjuntas y dos independientes. Resultados globales de los estadísticos de calidad en modelos de orden cuatro incluyendo las variable diámetro-material-velocidad mínima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	141
Figura 86. Modelos de dos variables conjuntas y dos independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-año-presión máxima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	142
Figura 87. Modelos de dos variables conjuntas y dos independientes. Resultados globales de los estadísticos de calidad modelos de orden cuatro incluyendo las variable diámetro-año-presión máxima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	143
Figura 88. Modelos de dos variables conjuntas y dos independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-material-profundidad en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	144
Figura 89. Modelos de dos variables conjuntas y dos independientes. Resultados globales de los estadísticos de calidad modelos de orden cuatro incluyendo las variable diámetro-material-profundidad. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	145
Figura 90. Modelos de dos variables conjuntas y dos independientes. Resultados globales de los estadísticos de calidad mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	146

Figura 91. Modelos de cuatro variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro – terreno – presión mínima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	147
Figura 92. Modelos de cuatro variables independientes. Resultados globales de los estadísticos de calidad modelos de orden cuatro incluyendo las variables diámetro – terreno – presión mínima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	148
Figura 93. Modelos de cuatro variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro – terreno – presión media en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	149
Figura 94. Modelos de cuatro variables independientes. Resultados globales de los estadísticos de calidad modelos de orden cuatro incluyendo las variable diámetro – terreno – presión media. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación.....	150
Figura 95. Modelos de cuatro variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro – terreno – presión máxima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	151
Figura 96. Modelos de cuatro variables independientes. Resultados globales de los estadísticos de calidad modelos de orden cuatro incluyendo las variables diámetro – terreno – presión máxima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	152
Figura 97. Modelos de cuatro variables independientes. Resultados globales de los estadísticos de calidad mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	153
Figura 98. Evolución del radio medio con el orden del modelo con dos variables conjuntas y el resto independientes.	154
Figura 99. Evolución del radio medio con el orden del modelo con todas las variables independientes.	155
Figura 100. Evolución del radio medio con el orden del modelo comparativa entre modelos con dos variables conjuntas y el resto independientes (rojo) y modelos con todas las variables independientes (azul)	155
Figura 101. Modelos de dos variables conjuntas y tres independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-año con presión media y profundidad en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.....	157
Figura 102. Modelos de dos variables conjuntas y tres independientes. Resultados globales de los estadísticos de calidad modelos de orden cinco incluyendo las variable diámetro-año con presión media y profundidad. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación.....	158
Figura 103. Modelos de dos variables conjuntas y tres independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-año con presión mínima y profundidad en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.....	159
Figura 104. Modelos de dos variables conjuntas y tres independientes. Resultados globales de los estadísticos de calidad modelos de orden cinco incluyendo las variable diámetro-año con presión mínima y profundidad. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación.....	160
Figura 105. Modelos de dos variables conjuntas y tres independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro-año con presión máxima y profundidad en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.....	161
Figura 106. Modelos de dos variables conjuntas y tres independientes. Resultados globales de los estadísticos de calidad modelos de orden cinco incluyendo las variable diámetro-año con presión máxima y profundidad. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación.....	162

Figura 107. Modelos de dos variables conjuntas y tres independientes. Resultados globales de los estadísticos de calidad mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	163
Figura 108. Modelos de cinco variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro – terreno - presión mínima – velocidad máxima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	164
Figura 109. Modelos de cinco variables independientes. Resultados globales de los estadísticos de calidad modelos de orden cinco incluyendo las variables diámetro – terreno – presión mínima – velocidad máxima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	165
Figura 110. Modelos de cinco variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro – terreno – presión media – velocidad máxima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	166
Figura 111. Modelos de cinco variables independientes. Resultados globales de los estadísticos de calidad modelos de orden cinco incluyendo las variable diámetro – terreno – presión media – velocidad máxima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	167
Figura 112. Modelos de cinco variables independientes. Resultados de los estadísticos de calidad para cada variable combinada con diámetro – terreno – presión máxima – velocidad máxima en cada validación. (a) Pendiente de ajuste. (b) Dispersión. (c) Radio.	168
Figura 113. Modelos de cinco variables independientes. Resultados globales de los estadísticos de calidad modelos de orden cinco incluyendo las variables diámetro – terreno – presión máxima – velocidad máxima. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	169
Figura 114. Modelos de cinco variables independientes. Resultados globales de los estadísticos de calidad mejores modelos. (a) Valor medio de los radios en función de su desviación estándar. (b) Pendiente y coeficiente correlación	171
Figura 115. Evolución del radio medio con el orden del modelo con dos variables conjuntas y el resto independientes.	172
Figura 116. Evolución del radio medio con el orden del modelo con todas las variables independientes.	172
Figura 117. Evolución del radio medio con el orden del modelo comparativa entre modelos con dos variables conjuntas y el resto independientes (rojo) y modelos con todas las variables independientes (azul)	173
Figura 118. Mapa con la distribución de los tipos de terreno propuesto	212
Figura 119. Función de distribución acumulada genérica y condicionada a las roturas, según el año de instalación (1900-2015)	218
Figura 120. Detalle función de distribución acumulada genérica y condicionada a las roturas, según el año de instalación (1950-2015).....	218
Figura 121. Función de distribución genérica y condicionada a las roturas, según el material	219
Figura 122. Función de distribución acumulada genérica y condicionada a las roturas, según el diámetro.....	219
Figura 123. Detalle función de distribución acumulada genérica y condicionada a las roturas, según el diámetro.....	220
Figura 124. Función de distribución genérica y condicionada a las roturas, según el tipo de ubicación	220
Figura 125. Función de distribución genérica y condicionada a las roturas, según el tipo de terreno	221
Figura 126. Función de distribución acumulada genérica y condicionada a las roturas, según la profundidad de enterramiento.....	221

Figura 127. Detalle función de distribución acumulada genérica y condicionada a las roturas, según la profundidad de enterramiento	222
Figura 128. Función de distribución genérica y condicionada a las roturas, según el uso del suelo	222
Figura 129. Función de distribución genérica y condicionada a las roturas, según la presión media en la tubería	223
Figura 130. Función de distribución genérica y condicionada a las roturas, según la presión máxima en la tubería	223
Figura 131. Función de distribución genérica y condicionada a las roturas, según la presión mínima en la tubería	224
Figura 132. Función de distribución genérica y condicionada a las roturas, según la velocidad media en la tubería	224
Figura 133. Función de distribución genérica y condicionada a las roturas, según la velocidad máxima en la tubería ...	225
Figura 134. Función de distribución genérica y condicionada a las roturas, según la velocidad mínima en la tubería	225

ÍNDICE DE TABLAS

Tabla 1. Modelos de regresión lineal y exponencial para la predicción de roturas (Kabir, 2016)	17
Tabla 2. Modelos de supervivencia (Kabir, 2016)	19
Tabla 3. Modelos de regresión lineal y exponencial para la predicción de roturas (Kabir, 2016)	21
Tabla 4. Variables secundarias.	31
Tabla 5. Clasificación numérica de las variables.....	38
Tabla 6. Periodos de ajuste del modelo y validación	42
Tabla 7. Variables consideradas.	59
Tabla 8. Comparación de los estadísticos de calidad del modelo de orden uno en función del tamaño de la zona de validación	63
Tabla 9. Comparación de los estadísticos de calidad del modelo de orden dos en función del tamaño de la zona de validación	64
Tabla 10. Comparación de los estadísticos de calidad del modelo de orden uno en función del tamaño de muestra.....	73
Tabla 11. Comparación de los estadísticos de calidad del modelo de orden dos en función del tamaño de muestra	74
Tabla 12. Análisis periodo de validación – comparación en dos periodos. Datos observados en cada periodo.	82
Tabla 13. Análisis periodo de validación – comparación en dos periodos. Comparación de los estadísticos de calidad de los modelos de orden uno en función del ajuste del periodo de validación – dos periodos.	84
Tabla 14. Análisis periodo de validación – comparación en dos periodos. Comparación de los estadísticos de calidad de los modelos de orden dos en función del ajuste del periodo de validación – dos periodos.	85
Tabla 15. Análisis periodo de validación – comparación cuantitativa. Datos de roturas por periodo.	90
Tabla 16. Periodos de ajuste del modelo y validación	91
Tabla 17. Modelo de orden cero. Estadísticos de calidad.	99
Tabla 18. Modelo de orden cero. Estadísticos de calidad.	100
Tabla 19. Modelo de dos variables conjuntas. Estadísticos de calidad, mejores modelos.	110
Tabla 20. Modelo de dos variables independientes. Estadísticos de calidad, mejores modelos.	118
Tabla 21. Modelo de dos variables conjuntas y una independiente. Estadísticos de calidad, mejores modelos.....	128
Tabla 22. Modelo de tres variables independientes. Estadísticos de calidad, mejores modelos.....	136
Tabla 23. Modelo de dos variables conjuntas y dos independientes. Estadísticos de calidad, mejores modelos.	146
Tabla 24. Modelo de cuatro variables independientes. Estadísticos de calidad, mejores modelos.	153
Tabla 25. Modelo de dos variables conjuntas y tres independientes. Estadísticos de calidad, mejores modelos.	163
Tabla 26. Modelo de cinco variables independientes. Estadísticos de calidad, mejores modelos.	170
Tabla 27. Mejores modelos por cada orden y grupo.	174
Tabla 28. Resultados de los mejores modelos	175

Tabla 29. Desagregación de variables para modelos con dos variables analizadas conjuntamente	177
Tabla 30. Desagregación de variables para modelos con todas las variables independientes.....	177
Tabla 31. Diámetros de los elementos de la red	209
Tabla 32. Tipos de material	210
Tabla 33. Rangos de velocidades.....	211
Tabla 34. Tipos de terreno	212
Tabla 35. Tipos de ubicación	213
Tabla 36. Categorías de usos del suelo.....	213

ANEXO 1: RANGOS DE LAS VARIABLES EXPLICATIVAS EN EL CASO DE ESTUDIO

ANEXO 1: Rangos de las variables explicativas en el caso de estudio

A1.1 Introducción

En las tablas siguientes se recoge la definición de los rangos observados en las diferentes variables explicativas consideradas en el caso de estudio de aplicación de la metodología a la red de Canal de Isabel II.

A1.1 Diámetro

Diámetro [mm]						
20	45	75	150	350	900	1400
22	47	78	160	400	915	1500
25	49	80	170	450	1000	1600
26	50	85	175	500	1100	1750
30	55	90	180	550	1150	1800
31	59	100	200	600	1200	2000
32	60	110	201	700	1224	2200
33	62	120	210	712	1228	2500
35	63	125	250	750	1250	3000
40	65	130	300	800	1275	
42	70	140	310	850	1300	

Tabla 31. Diámetros de los elementos de la red

A1.2 Tipos de material

Material	Código
Acero	AC
Caña	CA
Fibro cemento	FC
Fundición dúctil	FD
Fundición gris	FG
Hormigón	HC
Polietileno dentro de norma	PE
Polietileno fuera de norma	PO
PVC dentro de norma	PV
PVC fuera de norma	PN

Tabla 32. Tipos de material

A1.3 Rangos de velocidades

Rango de velocidades (m/s)	
V=0	0.1 < V ≤ 0.12
0 < V ≤ 0.005	0.12 < V ≤ 0.14
0.005 < V ≤ 0.01	0.14 < V ≤ 0.16
0.01 < V ≤ 0.02	0.16 < V ≤ 0.18
0.02 < V ≤ 0.03	0.18 < V ≤ 0.2
0.03 < V ≤ 0.04	0.2 < V ≤ 0.25
0.04 < V ≤ 0.05	0.25 < V ≤ 0.3
0.05 < V ≤ 0.06	0.3 < V ≤ 0.4
0.06 < V ≤ 0.07	0.4 < V ≤ 0.5
0.07 < V ≤ 0.08	0.5 < V ≤ 1
0.08 < V ≤ 0.09	V < 1
0.09 < V ≤ 0.1	

Tabla 33. Rangos de velocidades

A1.4 Tipos de terreno

Código	Tipo de terreno
MA	Masas de agua
YE	Yesos
CL	Arcillas
LI	Limos
AR	Arenas
GR	Gravas
RO	Rocas

Tabla 34. Tipos de terreno

Figura 118. Mapa con la distribución de los tipos de terreno propuesto

A1.5 Tipos de ubicación

Código	Tipo de ubicación
A	Aérea
AR	Arqueta
B	Bypass
E	Enterrada
G	Galería
H	Hinca
S	Superficie
TA	Tubería de aspiración
TI	Tubería de impulsión

Tabla 35. Tipos de ubicación

A1.6 Categorías de usos del suelo

Código	Categorías de usos del suelo
ZT	Zonas de transportes
ZU	Zonas urbanas e industriales
ZL	Zonas forestales, agrícolas y con poca o ninguna vegetación

Tabla 36. Categorías de usos del suelo.

ANEXO 2: ANÁLISIS DE DEPENDENCIA DE VARIABLES EXPLICATIVAS

ANEXO 2: Análisis de dependencia de variables explicativas

A2.1 Introducción

En este anexo se recoge el análisis de dependencia estadística realizado sobre las variables inicialmente. Según se describe en la metodología propuesta, todas las potenciales variables seleccionadas de manera inicial deben ser estudiadas para confirmar si existe una dependencia entre dichas variables y la ocurrencia de roturas. Para realizar esta verificación se realiza el test K-S analizando las distribuciones de probabilidad genérica y la condicionada a las roturas.

A2.2 Análisis de las variables propuestas

A continuación, se recoge para cada una de las variables propuestas los resultados obtenidos del test K-S realizado. Para ello se muestra en primer lugar la función de distribución acumulada (FDA) genérica y en segundo lugar la función de distribución acumulada condicionada a las roturas. Siguiendo la metodología del test K-S, se analiza la máxima diferencia entre ambas en valor absoluto, dicho valor es el denominado estadístico de contraste en el test y es el que se compara con el umbral de significancia considerado en este estudio. Dicho umbral se establece en un nivel de $\alpha = 0,05$. En función del número de grado de libertad se obtiene el valor crítico de Kolmogorov-Smirnov mediante la siguiente fórmula:

$$D_{\alpha=0,05} = \frac{1,36}{\sqrt{n}}$$

Según el test K-S, cuando $D > D_{\alpha=0,05}$ se acepta la hipótesis H_0 que significa que los datos no siguen la misma distribución y por tanto se afirma que la variable explicativa en cuestión tiene influencia en la ocurrencia de roturas.

A2.2.1 Año de instalación

Figura 119. Función de distribución acumulada genérica y condicionada a las roturas, según el año de instalación (1900-2015)

Figura 120. Detalle función de distribución acumulada genérica y condicionada a las roturas, según el año de instalación (1950-2015)

A2.2.2 Material

Figura 121. Función de distribución genérica y condicionada a las roturas, según el material

A2.2.3 Diámetro

Figura 122. Función de distribución acumulada genérica y condicionada a las roturas, según el diámetro

Figura 123. Detalle función de distribución acumulada genérica y condicionada a las roturas, según el diámetro

A2.2.4 Tipo de ubicación

Figura 124. Función de distribución genérica y condicionada a las roturas, según el tipo de ubicación

A2.2.5 Tipo de terreno

Figura 125. Función de distribución genérica y condicionada a las roturas, según el tipo de terreno

A2.2.6 Profundidad

Figura 126. Función de distribución acumulada genérica y condicionada a las roturas, según la profundidad de enterramiento

Figura 127. Detalle función de distribución acumulada genérica y condicionada a las roturas, según la profundidad de enterramiento

A2.2.7 Uso del suelo

Figura 128. Función de distribución genérica y condicionada a las roturas, según el uso del suelo

A2.2.8 Presión

Figura 129. Función de distribución genérica y condicionada a las roturas, según la presión media en la tubería

Figura 130. Función de distribución genérica y condicionada a las roturas, según la presión máxima en la tubería

Figura 131. Función de distribución genérica y condicionada a las roturas, según la presión mínima en la tubería

A2.2.9 Velocidad

Figura 132. Función de distribución genérica y condicionada a las roturas, según la velocidad media en la tubería

Figura 133. Función de distribución genérica y condicionada a las roturas, según la velocidad máxima en la tubería

Figura 134. Función de distribución genérica y condicionada a las roturas, según la velocidad mínima en la tubería

A2.3 Conclusión

Del análisis realizado se concluye que todas las variables estudiadas salvo el tipo de ubicación presentan incidencia en la ocurrencia de roturas. Para esta variable el test K-S arroja un estadístico de contraste inferior al umbral considerado por lo que se considera que dicha variable no tiene influencia en la ocurrencia de roturas en la red de distribución analizada.

De esta forma se concluye que las siguientes variables propuestas tienen influencia en la ocurrencia de roturas y son consideradas como variables explicativas en el estudio desarrollado: año de instalación, material, diámetro, tipo de terreno, profundidad, uso del suelo, presión y velocidad.

ANEXO 3: RESULTADOS COMPLETOS DE TODOS LOS MODELOS PROPUESTOS

ANEXO 3: Resultados completos de todos los modelos propuestos

A3.1 Introducción

Según la metodología propuesta se plantean todas las combinaciones posibles de modelos en cada orden analizado con ciertas limitaciones. En este sentido, se toman los diez mejores modelos del orden anterior y sobre ellos se introduce una variable explicativa adicional, con estos modelos se realizan todas las combinaciones posibles con las variables restantes. De esta forma, para cada uno de los diez mejores modelos, se tienen múltiples modelos a estudiar. En el cuerpo del documento se analizan tres de estos modelos en detalle y se recogen los resultados globales de todas las combinaciones posibles.

En este anexo se incluyen todas las figuras complementarias correspondientes a los modelos no analizados de manera individual en el cuerpo del documento. Se estructura el anexo en 4 subcapítulos recogiendo los modelos de orden 2 hasta orden 5. Cada uno de ellos se divide entre las dos familias analizadas, dos variables conjuntas y el resto independientes o todas las variables independientes.

Para cada uno de los modelos se muestran cinco gráficos de resultados:

- Análisis de la pendiente de ajuste para cada una de las doce validaciones realizadas.
- Análisis de la dispersión para cada una de las doce validaciones realizadas.
- Análisis del radio obtenido para cada una de las doce validaciones realizadas.
- Resultados globales (R1).
- Resultados globales (R2).

A3.2 Modelos de orden dos

A3.2.1 Modelos de dos variables conjuntas (2,0)

A3.2.1.1 Año

A3.2.1.2 Diámetro

A3.2.1.3 Material

A3.2.1.4 Presión Máxima

A3.2.1.5 Presión Media

A3.2.1.6 Presión Mínima

A3.2.1.7. Profundidad

A3.2.1.8 Terreno

A3.2.1.9 Uso del suelo

A3.2.1.10 Velocidad Máxima

A3.2.1.11 Velocidad Media

A3.2.1.12 Velocidad Mínima

A3.2.2 Modelos de dos variables independientes (0,2)

A3.2.2.1 Año

A3.2.2.2 Diámetro

A3.2.2.3 Material

A3.2.2.4 Presión Máxima

A3.2.2.5 Presión Media

A3.2.2.6 Presión Mínima

A3.2.2.7. Profundidad

A3.2.2.8 Terreno

A3.2.2.9 Uso del suelo

A3.2.2.10 Velocidad Máxima

A3.2.2.11 Velocidad Media

A3.2.2.12 Velocidad Mínima

A3.3 Modelos de orden tres

A3.3.1 Modelos de dos variables conjuntas y una independiente (2,1)

A3.3.1.1 Diámetro-Año

A3.3.1.2 Diámetro-Material

A3.3.1.3 Diámetro-Profundidad

A3.3.1.4 Diámetro-Terreno

A3.3.1.5 Diámetro-Uso del suelo

A3.3.1.6 Diámetro-Velocidad Máxima

A3.3.1.7. Diámetro-Velocidad Media

A3.3.1.8 Diámetro-Velocidad Mínima

A3.3.1.9 Material-Uso del suelo

A3.3.1.10 Terreno-Material

A3.3.2 Modelos de tres variables independientes (0,3)

A3.3.2.1 Año-Profundidad

A3.3.2.2 Diámetro-Presión Mínima

A3.3.2.3 Diámetro- Profundidad

A3.3.2.4 Diámetro-Terreno

A3.3.2.5 Diámetro-Uso del suelo

A3.3.2.6 Diámetro-Velocidad Máxima

A3.3.2.7. Diámetro-Velocidad Media

A3.3.2.8 Diámetro-Velocidad Mínima

A3.3.2.9 Material-Profundidad

A3.3.2.10 Material-Usó

A3.4 Modelos de orden cuatro

A3.4.1 Modelos de dos variables conjuntas y dos independientes (2,2)

A3.4.1.1 Diámetro-Año-Presión Máxima

A3.4.1.2 Diámetro-Año-Presión Media

A3.4.1.3 Diámetro-Año-Presión Mínima

A3.4.1.4 Diámetro-Año-Profundidad

A3.4.1.5 Diámetro-Material-Presión Mínima

A3.4.1.6 Diámetro-Material-Profundidad

A3.4.1.7. Diámetro-Material-Velocidad Máxima

A3.4.1.8 Diámetro-Material-Velocidad Media

A3.4.1.9 Diámetro-Material-Velocidad Mínima

A3.4.1.10 Diámetro-Terreno-Presión Mínima

A3.4.2 Modelos de cuatro variables independientes (0,4)

A3.4.2.1 Año-Profundidad-Presión Mínima

A3.4.2.2 Diámetro-Terreno-Presión Máxima

A3.4.2.3 Diámetro-Terreno-Presión Media

A3.4.2.4 Diámetro-Terreno-Presión Mínima

A3.4.2.5 Diámetro-Terreno-Velocidad Máxima

A3.4.2.6 Diámetro-Terreno-Velocidad Media

A3.4.2.7. Diámetro-Uso-Presión Máxima

A3.4.2.8 Diámetro-Usa-Presión Media

A3.4.2.9 Diámetro-Uso-Presión Mínima

A3.4.2.10 Material-Uso-Velocidad Mínima

A3.5 Modelos de orden cinco

A3.5.1 Modelos de dos variables conjuntas y tres independientes (2,3)

A3.5.1.1 Diámetro-Año-Presión Máxima-Profundidad

A3.5.1.2 Diámetro-Año-Presión Media-Profundidad

A3.5.1.3 Diámetro-Año-Presión Mínima-Profundidad

A3.5.1.4 Diámetro-Año-Profundidad-Velocidad Máxima

A3.5.1.5 Diámetro-Material-Profundidad-Presión Máxima

A3.5.1.6 Diámetro-Material-Profundidad-Presión Media

A3.5.1.7. Diámetro-Material-Profundidad-Presión Mínima

A3.5.1.8 Diámetro-Material-Profundidad-Velocidad Máxima

A3.5.1.9 Diámetro-Material-Profundidad-Velocidad Media

A3.5.1.10 Diámetro-Material-Velocidad Mínima-Profundidad

A3.5.2 Modelos de cinco variables independientes (0,5)

A3.5.2.1 Diámetro-Terreno-Presión Máxima-Velocidad Máxima

A3.5.2.2 Diámetro-Terreno-Presión Máxima-Velocidad Media

A3.5.2.3 Diámetro-Terreno-Presión Máxima-Velocidad Mínima

A3.5.2.4 Diámetro-Terreno-Presión Media-Velocidad Máxima

A3.5.2.5 Diámetro-Terreno-Presión Media-Velocidad Mínima

A3.5.2.6 Diámetro-Terreno-Presión Mínima-Velocidad Máxima

A3.5.2.7. Diámetro-Terreno-Presión Mínima-Velocidad Mínima

A3.5.2.8 Diámetro-Terreno-Velocidad Media-Presión Media

A3.5.2.9 Diámetro-Terreno-Velocidad Media-Presión Mínima

A3.5.2.10 Diámetro-Uso-Presión Mínima-Velocidad Máxima