



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA CIVIL

**MODELO DE DETERIORO EN TUBERÍAS DE AGUA POTABLE CON
MÉTODOS PROBABILÍSTICOS Y DE INTELIGENCIA ARTIFICIAL**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL

MIGUEL IGNACIO CUBILLOS JIMÉNEZ

PROFESOR GUÍA:

YOLANDA ALBERTO HERNÁNDEZ

PROFESOR CO-GUÍA:

YARKO NIÑO CAMPOS

COMISIÓN:

NITZA MIRANDA VALLEJOS

Powered@NLHPC: Esta investigación/tesis fue parcialmente apoyada por la
infraestructura de supercómputo del NLHPC (ECM-02)

SANTIAGO DE CHILE
2022

**RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL**

POR: MIGUEL IGNACIO CUBILLOS JIMÉNEZ

FECHA: 2022

PROFESOR GUÍA: YOLANDA ALBERTO
HERNÁNDEZ

**MODELO DE DETERIORO EN TUBERÍAS DE AGUA POTABLE CON
MÉTODOS PROBABILÍSTICOS Y DE INTELIGENCIA ARTIFICIAL**

El *agua no facturada* (ANF) corresponde a la diferencia entre el volumen de agua producida y el volumen de agua facturada a los clientes y representa el 33,4% del volumen total de agua potable producida. Se estima que del total de ANF, el 75% corresponde a pérdidas físicas en las redes de distribución de agua potable (Superintendencia de Servicios Sanitarios, 2020).

En el escenario actual de sequía en la zona central de Chile, se hace fundamental la protección y buen manejo de los recursos hídricos, lo que incluye un correcto manejo y operación de los sistemas de distribución de agua potable, dentro de los que se encuentran tanto la predicción y detección de fugas como las buenas prácticas de mantención de redes.

El envejecimiento en redes de distribución y el deterioro natural asociado a este produce un aumento en las fallas de tuberías, aumentando las pérdidas físicas con los años. Este deterioro puede ser modelado con métodos estadísticos y métodos de inteligencia artificial con el fin de predecir que tuberías serán más propensas a fallar dadas sus condiciones actuales.

En el presente trabajo se utilizaron tres algoritmos de inteligencia artificial (*Gradient Boosting Trees*, *Multi Layer Perceptron* y *Support Vector Machine*) para predecir el deterioro en tuberías de agua potable, los que fueron aplicados a un caso de estudio. El algoritmo con un mejor desempeño fue el *Gradient Boosting Trees*. Sin embargo, el *Multi Layer Perceptron* presentó un desempeño levemente inferior con un tiempo de entrenamiento menor.

Los algoritmos se aplicaron en redes experimentales en conjunto con un análisis probabilístico de riesgo sísmico, considerando las variaciones producto del envejecimiento en el coeficiente de pérdida de carga de las tuberías, permitiendo evaluar los efectos de escenarios sísmicos en el contexto del deterioro de las redes.

Se concluye que este tipo de algoritmos puede ser útil siempre que se cuente con registros lo suficientemente extensos como para hacer el entrenamiento. Además, los resultados extraíbles de estos son fáciles de complementar con modelos multiamenazas que consideren el deterioro como una de sus variables.

*A mi familia, en el sentido
más amplio de la palabra.*

AGRADECIMIENTOS

Primero que todo, quisiera agradecer a mi profesora guía, Yolanda Alberto, quien siempre fue un oasis de esperanza y tuvo una palabra de aliento pese a mi infinita negatividad.

Junto a ella me gustaría destacar el enorme apoyo que significó el grupo de CAPTA. Desde el comienzo de mi trabajo junto a Nitza a la eterna e infinita disposición de Diego y Ariel, sinceramente muchas gracias por su apoyo y ayuda.

Además, agradecer al NLHPC por facilitarnos su infraestructura para hacer posible este trabajo.

A mis compañeros de batallas a lo largo de estos años. A Benja, Villa, mi compañero en mi aventura en estructuras, Bea, Graci, y su eterna preocupación por mí. Gracias por esas largas mañanas de clases, que se convertían en largas tardes/noches de estudio y tareas. Gracias por hacer más ameno este proceso, gracias por estar ahí siempre que fue necesario.

A mi familia. A mi mamá siempre presente, a mi papá y su manera tan especial de mostrar cariño, al Nico y a la mami, que siempre esperó el día en que terminara este texto. Gracias. Gracias por todo. Por la preocupación, por el cariño, por los días buenos, por los días malos. Gracias por hacer de mí quien soy.

Por último, a la persona probablemente más importante en este proceso, la Tamiii. Me es imposible escribir lo importante que has sido en estos últimos siete años. Gracias Tamara. Gracias por la compañía, por la paciencia, por ser mi refugio, por ir lado a lado conmigo en este camino. Gracias por ser quien eres, y por decidir serlo junto a mí.

Tabla de contenido

1	Introducción.....	1
1.1	Objetivos.....	2
1.1.1	Objetivo general	2
1.1.2	Objetivos específicos.....	2
1.1.3	Alcances.....	2
1.1.4	Contenidos del informe	2
2	Marco teórico.....	3
2.1	Factores que influyen en los procesos de deterioro	3
2.2	Efectos del envejecimiento en redes de agua potable	6
2.3	Modelos de deterioro: estado del arte.....	8
3	Metodología.....	12
3.1	Modelo con algoritmos de Inteligencia Artificial.....	12
3.1.1	Preprocesamiento de la base de datos.....	13
3.1.2	Modelo de deterioro con Gradient Boosting	16
3.1.3	Modelo de deterioro con Perceptrón Multicapa.....	17
3.1.4	Modelo de deterioro con Máquina de Vector de Apoyo.....	18
3.1.5	Métricas para el análisis de modelos	20
3.1.6	Caso de estudio.....	22
3.2	Incorporación del deterioro dentro de un modelo hidráulico	23
3.2.1	Paquete WNTR y redes experimentales	23
3.2.2	Incorporación de los modelos de deterioro en el análisis hidráulico	25
3.2.3	Incorporación de cambios físicos de la red	26
3.2.4	Incorporación de amenazas naturales	29
3.2.5	Simulación.....	30
4	Resultados.....	32
4.1	Modelo de deterioro: caso de estudio	32
4.1.1	Modelo de deterioro con Gradient Boosting (GBT)	36
4.1.2	Modelo de deterioro con Perceptrón Multicapa (MLP)	42
4.1.3	Modelo de deterioro con Máquina de Vector de Apoyo (SVM).....	48
4.1.4	Comparación de resultados	52
4.2	Modelo hidráulico.....	53

4.2.1	Incorporación del deterioro en un modelo hidráulico.....	53
4.2.2	Incorporación del deterioro en un modelo hidráulico, considerando efectos sísmicos.....	62
5	Análisis de resultados.....	77
6	Conclusiones.....	82
7	Bibliografía.....	83

Índice de Figuras

Figura 2.1: Patrones presentes en Kettler & Goulter (1985). Figura a): Relación entre la tasa de fallas y el diámetro en tuberías de hierro fundido. Figura b): Relación entre el total de fallas y la edad en tuberías de asbesto cemento. Modificado de Kettler & Goulter (1985).	4
Figura 2.2: Ejemplo de rotura circunferencial. Modificado de Rajani & Kleiner (2001).	4
Figura 3.1: Transformado de variables categóricas.	14
Figura 3.2: Random Undersampling.	15
Figura 3.3: Diagrama de flujo optimización de hiperparámetros. Modificado de la documentación de Scikit-learn.	16
Figura 3.4: Variación de la desviación del set de testeo con el número de iteraciones de boosting. Modificado de la documentación de Scikit-learn.	17
Figura 3.5: Arquitectura de una red neuronal.	18
Figura 3.6: Ejemplo de hiperplano de separación y margen. Extraído de Nefedov (2016).	19
Figura 3.7: Ejemplo función kernel. Extraído de Nefedov (2016).	19
Figura 3.8: Matriz de confusión tipo.	20
Figura 3.9: Ejemplo de curva ROC. Modificado de Giraldo-González & Rodríguez, 2020.	22
Figura 3.10: Ejemplo de la asociación de fallas.	23
Figura 3.11: Modelo basado en presiones. Modificado de la documentación de WNTR.	24
Figura 3.12: Diagrama de flujo para la creación de un historial de fallas.	26
Figura 3.13: Adición de fugas. a) Tubería sin fugas, b) Metodología WNTR, c) Metodología usada.	27
Figura 3.14: Variación de la rugosidad en el tiempo para tuberías de asbesto cemento y PVC. Modificado de Abdelmonem et al. (2020).	27
Figura 3.15: Curva de fragilidad para tuberías de asbesto cemento según Garvs (2020).	30
Figura 4.1: Materialidad de la red.	32
Figura 4.2: Distribución de características de la red.	33
Figura 4.3: Distribución de las roturas según cada variable explicativa.	34
Figura 4.4: Distribución roturas según materialidad.	34
Figura 4.5: Matriz de confusión, GBT, primera configuración.	36
Figura 4.6: Curva ROC, GBT, primera configuración.	36
Figura 4.7: Matriz de confusión, GBT, segunda configuración.	37

Figura 4.8: Curva ROC, GBT, segunda configuración.....	37
Figura 4.9: Matriz de confusión, GBT, tercera configuración.....	38
Figura 4.10: Curva ROC, GBT, tercera configuración.....	38
Figura 4.11: Matriz de confusión, GBT, cuarta configuración.....	39
Figura 4.12: Curva ROC, GBT, cuarta configuración.....	39
Figura 4.13: Matriz de confusión, GBT, quinta configuración.....	40
Figura 4.14: Curva ROC, GBT, quinta configuración.....	40
Figura 4.15: Matriz de confusión, GBT, sexta configuración.....	41
Figura 4.16: Curva ROC, GBT, sexta configuración.....	41
Figura 4.17: Matriz de confusión, MLP, primera configuración.....	42
Figura 4.18: Curva ROC, MLP, primera configuración.....	42
Figura 4.19: Matriz de confusión, MLP, segunda configuración.....	43
Figura 4.20: Curva ROC, MLP, segunda configuración.....	43
Figura 4.21: Matriz de confusión, MLP, tercera configuración.....	44
Figura 4.22: Curva ROC, MLP, tercera configuración.....	44
Figura 4.23: Matriz de confusión, MLP, cuarta configuración.....	45
Figura 4.24: Curva ROC, MLP, cuarta configuración.....	45
Figura 4.25: Matriz de confusión, MLP, quinta configuración.....	46
Figura 4.26: Curva ROC, MLP, quinta configuración.....	46
Figura 4.27: Matriz de confusión, MLP, sexta configuración.....	47
Figura 4.28: Curva ROC, MLP, sexta configuración.....	47
Figura 4.29: Matriz de confusión, SVM, tercera configuración.....	48
Figura 4.30: Curva ROC, SVM, tercera configuración.....	48
Figura 4.31: Matriz de confusión, SVM, cuarta configuración.....	49
Figura 4.32: Curva ROC, SVM, cuarta configuración.....	49
Figura 4.33: Matriz de confusión, SVM, quinta configuración.....	50
Figura 4.34: Curva ROC, SVM, quinta configuración.....	50
Figura 4.35: Matriz de confusión, SVM, sexta configuración.....	51
Figura 4.36: Curva ROC, SVM, sexta configuración.....	51
Figura 4.37: Elevación red experimental, caso sólo deterioro.....	53

Figura 4.38: Descripción de la red experimental, caso sólo deterioro. (a) Edad, (b) Diámetros, (c) Materialidad.....	53
Figura 4.39: Probabilidades de falla, red experimental, situación sólo deterioro. (a) Condición actual, (b) Condición a 5 años, (c) Condición a 10 años.....	54
Figura 4.40: Fallas por deterioro, red experimental, situación sólo deterioro. (a) Condición actual, (b) Condición a 5 años, (c) Condición a 10 años.	54
Figura 4.41: Presión promedio en condiciones normales, red experimental, caso sólo deterioro.	55
Figura 4.42: Presión promedio en condiciones actuales, red experimental, caso sólo deterioro.	55
Figura 4.43: Fracción de demanda cubierta en condiciones actuales, red experimental, caso sólo deterioro.	56
Figura 4.44: Porcentaje de cumplimiento de presiones en condiciones actuales, red experimental, caso sólo deterioro.	56
Figura 4.45: Distribución espacial nodos que no cumplen presión requerida en condiciones actuales, red experimental, caso sólo deterioro.	57
Figura 4.46: Presión promedio escenario futuro año 5, red experimental, caso sólo deterioro.	57
Figura 4.47: Fracción de demanda cubierta escenario futuro año 5, red experimental, caso sólo deterioro.	58
Figura 4.48: Porcentaje de cumplimiento de presiones escenario futuro año 5, red experimental, caso sólo deterioro.	58
Figura 4.49: Distribución espacial nodos que no cumplen presión requerida escenario futuro año 5, red experimental, caso sólo deterioro.....	59
Figura 4.50: Presión promedio escenario futuro año 10, red experimental, caso sólo deterioro.	59
Figura 4.51: Fracción de demanda cubierta escenario futuro año 10, red experimental, caso sólo deterioro.	60
Figura 4.52: Porcentaje de cumplimiento de presiones escenario futuro año 10, red experimental, caso sólo deterioro.	60
Figura 4.53: Distribución espacial nodos que no cumplen presión requerida escenario futuro año 10, red experimental, caso sólo deterioro.....	61
Figura 4.54: Descripción de la red experimental, caso con deterioro y sismo. (a) Diámetros, (b) Edad.....	62
Figura 4.55: Elevación red experimental, caso con deterioro y sismo.	62

Figura 4.56: Probabilidades de falla, red experimental, situación con deterioro y sismo. (a) Condición actual, (b) Condición a 5 años, (c) Condición a 10 años.....	63
Figura 4.57: Fallas por deterioro, red experimental, situación con deterioro y sismo. (a) Condición actual, (b) Condición a 5 años, (c) Condición a 10 años.....	63
Figura 4.58: Fallas por sismo, red experimental, situación con deterioro y sismo. (a) Escenario sísmico 1, (b) Escenario sísmico 2.	64
Figura 4.59: Presión promedio en condiciones normales, red experimental, situación con deterioro y sismo.	64
Figura 4.60: Presión promedio en condiciones actuales, red experimental, escenario sísmico 1.	65
Figura 4.61: Fracción de demanda cubierta en condiciones actuales, red experimental, escenario sísmico 1.	65
Figura 4.62: Porcentaje de cumplimiento de presiones en condiciones actuales, red experimental, escenario sísmico 1.	66
Figura 4.63: Distribución espacial nodos que no cumplen presión requerida en condiciones actuales, red experimental, escenario sísmico 1.	66
Figura 4.64: Presión promedio escenario futuro año 5, red experimental, escenario sísmico 1.	67
Figura 4.65: Fracción de demanda cubierta escenario futuro año 5, red experimental, escenario sísmico 1.	67
Figura 4.66: Porcentaje de cumplimiento de presiones escenario futuro año 5, red experimental, escenario sísmico 1.	68
Figura 4.67: Distribución espacial nodos que no cumplen presión requerida escenario futuro año 5, red experimental, escenario sísmico 1.	68
Figura 4.68: Presión promedio escenario futuro año 10, red experimental, escenario sísmico 1.	69
Figura 4.69: Fracción de demanda cubierta escenario futuro año 10, red experimental, escenario sísmico 1.	69
Figura 4.70: Porcentaje de cumplimiento de presiones escenario futuro año 10, red experimental, escenario sísmico 1.	70
Figura 4.71: Distribución espacial nodos que no cumplen presión requerida escenario futuro año 10, red experimental, escenario sísmico 1.	70
Figura 4.72: Presión promedio en condiciones actuales, red experimental, escenario sísmico 2.	71
Figura 4.73: Fracción de demanda cubierta en condiciones actuales, red experimental, escenario sísmico 2.	71

Figura 4.74: Porcentaje de cumplimiento de presiones en condiciones actuales, red experimental, escenario sísmico 2.	72
Figura 4.75: Distribución espacial nodos que no cumplen presión requerida en condiciones actuales, red experimental, escenario sísmico 2.	72
Figura 4.76: Presión promedio escenario futuro año 5, red experimental, escenario sísmico 2.	73
Figura 4.77: Fracción de demanda cubierta escenario futuro año 5, red experimental, escenario sísmico 2.	73
Figura 4.78: Porcentaje de cumplimiento de presiones escenario futuro año 5, red experimental, escenario sísmico 2.	74
Figura 4.79: Distribución espacial nodos que no cumplen presión requerida escenario futuro año 5, red experimental, escenario sísmico 2.	74
Figura 4.80: Presión promedio escenario futuro año 10, red experimental, escenario sísmico 2.	75
Figura 4.81: Fracción de demanda cubierta escenario futuro año 10, red experimental, escenario sísmico 2.	75
Figura 4.82: Porcentaje de cumplimiento de presiones escenario futuro año 10, red experimental, escenario sísmico 2.	76
Figura 4.83: Distribución espacial nodos que no cumplen presión requerida escenario futuro año 10, red experimental, escenario sísmico 2.	76
Figura 5.1: Evolución de la instalación de tuberías de asbesto cemento y PVC en la red estudiada.	78

Índice de Tablas

Tabla 3.1: Parámetros regresión de Poisson. Extraídos de Giraldo-González & Rodríguez (2020).	25
Tabla 3.2: Coeficientes de regresión para asbesto cemento (AC) y PVC. Extraído de Abdelmonem et al. (2020).	28
Tabla 3.3: Parámetros históricos de sismos en Valparaíso. Extraído de Abarca (2022)...	29
Tabla 4.1: Comparación de los tiempos de ejecución de los modelos de deterioro para las distintas configuraciones.	52
Tabla 4.2: Comparación métricas algoritmos de inteligencia artificial en cada configuración.	52
Tabla 4.3: Comparación métricas algoritmos de inteligencia artificial, caso sólo deterioro.	54
Tabla 4.4: Evolución probabilidad de falla promedio, red experimental, situación sólo deterioro.	55
Tabla 4.5: Comparación métricas algoritmos de inteligencia artificial, caso con deterioro y sismo.	62
Tabla 4.6: Evolución probabilidad de falla promedio, red experimental, situación con deterioro y sismo.	63

1 Introducción

Según el Informe de Gestión del Sector Sanitario de 2020 realizado por la Superintendencia de Servicios Sanitarios (SISS), el *agua no facturada* (ANF) representa el 33,4% del volumen total de agua potable producida en el país, la cual alcanza valores en torno a los 1800 millones de metros cúbicos. Se estima que del total de ANF, el 75% corresponde a pérdidas físicas en las redes de distribución de agua potable

El porcentaje de pérdidas no es muy distinto en otros lugares del mundo. De acuerdo a Ress & Roberson (2016), los reportes de pérdidas en los sistemas de distribución de agua potable alrededor del mundo oscilan entre un 10 y un 30%, llegando a más del 50% en los casos más extremos.

En este contexto, y en el escenario actual de escasez hídrica en Chile, se hace fundamental la protección y buen manejo de los recursos hídricos, lo que incluye un correcto manejo y operación de los sistemas de distribución de agua potable, dentro de los que se encuentran tanto la predicción y detección de fugas como las buenas prácticas de mantención de redes.

Una parte fundamental de la operación y mantención de redes incluye la determinación de los tiempos óptimos de reemplazo, teniendo en cuenta las condiciones actuales de las tuberías de la red y priorizando según la importancia de cada una de ellas en la distribución.

Tanto el estado actual de la red como la importancia de cada tubería dentro de ésta se pueden determinar con la construcción de modelos de deterioro (predicción de fugas) en interacción con modelos de simulación hidráulicos (impacto de una falla en la red). Los modelos de deterioro permiten, además, predecir las condiciones futuras de la red y determinar su funcionamiento de seguir las condiciones en las que se encuentra actualmente.

La presente memoria muestra los resultados de la construcción de un modelo de deterioro usando algoritmos de inteligencia artificial y su interacción con un modelo de simulación hidráulico que permite incorporar amenazas naturales. Los modelos de deterioro se aplican a un caso de estudio y se analiza su capacidad predictiva.

El texto se inicia con una descripción de los objetivos y alcances del proyecto; siguiendo con el marco teórico sobre los factores que impactan en el deterioro de tuberías, junto al estudio de los cambios físicos en redes producto de este y una revisión de distintos tipos de modelos de deterioro presentes en la literatura. Luego, se describe la metodología de trabajo y se presentan los resultados, análisis y conclusiones de este.

El presente trabajo se enmarca en el proyecto de Riesgo de Pérdidas en Redes de Agua Potable, RIPRA, del Centro Avanzado para Tecnologías del Agua (CAPTA), que busca desarrollar un sistema que permite analizar el riesgo de pérdidas de agua potable y resiliencia del sistema, debido a amenazas naturales o al deterioro de los elementos.

1.1 Objetivos

1.1.1 Objetivo general

El objetivo del trabajo es desarrollar un modelo de deterioro para redes de distribución de agua potable con tuberías de diversos materiales, que contemple el desgaste gradual al que están sometidas las redes y permita incorporar amenazas naturales y/u operacionales a las que estas están sujetas durante su vida útil.

1.1.2 Objetivos específicos

Se presentan a continuación los objetivos específicos del trabajo:

- Definición de los parámetros que permiten incorporar el modelo de deterioro en el análisis hidráulico (cambios de diámetro, rugosidad, fugas)
- Estudiar impactos del deterioro gradual y de las amenazas naturales en el funcionamiento de las redes de distribución de agua potable.

1.1.3 Alcances

Los alcances del tema de memoria son los siguientes:

- Construcción del modelo de deterioro.
- Incorporación de este en el análisis hidráulico de redes experimentales (*Toy Networks*).
- Incorporación de amenazas naturales en el análisis hidráulico de redes experimentales.
- Estimación de fallas en la red.

1.1.4 Contenidos del informe

El presente trabajo se estructura en seis capítulos principales, cuyo contenido se describe a continuación:

1. Introducción. Se describe la motivación del trabajo, objetivos y alcances de este.
2. Marco teórico. Se presentan las principales causas y efectos del deterioro en tuberías, además del estado del arte de los modelos de deterioro.
3. Metodología. Explica los procedimientos utilizados en cada una de las secciones del trabajo.
4. Resultados. Se exponen los resultados para el caso de estudio y para el modelo hidráulico en redes experimentales con y sin escenario sísmico.
5. Análisis de resultados. Se realiza el análisis y discusión de los resultados de la sección anterior.
6. Conclusiones. Presenta los aspectos a resaltar del trabajo además del cumplimiento de objetivos y posibles orientaciones futuras del trabajo.

2 Marco teórico

En esta sección se presentan los conceptos generales asociados al deterioro en tuberías de agua potable. El capítulo se divide en tres temas principales: factores que influyen en los procesos de deterioro, efectos del envejecimiento en redes de agua potable y modelos de deterioro presentes en la literatura.

2.1 Factores que influyen en los procesos de deterioro

Un modelo es una simplificación de la realidad, donde a través de una serie de variables se busca obtener un comportamiento similar a los sistemas reales. La definición de estas variables es un proceso fundamental en el desarrollo de un modelo. Por ello, entender cuáles son los principales factores que influyen en los procesos de deterioro es una pieza clave en el desarrollo de los modelos.

De acuerdo con Zangenehmadar & Moselhi (2016) estos factores pueden ser agrupados en físicos, ambientales y operacionales. En este estudio los autores utilizan el método de Delphi (método de decisiones basado en opinión experta) para definir cuáles son los factores más relevantes en el deterioro de tuberías, obteniendo los siguientes factores para cada una de las categorías señaladas:

- Físicos: material, instalación, edad, revestimiento, espesor, uso de distintos metales y tipo de uniones.
- Ambientales: tipo de suelo, tipo de material de relleno, pH del suelo, actividad sísmica, perturbaciones.
- Operacionales: presión hidráulica, prácticas de operación y mantenimiento, fugas, pH del agua

Además de los señalados, estudios de predicción de fallas, como los de Debón et al. (2010), Giraldo-González & Rodríguez (2020) y Winkler et al. (2018), muestran que algunos de los factores más relevantes en el deterioro son la edad, el material, el largo o el diámetro.

Entre estos factores, especialmente en las características físicas, se puede observar una serie de patrones. En particular, se observa una relación entre el diámetro de las tuberías y la tasa de fallas por kilómetro por año (Kettler & Goulter, 1985), la que decrece linealmente con el diámetro en tuberías de hierro fundido (Figura 2.1, a). Este decaimiento cambia de ciudad en ciudad y su variación es atribuible a las características propias de cada sitio. Sin embargo, esta relación se volvería no lineal para diámetros muy grandes.

Este fenómeno también fue observado por Guzmán (2011), donde se hizo un análisis de las fallas en tuberías de agua potable para las ciudades de Concepción, Talcahuano y Chiguayante para los años 2009 y 2010, cuyas redes en conjunto se componen en un 49% por tuberías de PVC y en un 36% por tuberías de asbesto cemento, obteniendo que entre el 70 y el 85% de las fallas corresponden a ductos de diámetros pequeños (de 30 a 110

mm), entre 14 y 29% a diámetros medianos (110 a 500 mm) y menos de un 1% a diámetros grandes (mayores a 500 mm).

La disminución de la tasa de fallas con el diámetro está directamente relacionada con el espesor de las tuberías de mayor tamaño dado que, a mayor espesor, hay un aumento en la resistencia estructural y una mayor resistencia a la corrosión (se debe corroer una mayor cantidad de material antes de la falla). Cabe mencionar que, pese a que en Kettler & Goulter (1985) no encontraron variaciones de la tasa de roturas con el diámetro para tuberías de asbesto cemento, este patrón sí es mencionado por Hu & Hubble (2007), donde aseguran que, según Guan (1995) y Mordak & Wheeler (1988) entre otros, la tasa de roturas en tuberías de asbesto cemento está ligada al diámetro, habiendo una menor tasa en elementos de mayor diámetro, resultado que es atribuible a un mayor espesor de los elementos y que implica una mayor inercia a la flexión.

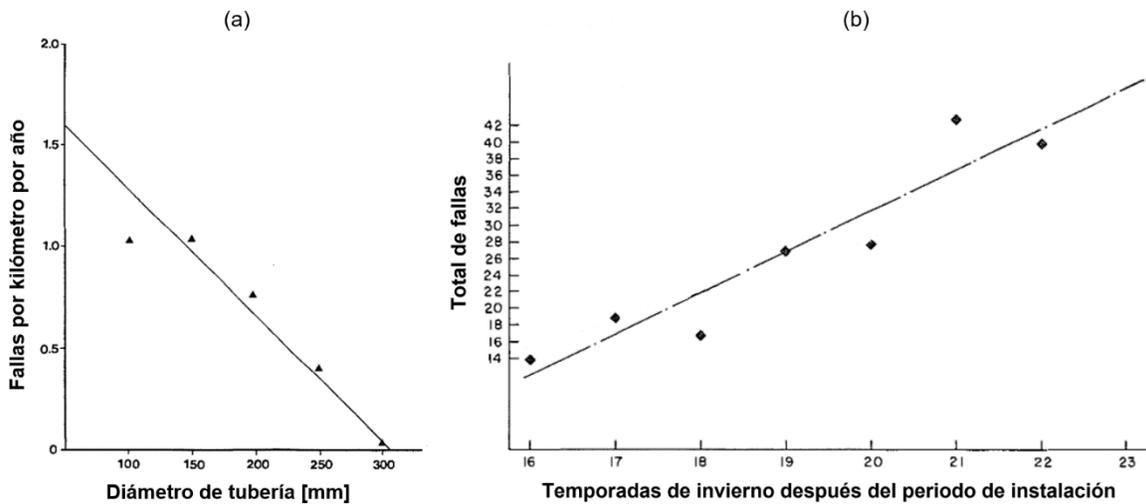


Figura 2.1: Patrones presentes en Kettler & Goulter (1985). Figura a): Relación entre la tasa de fallas y el diámetro en tuberías de hierro fundido. Figura b): Relación entre el total de fallas y la edad en tuberías de asbesto cemento. Modificado de Kettler & Goulter (1985).

Otro patrón que se identifica es la relación entre la edad de la tubería y el total de fallas (Kettler & Goulter, 1985), encontrando un aumento lineal en el tiempo para tuberías de asbesto cemento y de hierro fundido, como se muestra en la Figura 2.1, b). Además, encuentran patrones en la evolución de los tipos de falla en el tiempo, destacando que, en tuberías de hierro fundido, las fallas en uniones aumentan y las fallas por rotura circunferencial disminuyen. Las fallas por rotura circunferencial se refieren a aquellas ocurridas en la sección transversal de la tubería por esfuerzos longitudinales (Figura 2.2).

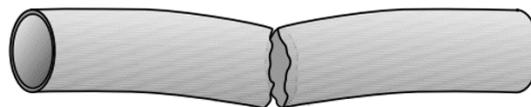


Figura 2.2: Ejemplo de rotura circunferencial. Modificado de Rajani & Kleiner (2001).

En línea con lo anterior, al analizar la variación de las fallas según el material de las tuberías, Guzmán (2011) observa que las fallas en los ductos de PVC se dan en las piezas especiales (coplas, reducciones, etc.), en los de asbesto cemento se deben a la fragilidad del material en la sección transversal, en las tuberías de acero se observaron casos de aplastamiento y torsión, y en las de hierro fundido casos de torsión y compresión.

Este estudio, al abarcar los años 2009 y 2010, incluye los efectos del terremoto del Maule de 2010. Si se estudia las fallas según tipo de material, se observa que la mayoría ocurren en tuberías de asbesto cemento (61,5% de las fallas), seguido del PVC (27,5%) que tuvo la mayoría de sus fallas en las coplas. Se destaca que no hubo fallas en las tuberías de HDPE debido a la capacidad de este material de deformarse sin perder rigidez ni resistencia y que la mayor cantidad de fallas ocurren en tuberías de asbesto cemento pese a que el 49% de la red es de PVC.

El tipo de suelo es otro factor a considerar en el deterioro en redes. La expansión o contracción del suelo podría generar situaciones donde no toda la tubería esté siendo soportada, haciendo que ésta se comporte como una viga sometida a un momento flector, lo que induce fallas del tipo rotura circunferencial en tuberías de asbesto cemento. La reducción de las fallas con el diámetro puede estar ligada a este factor, dado que un diámetro mayor significa una mayor inercia y menores tensiones por el momento flector. Otro efecto posible es un movimiento diferencial de suelo que provocaría esfuerzos de corte además de los momentos flectores señalados (Hu & Hubble, 2007).

Más efectos del tipo de suelo se pueden observar en Guzmán et al. (2012), donde a partir de los datos de fallas de tuberías en Concepción producto del terremoto de 2010 observan que el 93,5% se produjo en suelos arenosos (arenas finas y gruesas de granulometrías uniformes SP y arenas limosas SM) y que el 6,5% restante se produjo en suelos arcillosos y limosos. Además, concluyen que las tuberías enterradas en suelos saturados o en suelos con alta humedad sufrieron más daño.

Por otra parte, Hu & Hubble (2007) estudiaron los principales factores que afectan el deterioro de las tuberías de asbesto cemento en la Ciudad de Regina, Canadá, a partir de los datos de roturas desde 1980 a 2004. Adicional a los factores ya señalados, presentan los siguientes:

- Químicos:

Plantean que la química del agua subterránea, del agua transportada o a las propiedades químicas del suelo pueden ser factores relevantes. Según la AWWA y ASTM se pueden distinguir dos tipos de ambientes de suelo agresivos para las tuberías de asbesto cemento: suelos ácidos y suelos sulfatados.

- Climáticos:

El clima lo abordan desde dos aspectos: el déficit de lluvias y el índice de congelamiento. En cuanto al déficit de precipitación, lo definen como la diferencia entre la precipitación y la

humedad necesaria para los procesos de evapotranspiración. El patrón que ellos observaron sugiere que los peaks en la tasa de roturas coinciden con peaks en el déficit de precipitación. Estas observaciones coinciden con los resultados de los modelos de Giraldo-González & Rodríguez (2020), donde la precipitación aparece como una de las variables explicativas más relevantes para la predicción de roturas en tuberías de PVC y asbesto cemento.

El índice de congelamiento lo definen como la acumulación de las temperaturas medias diarias bajo 0 °C, para los meses en los que el promedio histórico mensual está bajo 0°C, observándose que los peaks del índice de congelamiento coinciden con peaks en la tasa de rotura, lo que indica una alta influencia del frío en las roturas de tuberías de asbesto cemento. Además, mencionan que, según Kleiner & Rajani (2002), un suelo seco tendrá menor capacidad de calor latente, por lo que la penetración del frío será mayor y habrá una mayor carga por congelamiento en las tuberías.

- Clustering de roturas:

Plantean que, de acuerdo con Goulter & Kazemi (1988), una falla induce otras tanto espacial como temporalmente, por lo que desarrollaron un modelo donde la primera rotura es independiente y las subsecuentes adyacentes son dependientes. En el caso de la ciudad de Regina, Canadá se detectó que, del total de roturas, el 21,6% corresponde a nuevas roturas del mismo elemento en el mismo lugar, sugiriendo que una rotura inicial es un indicador de más roturas en el mismo sitio.

2.2 Efectos del envejecimiento en redes de agua potable

Una serie de cambios físicos ocurren en las redes de agua potable producto del envejecimiento, los que pueden ir desde las típicas fugas (producto del desgaste natural del material, corrosión, eventos sísmicos, expansión del suelo, etc.) hasta cambios en el diámetro y en la rugosidad asociados a la calidad del agua transportada. Estos cambios pueden alterar los patrones de flujo, los tiempos de llegada al final de la red o el gasto de energía (Shahzad & James, 2002).

Sharp & Walski (1988) plantean una metodología para la predicción de la rugosidad en el tiempo para tuberías metálicas no revestidas de diámetros entre 100 y 400 mm, cuya base está en la hipótesis de crecimiento lineal de esta. Dicho crecimiento depende directamente de la calidad del agua transportada, ya sea a partir del pH o a partir del índice de Langelier (LI), que es un indicador de la saturación por carbonato de calcio en el agua.

Para determinar las tasas de crecimiento, plantean dos posibles caminos. El primero es hacer las regresiones correspondientes en caso de poseer mediciones históricas de la rugosidad. Luego, con la tasa de crecimiento se puede estimar el valor de la rugosidad para un año dado. El segundo camino es estimar las tasas de crecimiento a partir de la calidad del agua con el índice de Langelier.

Por otro lado, Echávez (1997) hace un estudio de la evolución de la rugosidad para tuberías de hierro galvanizado y cobre de hasta 50.8 mm (dos pulgadas), concluyendo que las tuberías de hierro galvanizado no tienen un cambio lineal, sino que la evolución de la rugosidad sigue una tendencia parabólica. Por otro lado, en las tuberías de cobre no se observaron cambios en el tiempo.

La hipótesis de crecimiento parabólico es recogida por Abdelmonem et al. (2020), donde experimentan con tuberías de 100 mm de distintos materiales, entre los que se incluyen hierro fundido (revestido y sin revestir), concreto reforzado, asbesto cemento, PVC y acero, haciendo una distinción entre aquellas que han transportado agua filtrada y aquellas que no.

El análisis muestra que, por ejemplo, en caso de llevar agua filtrada, las tuberías de PVC tienen un aumento en la rugosidad de 1,5 veces en 50 años mientras que las de acero tienen un aumento de 40 veces. Para el caso de agua sin filtrar, el PVC aumenta 15 veces su valor y el acero más de 100 veces.

Los resultados de este estudio en cuanto a las tendencias de crecimiento coinciden de cierto modo con los de Echávez (1997), observándose una variación parabólica de la rugosidad. Sin embargo, en el estudio de Echávez el crecimiento se va atenuando en el tiempo mientras que en Abdelmonem et al. (2020) el crecimiento aumenta en el tiempo.

Además, ambos estudios presentan diferencias metodológicas en la forma de estimar la rugosidad. En el caso de Echávez (1997) la medición se hace estudiando la pérdida de carga en una tubería de 11 m de largo, mientras que en Abdelmonem et al. (2020) se define como la mitad de la diferencia entre el diámetro inicial y final.

Pese a que las metodologías mostradas son distintas y obtienen diferentes parametrizaciones para el mismo fenómeno, los resultados de crecimiento a 50 años se encuentran dentro del mismo rango.

Por ejemplo, Sharp & Walski (1988) reportan tasas de crecimiento que van desde 0,015 mm/año (San Antonio, Texas, datos de Hudson) a 0,63 mm/año (máximo del rango reportado por Colebrook – White) lo que se traduce en un crecimiento de entre 0,75 y 31,5 mm. Considerando una rugosidad inicial de 0,18 mm se tiene un aumento que va desde 5 hasta 176 veces en tuberías metálicas.

En el caso de Echávez (1997), el valor de la rugosidad absoluta para tuberías de hierro galvanizado va desde un valor inicial de 0,0062 mm a 1,02 mm en 50 años, lo que supone un aumento de 164 veces su valor inicial.

Ambos valores están dentro de los rangos que se muestran en Abdelmonem et al. (2020), en tuberías de hierro fundido se tiene un aumento de entre 20 y 40 veces; y las de acero entre 80 y 180 veces su valor inicial dependiendo del tipo de agua transportada.

2.3 Modelos de deterioro: estado del arte

Desde la segunda mitad del siglo XX el deterioro en tuberías ha sido ampliamente estudiado con la idea de predecir las fallas en redes de agua potable. En el transcurso de estos años una serie de modelos han sido desarrollados utilizando una amplia variedad de técnicas, las que van desde modelos físicos que determinan la falla mecánica en función de las cargas a las que están sometidos los ductos, hasta el uso de estadística avanzada o de modelos con inteligencia artificial.

Los modelos físicos apuntan a entender el comportamiento estructural de las tuberías enterradas, incorporando los mecanismos físicos que llevan a la rotura (Rajani & Kleiner, 2001). Estos mecanismos pueden ser:

- a) Propiedades estructurales de las tuberías, tipo de material, interacción suelo-tubería, instalación.
- b) Cargas de operación internas y externas: presión de operación, cargas por tráfico, cargas por congelamiento.
- c) Deterioro del material por el ambiente químico.

Los modelos de base física suelen tener como atributos de entrada las características de la tubería, propiedades del material, edad, condiciones de corrosión, factores ambientales (Nishiyama & Fillion, 2013; Rajani & Kleiner, 2001).

Si bien el tipo de datos requeridos es, en general, similar al que requieren los modelos estadísticos, esta clase de modelos requieren muchas veces datos difíciles de conseguir, además de que están orientados al comportamiento estructural de cada una de las unidades de la red, aumentando el requerimiento de datos a la vez que el tiempo de cómputo que necesitan los modelos. (Kleiner & Rajani, 2001; Nishiyama & Fillion, 2013).

Los modelos estadísticos son usados para predecir el tiempo de vida o el tiempo a la falla de la infraestructura (St. Clair & Sinha, 2012) y pueden ser clasificados en modelos deterministas y modelos probabilísticos (Kleiner & Rajani, 2001; Nishiyama & Fillion, 2013). La principal diferencia entre ambas clasificaciones es que los modelos probabilísticos manejan cierta aleatoriedad y probabilidades mientras que los deterministas asumen datos sin aleatoriedad (Kleiner & Rajani, 2001; Nishiyama & Fillion, 2013).

Por otro lado, el problema del deterioro en tuberías también ha sido abordado con técnicas más recientes tales como los algoritmos de inteligencia artificial. El concepto de *inteligencia artificial* (IA) surge de la posibilidad de hacer que las computadoras piensen por sí mismas. Dentro de las muchas definiciones que se le han dado, Kok et al. (2010) destaca que una de las más fuertes es la de “imitar el comportamiento humano inteligente”. En síntesis, un sistema de inteligencia artificial es un sistema que piensa (actúa) racionalmente.

Para que se genere esta imitación del “comportamiento humano inteligente” es necesaria la habilidad de “adquirir conocimiento”, la que se produce al extraer patrones de los datos.

El *machine learning* es una colección de técnicas que se usan para diseñar sistemas que aprenden de los datos. Aquí, un modelo es construido a partir de los datos de entrada, seguido de un aprendizaje y reconocimiento de patrones entre las entradas y las salidas del modelo (Dawood et al., 2019).

Este es un paradigma distinto al de la programación clásica, donde con un conjunto de reglas y datos generan respuestas. En los sistemas de machine learning los datos de entrada son el conjunto de datos y sus respuestas asociadas, de modo que el programa genere el conjunto de reglas que mejor se ajusten a los datos entregados.

Dentro de las áreas del machine learning se encuentra el *deep learning*, un tipo de aprendizaje que se enfoca en el aprendizaje a través de capas sucesivas, las que usualmente se estructuran en modelos llamados *redes neuronales*.

Los distintos algoritmos utilizados en machine learning se dividen en tres categorías principales:

- Aprendizaje supervisado, donde los datos utilizados para el entrenamiento están etiquetados.
- Aprendizaje no supervisado, donde los datos de entrenamiento no tienen etiquetas y lo que se busca es encontrar relaciones entre los datos entregados.
- Aprendizaje por refuerzo, donde el algoritmo aprende de su propia experiencia en un proceso de ensayo y error que recompensa las decisiones correctas.

De acuerdo a Dawood et al. (2020), el uso de algoritmos de inteligencia artificial tiene una serie de ventajas por sobre enfoques estadísticos dado que puede hacer frente a las imprecisiones, incertezas o ambigüedades de estos modelos. El mismo autor señala cinco ventajas de estos modelos por sobre los modelos estadísticos:

1. La precisión de los modelos estadísticos está limitada a los datos usados para crearlos.
2. Los modelos no son generales dado que aplican sólo a la región desde donde los datos fueron recolectados.
3. Los modelos estadísticos requieren un largo tiempo de observación para obtener la información necesaria para construirlos.
4. Tal como se señaló en la sección de factores, las fallas en las tuberías disminuyen con el diámetro, lo que implica que la disponibilidad de datos sobre roturas para tuberías de mayor tamaño es escasa, limitando la aplicabilidad de estos modelos.
5. En concordancia con lo anterior, debido a que los modelos son más bien locales y a que se necesita de un largo tiempo de observación para obtener los datos, la aplicabilidad de los modelos estadísticos se limita a redes antiguas.

Algunos de los enfoques usados por los algoritmos de inteligencia artificial son: Redes Neuronales Artificiales (ANN), Sistemas de Inferencia Difusos (FIS), Sistemas Neuro-Difusos (NFS), Algoritmos Genéticos, Máquina de Vector de Apoyo (SVM), etc. El objetivo de esta sección es presentar una revisión de algunos de los modelos de deterioro que

utilizan métodos de inteligencia artificial y señalar cuales de los algoritmos y variables de entrada presentan un mejor comportamiento.

En cuanto a los modelos de predicción de fallas, Kutylowska (2015) utiliza ANN para predecir la tasa de roturas en una ciudad polaca. El método empleado fue una ANN perceptrón multicapa, entrenada con el algoritmo *quasi – Newton*, obteniendo una correlación aceptable entre datos experimentales y predichos en la fase de aprendizaje ($R^2 = 0.951$ en conexiones domiciliarias y $R^2 = 0.9268$ en la red de distribución) y usando los siguientes parámetros de entrada: material, largo total, diámetro y año de instalación (para conexiones domiciliarias). Sin embargo, la correlación falla en la etapa de pronóstico. Para mejorar los resultados de la etapa de pronóstico, se sugiere agregar más parámetros de entrada al modelo. El autor concluye que las ANN son una buena herramienta para la predicción de la tasa de fallas debido a su corto tiempo de cómputo y a la rapidez con la que se pueden hacer cambios en la red, aunque juega un papel fundamental la disponibilidad de datos y la calidad de estos.

Bubtiena et al. (2011) también utiliza MLP – ANN pero el entrenamiento lo hace con un algoritmo de retro propagación, con la que modelan la rotura de las tuberías de la red de agua potable de Bengasi y con ello el tiempo óptimo de reparación. Los parámetros de entrada del modelo son: diámetro, largo, material, suelo, edad, profundidad, espesor, rugosidad, presión hidráulica y localización de la tubería. Se obtuvieron errores menores al 3% para la fase de entrenamiento y menores al 10% para la fase de testeo.

Por otro lado, Giraldo-González & Rodríguez (2020) realizan una comparación de distintos modelos estadísticos y de machine learning en el modelamiento de fallas de tuberías. En particular, para los modelos con machine learning, los algoritmos estudiados son el *Gradient Boosting Tree (GBT)*, *Support Vector Machine (SVM)*, *Artificial Neural Network (ANN)* y *Bayes*. Acorde con su análisis, el desempeño de los algoritmos depende de la métrica con la que se los evalúe. En particular para sus resultados, si se utiliza la *F-measure* (media armónica entre el número de predicciones positivas correctas y el número de eventos positivos predichos correctamente), GBT y SVM presentan un desempeño aceptable; si se usa la tasa de verdaderos positivos (TPR), Bayes y GBT tienen mejor desempeño y si se mide según la tasa de falsos positivos, SVM y GBT son los mejores. Pese a que Bayes presenta buenas métricas, tiene una alta tasa de falsos negativos, lo que se traduce en que este algoritmo es el que más falla al predecir que una tubería está en buena condición, lo que supone un aumento de los costos. Además, el estudio concluye que alguna de las variables más relevantes en la rotura son la precipitación, el número de fallas previas y el largo.

Winkler et al. (2018) estudian en profundidad la aplicación de métodos de *Boosting* con árboles de decisión para la predicción de fugas en redes de agua potable, los que son testeados en una red de más de 850 km en una ciudad austriaca. De los algoritmos estudiados, el RUSBoost (Random UnderSampling Boost) fue el que presentó un mejor desempeño debido a su baja tasa de falsos positivos (FPR). En este mismo estudio,

además de las predicciones de roturas, determinan una probabilidad de falla para cada elemento de la red y realizan una predicción de esta probabilidad a cinco y diez años. Sin embargo, los resultados obtenidos de este proceso indican que, contrario a lo esperado, la probabilidad de falla puede ir disminuyendo en el tiempo, fenómeno que los autores atribuyen a dos posibles razones. La primera es que, al aumentar la edad de la tubería, los predictores quedan fuera del dominio de entrenamiento del modelo, forzando la extrapolación de predicciones. La segunda es que en los procesos de instalación de las tuberías se pueden producir daños que aumentan el número de roturas en edades tempranas, las que luego disminuyen durante la vida útil y vuelven a aumentar al final de su periodo de uso (lo que se conoce como “diagrama de tina de baño”).

3 Metodología

De acuerdo con lo estudiado en la sección 2.3 se opta por trabajar con algoritmos de inteligencia artificial por sobre modelos probabilísticos sean estos físicos o estadísticos. Esta decisión se basa en dos aspectos fundamentales. El primero es la disponibilidad y manejo de los datos. Los modelos físicos tienen un alto requerimiento de datos y de parámetros que son difíciles y costosos de obtener para la empresa sanitaria, además de tener un alto requerimiento computacional al tener que calcular cada tubería de la red (Kleiner & Rajani, 2001; Nishiyama & Filion, 2013). En el caso de los modelos estadísticos, si bien el requerimiento de datos es menor, se debe tener extrema precaución en el tratamiento de los datos, realizando una cuidadosa *clusterización* de estos, de modo de obtener grupos de tuberías lo más homogéneos posibles a la hora de modelar.

El segundo aspecto tiene que ver con la generalidad del modelo. El tipo de programación de los modelos con algoritmos de inteligencia artificial permite la construcción de un único código independiente del número y tipo de variables explicativas que se quiera ingresar, permitiendo un programa más general aplicable a cualquier red siempre y cuando se ingrese una base de datos debidamente revisada.

En la presente sección se presenta la metodología elegida para los modelos con algoritmos de inteligencia artificial y para el modelo hidráulico.

3.1 Modelo con algoritmos de Inteligencia Artificial

Dentro de los modelos revisados en la sección 2.3, algunos de los algoritmos que presentan un mejor desempeño en la predicción de fallas son las redes neuronales, las máquinas de vector de apoyo, los sistemas neuro-difusos y los árboles con *gradient boosting* (Dawood, Elwakil, Novoa, & Gárate Delgado, 2020; Giraldo-González & Rodríguez, 2020).

Según lo anterior, se define para el desarrollo de este trabajo la programación de tres algoritmos distintos dentro de los antes mencionados: *Support Vector Machine (SVM)*, *Artificial Neural Network – Multi Layer Perceptron (ANN – MLP)* y *Gradient Boosting Trees (GBT)*.

Se define, además, la predicción de fugas como un problema de clasificación binaria de aprendizaje supervisado, donde la clase positiva (1) es la condición de *falla* y la negativa (0) la de *no falla*. Los tres algoritmos señalados son comparados para determinar cuál de ellos presenta el mejor desempeño para la base de datos estudiada.

En las subsecciones siguientes se detalla el proceso de pretratamiento de los datos y se hace una descripción sencilla del funcionamiento de cada uno de los algoritmos. Además, se describen las métricas estudiadas para la comparación de los métodos.

La programación de todos los algoritmos y del preprocesado de datos se hace en Python utilizando la librería *Scikit-learn* (Pedregosa et al., 2011).

3.1.1 Preprocesamiento de la base de datos

Como se ha señalado en secciones anteriores, los modelos de Machine Learning son algoritmos que aprenden de los datos, por lo que el proceso de pretratamiento es fundamental para garantizar que la base no presente errores y esté en el formato apropiado.

La metodología de construcción y tratamiento de la base de datos se modifica a partir de Winkler et al. (2018) y se detalla a continuación.

3.1.1.1 Construcción de la base de datos

La base de los datos usados para construir el modelo es la red completa en su estado actual, es decir, se tiene una matriz donde cada fila corresponde a una tubería del sistema bajo el supuesto de que están todas operativas (no hay fallas en la red). Cada una de las filas de esta matriz se describe según el par (x_p, y_p) , donde x_p es el conjunto de características del elemento p del sistema e y_p es la condición de dicho elemento. Como se mencionó previamente, en estos modelos se considera que y_p toma un valor de cero cuando la tubería está en buena condición y un valor de uno cuando falla.

Dado que con esto sólo se tienen datos de tuberías en buen estado, para generar el aprendizaje de los modelos se deben agregar los registros históricos de falla. Por lo tanto, para cada tubería x_p de la red que haya fallado según el historial, se agrega una fila (x_{f_i}, y_f) , donde i corresponde a cada una de las fallas que ha tenido el mismo tubo. A modo de ejemplo, si una tubería x_p ha fallado cinco veces en su historia, se agregan cinco filas (x_{f_i}, y_f) que representan cada una de estas fallas y donde $y_f = 1$.

Las características asociadas a cada vector x_{f_i} están relacionadas al vector x_p al que hacen referencia. Las columnas asociadas a la ubicación se mantienen (no hay variación en la ubicación con o sin fallas) y las características físicas serán las mismas a menos que haya un reemplazo de tubería, donde estas se deben sustituir por aquellas de la nueva tubería.

Además de las características físicas y geográficas de cada fila, se agrega una columna con el número total de roturas previas, el que parte en cero si la tubería nunca se ha roto y suma uno por cada falla. Se agrega una columna con el número de fallas desde el último reemplazo, el cual crece igual que el anterior, pero al momento del reemplazo se devuelve a cero.

La edad de la tubería se incorpora como la diferencia entre el año de falla y el año de instalación. Para las filas de la línea base se considera la diferencia entre el año de instalación y el año actual.

3.1.1.2 Curación de los datos

Previo a la modelación se debe limpiar la base de datos, dejando sólo las columnas que son relevantes en la predicción de fallas, es decir, las variables explicativas del modelo. Ejemplos de estas variables pueden ser diámetro, longitud, edad, tipo de suelo,

profundidad, pH del agua transportada, fallas previas, etc. Se deben remover todas las columnas asociadas a algún tipo de ubicación (calles, coordenadas, comunas, etc.).

Las columnas escogidas para la modelación son separadas en variables categóricas (ejemplo: material) y variables numéricas (ejemplo: diámetro).

Las variables categóricas son transformadas en un set de valores booleanos. Por ejemplo, se tiene una red con dos materiales (PVC y asbesto cemento). Luego, la columna “Material” será transformada en una “Material: PVC” y otra “Material: Asbesto Cemento”. Si la tubería es de PVC, la transformación queda según se muestra en la Figura 3.1:

Tubería	Diámetro [mm]	Largo [m]	Material
51373	75	7.5	PVC



Tubería	Diámetro [mm]	Largo [m]	Material: PVC	Material: Asbesto Cemento
51373	75	7.5	1	0

Figura 3.1: Transformado de variables categóricas.

Por otro lado, dependiendo del algoritmo utilizado las variables numéricas deben ser escaladas para prevenir que las variables que se miden en escalas más grandes tomen un rol demasiado preponderante en el modelo. En el desarrollo del presente trabajo se decide usar el StandardScaler provisto por Scikit-learn, donde la media se centra en cero y la varianza en uno. Los datos se escalan según la siguiente expresión:

$$X_{escalado} = \frac{X - \mu_X}{\sigma_X}$$

3.1.1.3 Remuestreo de los datos en muestras desbalanceadas

Se dice que una muestra está desbalanceada cuando hay una desproporción en el número de datos asociados a cada clase dentro de la muestra. A partir de ello, se puede definir una “clase minoritaria” y una “clase mayoritaria” (clasificación binaria).

Como ejemplo, sea una muestra donde la clase mayoritaria posee el 95% de los datos y la minoritaria el 5% restante. En este caso el modelo construido podría indicar siempre que el resultado pertenece a la clase mayoritaria obteniendo en el global un 95% de acierto pero con un 100% de error en la predicción de la clase minoritaria.

En caso de que la base de datos construida presente esta situación, se deben aplicar medidas que apunten a corregirlo, de modo de conseguir un modelo que sea capaz de detectar correctamente la clase minoritaria.

Existen dos enfoques principales para abordar el problema del desbalanceo (Arnejo, 2017; Miravet, 2021), uno es el refinamiento de los algoritmos para mejorar la predicción y el otro es el uso de técnicas de remuestreo.

En particular, las técnicas de remuestreo consisten en modificar la distribución de los datos a modo de equilibrar la muestra, objetivo que se puede conseguir reduciendo las instancias

de la clase mayoritaria al mismo número de la clase minoritaria (*Undersampling*) o replicando la clase minoritaria (*Oversampling*).

Para el caso estudiado, se decide hacer un *Random Undersampling* (Figura 3.2), donde se eligen aleatoriamente tantas instancias de la clase mayoritaria como las que tenga la clase minoritaria. Sin embargo, este procedimiento podría suponer la pérdida de información relevante para el entrenamiento del modelo y afectar en su desempeño.

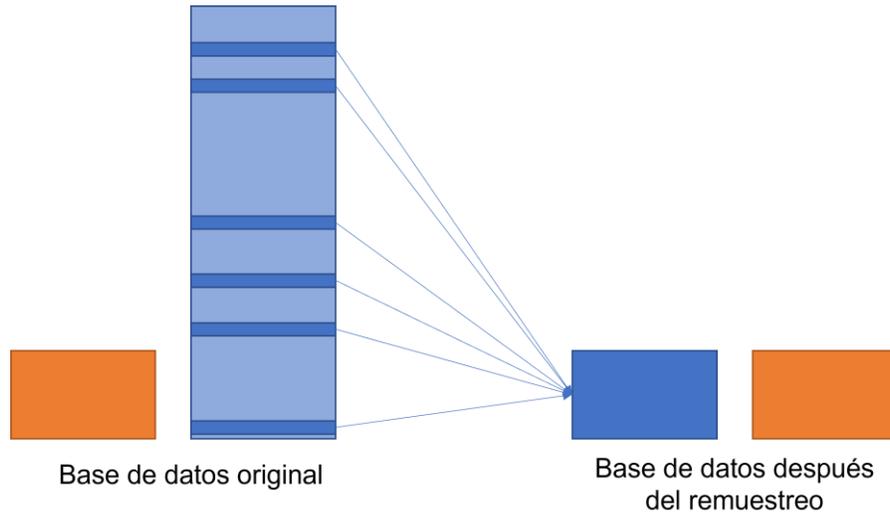


Figura 3.2: Random Undersampling.

3.1.1.4 Construcción de los sets de entrenamiento y testeo

Previo al modelado se definen los sets de entrenamiento y testeo. Se usa el 75% de los datos para entrenar el modelo y el 25% para testearlo. La división de los datos se estratifica según la condición de falla para prevenir que los sets estén compuestos sólo de tuberías en buen estado de haber una base de datos desbalanceada.

El resultado de este proceso son cuatro sets: las variables de entrenamiento (X_{train}) con sus resultados (y_{train}) y las variables de testeo (X_{test}) con sus resultados (y_{test}).

3.1.1.5 Selección de hiperparámetros del modelo

Cada algoritmo tiene asociado una serie de hiperparámetros que finalmente son los que determinan la capacidad predictiva de cada modelo. Algunos ejemplos de estos hiperparámetros son, en una red neuronal, el número de capas ocultas o el número de neuronas en estas capas o el kernel en una máquina de vector de apoyo.

Para cada algoritmo estudiado se define una grilla de hiperparámetros y sus posibles valores. El proceso de selección consiste en el ajuste de dicho algoritmo con los sets de entrenamiento definidos previamente y cada una de las combinaciones de hiperparámetros de esta grilla. Cada ajuste de los datos se hace con un proceso de validación cruzada, la cual consiste en dividir el set de entrenamiento en k subconjuntos y entrenarlo con $k - 1$. El conjunto restante se usa para evaluar el modelo entrenado. El proceso se repite k veces, y en cada ocasión se deja un subconjunto distinto para la evaluación. El valor final de la

métrica objetivo se calcula como el promedio del valor de esta para cada uno de los modelos intermedios. Lo descrito se repite n veces.

Cabe señalar que el proceso de selección de hiperparámetros y ajuste del modelo es el que supone la mayor parte del tiempo de cómputo, el que está determinado en gran medida por los valores elegidos de k y n .

El procedimiento, en general, se muestra en la Figura 3.3:

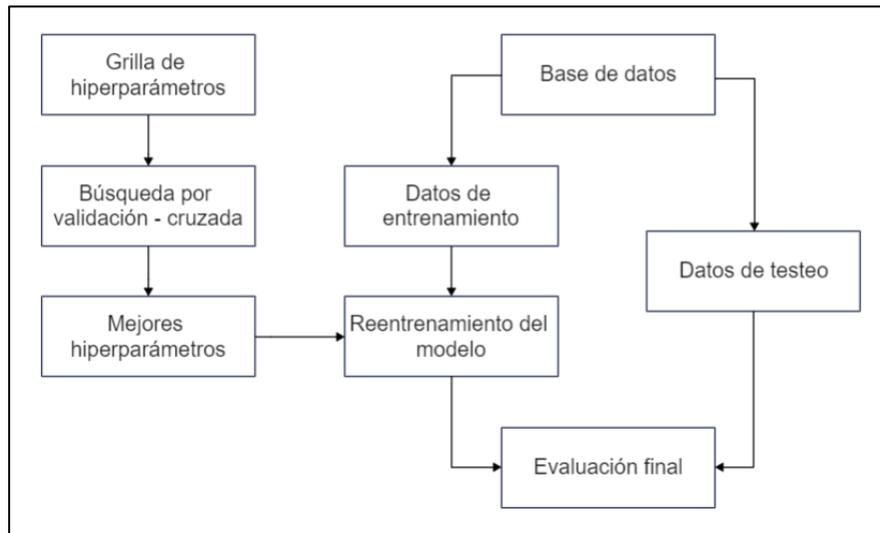


Figura 3.3: Diagrama de flujo optimización de hiperparámetros. Modificado de la documentación de Scikit-learn.

3.1.2 Modelo de deterioro con Gradient Boosting

Gradient Boosting es uno de los métodos de *ensemble* (estrategias de combinación de modelos) que existen dentro del *machine learning*. Este consiste en construir una regresión aditiva ajustando secuencialmente una función simple (*base learner*) a los pseudo-residuales actuales, los que corresponden al gradiente de la función de pérdida que está siendo minimizada (Friedman, 2002).

En este caso, el *base learner* son árboles de decisión con pocas ramificaciones, donde cada árbol mejorará la desviación del árbol anterior. Un ejemplo de la variación de la desviación con el progreso de las iteraciones se muestra en la Figura 3.4.

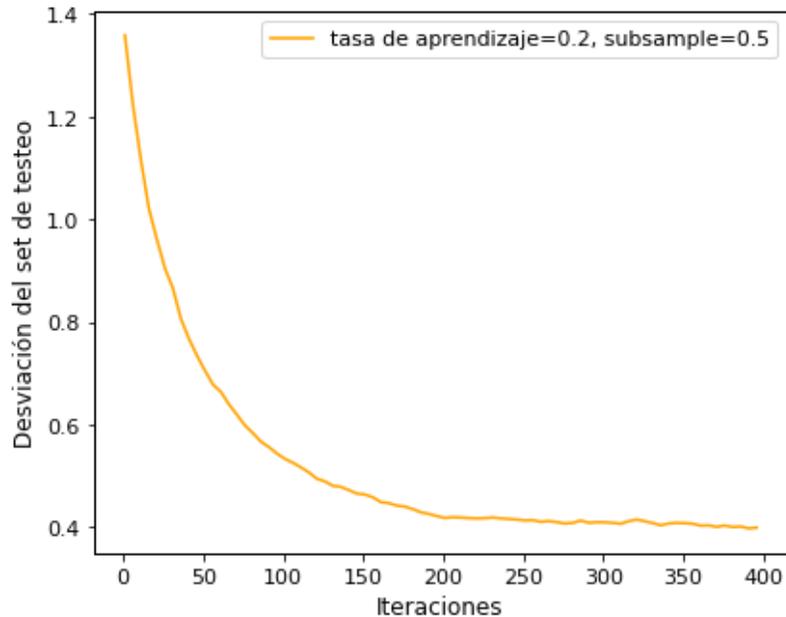


Figura 3.4: Variación de la desviación del set de testeo con el número de iteraciones de boosting. Modificado de la documentación de Scikit-learn.

Tres de los hiperparámetros más relevantes en este algoritmo son:

- Número de *base learners*: cantidad de iteraciones utilizadas en el *boosting*
- Tasa de aprendizaje (*learning rate*): determina la influencia de cada árbol en el modelo. A menor tasa de aprendizaje, mayor número de árboles necesarios y menor riesgo de *overfitting*.
- Máxima profundidad de los árboles

Previo al modelado se hace el preprocesamiento señalado en 3.1.1 (para este algoritmo no necesario realizar el escalado de variables). En el proceso de modelado se definen los hiperparámetros a encontrar en la grilla, en este caso se consideran el número de estimadores, la cantidad máxima de características a considerar para hacer las divisiones en los árboles, la tasa de aprendizaje y la fracción de la muestra usada para ajustar cada *base learner* (*subsample*). Algunos de los posibles valores de estos parámetros se definen en función de otros estudios que hayan usado esta metodología.

3.1.3 Modelo de deterioro con Perceptrón Multicapa

Las redes neuronales son modelos de aprendizaje estructurados en capas y su nombre es una referencia a la neurobiología.

Estos modelos tienen una arquitectura particular en la que se distinguen tres tipos de capas:

- *Input layer* o capa de entrada: es la que recibe los datos de entrada al modelo.
- *Hidden layers* o capas intermedias: reciben información desde la capa de entrada o desde otras capas intermedias.

- *Output layer* o capa de salida: combina la información entregada por la última capa intermedia para generar la predicción.

Cada una de las capas que conforman la estructura está compuesta por neuronas o unidades, las que se conectan con las unidades de la capa anterior y de la capa siguiente. Estas conexiones están reguladas por valores conocidos como “pesos”. Un esquema de la estructura de una red neuronal tipo se muestra en la Figura 3.5.

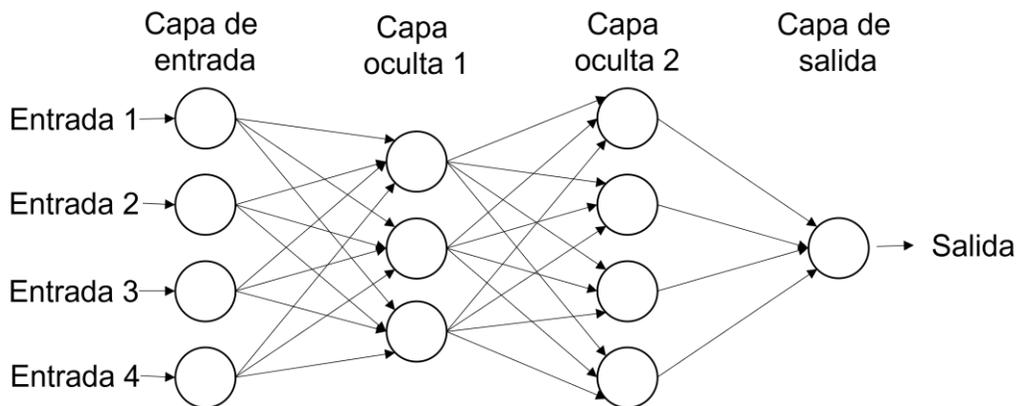


Figura 3.5: Arquitectura de una red neuronal.

Dentro de las unidades se calcula el valor de entrada como la suma ponderada de los *inputs* y sus pesos. Posteriormente, dicho valor es ingresado a una función de activación. El resultado de este proceso es el valor de salida de la neurona.

Previo al modelado se hace el preprocesamiento señalado en 3.1.1 (para este algoritmo se debe considerar el escalado de las variables). En el proceso de modelado se definen los hiperparámetros a encontrar en la grilla, en este caso se consideran el número de capas/neuronas ocultas, la función de activación, la tasa de aprendizaje inicial y la regularización. Algunos de los valores de estos parámetros se definen en función de otros estudios que hayan usado esta metodología.

3.1.4 Modelo de deterioro con Máquina de Vector de Apoyo

Una Máquina de Vector de Apoyo (SVM por sus siglas en inglés) es un algoritmo de inteligencia artificial que tiene por objetivo clasificar objetos en dos o más clases con el menor error posible. Para ello, cada objeto de la muestra se representa como un vector de n dimensiones, $x = \{x_1, x_2, \dots, x_n\}$, con una etiqueta “ y ” que define la clase a la que pertenece el objeto.

Un hiperplano, en el caso del espacio de n dimensiones, es un conjunto de vectores en \mathbb{R}^{n-1} definidos según la ecuación:

$$\langle w, x \rangle + b = 0$$

dividiendo el espacio en dos partes, una a cada lado de este. Si se tienen dos clases de vectores, el margen se define como la suma de la mínima distancia perpendicular entre

cada clase y el hiperplano (Figura 3.6). El objetivo de la SVM es encontrar un hiperplano capaz de separar las clases del problema obteniendo el mayor margen, donde los puntos de cada clase que recaen en los bordes del margen se conocen como vectores de apoyo.

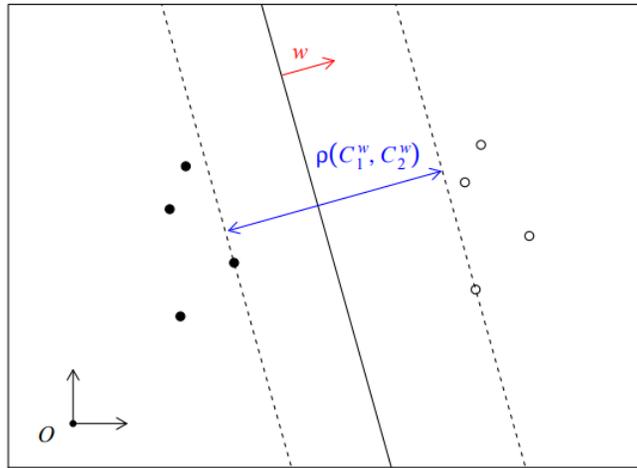


Figura 3.6: Ejemplo de hiperplano de separación y margen. Extraído de Nefedov (2016).

El *kernel* en las SVM es una función usada en la función de decisión en lugar del producto interno. Esta función es particularmente útil cuando dos clases no son linealmente separables, ya que transforma linealmente el espacio a uno donde las clases son separables (Figura 3.7).

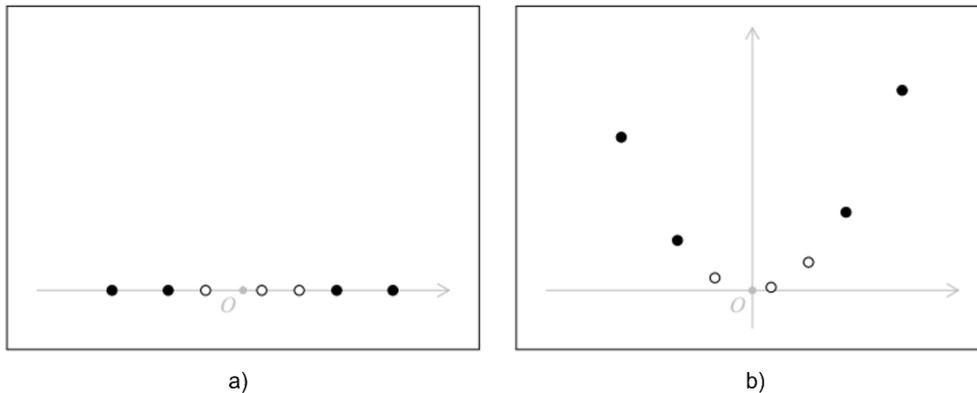


Figura 3.7: Ejemplo función kernel. Extraído de Nefedov (2016).

Previo al modelado se hace el preprocesamiento señalado en 3.1.1 (para este algoritmo se hace necesario el escalado de variables). En el proceso de modelado se definen los hiperparámetros a encontrar en la grilla, en este caso se consideran la función del *kernel* y los parámetros C y γ , donde el primero controla el equilibrio entre la uniformidad de los límites y la clasificación correcta de los puntos, mientras que el segundo define hasta donde llega la influencia de un solo punto (Lee, 2019). Algunos de los posibles valores de estos parámetros se definen en función de otros estudios que hayan usado esta metodología.

3.1.5 Métricas para el análisis de modelos

Para definir el algoritmo que presenta un mejor desempeño predictivo, se construyen las matrices de confusión, las cuales tienen la forma mostrada en la Figura 3.8:

		Roturas predichas	
		Sin falla (0)	Falla (1)
Roturas observadas	Sin falla (0)	TN	FP
	Falla (1)	FN	TP

Figura 3.8: Matriz de confusión tipo.

Donde:

- TN (verdadero negativo): es el número de veces que el modelo predijo correctamente la salida como negativa. En este caso, es la cantidad de veces que el modelo predijo correctamente que una tubería no estaba rota.
- FN (falso negativo): es el número de veces que el modelo predijo incorrectamente la salida como negativa. En este caso, es la cantidad de veces que el modelo predijo que una tubería NO estaba rota cuando si lo estaba.
- FP (falso positivo): es el número de veces que el modelo predijo incorrectamente la salida como positiva. En este caso, es la cantidad de veces que el modelo predijo que una tubería estaría rota cuando no lo estaba.
- TP (verdadero positivo): es el número de veces que el modelo predijo correctamente la salida como negativa. En este caso, es la cantidad de veces que el modelo predijo correctamente que una tubería estaba rota.

Con los valores de la matriz de confusión se calculan las siguientes métricas:

- *Accuracy*: es la suma del total de predicciones correctas sobre el total de predicciones realizadas.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- *Precision*: relativa al número de predicciones positivas correctas. Se calcula como:

$$Precision = \frac{TP}{TP + FP}$$

- *Recall* o *True Positive Rate (TPR)*: relativa al número de eventos positivos predichos correctamente. Se calcula como:

$$TPR = \frac{TP}{TP + FN}$$

- *True Negative Rate (TNR)*: relativa al número de eventos negativos predichos correctamente. Se calcula como:

$$TNR = \frac{TN}{TN + FP}$$

- *F1 Score*: compara el comportamiento del modelo a nivel de *Precision* y *TPR*. Se calcula como:

$$F1\ Score = \frac{2 \cdot Precision \cdot TPR}{Precision + TPR}$$

- *False Positive Rate (FPR)*: corresponde a la proporción de datos negativos que son tomados como positivos con respecto a todos los datos negativos. Se calcula como:

$$FPR = \frac{FP}{FP + TN}$$

- *False Negative Rate (FNR)*: corresponde a la proporción de datos positivos que son tomados como negativos con respecto a todos los datos positivos. Se calcula como:

$$FNR = \frac{FN}{TP + FN}$$

- *Balanced Accuracy*: métrica usada en bases de datos desbalanceadas que consiste en el promedio de la TPR con la TNR y suele ser usada cuando se trabaja con muestras desbalanceadas.

$$Balanced\ Accuracy = \frac{TPR + TNR}{2}$$

Además de las métricas señaladas, se construye la curva de *Receiver Operating Characteristic* (curva ROC), que es un gráfico bidimensional donde en el eje de las ordenadas se traza la TPR y en el eje de las abscisas la FPR (Debón et al., 2010). Para determinar los distintos valores de la TPR y de la FPR, se cambia el umbral de probabilidad que define si una predicción pertenece a una clase u otra. Es decir, si originalmente una predicción era clasificada como “falla” cuando la probabilidad calculada por el modelo para pertenecer a esa clase estaba por sobre el 50%, ahora se genera una lista de umbrales entre 0 y 100% y para cada umbral se determina una TPR y una FPR.

En la Figura 3.9 se muestra un ejemplo de curva ROC. Mientras más cercana esté la curva del modelo (curva gris claro) a la curva del modelo perfecto, mejor será la capacidad predictiva. Por otro lado, mientras más cercana esté la curva del modelo a la recta no predictiva (gris oscuro) peor será la capacidad predictiva. Este análisis se traduce en que, a mayor área bajo la curva (AUC, por sus siglas en inglés), se tiene a priori un mejor modelo.

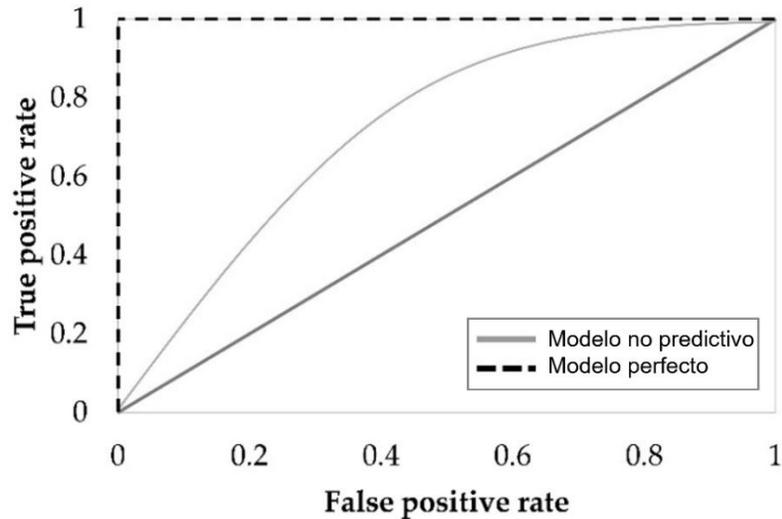


Figura 3.9: Ejemplo de curva ROC.
Modificado de Giraldo-González & Rodríguez, 2020.

3.1.6 Caso de estudio

Las metodologías descritas previamente se aplican a un caso de estudio. Los datos de falla de la red de distribución de agua potable de una ciudad grande de Chile fueron solicitados por el sistema de transparencia a la Superintendencia de Servicios Sanitarios, institución desde la cual se entregaron los reportes de indicadores de calidad del servicio (PR013) que contiene los eventos de discontinuidad para el periodo 2017-2021. Junto a esta información se hizo entrega del archivo shape de la red de distribución de agua potable que la empresa entrega a la SISS anualmente en su reporte de infraestructura (PR012).

Debido a que los registros de discontinuidad señalados no contienen ninguna referencia a qué tubería de la red es la dañada y sólo contienen coordenadas del sitio, se programa un algoritmo que asocia cada falla a una tubería de la red, buscando la menor distancia ortogonal entre la falla y las tuberías del mismo diámetro ubicadas dentro de un radio de búsqueda proporcional a la longitud de la tubería más larga de la red. Como se muestra en la Figura 3.10, si bien hay eventos que son fácilmente asignables a una tubería (sector izquierdo de la figura), también existen eventos que no son fácilmente atribuibles a una u otra, como el del sector derecho de la imagen, donde se señala con una flecha color naranja la tubería a la cual finalmente corresponde la falla.



Figura 3.10: Ejemplo de la asociación de fallas.

En cuanto a la construcción de las columnas asociadas a fallas previas en la base de datos, dado que los eventos de discontinuidad no señalan el tipo de trabajo que se hizo en la tubería afectada (reparación o reemplazo), sólo se usa una columna de fallas previas, sin considerar reemplazos. Cabe mencionar que la red presenta inconsistencias en cuanto a la materialidad de las tuberías, señalando un gran número de instalaciones de tuberías de asbesto cemento posterior a 2001, año en que fue prohibida su instalación. Además, según el archivo shape recibido, el mayor número de instalaciones de tuberías de PVC se produjo en 1957, siendo que este es un material de uso mucho más reciente.

Tomando en consideración los datos con los que se cuenta, sólo se consideran las siguientes variables explicativas: diámetro, longitud, material, edad y el número de fallas previas.

3.2 Incorporación del deterioro dentro de un modelo hidráulico

En esta sección se detalla la metodología para integrar los algoritmos descritos en la sección 3.1 en el modelamiento hidráulico de la red.

3.2.1 Paquete WNTR y redes experimentales

La modelación hidráulica de la red se realiza utilizando el paquete *Water Network Tool for Resilience* (WNTR), el cual fue desarrollado por la Agencia de Protección Ambiental de los Estados Unidos, EPA, para simular y analizar la resiliencia de los sistemas de distribución de agua potable.

Este módulo permite realizar simulaciones hidráulicas similares a las realizadas en EPANET, pero además incluye una serie de características útiles en el estudio de amenazas, como lo son la adición de eventos disruptivos, estrategias de reparación, simulaciones probabilísticas a través de curvas de fragilidad o hacer análisis basado en presiones.

El análisis basado en presiones considera que el caudal final de salida en un nodo será una fracción de la demanda, la que depende de la presión en dicho nodo. Para ello se definen dos umbrales de funcionamiento: uno de presión mínima y otro de presión requerida. La presión mínima es aquella mínima necesaria para que haya una salida de caudal en dicho nodo, es decir, bajo esa presión no hay flujo. Por otro lado, la presión requerida es aquella a partir de la cual se satisface el 100% de la demanda. Para cualquier presión entre la mínima y la requerida se satisface una fracción de la demanda (Figura 3.11). En este trabajo se usa como presión mínima de funcionamiento el límite normativo en caso de incendio (5 m.c.a.) y como presión requerida el mínimo en condiciones normales (15 m.c.a.).

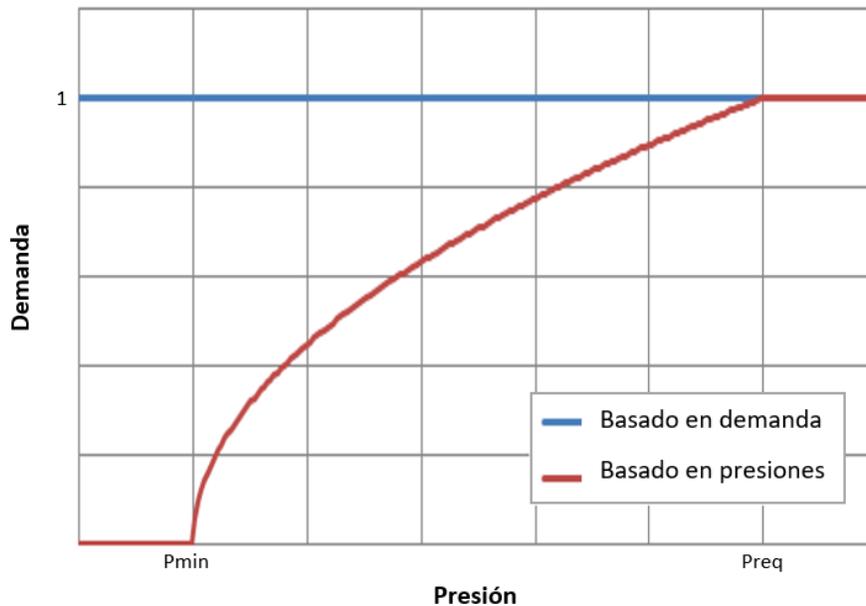


Figura 3.11: Modelo basado en presiones. Modificado de la documentación de WNTR.

Para la aplicación de la metodología se construye una red experimental cuadrada, con 64 nodos que están espaciados cada 50 metros. La red cuenta con tres fuentes y, para el análisis, se hace el supuesto de que las tuberías que alimentan la red desde estas fuentes no fallan.

Debido al tipo de problema que se busca simular, a cada tubería se le asocia aleatoriamente una edad, un material y un diámetro distinto, los cuales son proporcionales a los datos presentes en el caso de estudio.

3.2.2 Incorporación de los modelos de deterioro en el análisis hidráulico

Debido a que el análisis se está haciendo sobre una red experimental que no cuenta con ningún tipo de registro, se construye un historial de fallas sobre el cual aplicar los modelos de deterioro. Para ello, se toman los parámetros obtenidos en Giraldo-González & Rodríguez (2020) para un modelo de regresión de Poisson, con el que se calcula una probabilidad de falla para cada tubería y, según la edad asignada previamente, se estima el número de fallas. La probabilidad de falla según el modelo de regresión de Poisson es:

$$P(k_{i,t}) = \lambda_{i,t}^{k_{i,t}} \cdot \frac{\exp(-\lambda_{i,t})}{k_{i,t}!}$$

donde $P(k_{i,t})$ es la probabilidad de tener k fallas en la tubería i durante el año t y $\lambda_{i,t} = \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_j X_j)$, con X_j las variables independientes y β_j los coeficientes de la Tabla 3.1:

Tabla 3.1: Parámetros regresión de Poisson. Extraídos de Giraldo-González & Rodríguez (2020).

	$\beta_{diámetro}$	β_{largo}	β_{edad}	β_0
Asbesto cemento	-0,074	0,034	-0,001	4,466
PVC	-0,009	0,002	-0,001	5,810

Como ejemplo, sea una tubería de t años. Para cada uno de esos años se estima una probabilidad de falla y con dicha probabilidad se estima si aquella tubería falla en el año en cuestión. En caso de falla, se siguen los procedimientos señalados en la sección 3.1.1.1. El diagrama de flujo para la creación del historial se muestra en la Figura 3.12:

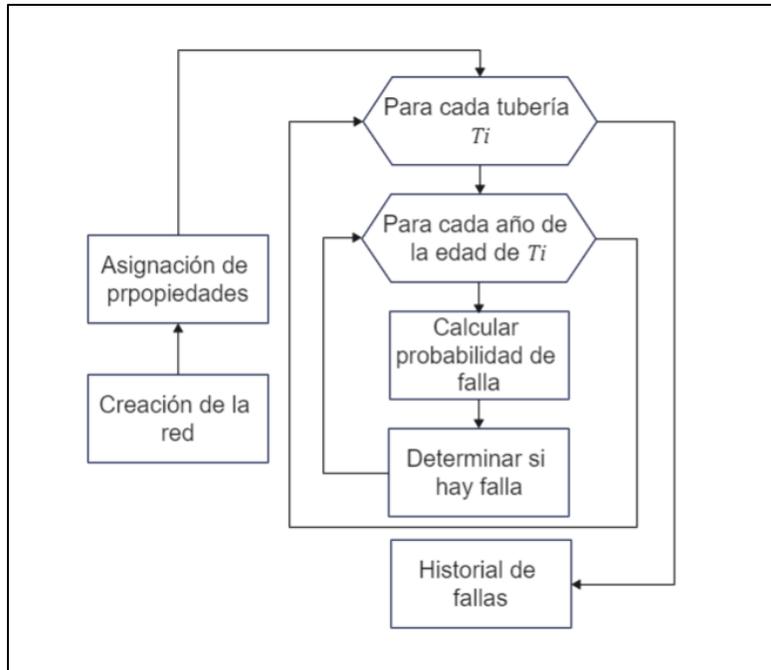


Figura 3.12: Diagrama de flujo para la creación de un historial de fallas.

Con el historial creado, se generan los modelos utilizando los algoritmos detallados en la sección 3.1 y se extraen las métricas asociadas a cada uno de ellos. Se elige aquel que presente un mejor desempeño de acuerdo con las métricas mostradas en la misma sección.

3.2.3 Incorporación de cambios físicos de la red

Para la modelación hidráulica se definen dos cambios físicos principales dentro del funcionamiento de la red. El primero son las fugas propias producto del deterioro y el segundo es el cambio en la rugosidad de las tuberías producto del envejecimiento. Ambos se detallan a continuación.

3.2.3.1 Adición de fugas

El paquete WNTR propone una metodología para añadir fugas, la que consiste en el uso de nodos auxiliares, donde en el lugar de la fuga se agrega un nodo (esquema b, Figura 3.13) con una pérdida de caudal que se estima según:

$$d_{leak} = C_d A \sqrt{2gh} \quad (1)$$

donde C_d es el coeficiente de descarga, A es el área de fuga en m^2 , g es la aceleración de gravedad en m/s y h es la carga hidráulica en metros.

Sin embargo, esta metodología supone problemas para el cálculo de métricas de desempeño de la red ya que añade nodos y tuberías extra, además de dificultar la comparación entre los distintos escenarios analizados. Una alternativa es distribuir la fuga en los nodos adyacentes (esquema b, Figura 3.13), asignando a cada uno de estos la mitad del área total.

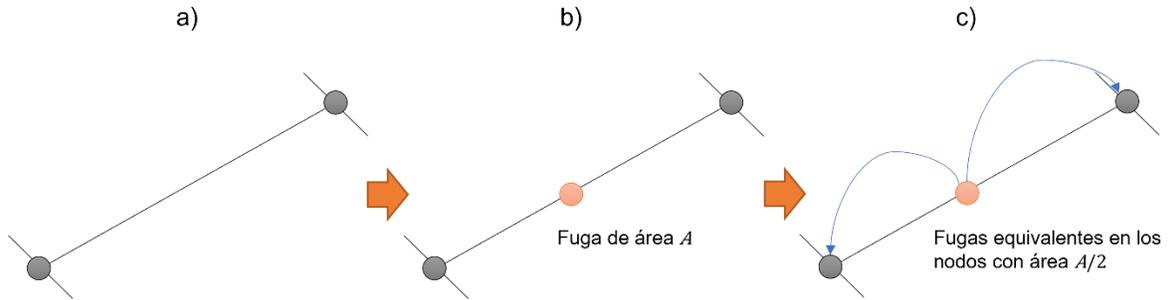


Figura 3.13: Adición de fugas. a) Tubería sin fugas, b) Metodología WNTR, c) Metodología usada.

Dado que los estudios sobre la predicción del tamaño (área) de fuga son limitados, se asume un área de fuga equivalente al 30% de la sección transversal de la tubería en cuestión.

3.2.3.2 Incorporación de cambios en la rugosidad de las tuberías

Como en la red experimental sólo se consideran materiales no metálicos, los cambios de la rugosidad se considera lo propuesto por Abdelmonem et al. (2020).

La fórmula general definida en este estudio para el cambio de la rugosidad es:

$$\frac{e_t}{e_0} = C_1 \left(\frac{t}{T}\right)^2 + C_2 \left(\frac{t}{T}\right) + C_3 \quad (2)$$

donde e_t es la rugosidad en el tiempo t , e_0 es la rugosidad inicial de la red, t es el tiempo de evaluación, T es el total de tiempo del estudio (50 años en este caso) y C_i son coeficientes que dependen del material y del tipo de agua que se esté evaluando. El comportamiento de la rugosidad en el tiempo se muestra en la Figura 3.14:

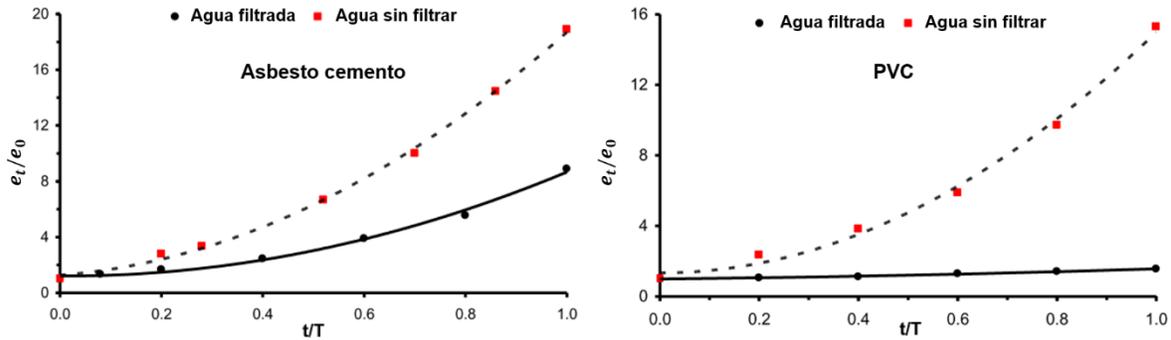


Figura 3.14: Variación de la rugosidad en el tiempo para tuberías de asbesto cemento y PVC. Modificado de Abdelmonem et al. (2020).

Como para crear el historial de fallas de la red modelo se consideran los parámetros de la regresión de Poisson presentes en Giraldo-González & Rodríguez (2020). Los materiales que componen la red se limitan a los que están presentes en este estudio, es decir, PVC y asbesto cemento (AC). En ambos casos se supone que el agua transportada ha sido filtrada. Los coeficientes asociados a estos materiales se presentan en la Tabla 3.2:

Tabla 3.2: Coeficientes de regresión para asbesto cemento (AC) y PVC. Extraído de Abdelmonem et al. (2020).

Material	C_1	C_2	C_3	r^2
AC	7,627	-0,171	1,211	0,994
PVC	0,302	0,274	0,993	0,991

Los cambios de diámetro se estiman con la rugosidad obtenida a partir de:

$$D_t = D_0 - 2e_t \quad (3)$$

Sin embargo, la ecuación (2) permite estimar las variaciones de la altura rugosidad en el tiempo, pero los softwares de modelamiento hidráulico como EPANET o el paquete WNTR realizan los cálculos de pérdidas de carga en redes con la ecuación de Hazen – Williams. Es por ello que se debe hacer una transformación para llegar al factor necesario.

El vínculo entre el factor C de Hazen – Williams y la altura rugosidad e se hace siguiendo la metodología expuesta en Sharp & Walski (1988) para este mismo propósito.

La ecuación de Hazen – Williams se define según:

$$V = 0,55 \cdot CD^{0,63}S^{0,54} \quad (4)$$

donde V es la velocidad en fps, D es el diámetro en pies y S es el gradiente hidráulico.

Resolviendo la ecuación de Darcy – Weisbach para la velocidad se tiene:

$$V = \left(\frac{2gSD}{f} \right)^{0,5} \quad (5)$$

Combinando las ecuaciones (4) y (5), se obtiene una expresión donde el factor C es función del factor f :

$$C = \frac{17,25}{f^{0,54}(VD)^{0,081}} \quad (6)$$

Finalmente, como la ecuación de Colebrook – White no tiene una solución explícita para f en régimen liso ni en régimen de transición liso-rugosa, se usa la ecuación de Swamee & Jain:

$$f = \frac{0,25}{\left(\log_{10} \left[\left(\frac{0,27e}{D} \right) + \left(\frac{5,74}{N^{0,9}} \right) \right] \right)^2} \quad (7)$$

donde N es el número de Reynolds. Así, combinando las ecuaciones (6) y (7) se determina el valor del factor C en función de la altura rugosidad e .

3.2.4 Incorporación de amenazas naturales

La incorporación de amenazas naturales al análisis implica necesariamente la aplicación de modelos multiamenazas, es decir, considerar el deterioro y una o más amenazas naturales. Para ello se considera lo propuesto por Abarca (2022), donde se realiza un análisis multiamenaza que incluye un modelo de amenaza sísmica, los efectos del deterioro (segunda amenaza) y de los incendios post sismo. Para este trabajo sólo se implementan las amenazas por sismo y deterioro. A continuación se describe a grandes rasgos la metodología propuesta.

El modelo de amenaza sísmica se basa en un análisis probabilístico de amenaza sísmica (PSHA) y busca cuantificar la incertidumbre de la ubicación, magnitud e intensidad de movimiento.

La metodología consiste en determinar la curva de amenaza, que relaciona medidas de intensidad con tasas anuales de excedencia, para luego asociar dichas magnitudes a un nivel de daño, obtenido de una curva de fragilidad.

Para la simulación se consideran los mismos parámetros presentados en Abarca (2022), donde la ciudad de análisis corresponde a Valparaíso y se utilizan todos los sismos con magnitudes mayores a M_w 6,5 ocurridos desde 1950 en un radio de 200 km. Los parámetros históricos obtenidos de este análisis se muestran en la Tabla 3.3.

Tabla 3.3: Parámetros históricos de sismos en Valparaíso. Extraído de Abarca (2022).

$M_{mín}$ [-]	$M_{máx}$ [-]	$R_{mín}$ [km]	$R_{máx}$ [km]	R_{mean} [km]	R_{cov} [-]	Z_{mean} [km]
6,5	8,3	32	198	111	0,4834	32

Según lo señalado en dicho trabajo se determina la curva de amenaza, a partir de la cual se extraen los escenarios más extremos:

- Escenario sísmico 1: correspondiente a un sismo con una probabilidad de excedencia de 43,7%.
- Escenario sísmico 2: correspondiente a un sismo con una probabilidad de excedencia de $4,37 \cdot 10^{-3}\%$.

Como se señaló previamente, para obtener el nivel de daño de la tubería dada la aceleración de suelo obtenida de la curva de amenaza, se debe usar una curva de fragilidad. Siguiendo la metodología del trabajo antes mencionado, se usan las curvas de fragilidad desarrolladas por Garvs (2020) para tuberías de asbesto cemento en suelos chilenos (Figura 3.15). Es por esto que, para los modelos de deterioro que incorporen amenaza sísmica, se utiliza una red experimental completa de asbesto cemento (el resto de las características de la red mantienen lo señalado en 3.2.1).

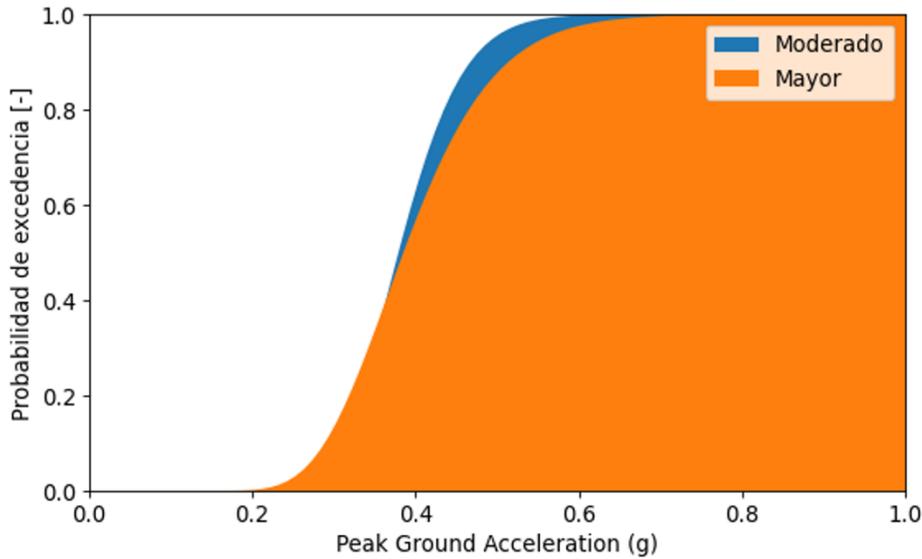


Figura 3.15: Curva de fragilidad para tuberías de asbesto cemento según Garvs (2020).

Según la probabilidad de falla determinada se obtiene el nivel de daño por sismo, el que se adiciona al daño obtenido por deterioro. Para determinar el área de fuga según el nivel de daño y en línea con lo presentado en 3.2.3.1, cada nivel de daño supone un aumento del área de fuga equivalente al 30% de la sección de la tubería, es decir, si una tubería falla por deterioro y tiene un daño sísmico moderado tendrá una fuga equivalente al 60% de su sección transversal.

3.2.5 Simulación

La simulación hidráulica, como se señaló en 3.2.1, se realiza con un modelo basado en presiones, con $P_{\min} = 5$ mca y $P_{\text{req}} = 15$ mca. La red se simula por 96 horas, definiendo los tiempos de aplicación de las distintas amenazas. Se considera que los efectos del daño por deterioro están presentes durante toda la simulación y que los efectos del daño por sismo, para el caso correspondiente, están activos durante las segundas 24 horas de simulación.

Luego, se extraen métricas de cumplimiento de presiones y de demanda, las que se definen según lo planteado en Abarca (2022).

La primera es la fracción de demanda cubierta o *Water Service Availability* (WSA, por sus siglas en inglés), que es la proporción entre la demanda entregada y la demanda esperada para un nodo. Para la red completa, se calcula la WSA como el promedio de la WSA de los nodos.

$$WSA_{\text{red}} = \sum \frac{\text{Demanda}_{\text{entregada}}}{\text{Demanda}_{\text{esperada}}} \cdot \frac{1}{\text{Número nodos}}$$

La segunda es el cumplimiento de la presión mínima y de la presión requerida, que se definen como el porcentaje de los nodos de la red que cumplen con estos límites para un determinado tiempo de análisis.

$$P_{req}[\%] = \frac{\text{Número nodos}(P \geq P_{req})}{\text{Número nodos}} \cdot 100$$

4 Resultados

En esta sección se muestran los resultados obtenidos en el trabajo. La presentación de resultados inicia con aquellos asociados al caso de estudio previamente señalado, para luego introducir los resultados asociados al análisis en redes modelo.

4.1 Modelo de deterioro: caso de estudio

Como se señala en la sección 3.1.6, el caso de estudio corresponde a una red de distribución de agua potable en una ciudad grande de Chile. El registro cuenta con las fallas desde 2017 hasta 2021.

La red estudiada se compone, de acuerdo con la información entregada por la SISS, de 298.696 ductos, cuya materialidad se presenta en la Figura 4.1, donde el apartado “Otros” engloba todos los materiales con una presencia menor al 5%, entre los que se cuentan acero revestido, cobre y hormigón armado y simple.

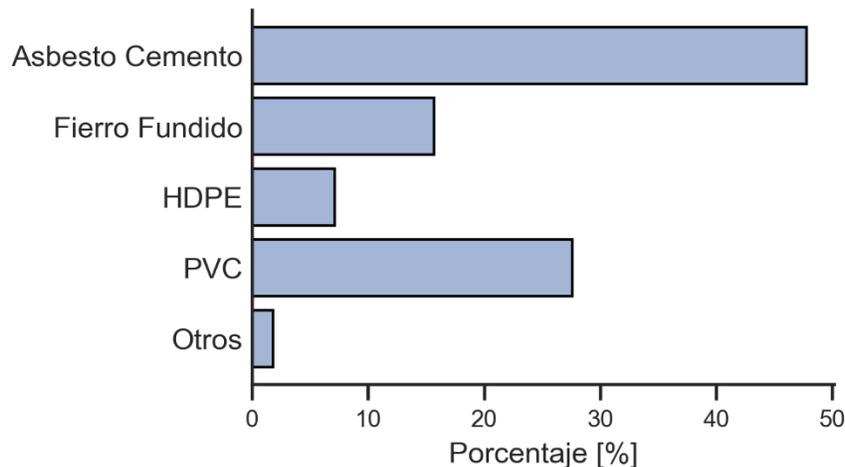


Figura 4.1: Materialidad de la red.

La distribución de los diámetros, largos y edades de los ductos se muestra en la Figura 4.2. En ella se aprecia que la mayor cantidad de tuberías instaladas son de 100, 110 y 75 [mm]. Además, se observa que las longitudes de los tramos se concentran en el rango entre cero y diez metros. Sin embargo, en el shape de la red utilizada se registran longitudes de hasta 4 [km], registros que fueron omitidos de la gráfica para privilegiar una buena visualización de los datos.

Por otro lado, las edades de las tuberías se concentran en el rango de 45 a 50 años, sugiriendo que es una red muy antigua sin reemplazos o que los registros del archivo shape no están actualizados. Cabe mencionar que hay un porcentaje marginal de registros de mayor edad que los presentados en la figura que fueron filtrados para favorecer la visualización de los datos. Para finalizar, se observa que casi el 100% de la red no ha tenido

fallas registradas en el periodo, y entre aquellas que han fallado se registra un máximo de diez fallas previas.

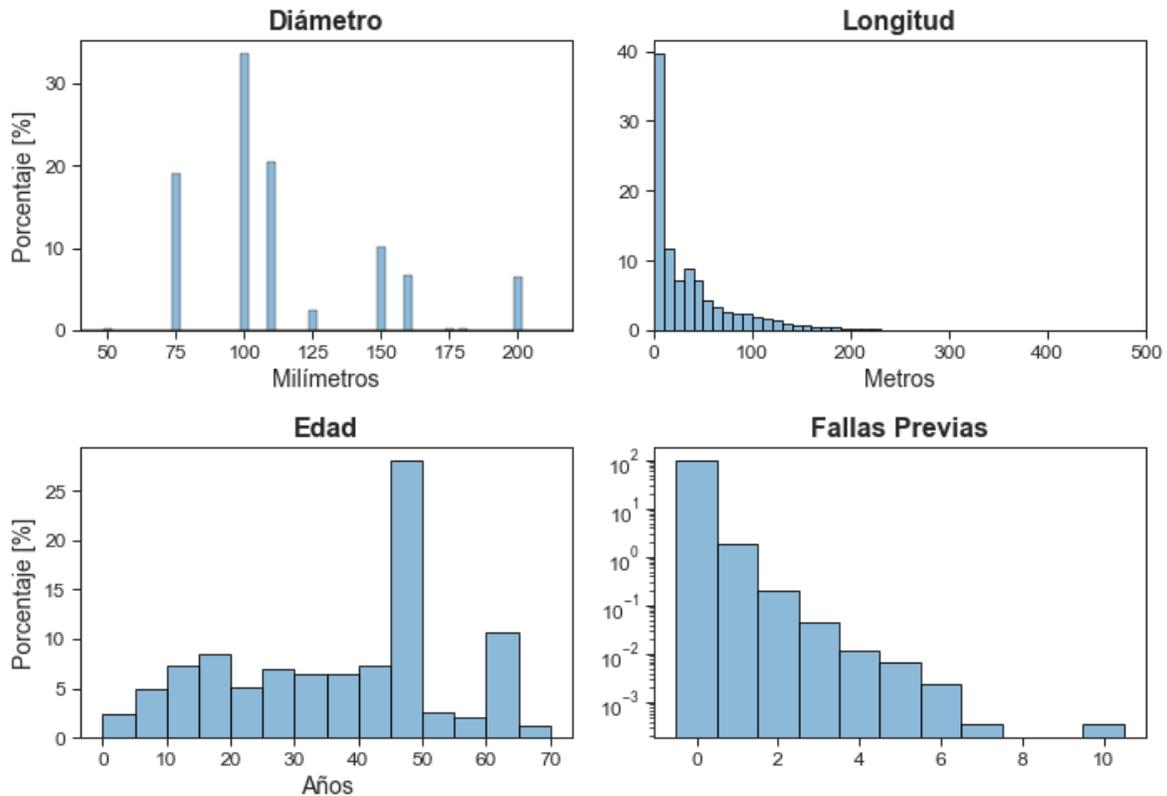


Figura 4.2: Distribución de características de la red.

La base de datos cuenta con 7099 fallas entre 2017 y 2021, las que de acuerdo con las variables explicativas del modelo se distribuyen según la Figura 4.3. En ella se observa que la mayor cantidad de fallas ocurren en tuberías de 75 y 100 [mm]. Para el caso de la longitud, la mayor cantidad de fallas se da en tuberías de entre 40 y 50 metros, aun cuando el mayor porcentaje de tuberías de la red se encuentra en el rango 0 – 10 metros.

Analizando la edad, se observa que la mayor cantidad de fallas se da en tuberías mayores a 40 años. En cuanto a las fallas previas en la red, se observa que la mayor cantidad de roturas ocurren en tuberías que no tienen fallas previas.

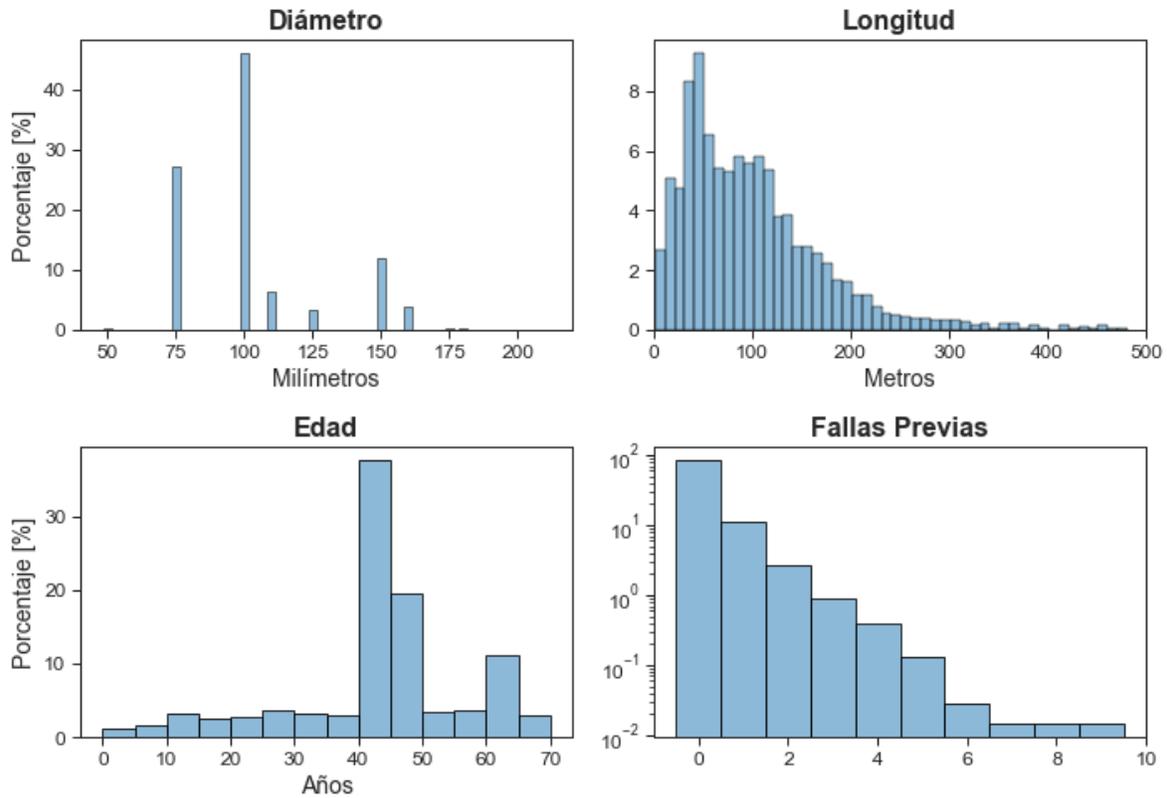


Figura 4.3: Distribución de las roturas según cada variable explicativa.

La distribución de fallas según materialidad se muestra en la Figura 4.4, donde en la categoría “Otros” se engloban todos aquellos materiales cuyas fallas representan menos del 5% de las fallas totales, entre los que se incluyen cobre, hierro dúctil, acero revestido y HDPE. Cabe señalar que el porcentaje de fallas correspondiente al PVC es muy bajo considerando la proporción que tiene este material en la red de distribución.

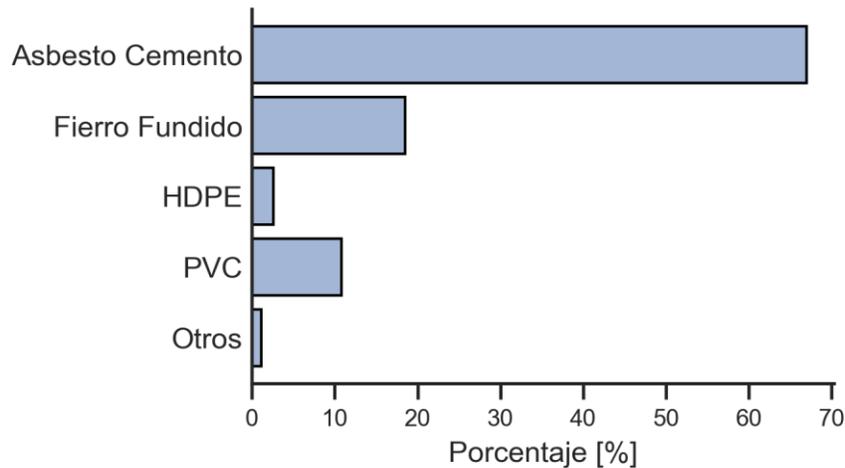


Figura 4.4: Distribución roturas según materialidad.

El procesamiento de los tres algoritmos se realiza usando la infraestructura de cómputo del NLHP, considerando para todos los casos un nodo con 43 núcleos funcionando en simultáneo.

Para cada algoritmo se probaron seis situaciones distintas, las que se detallan a continuación:

1. Red completa, sin remuestreo de datos y considerando *Accuracy* como métrica objetivo en la búsqueda por grilla.
2. Red completa, sin remuestreo de datos y considerando *Balanced Accuracy* como métrica objetivo en la búsqueda por grilla.
3. Red completa, con remuestreo de datos y considerando *Accuracy* como métrica objetivo en la búsqueda por grilla.
4. Red completa, con remuestreo de datos y considerando *Balanced Accuracy* como métrica objetivo en la búsqueda por grilla.
5. Comuna de La Florida, con remuestreo y considerando *Balanced Accuracy* como métrica objetivo en la búsqueda por grilla
6. Comuna de La Florida, con remuestreo y considerando *F1 Score* como métrica objetivo en la búsqueda por grilla

4.1.1 Modelo de deterioro con Gradient Boosting (GBT)

El algoritmo se prueba con distintas configuraciones a modo de obtener las mejores métricas posibles. La primera configuración otorga los resultados mostrados en la Figura 4.5 y en la Figura 4.6, donde se observa que la predicción recae en su mayoría en la clase “Nada”, mientras que la curva ROC se encuentra entre el modelo no predictivo y el modelo perfecto con un área bajo la curva del 90%.

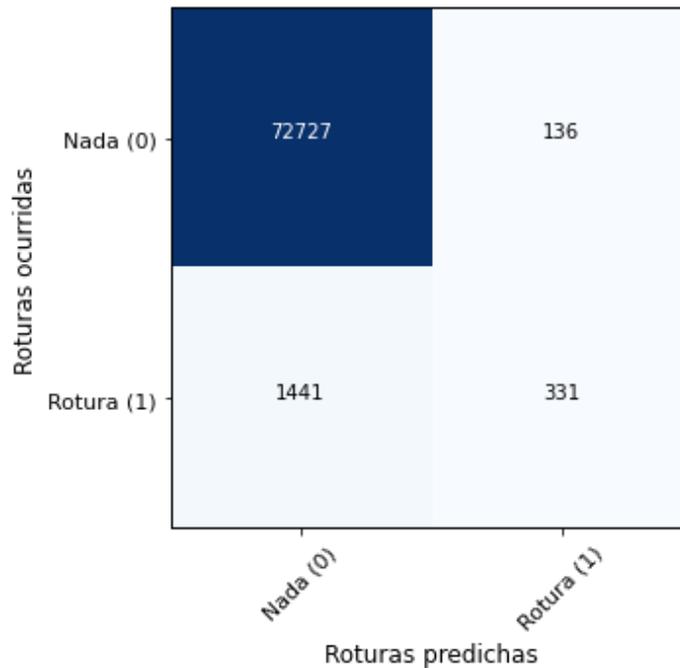


Figura 4.5: Matriz de confusión, GBT, primera configuración.

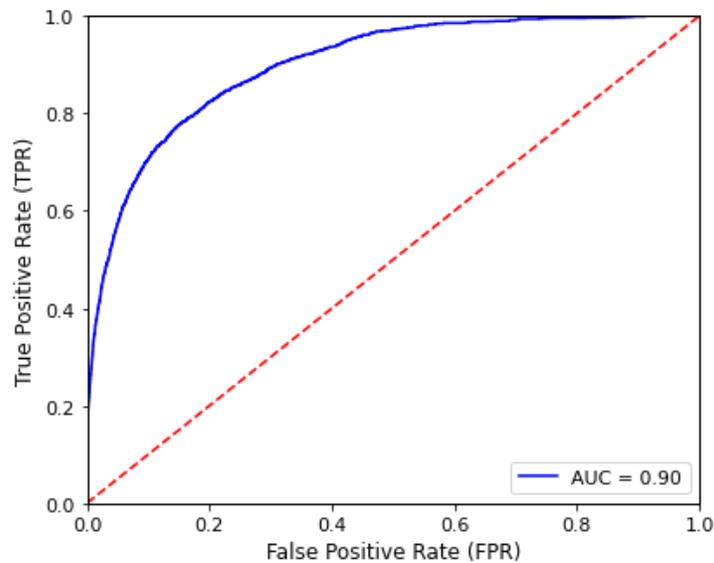


Figura 4.6: Curva ROC, GBT, primera configuración.

Para la segunda configuración (red completa, sin remuestreo, con *Balanced Accuracy* como métrica objetivo) se tienen los resultados de la Figura 4.7 y Figura 4.8, donde se observa que la predicción cae en su mayoría en la clase “Nada”, mientras que la curva ROC se encuentra entre el modelo no predictivo y el modelo perfecto.

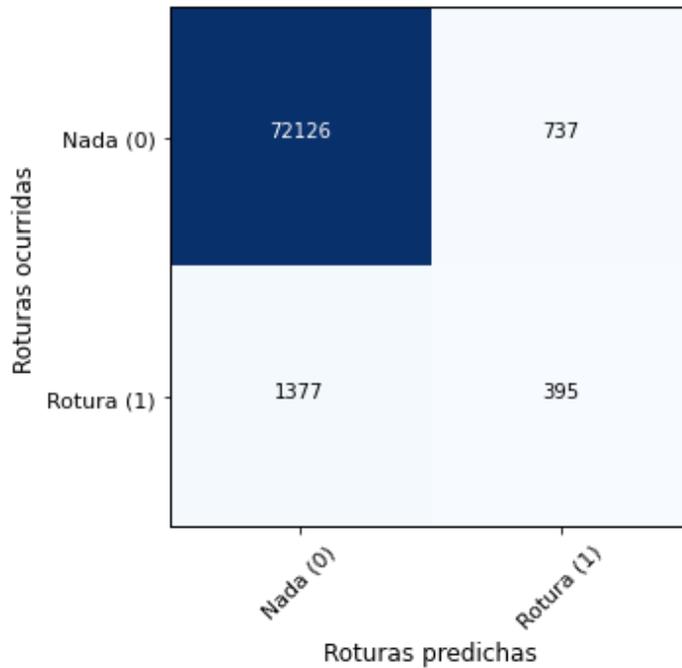


Figura 4.7: Matriz de confusión, GBT, segunda configuración.

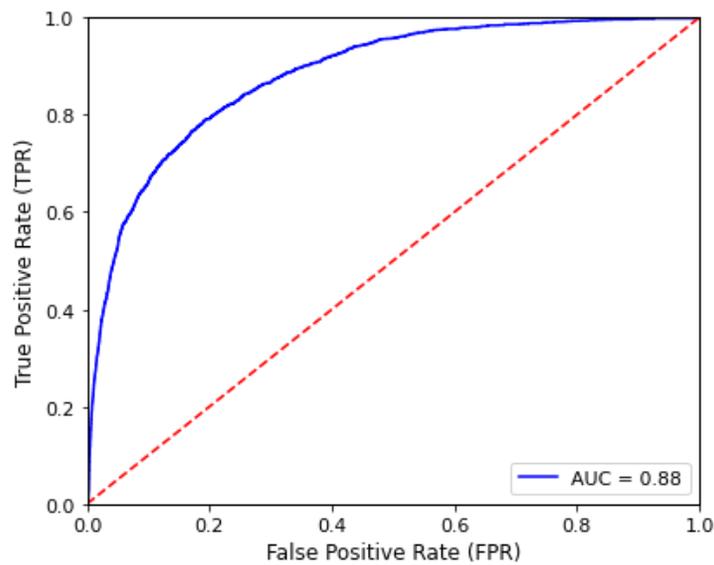


Figura 4.8: Curva ROC, GBT, segunda configuración.

Para la tercera configuración (red completa, con remuestreo, con *Accuracy* como métrica objetivo) se tienen los resultados de la Figura 4.10 y Figura 4.11, donde se observa un mayor número de aciertos en la clase positiva al comparar con los casos 1 y 2, mientras que la curva ROC se encuentra entre el modelo no predictivo y el modelo perfecto sin variaciones considerables respecto de los casos anteriores.

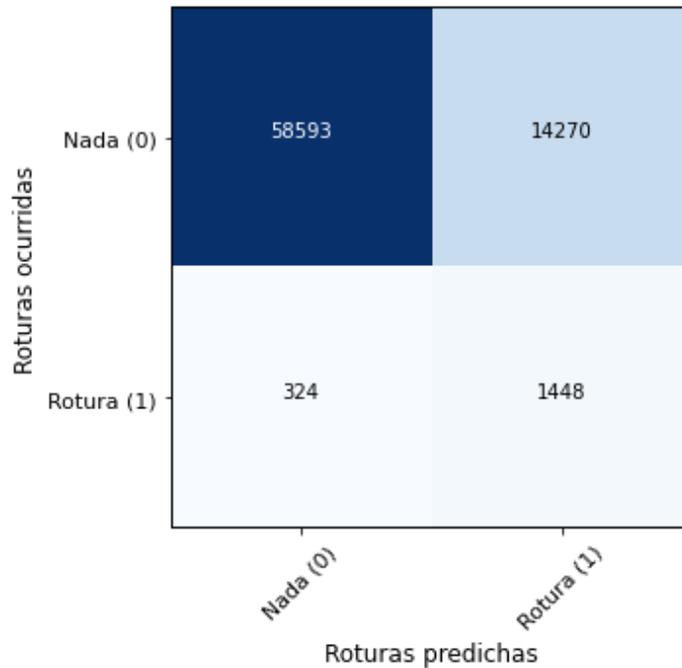


Figura 4.9: Matriz de confusión, GBT, tercera configuración.

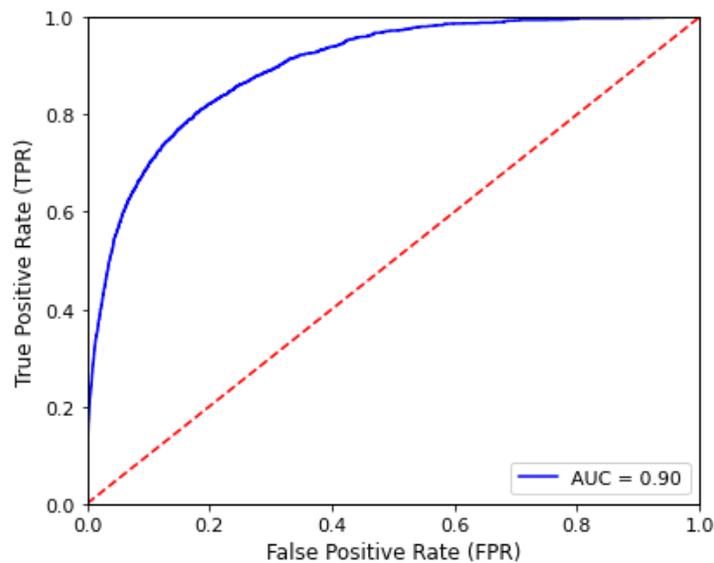


Figura 4.10: Curva ROC, GBT, tercera configuración.

Para la cuarta configuración (red completa, con remuestreo, con *Balanced Accuracy* como métrica objetivo) se tienen los resultados de la Figura 4.11 y Figura 4.12, donde se observa un mayor número de aciertos en la clase positiva al comparar con los casos 1 y 2, mientras que la curva ROC se encuentra entre el modelo no predictivo y el modelo perfecto.

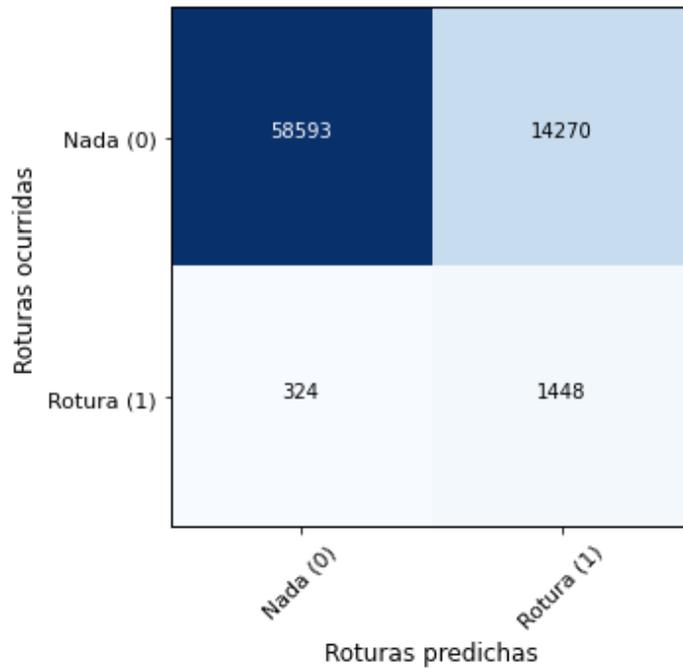


Figura 4.11: Matriz de confusión, GBT, cuarta configuración.

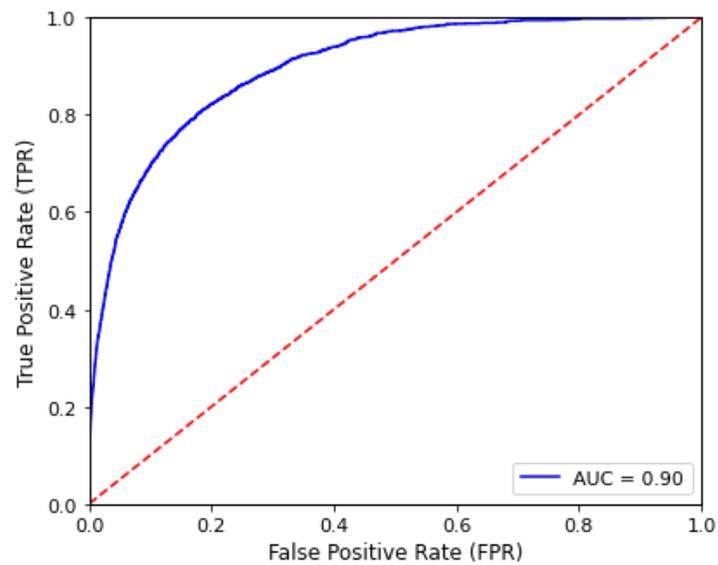


Figura 4.12: Curva ROC, GBT, cuarta configuración.

Para la quinta configuración (comuna de La Florida, con remuestreo, con *Balanced Accuracy* como métrica objetivo) se tienen los resultados de la Figura 4.13 y Figura 4.14, donde se observa que de cada 100 fallas predichas alrededor de cinco son correctas, mientras que la curva ROC se encuentra entre el modelo no predictivo y el modelo perfecto, con una menor área bajo la curva y un escalonamiento propio del número de datos utilizados para el testeo.

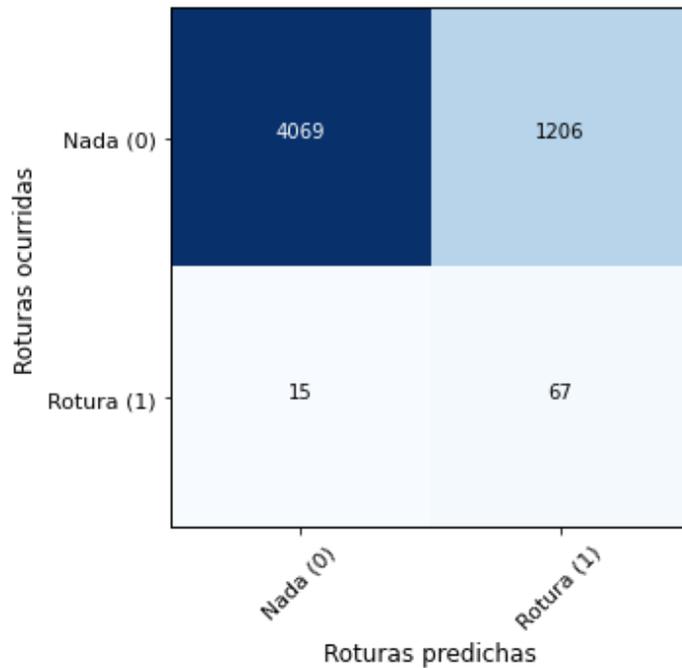


Figura 4.13: Matriz de confusión, GBT, quinta configuración.

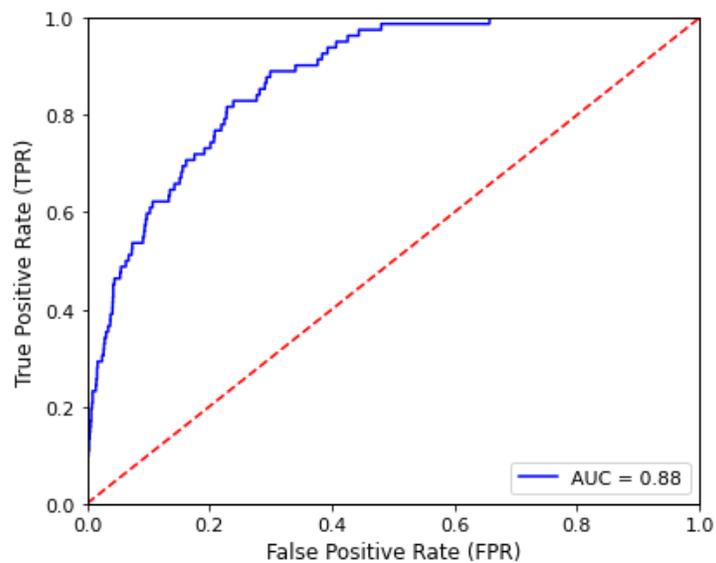


Figura 4.14: Curva ROC, GBT, quinta configuración.

Para la sexta configuración (comuna de La Florida, con remuestreo, con *F1 Score* como métrica objetivo) se tienen los resultados de la Figura 4.15 y Figura 4.16, donde, al igual que en el caso anterior, se observa que de cada 100 fallas predichas alrededor de cinco son correctas, mientras que la curva ROC se encuentra entre el modelo no predictivo y el modelo perfecto, con una menor área bajo la curva y un escalonamiento propio del número de datos utilizados para el testeo.

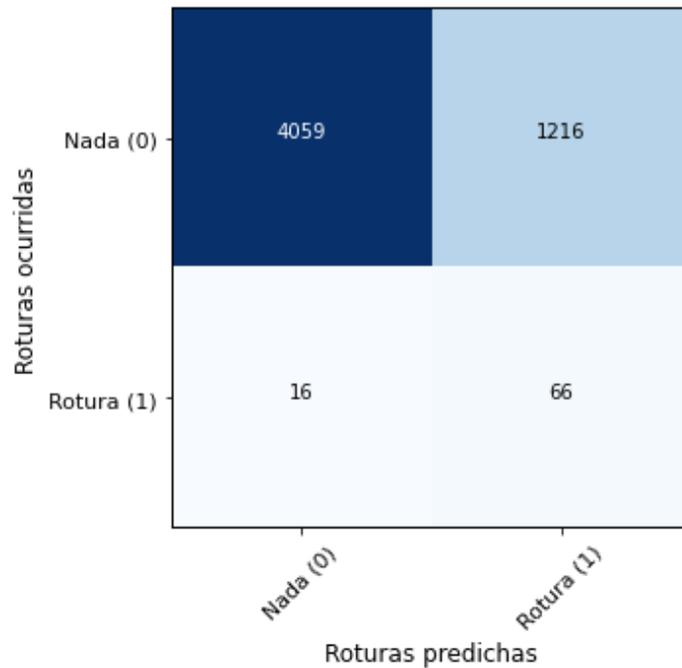


Figura 4.15: Matriz de confusión, GBT, sexta configuración.

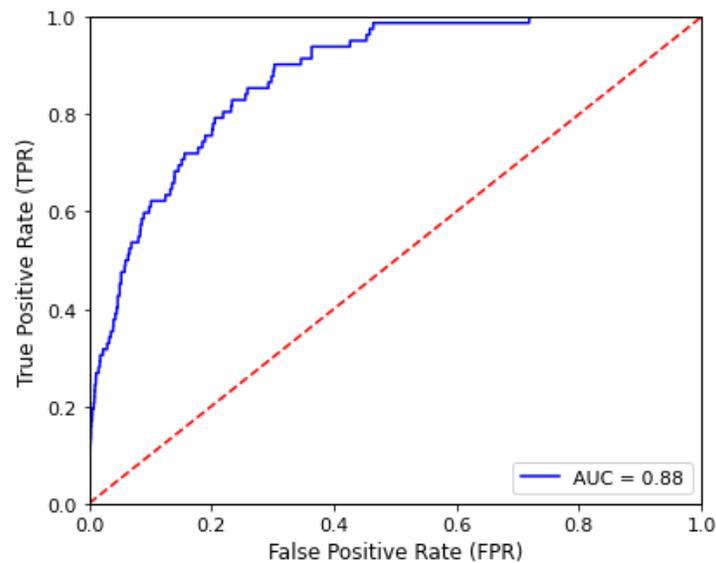


Figura 4.16: Curva ROC, GBT, sexta configuración.

4.1.2 Modelo de deterioro con Perceptrón Multicapa (MLP)

Para la primera configuración (red completa, sin remuestreo, con *Accuracy* como métrica objetivo) se tienen los resultados de la Figura 4.17 y Figura 4.18, donde se observa que la predicción cae en su mayoría en la clase “Nada”, mientras que la curva ROC se encuentra entre el modelo no predictivo y el modelo perfecto.

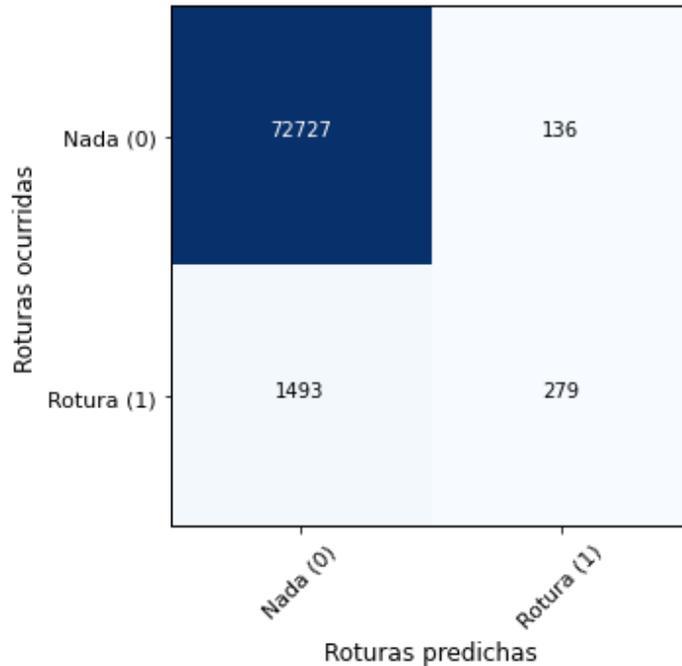


Figura 4.17: Matriz de confusión, MLP, primera configuración.

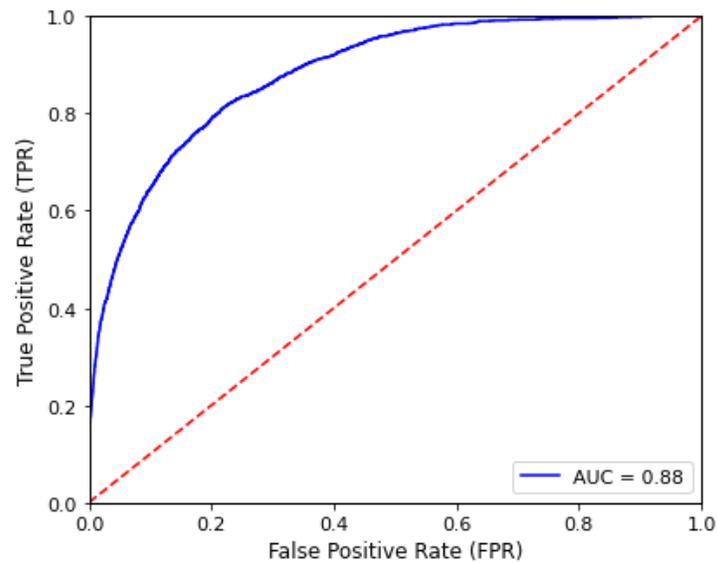


Figura 4.18: Curva ROC, MLP, primera configuración.

Para la segunda configuración (red completa, sin remuestreo, con *Balanced Accuracy* como métrica objetivo) se tienen los resultados de la Figura 4.19 y Figura 4.20, donde se observa, al igual que en el caso anterior, la predicción cae en su mayoría en la clase “Nada”, mientras que la curva ROC se encuentra entre el modelo no predictivo y el modelo perfecto.

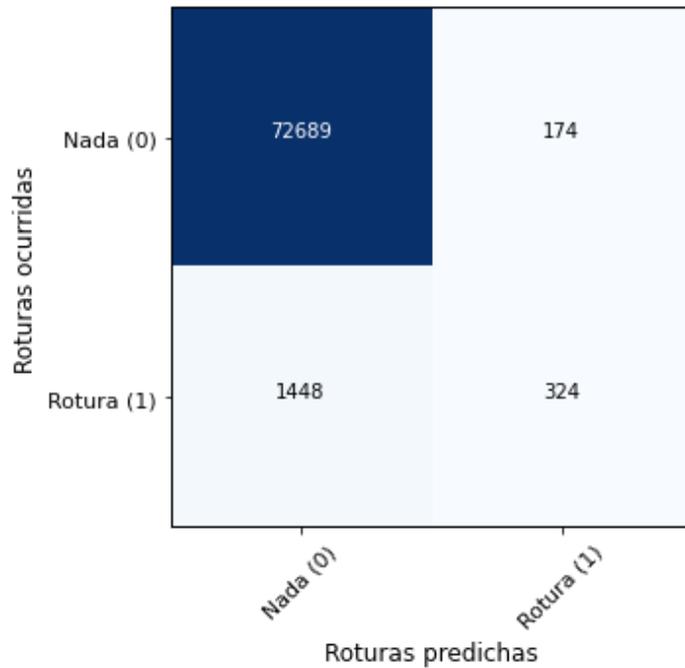


Figura 4.19: Matriz de confusión, MLP, segunda configuración.

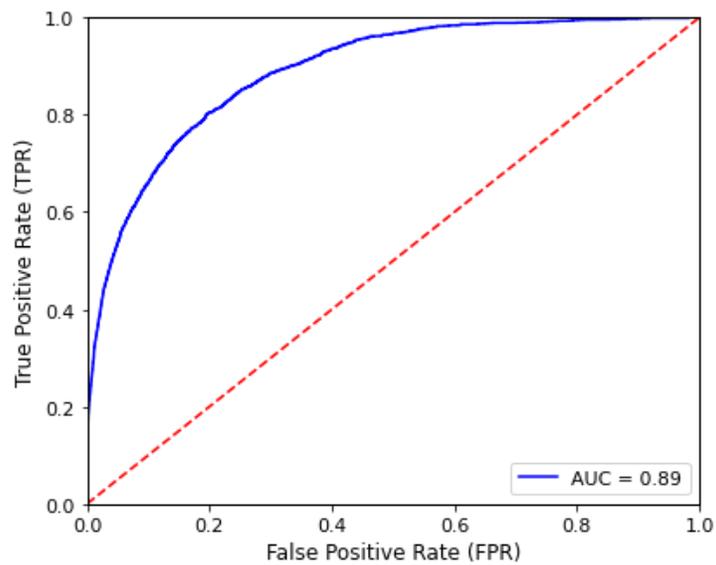


Figura 4.20: Curva ROC, MLP, segunda configuración.

Para la tercera configuración (red completa, con remuestreo, con *Accuracy* como métrica objetivo) se tienen los resultados de la Figura 4.21 y Figura 4.22, donde se observa un mayor número de aciertos en la clase positiva al comparar con los casos 1 y 2, mientras que la curva ROC se encuentra entre el modelo no predictivo y el modelo perfecto.

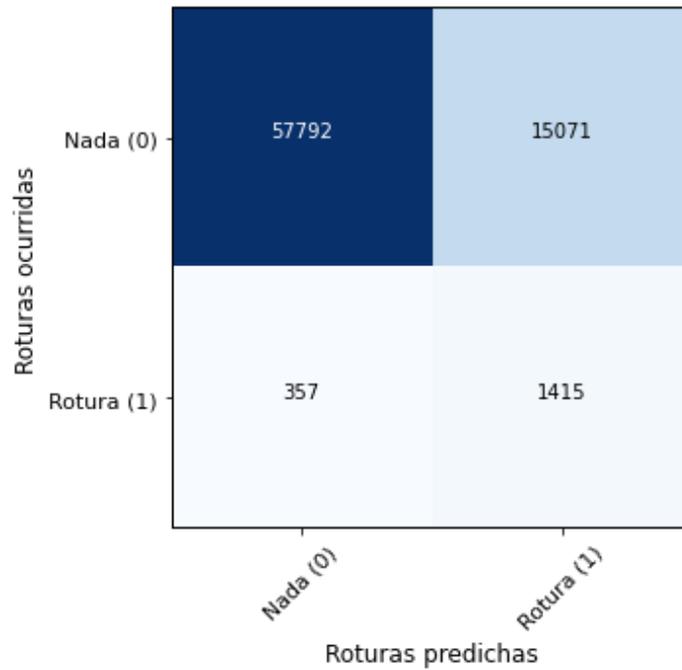


Figura 4.21: Matriz de confusión, MLP, tercera configuración.

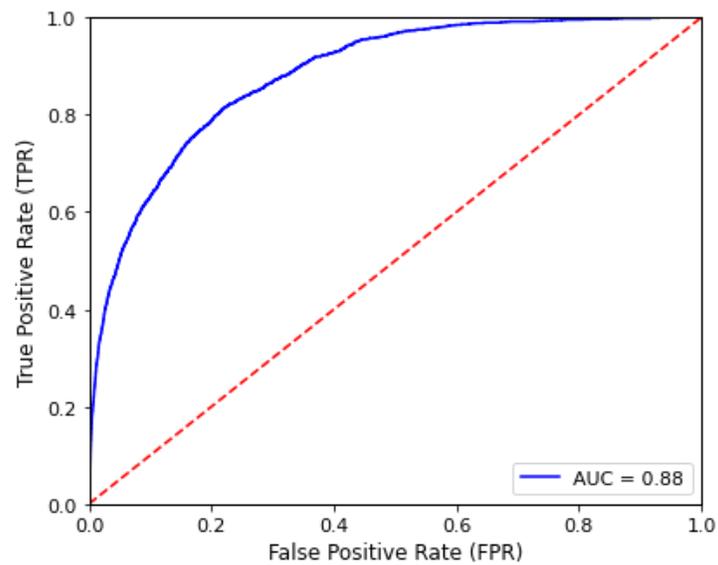


Figura 4.22: Curva ROC, MLP, tercera configuración.

Para la cuarta configuración (red completa, con remuestreo, con *Balanced Accuracy* como métrica objetivo) se tienen los resultados de la Figura 4.23 y Figura 4.24, donde, al igual que en el caso anterior, se observa un mayor número de aciertos en la clase positiva al comparar con los casos 1 y 2, mientras que la curva ROC se encuentra entre el modelo no predictivo y el modelo perfecto.

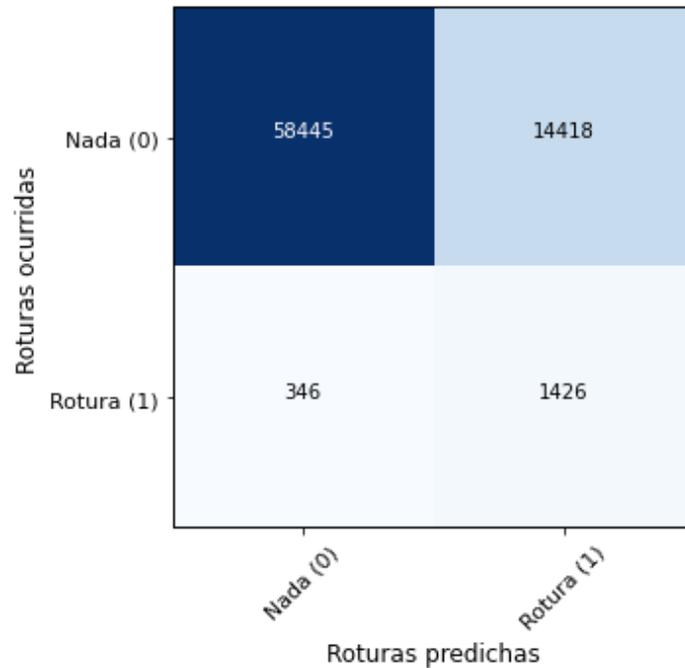


Figura 4.23: Matriz de confusión, MLP, cuarta configuración.

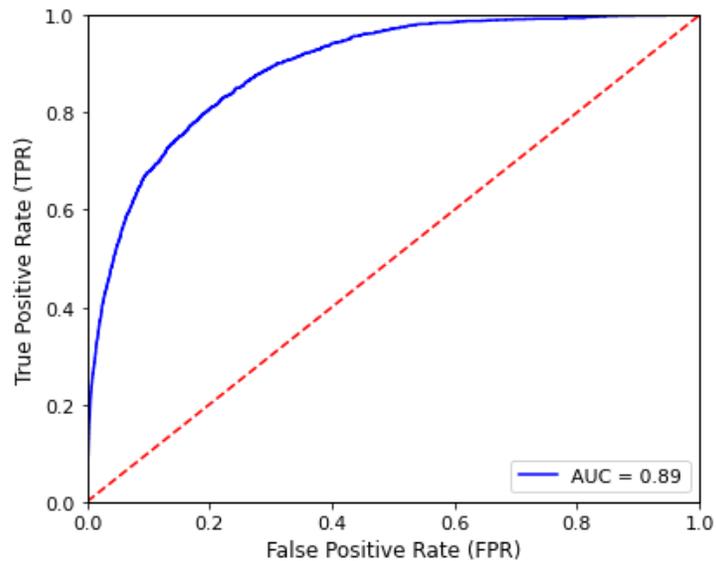


Figura 4.24: Curva ROC, MLP, cuarta configuración.

Para la quinta configuración (comuna de La Florida, con remuestreo, con *Balanced Accuracy* como métrica objetivo) se tienen los resultados de la Figura 4.25 y Figura 4.26, donde se observa que de cada 100 fallas predichas alrededor de cuatro son correctas, mientras que la curva ROC se encuentra entre el modelo no predictivo y el modelo perfecto, con una menor área bajo la curva y un escalonamiento propio del número de datos utilizados.

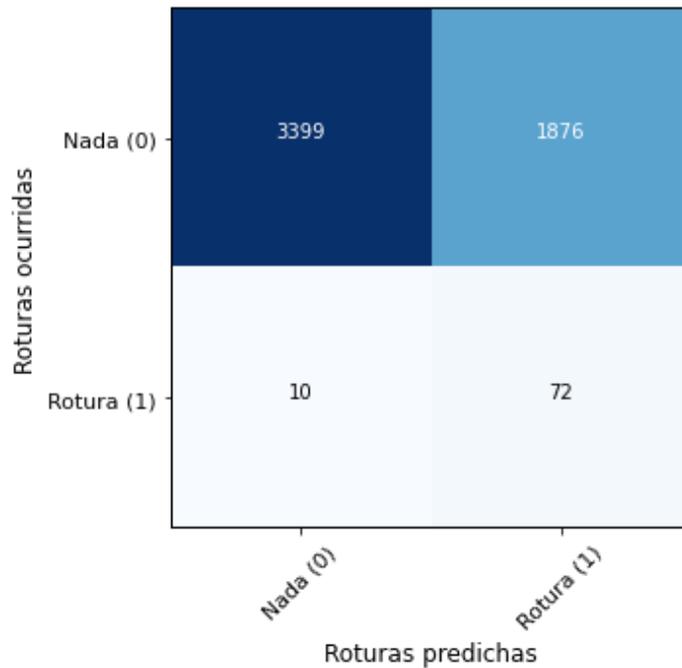


Figura 4.25: Matriz de confusión, MLP, quinta configuración.

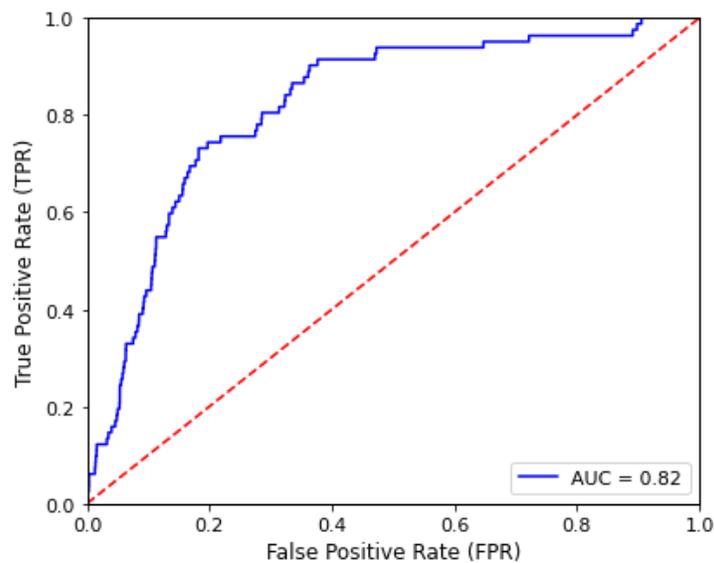


Figura 4.26: Curva ROC, MLP, quinta configuración.

Para la sexta configuración (comuna de La Florida, con remuestreo, con *F1 Score* como métrica objetivo) se tienen los resultados de la Figura 4.27 y Figura 4.28, donde se observa que de cada 100 fallas predichas alrededor de tres son correctas, mientras que la curva ROC se encuentra entre el modelo no predictivo y el modelo perfecto, con una menor área bajo la curva y un escalonamiento propio del número de datos utilizados.

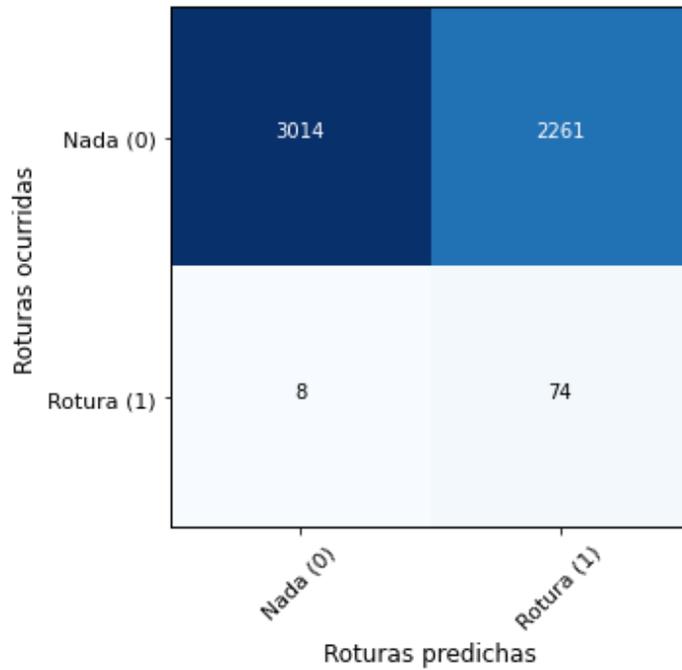


Figura 4.27: Matriz de confusión, MLP, sexta configuración.

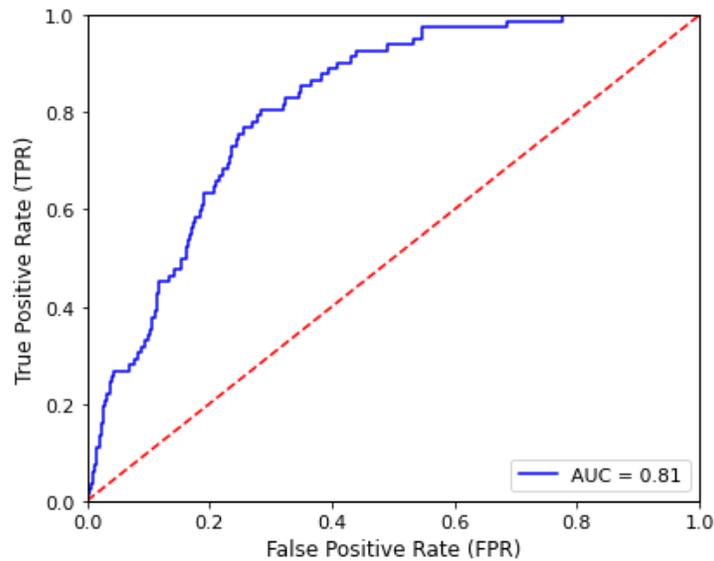


Figura 4.28: Curva ROC, MLP, sexta configuración.

4.1.3 Modelo de deterioro con Máquina de Vector de Apoyo (SVM)

Para las primeras dos configuraciones no se pudo obtener resultados en menos de cinco días de ejecución del código en las condiciones señaladas previamente, por lo que se omiten dichos casos para este algoritmo.

Para la tercera configuración (red completa, con remuestreo, con *Accuracy* como métrica objetivo) se tienen los resultados de la Figura 4.29 y Figura 4.30, las que presentan un comportamiento similar a los señalados en los algoritmos previos, pero con una menor área bajo la curva ROC (más cerca del modelo no predictivo).

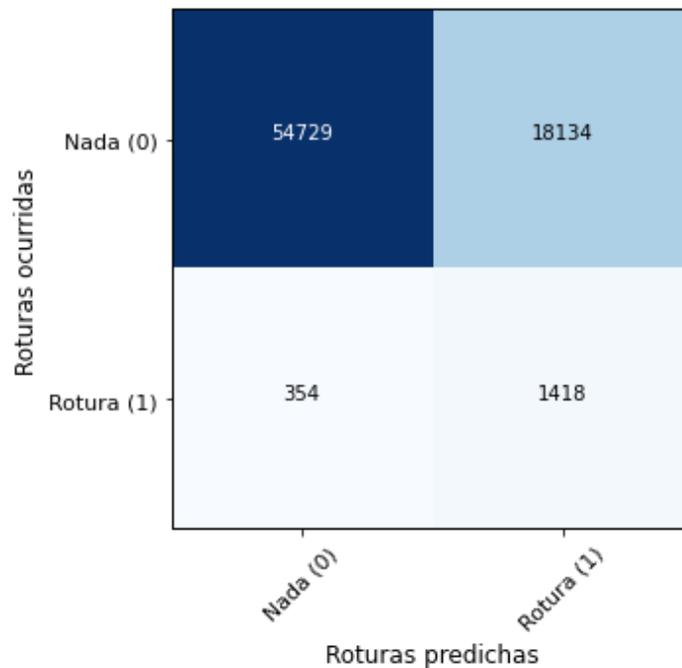


Figura 4.29: Matriz de confusión, SVM, tercera configuración.

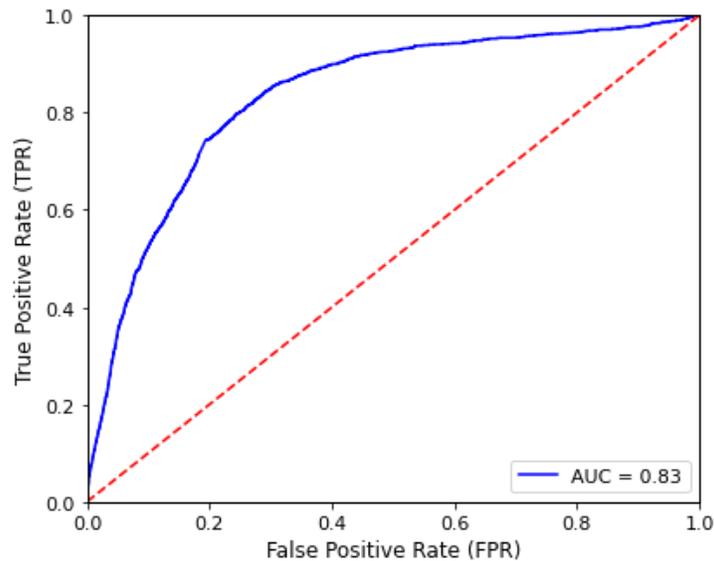


Figura 4.30: Curva ROC, SVM, tercera configuración.

Para la cuarta configuración (red completa, con remuestreo, con *Balanced Accuracy* como métrica objetivo) se tienen los resultados de la Figura 4.31 y Figura 4.32, con un comportamiento similar al de los algoritmos previamente analizados, donde siete de cada 100 fallas predichas están correctas. El área bajo la curva ROC presenta un valor levemente superior al del caso 3.

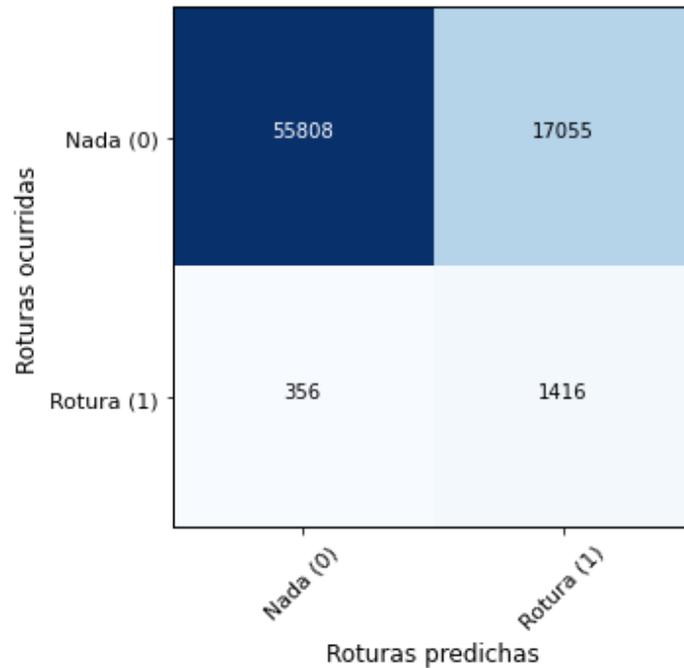


Figura 4.31: Matriz de confusión, SVM, cuarta configuración.

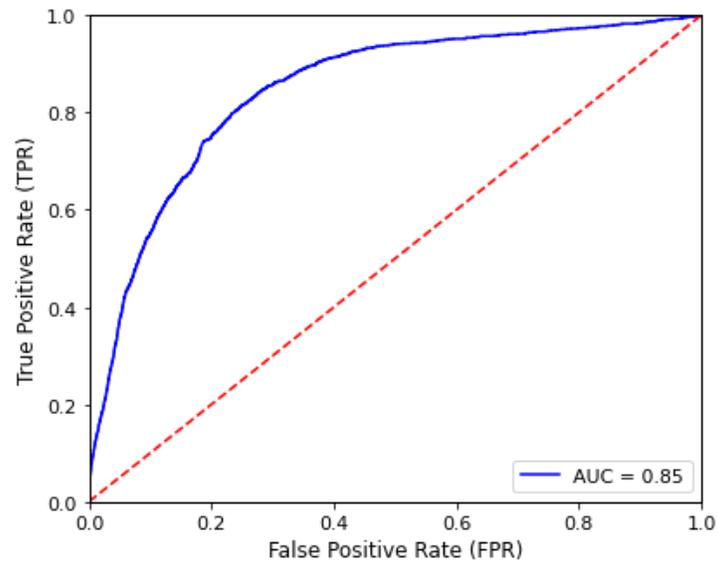


Figura 4.32: Curva ROC, SVM, cuarta configuración.

Para la quinta configuración (comuna de La Florida, con remuestreo, con *Balanced Accuracy* como métrica objetivo) se tienen los resultados de la Figura 4.33 y Figura 4.34, donde el modelo se inclina a predecir “Nada” en casi el total de observaciones analizadas. Este efecto se manifiesta también en la curva ROC, la que está muy cerca de un modelo no predictivo.

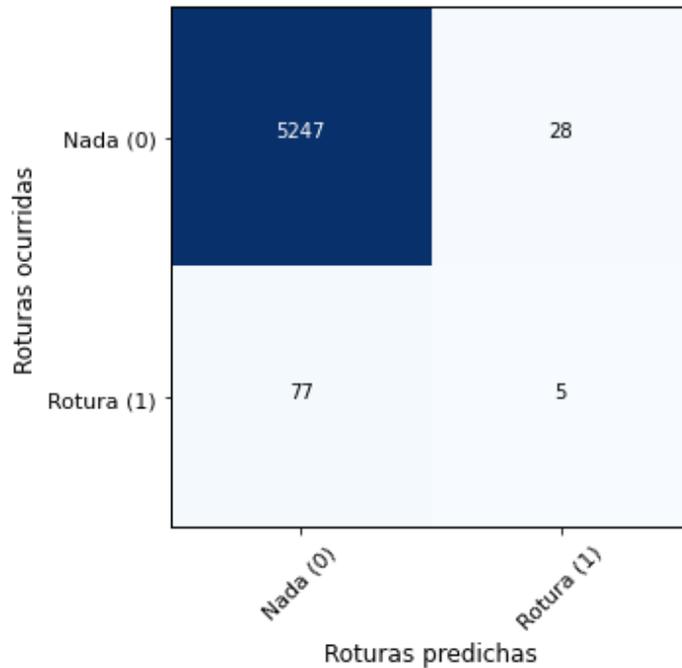


Figura 4.33: Matriz de confusión, SVM, quinta configuración.

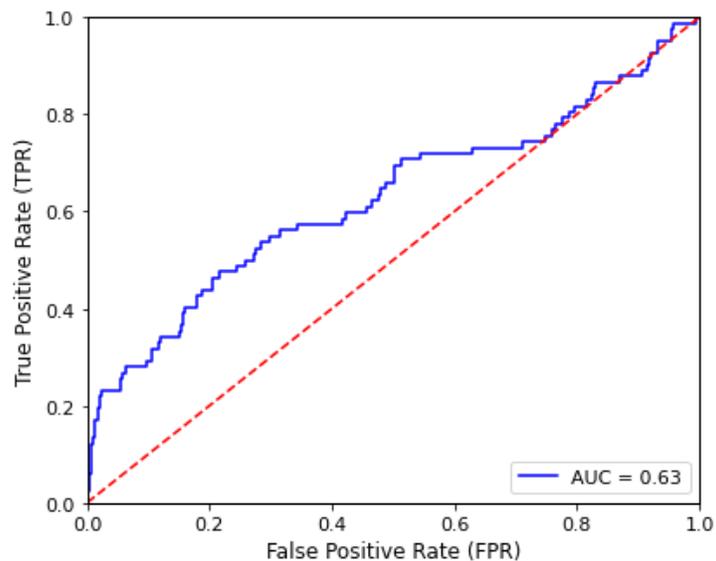


Figura 4.34: Curva ROC, SVM, quinta configuración.

Para la sexta configuración (comuna de La Florida, con remuestreo, con *F1 Score* como métrica objetivo) se tienen los resultados de la Figura 4.35 y Figura 4.36. Esta configuración presenta resultados similares al caso anterior, con casi el 100% de las predicciones en la clase “Nada” y una curva ROC cercana a un modelo no predictivo.

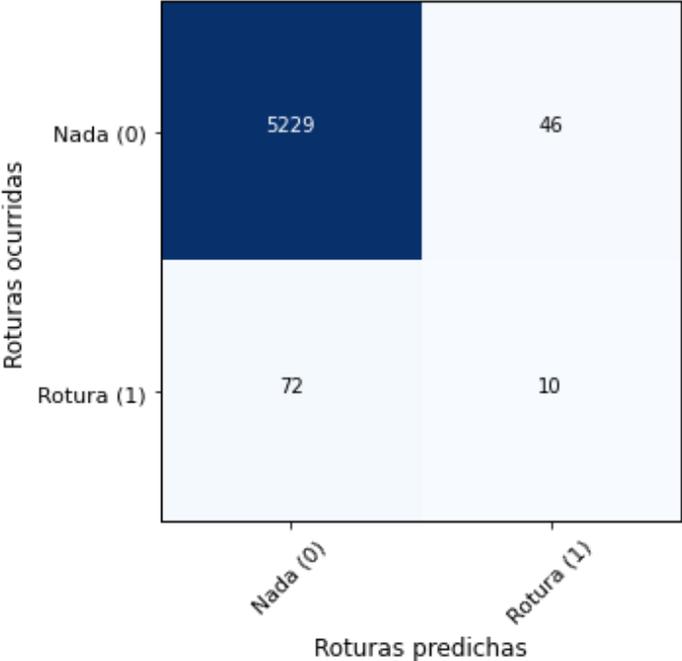


Figura 4.35: Matriz de confusión, SVM, sexta configuración.

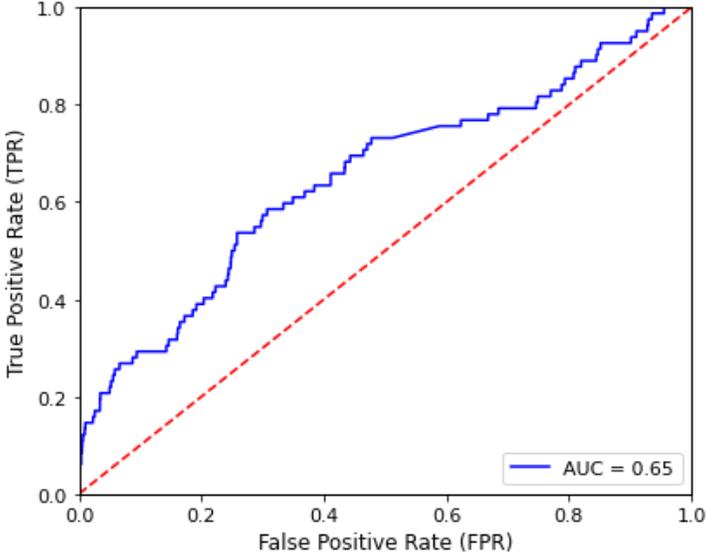


Figura 4.36: Curva ROC, SVM, sexta configuración.

4.1.4 Comparación de resultados

En la Tabla 4.1 se muestra una comparación de los tiempos de ejecución de los algoritmos en cada caso, los que fueron ejecutados en el clúster del NLHPC.

Tabla 4.1: Comparación de los tiempos de ejecución de los modelos de deterioro para las distintas configuraciones.

	Caso 1	Caso 2	Caso 3	Caso 4	Caso 5	Caso 6
GBT	33:39:57	33:48:22	01:22:22	01:22:06	00:12:50	00:12:45
MLP	06:05:26	06:17:22	00:55:59	00:56:24	00:07:00	00:06:58
SVM	-	-	22:25:22	32:48:56	00:04:21	00:10:34

En la Tabla 4.2 se muestra la comparación de las métricas de cada algoritmo en cada configuración:

Tabla 4.2: Comparación métricas algoritmos de inteligencia artificial en cada configuración.

	Caso	TPR	TNR	Precision	FPR	FNR	Accuracy	Balanced Accuracy	F1 Score	AUC
GBT	Caso 1	18.7%	99.8%	70.9%	0.2%	81.3%	97.9%	59.2%	29.6%	90.0%
	Caso 2	22.3%	99.0%	34.9%	1.0%	77.7%	97.2%	60.6%	27.2%	88.2%
	Caso 3	81.7%	80.4%	9.2%	19.6%	18.3%	80.4%	81.1%	16.6%	89.8%
	Caso 4	81.7%	80.4%	9.2%	19.6%	18.3%	80.4%	81.1%	16.6%	89.8%
	Caso 5	81.7%	77.1%	5.3%	22.9%	18.3%	77.2%	79.4%	9.9%	87.5%
	Caso 6	80.5%	76.9%	5.1%	23.1%	19.5%	77.0%	78.7%	9.7%	87.9%
MLP	Caso 1	15.7%	99.8%	67.2%	0.2%	84.3%	97.8%	57.8%	25.5%	88.3%
	Caso 2	18.3%	99.8%	65.1%	0.2%	81.7%	97.8%	59.0%	28.5%	89.0%
	Caso 3	79.9%	79.3%	8.6%	20.7%	20.1%	79.3%	79.6%	15.5%	88.2%
	Caso 4	80.5%	80.2%	9.0%	19.8%	19.5%	80.2%	80.3%	16.2%	89.2%
	Caso 5	87.8%	64.4%	3.7%	35.6%	12.2%	64.8%	76.1%	7.1%	82.1%
	Caso 6	90.2%	57.1%	3.2%	42.9%	9.8%	57.6%	73.7%	6.1%	81.4%
SVM	Caso 1	-	-	-	-	-	-	-	-	-
	Caso 2	-	-	-	-	-	-	-	-	-
	Caso 3	80.0%	75.1%	7.3%	24.9%	20.0%	75.2%	77.6%	13.3%	83.4%
	Caso 4	79.9%	76.6%	7.7%	23.4%	20.1%	76.7%	78.3%	14.0%	84.6%
	Caso 5	6.1%	99.5%	15.2%	0.5%	93.9%	98.0%	52.8%	8.7%	62.8%
	Caso 6	12.2%	99.1%	17.9%	0.9%	87.8%	97.8%	55.7%	14.5%	65.2%

Se observa de la Tabla 4.2 que los casos 1 y 2 presentan un alto *Accuracy* (sobre 97%), valor que está condicionado por el alto número de aciertos en las tuberías que no fallan producto del desbalance de la muestra y que se refleja en los bajos valores de *TPR* (entre 15 y 25%).

Por otro lado, para los casos 3 y 4, hay un aumento considerable en la *TPR*, alcanzando valores en torno al 80% pero sigue estando patente el desbalance en la muestra, que ahora se manifiesta en valores mucho más bajos de *Precision* (entre 7 y 10%).

Los casos 5 y 6 no presentan mejoras respecto a los casos anteriores.

4.2 Modelo hidráulico

En esta sección se presentan los resultados de la integración de los modelos de deterioro programados al modelamiento hidráulico de redes experimentales.

4.2.1 Incorporación del deterioro en un modelo hidráulico

Primero, se presentan los resultados de los efectos del deterioro en una red experimental de agua potable. En la Figura 4.37 se muestra la distribución de elevaciones en la red.

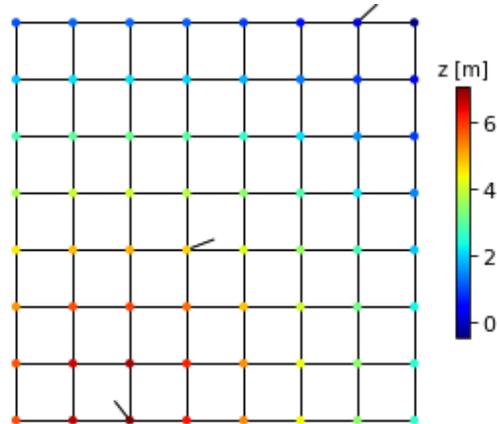


Figura 4.37: Elevación red experimental, caso sólo deterioro.

En la Figura 4.38 se muestra la distribución espacial de las características físicas de la red.

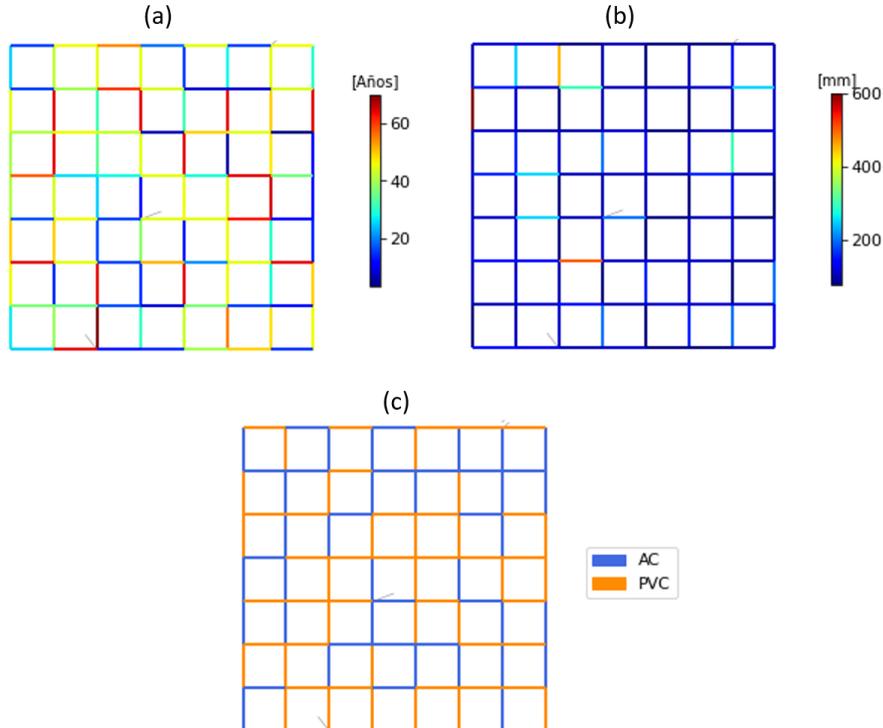


Figura 4.38: Descripción de la red experimental, caso sólo deterioro. (a) Edad, (b) Diámetros, (c) Materialidad

4.2.1.1 Resultados modelo de deterioro

En la Tabla 4.3 se presentan los resultados de los modelos de deterioro para la red descrita:

Tabla 4.3: Comparación métricas algoritmos de inteligencia artificial, caso sólo deterioro.

	TPR	TNR	Precision	FPR	FNR	Accuracy	Balanced Accuracy	F1 Score	AUC
GBT	86.4%	78.6%	76.0%	21.4%	13.6%	82.0%	82.5%	80.9%	90.3%
SVM	90.9%	75.0%	74.1%	25.0%	9.1%	82.0%	83.0%	81.6%	92.5%
MLP	81.8%	85.7%	81.8%	14.3%	18.2%	84.0%	83.8%	81.8%	94.0%

Según las métricas de la Tabla 4.3, el modelo con mejor desempeño (mayor TPR, menor FNR y resto de métricas en valores similares) es la SVM. Con dicho algoritmo se simulan las fallas y se desarrolla el modelamiento hidráulico.

Con el modelo obtenido se obtienen las probabilidades de falla (Figura 4.39) y las fallas efectivas en la red (Figura 4.40) para la condición actual y para la situación a cinco y diez años, asumiendo sólo un aumento de la edad:

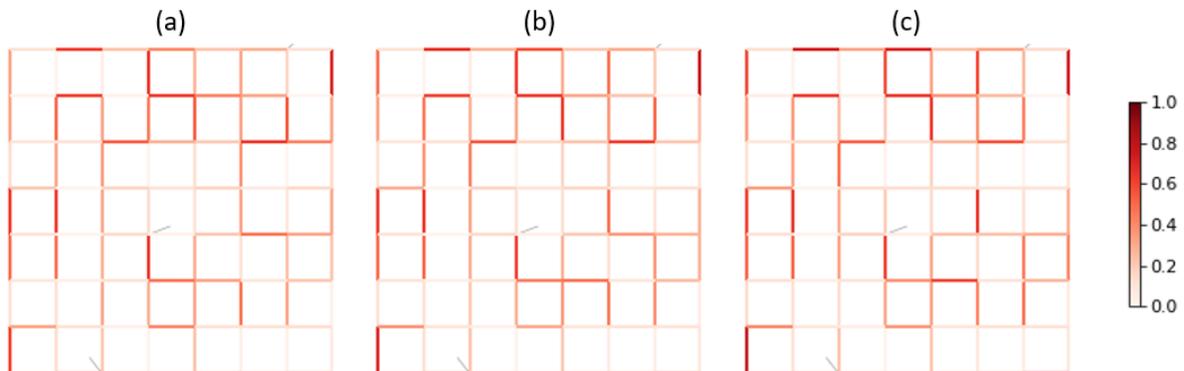


Figura 4.39: Probabilidades de falla, red experimental, situación sólo deterioro. (a) Condición actual, (b) Condición a 5 años, (c) Condición a 10 años.

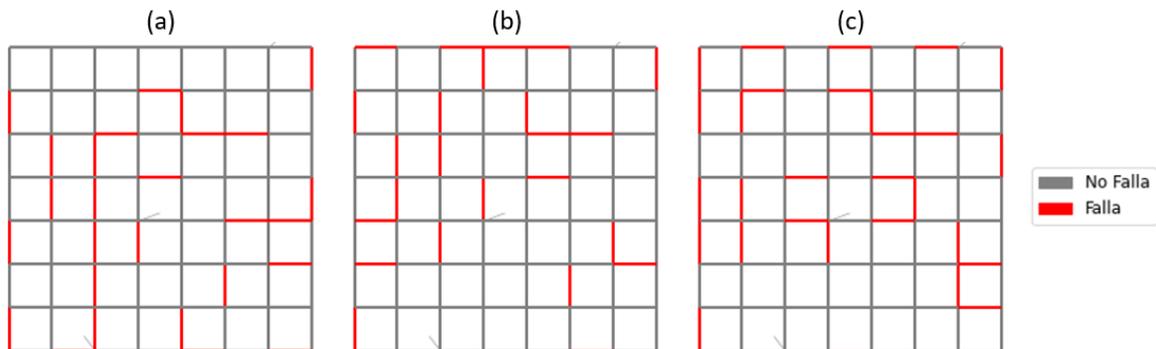


Figura 4.40: Fallas por deterioro, red experimental, situación sólo deterioro. (a) Condición actual, (b) Condición a 5 años, (c) Condición a 10 años.

Para observar la evolución de la probabilidad de falla en la red se muestra la probabilidad de falla promedio de la red para los tres años de estudio en la Tabla 4.4:

Tabla 4.4: Evolución probabilidad de falla promedio, red experimental, situación sólo deterioro.

$P_{falla}(t = 0)$	$P_{falla}(t = 5)$	$P_{falla}(t = 10)$
25,0%	25,3%	26,5%

4.2.1.2 Comportamiento base

Se presenta el comportamiento base de la red, es decir, sin considerar las fallas por deterioro. Todas las simulaciones hidráulicas consideran la variación temporal del factor C de Hazen – Williams según lo señalado en la sección 3.2.3.2. En la Figura 4.41 se muestran los valores medios de presión en los nodos de la red:

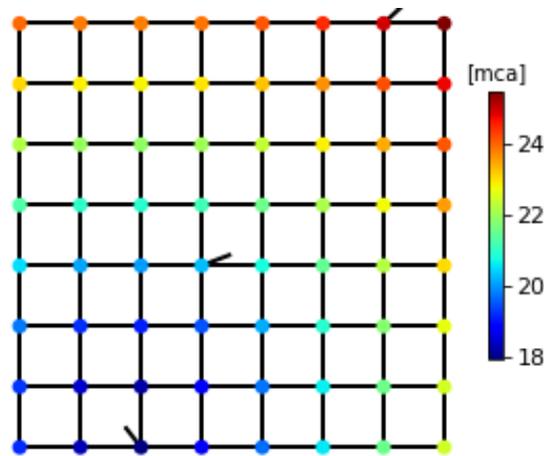


Figura 4.41: Presión promedio en condiciones normales, red experimental, caso sólo deterioro.

4.2.1.3 Escenario actual

Se muestran los resultados de la simulación hidráulica para el escenario actual, considerando sólo deterioro, según las fallas de la Figura 4.40 (a).

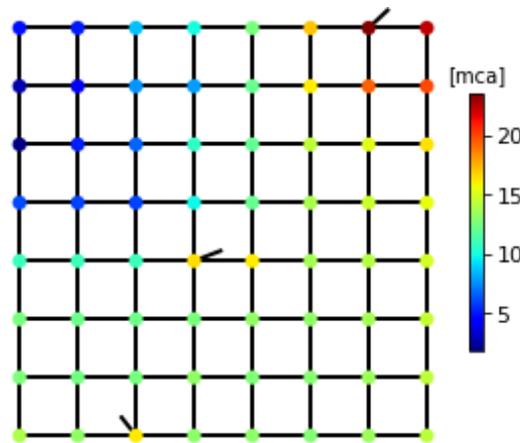


Figura 4.42: Presión promedio en condiciones actuales, red experimental, caso sólo deterioro.

A continuación se muestran los resultados de las métricas hidráulicas. En la Figura 4.43 se muestra la evolución de la disponibilidad de agua durante el tiempo de estudio, donde se observa que la demanda cubierta se encuentra por debajo del 80%.

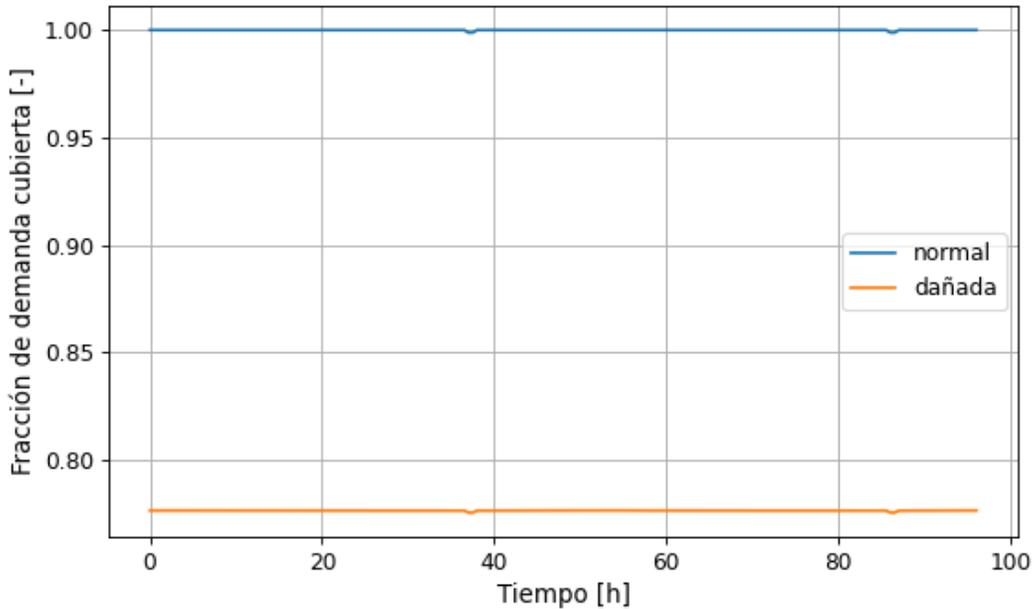


Figura 4.43: Fracción de demanda cubierta en condiciones actuales, red experimental, caso sólo deterioro.

En la Figura 4.44 se muestra el porcentaje de la red que cumple con la presión requerida, observándose que, en promedio, y con las fallas asociadas a este caso particular, aproximadamente un 20% de la red cumple con los 15 [m] de presión.

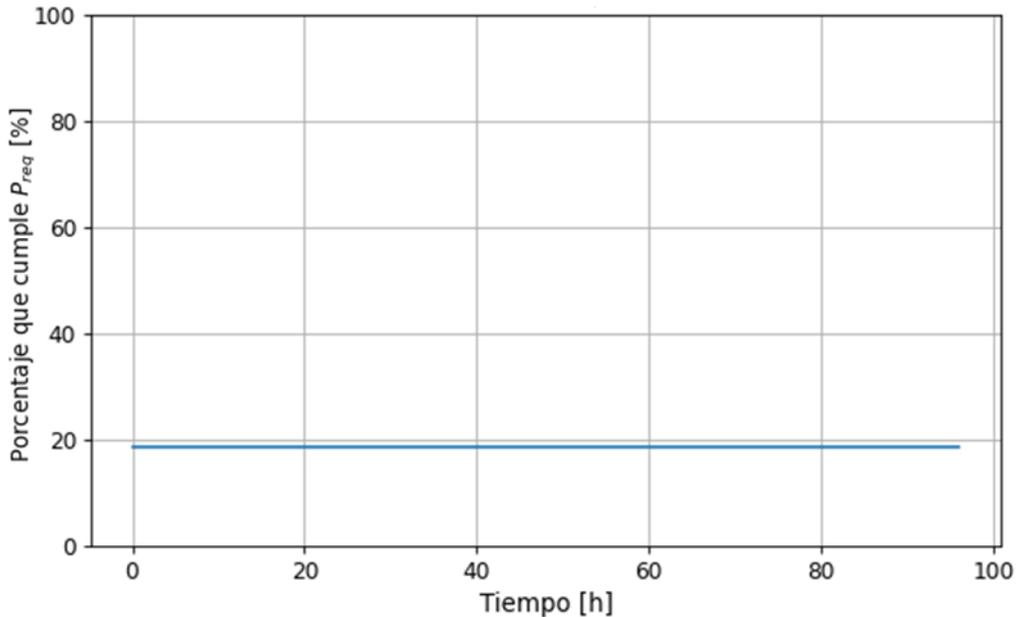


Figura 4.44: Porcentaje de cumplimiento de presiones en condiciones actuales, red experimental, caso sólo deterioro.

Finalmente, en la Figura 4.45 se muestra la distribución espacial de los nodos que no cumplen con la presión mínima:

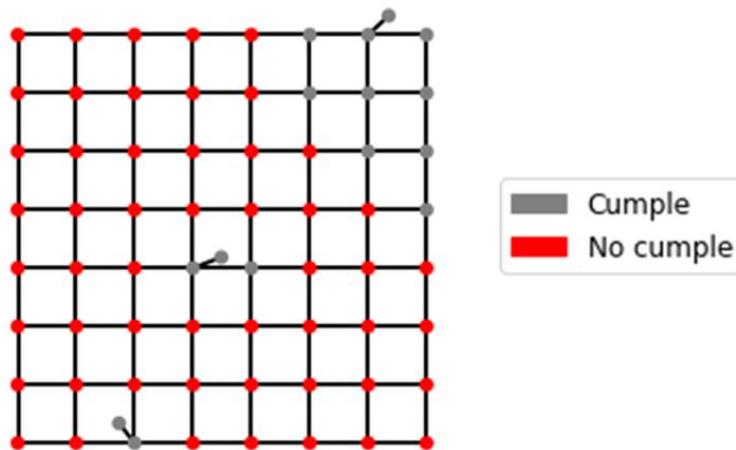


Figura 4.45: Distribución espacial nodos que no cumplen presión requerida en condiciones actuales, red experimental, caso sólo deterioro.

4.2.1.4 Escenarios futuros

Simulación a cinco años

En la Figura 4.46 se muestran los resultados de la simulación hidráulica para el escenario a cinco años, considerando sólo deterioro, según las fallas de la Figura 4.40 (b):

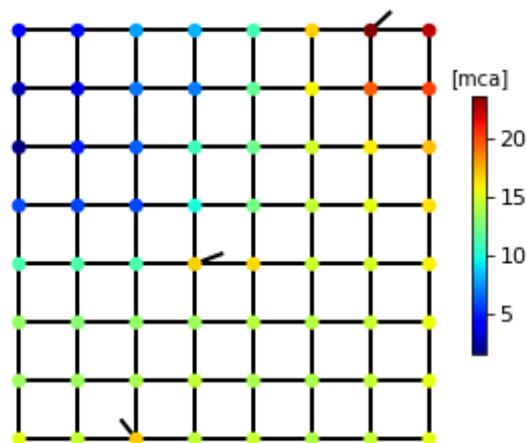


Figura 4.46: Presión promedio escenario futuro año 5, red experimental, caso sólo deterioro.

A continuación se muestran los resultados de las métricas hidráulicas. En la Figura 4.47 se muestra la evolución de la disponibilidad de agua durante el tiempo de estudio, observándose una cobertura similar al escenario actual pese al envejecimiento.

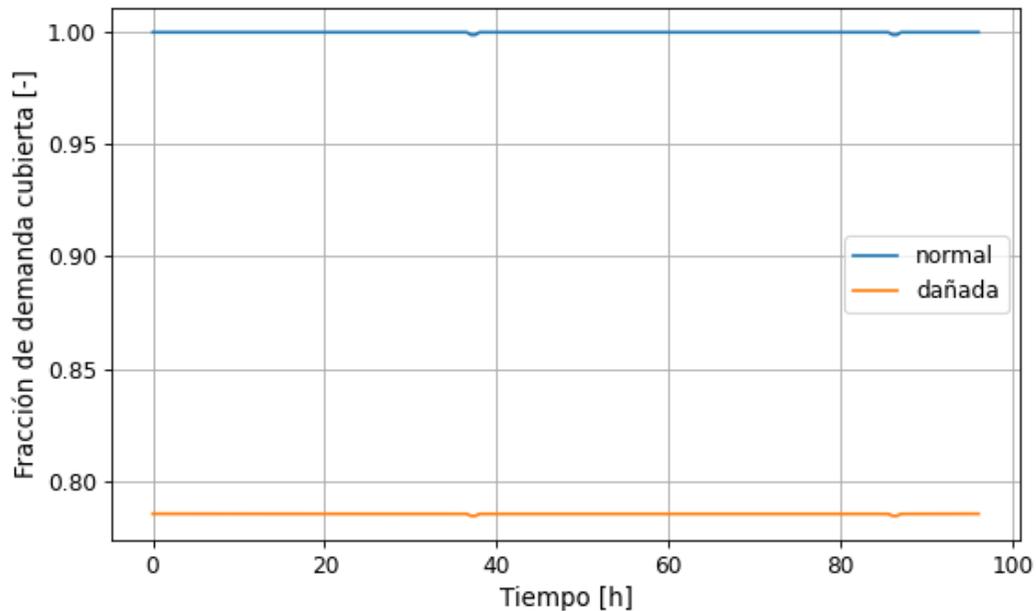


Figura 4.47: Fracción de demanda cubierta escenario futuro año 5, red experimental, caso sólo deterioro.

En la Figura 4.48 se muestra el porcentaje de la red que cumple con la presión requerida, presentando un cumplimiento mayor que el escenario actual pese al envejecimiento.

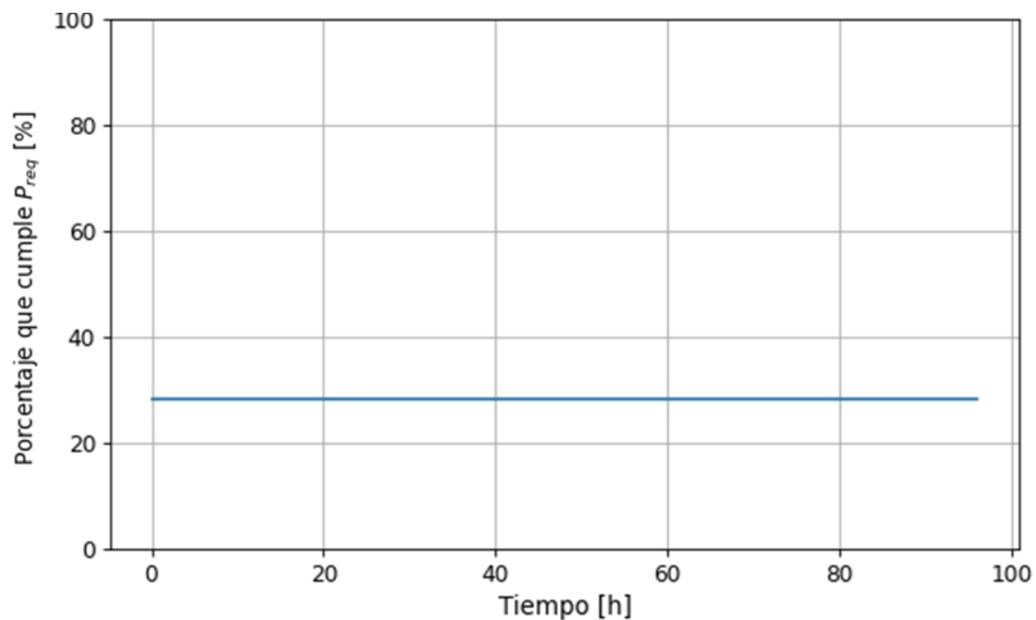


Figura 4.48: Porcentaje de cumplimiento de presiones escenario futuro año 5, red experimental, caso sólo deterioro.

En la Figura 4.49 se muestra la distribución espacial de los nodos que no cumplen con la presión mínima:

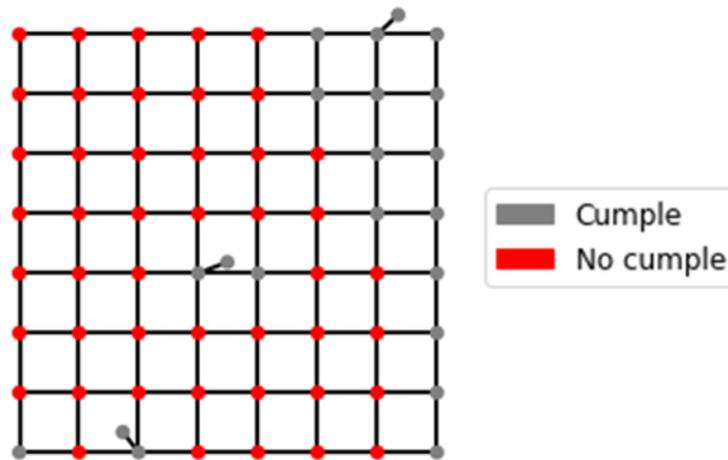


Figura 4.49: Distribución espacial nodos que no cumplen presión requerida escenario futuro año 5, red experimental, caso sólo deterioro.

Simulación a diez años

En la Figura 4.50 se muestran los resultados de la simulación hidráulica para el escenario actual, considerando sólo deterioro, según las fallas de la Figura 4.40 (c).

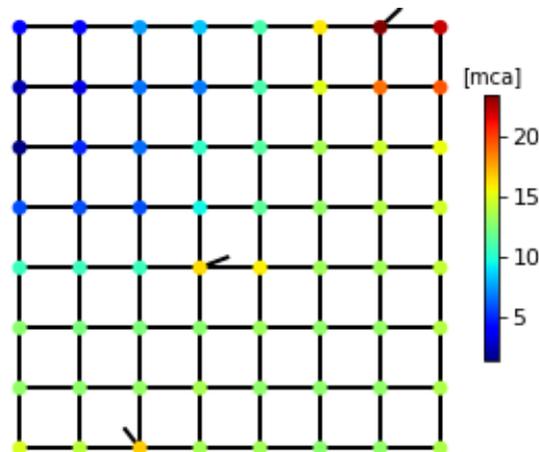


Figura 4.50: Presión promedio escenario futuro año 10, red experimental, caso sólo deterioro.

A continuación se muestran los resultados de las métricas hidráulicas. En la Figura 4.51 se muestra la evolución de la disponibilidad de agua durante el tiempo de estudio, donde se registra la menor cobertura si se compara con las situaciones anteriores, efecto atribuible al envejecimiento de la red.

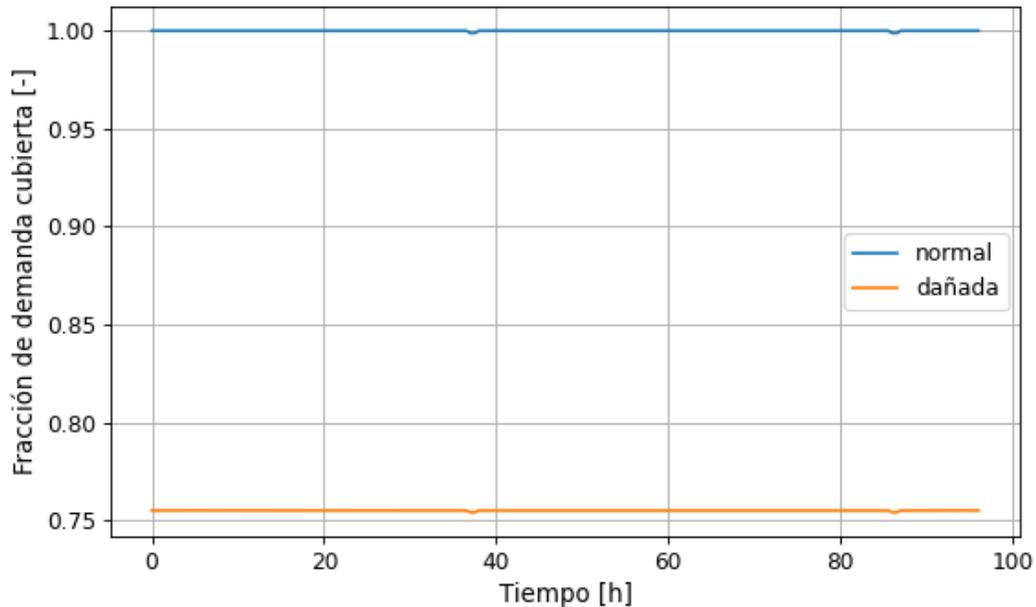


Figura 4.51: Fracción de demanda cubierta escenario futuro año 10, red experimental, caso sólo deterioro.

En la Figura 4.52 se muestra el porcentaje de la red que cumple con la presión requerida, observándose en este caso el menor porcentaje de cumplimiento.

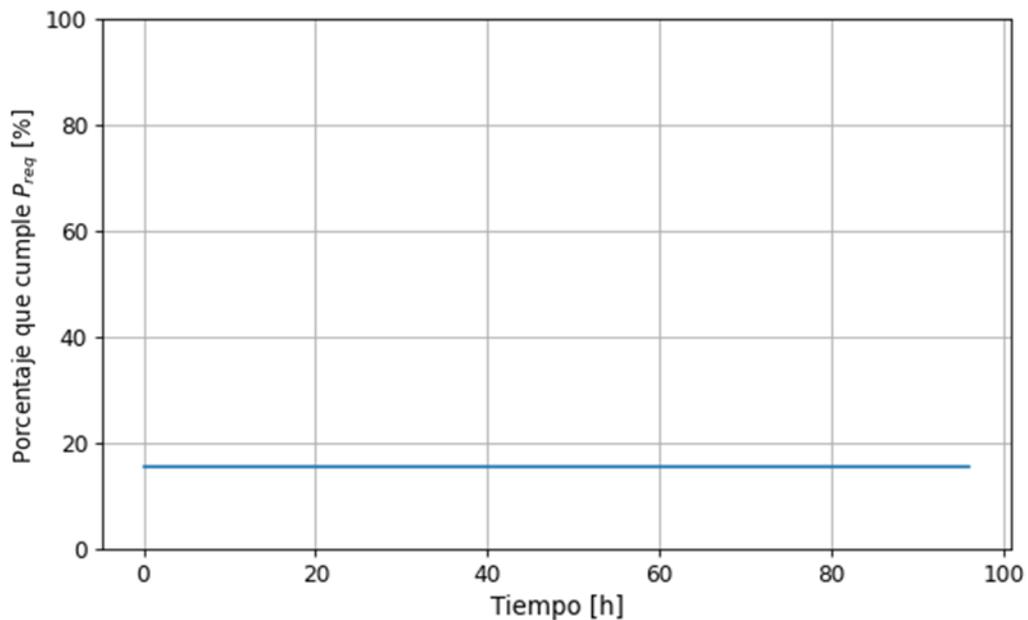


Figura 4.52: Porcentaje de cumplimiento de presiones escenario futuro año 10, red experimental, caso sólo deterioro.

Finalmente, en la Figura 4.53 se muestra la distribución espacial de los nodos que no cumplen con la presión mínima.

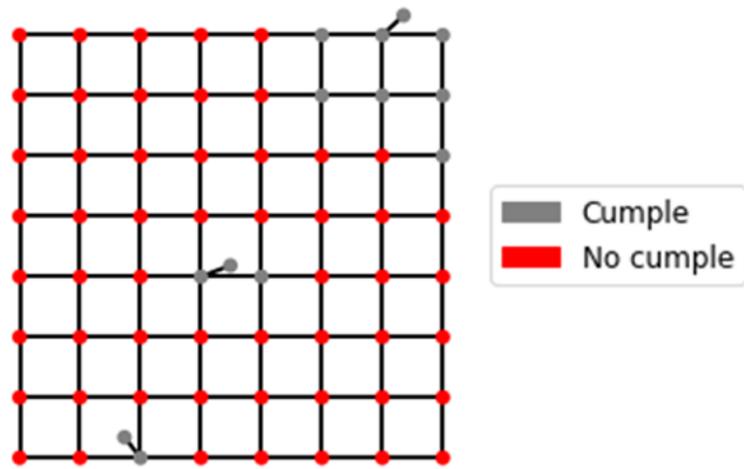


Figura 4.53: Distribución espacial nodos que no cumplen presión requerida escenario futuro año 10, red experimental, caso sólo deterioro.

4.2.2 Incorporación del deterioro en un modelo hidráulico, considerando efectos sísmicos

Se presentan los resultados de la integración del modelo de deterioro y sísmico en el análisis hidráulico. En la Figura 4.54 se muestra la distribución espacial de las características de la red (cabe recordar que para este caso la red es completa de asbesto cemento):

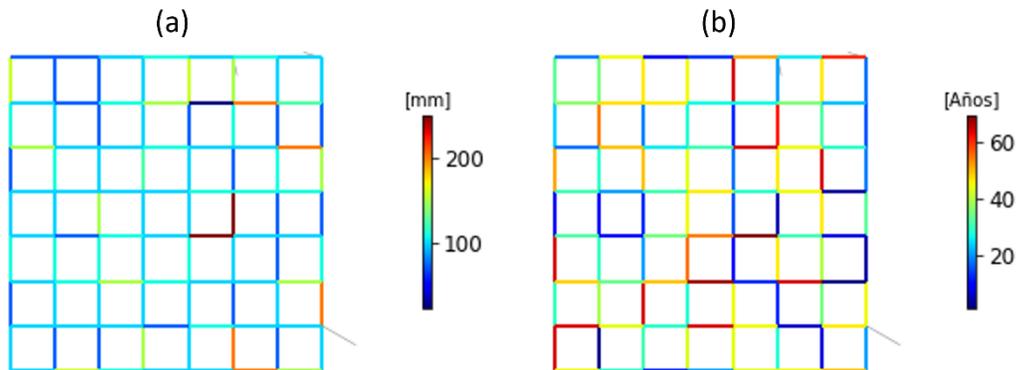


Figura 4.54: Descripción de la red experimental, caso con deterioro y sismo. (a) Diámetros, (b) Edad.

La distribución de elevaciones en la red experimental se muestra en la Figura 4.55:

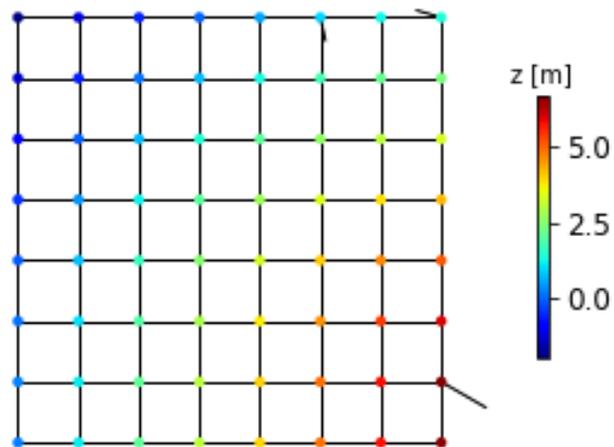


Figura 4.55: Elevación red experimental, caso con deterioro y sismo.

4.2.2.1 Resultados modelo de deterioro

En la Tabla 4.5 se presentan los resultados de los modelos de deterioro para la red descrita.

Tabla 4.5: Comparación métricas algoritmos de inteligencia artificial, caso con deterioro y sismo.

	TPR	TNR	Precision	FPR	FNR	Accuracy	Balanced Accuracy	F1 Score	AUC
GBT	68.7%	64.3%	82.1%	35.7%	31.3%	67.4%	66.5%	74.8%	74.8%
SVM	64.2%	64.3%	81.1%	35.7%	35.8%	64.2%	64.2%	71.7%	72.1%
MLP	64.2%	64.3%	81.1%	35.7%	35.8%	64.2%	64.2%	71.7%	72.9%

Según las métricas de la Tabla 4.5, el modelo con mejor desempeño (mayor TPR, menor FNR y mayor ROC AUC) es el GBT. Con dicho algoritmo se simulan las fallas y se desarrolla el modelamiento hidráulico.

Con los resultados del modelo, se obtienen las probabilidades de falla (Figura 4.56) y las fallas efectivas (Figura 4.57) para la condición actual, la condición a cinco años y la condición a diez años.

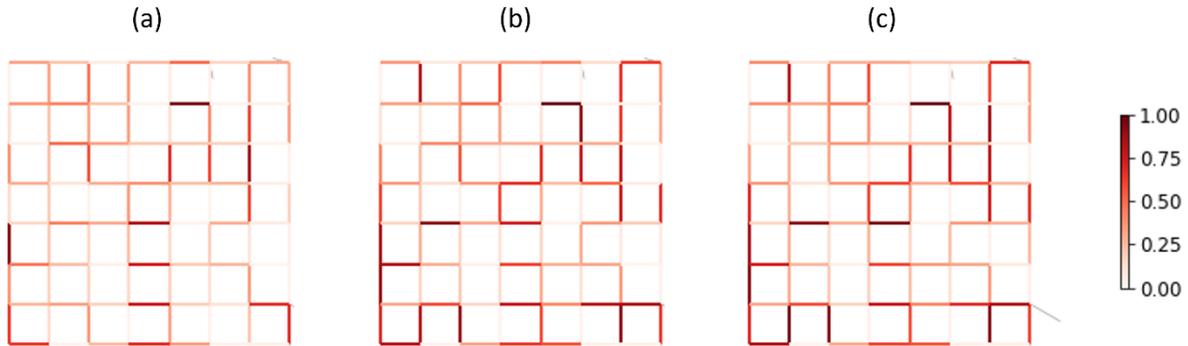


Figura 4.56: Probabilidades de falla, red experimental, situación con deterioro y sismo. (a) Condición actual, (b) Condición a 5 años, (c) Condición a 10 años.

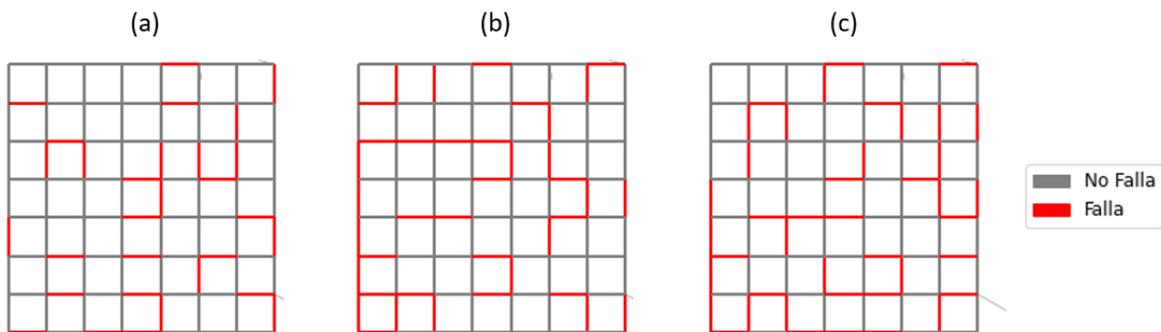


Figura 4.57: Fallas por deterioro, red experimental, situación con deterioro y sismo. (a) Condición actual, (b) Condición a 5 años, (c) Condición a 10 años.

Para observar la evolución de la probabilidad de falla en la red se muestra la probabilidad de falla promedio de la red para los tres años de estudio en la Tabla 4.6:

Tabla 4.6: Evolución probabilidad de falla promedio, red experimental, situación con deterioro y sismo.

$P_{falla}(t = 0)$	$P_{falla}(t = 5)$	$P_{falla}(t = 10)$
26,9%	35,8%	36,4%

4.2.2.2 Efectos del sismo

Para efectos del desarrollo de esta sección, se definen dos escenarios sísmicos. Se nombra al escenario con 43,7% Escenario sísmico 1, y al escenario con $4,37 \cdot 10^{-3}\%$ Escenario sísmico 2.

Se muestran en la Figura 4.58 los efectos de ambos escenarios sísmicos en la red:

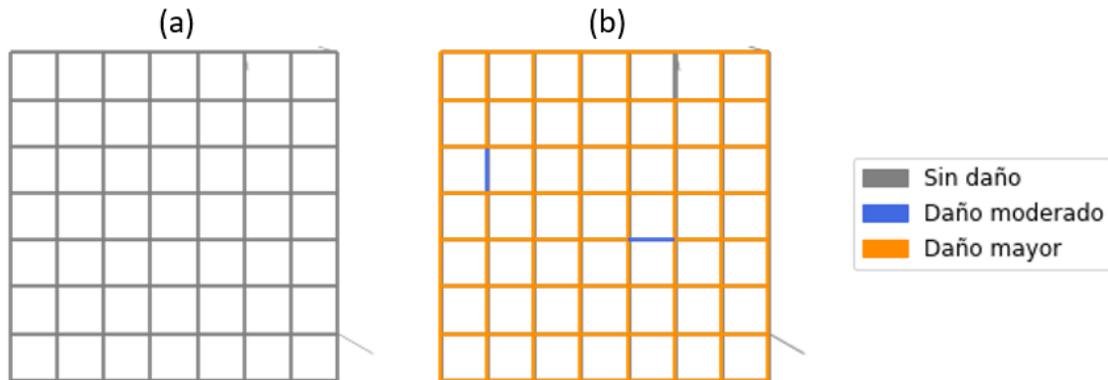


Figura 4.58: Fallas por sismo, red experimental, situación con deterioro y sismo. (a) Escenario sísmico 1, (b) Escenario sísmico 2.

4.2.2.3 Comportamiento base

Se presenta el comportamiento base de la red, es decir, sin considerar las fallas por deterioro. En la Figura 4.59 se muestran los valores medios de presión en los nodos de la red:

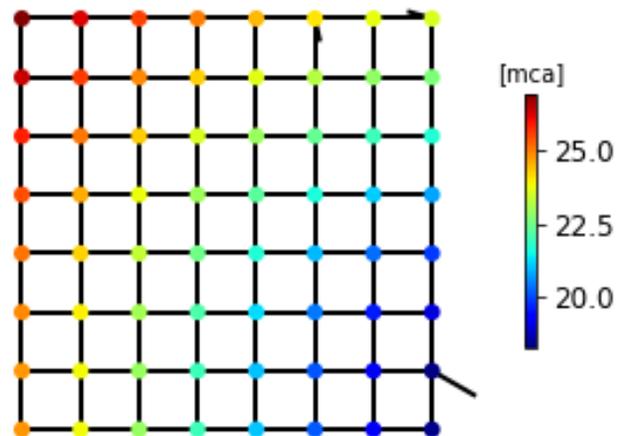


Figura 4.59: Presión promedio en condiciones normales, red experimental, situación con deterioro y sismo.

4.2.2.4 Escenario sísmico 1

Situación actual:

Se muestran los resultados de la simulación hidráulica para el escenario actual, considerando deterioro y el escenario sísmico 2, según las fallas de la Figura 4.57 (a) y Figura 4.58 (a):

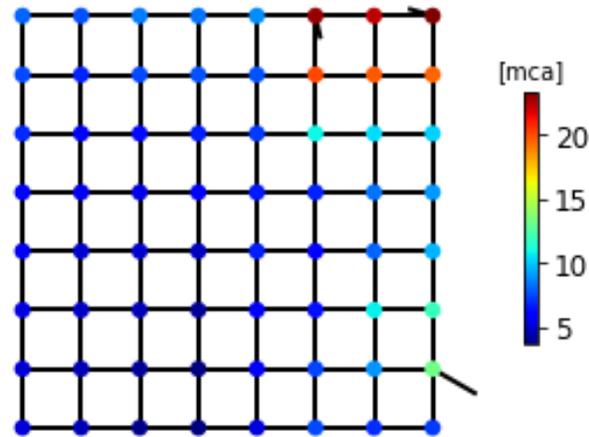


Figura 4.60: Presión promedio en condiciones actuales, red experimental, escenario sísmico 1.

A continuación se muestran los resultados de las métricas hidráulicas. En la Figura 4.61 se muestra la evolución de la disponibilidad de agua durante el tiempo de estudio, donde el 75% de la demanda está satisfecha.

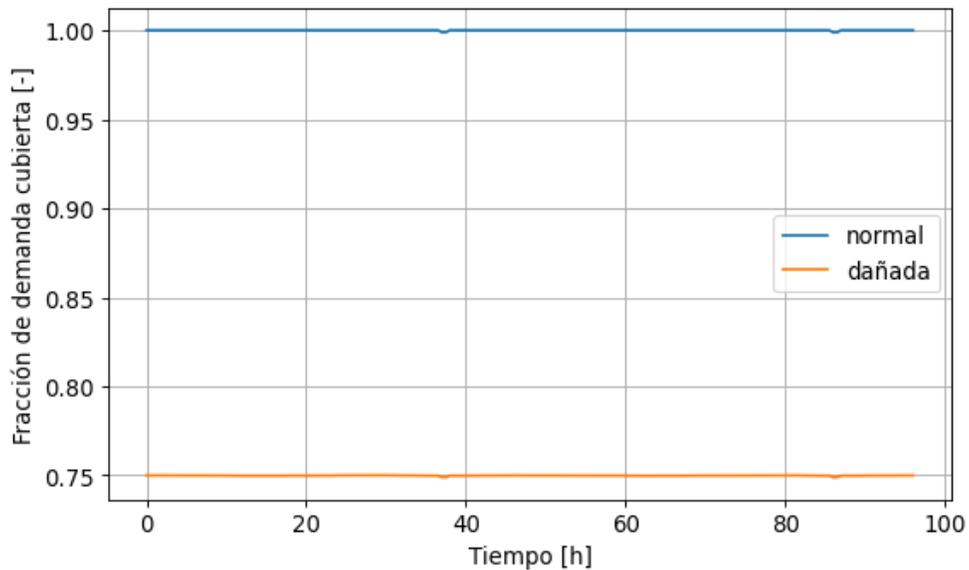


Figura 4.61: Fracción de demanda cubierta en condiciones actuales, red experimental, escenario sísmico 1.

En la Figura 4.62 se muestra el porcentaje de la red que cumple con la presión requerida, donde alrededor del 10% de los nodos cumplen con los 15 [m] de presión.

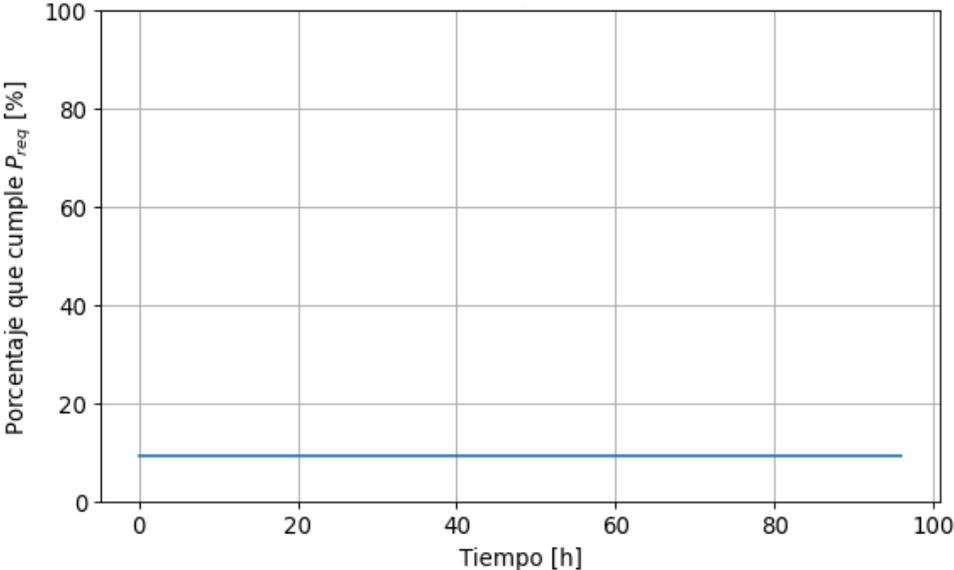


Figura 4.62: Porcentaje de cumplimiento de presiones en condiciones actuales, red experimental, escenario sísmico 1.

Finalmente, en la Figura 4.63 se muestra la distribución espacial de los nodos que no cumplen con la presión mínima:

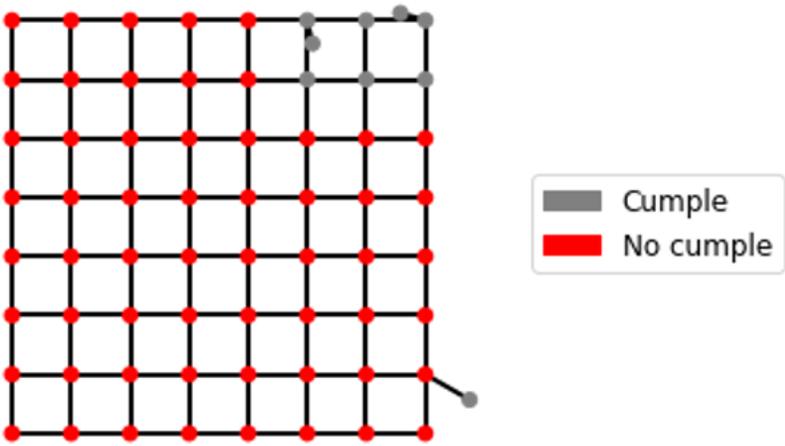


Figura 4.63: Distribución espacial nodos que no cumplen presión requerida en condiciones actuales, red experimental, escenario sísmico 1.

Situación a cinco años:

Se muestran los resultados de la simulación hidráulica para el escenario actual, considerando deterioro y el escenario sísmico 1, según las fallas de la Figura 4.57 (b) y Figura 4.58 (a):

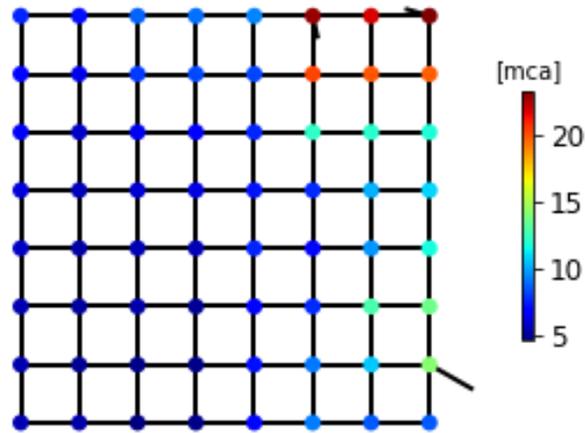


Figura 4.64: Presión promedio escenario futuro año 5, red experimental, escenario sísmico 1.

A continuación se muestran los resultados de las métricas hidráulicas. En la Figura 4.65 se muestra la evolución de la disponibilidad de agua durante el tiempo de estudio, con poco menos de un 50% de la demanda cubierta.

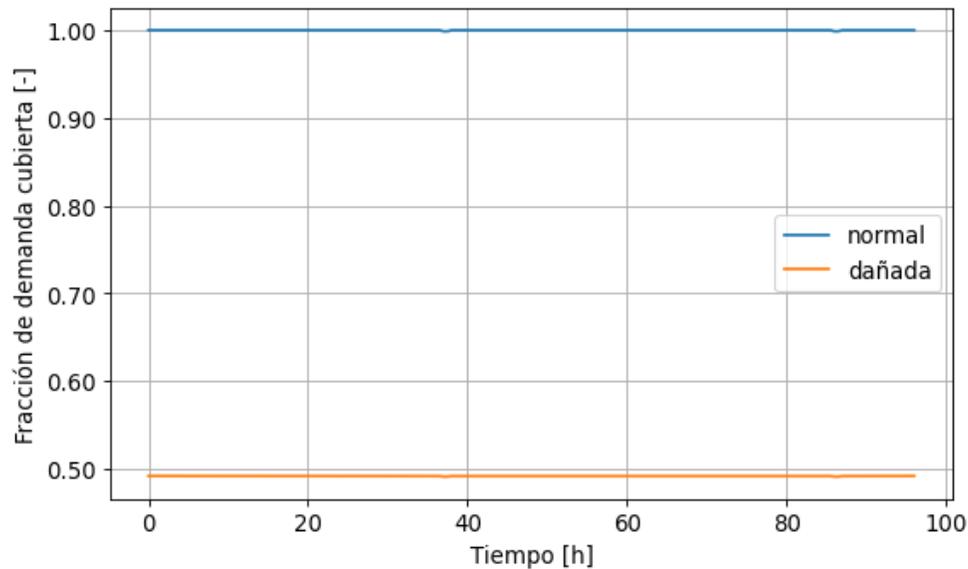


Figura 4.65: Fracción de demanda cubierta escenario futuro año 5, red experimental, escenario sísmico 1.

En la Figura 4.66 se muestra el porcentaje de la red que cumple con la presión requerida, donde alrededor del 10% de los nodos cumple con los 15 [m] de presión. Cabe señalar que para un porcentaje similar de cumplimiento, en este caso hay una menor fracción de la demanda cubierta.

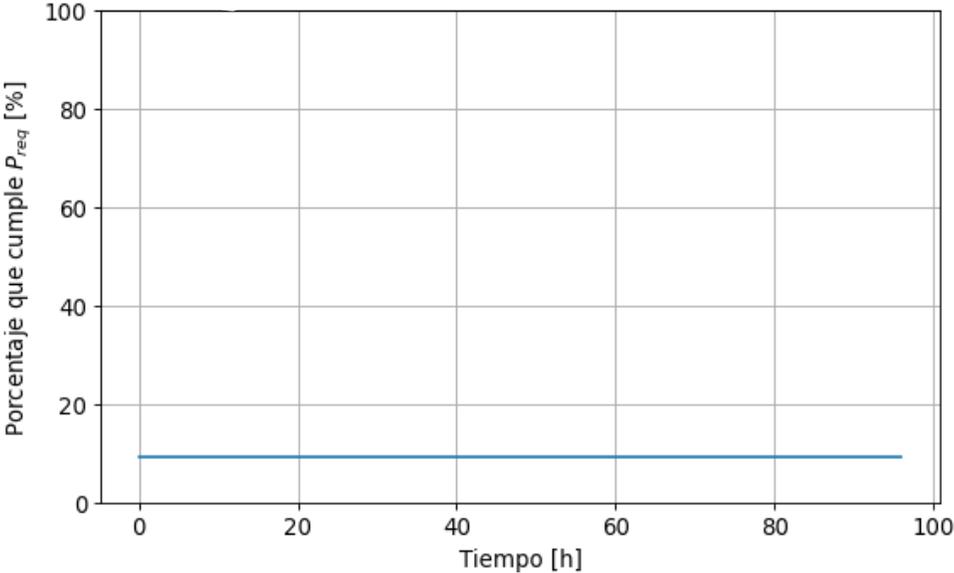


Figura 4.66: Porcentaje de cumplimiento de presiones escenario futuro año 5, red experimental, escenario sísmico 1.

Finalmente, en la Figura 4.67 se muestra la distribución espacial de los nodos que no cumplen con la presión mínima:

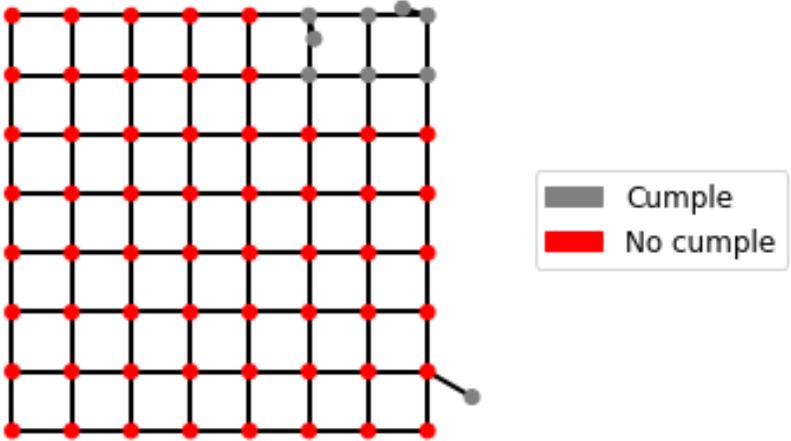


Figura 4.67: Distribución espacial nodos que no cumplen presión requerida escenario futuro año 5, red experimental, escenario sísmico 1.

Situación a diez años:

Se muestran los resultados de la simulación hidráulica para el escenario actual, considerando deterioro y el escenario sísmico 1, según las fallas de la Figura 61 (c) y Figura 4.58 (a):

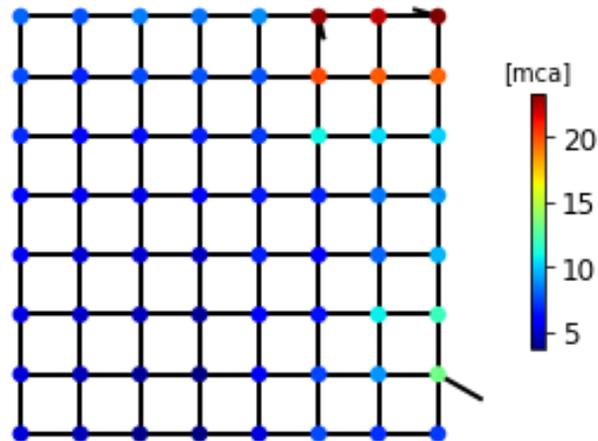


Figura 4.68: Presión promedio escenario futuro año 10, red experimental, escenario sísmico 1.

A continuación se muestran los resultados de las métricas hidráulicas. En la Figura 4.69 se muestra la evolución de la disponibilidad de agua durante el tiempo de estudio, donde se tiene la menor cobertura de los tres escenarios temporales analizados.

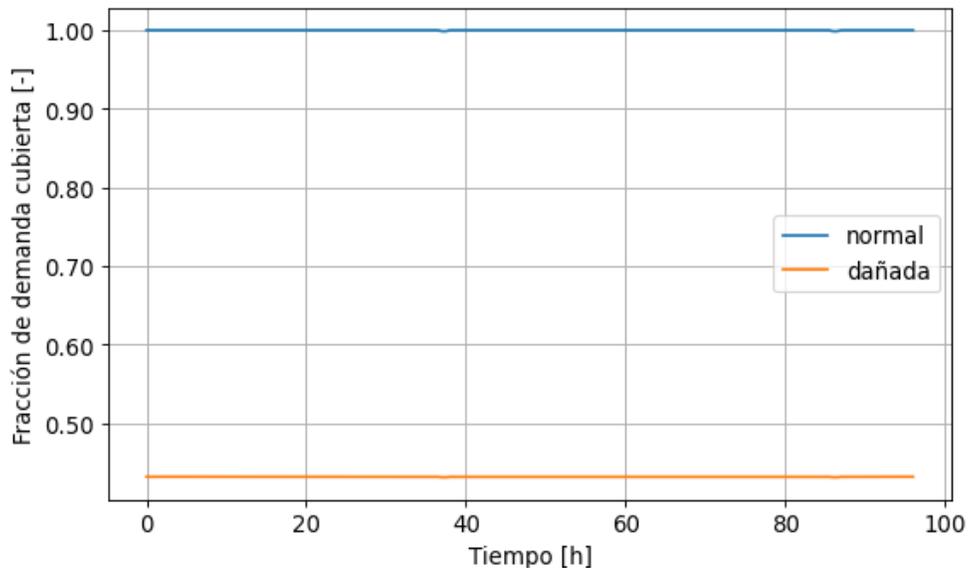


Figura 4.69: Fracción de demanda cubierta escenario futuro año 10, red experimental, escenario sísmico 1.

En la Figura 4.70 se muestra el porcentaje de la red que no cumple con la presión requerida, observándose un porcentaje de cumplimiento similar al de las situaciones actual y a cinco años.

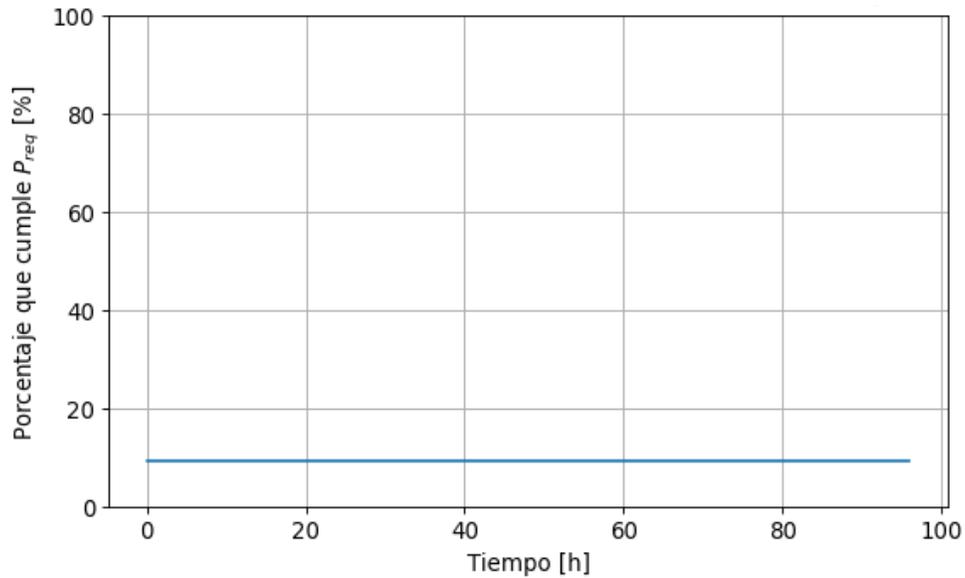


Figura 4.70: Porcentaje de cumplimiento de presiones escenario futuro año 10, red experimental, escenario sísmico 1.

Finalmente, en la Figura 4.71 se muestra la distribución espacial de los nodos que no cumplen con la presión mínima:

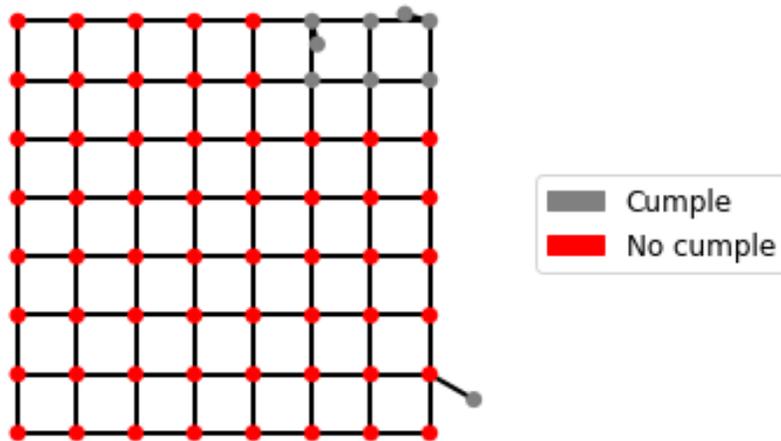


Figura 4.71: Distribución espacial nodos que no cumplen presión requerida escenario futuro año 10, red experimental, escenario sísmico 1.

En general, para el escenario sísmico 1, no se observan efectos de la sismicidad en la red dada la alta probabilidad de excedencia del sismo analizado.

4.2.2.5 Escenario sísmico 2

Situación actual:

Se muestran los resultados de la simulación hidráulica para el escenario actual, considerando deterioro y el escenario sísmico 2, según las fallas de la Figura 4.57 (a) y Figura 4.58 (b):

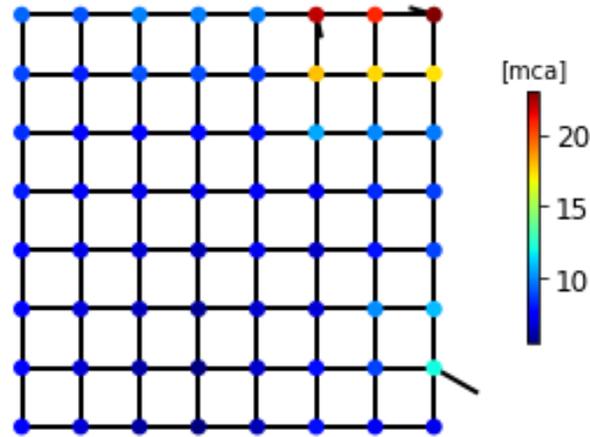


Figura 4.72: Presión promedio en condiciones actuales, red experimental, escenario sísmico 2.

A continuación se muestran los resultados de las métricas hidráulicas. En la Figura 4.73 se muestra la evolución de la disponibilidad de agua durante el tiempo de estudio, la cual se encuentra en general bajo el 80% y, cuando inician los efectos sísmicos, la cobertura disminuye casi a cero.

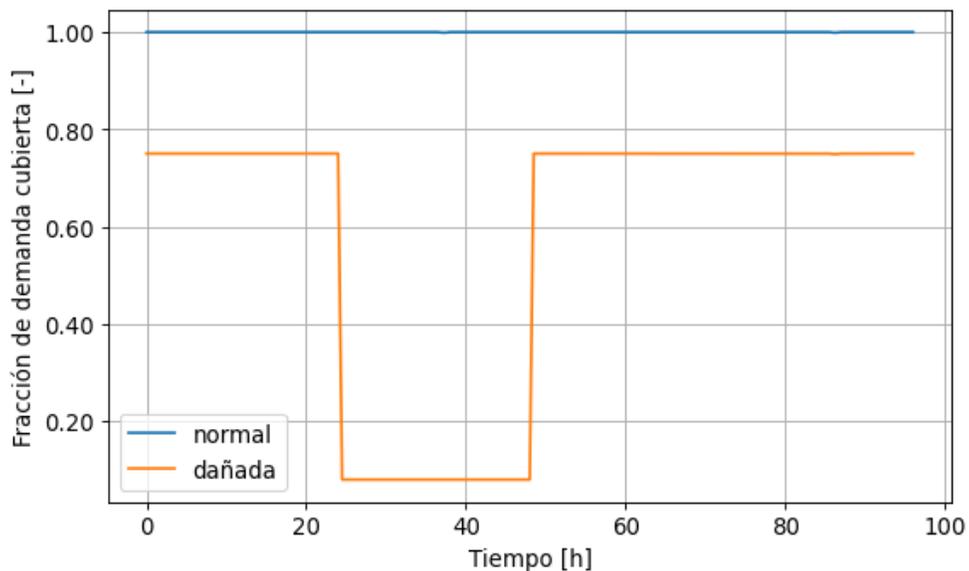


Figura 4.73: Fracción de demanda cubierta en condiciones actuales, red experimental, escenario sísmico 2.

En la Figura 4.74 se muestra el porcentaje de la red que cumple con la presión requerida, observándose que en el momento en que inician los efectos del sismo, al igual que con la demanda cubierta, el cumplimiento de presiones cae casi a cero.

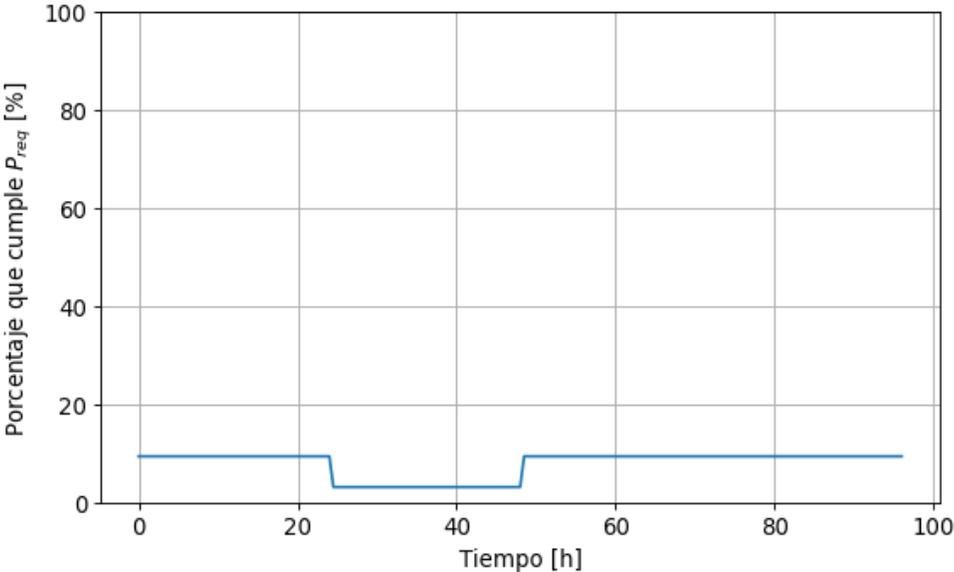


Figura 4.74: Porcentaje de cumplimiento de presiones en condiciones actuales, red experimental, escenario sísmico 2.

Finalmente, en la Figura 4.75 se muestra la distribución espacial de los nodos que no cumplen con la presión mínima:

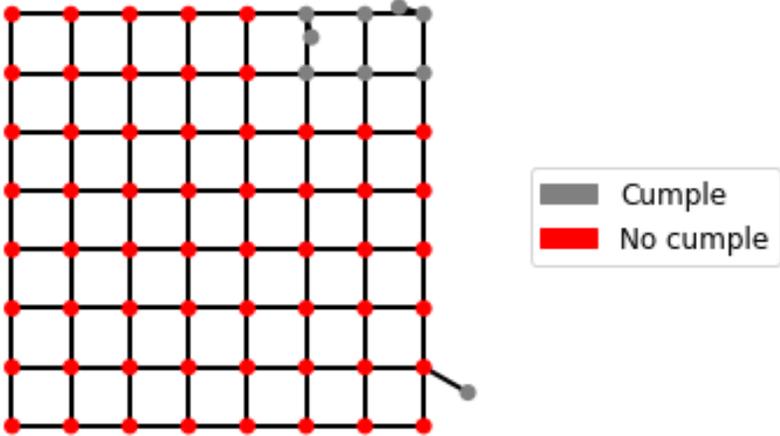


Figura 4.75: Distribución espacial nodos que no cumplen presión requerida en condiciones actuales, red experimental, escenario sísmico 2.

Situación a cinco años:

Se muestran los resultados de la simulación hidráulica para el escenario actual, considerando deterioro y el escenario sísmico 2, según las fallas de la Figura 4.57 (b) y Figura 4.58 (b):

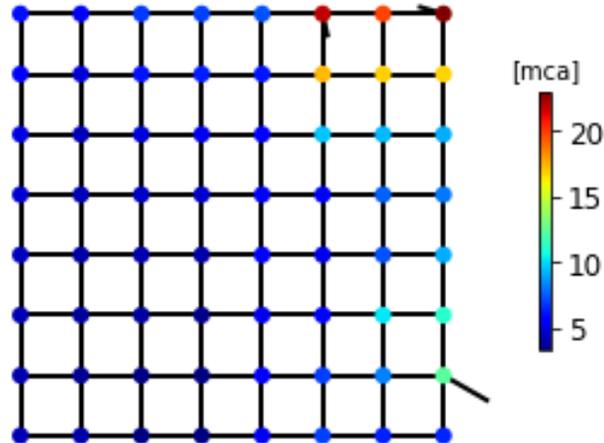


Figura 4.76: Presión promedio escenario futuro año 5, red experimental, escenario sísmico 2.

A continuación se muestran los resultados de las métricas hidráulicas. En la Figura 4.77 se muestra la evolución de la disponibilidad de agua durante el tiempo de estudio, la cual se encuentra en general bajo el 80% y, cuando inician los efectos sísmicos, la cobertura disminuye casi a cero.

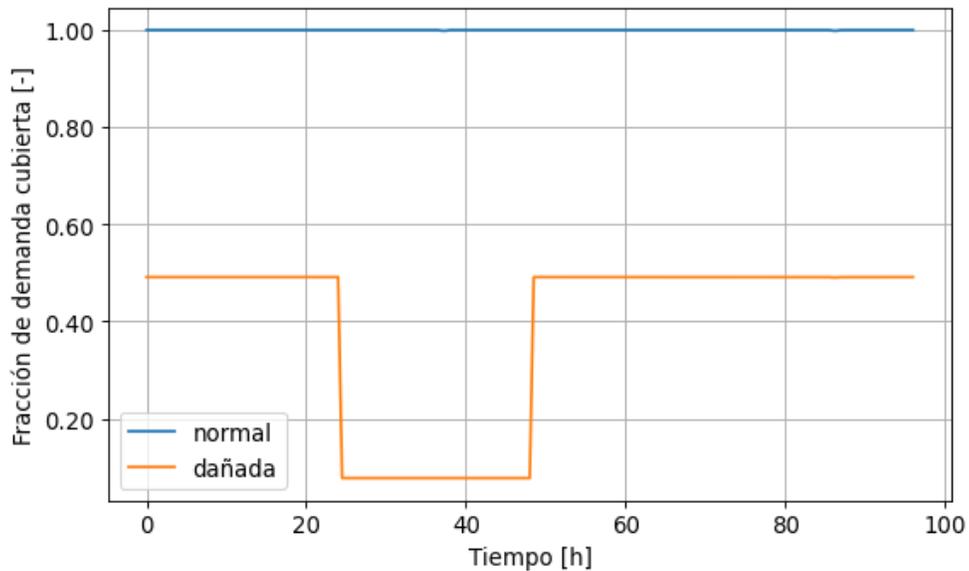


Figura 4.77: Fracción de demanda cubierta escenario futuro año 5, red experimental, escenario sísmico 2.

En la Figura 4.78 se muestra el porcentaje de la red que cumple con la presión requerida, observándose que en el momento en que inician los efectos del sismo, al igual que con la demanda cubierta, el cumplimiento de presiones cae casi a cero.

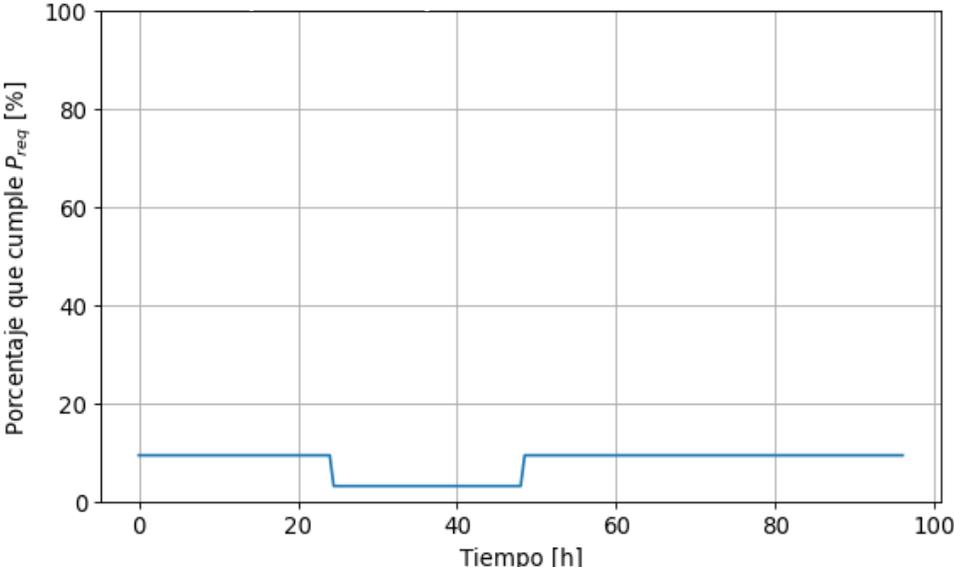


Figura 4.78: Porcentaje de cumplimiento de presiones escenario futuro año 5, red experimental, escenario sísmico 2.

Finalmente, en la Figura 4.79 se muestra la distribución espacial de los nodos que no cumplen con la presión mínima:

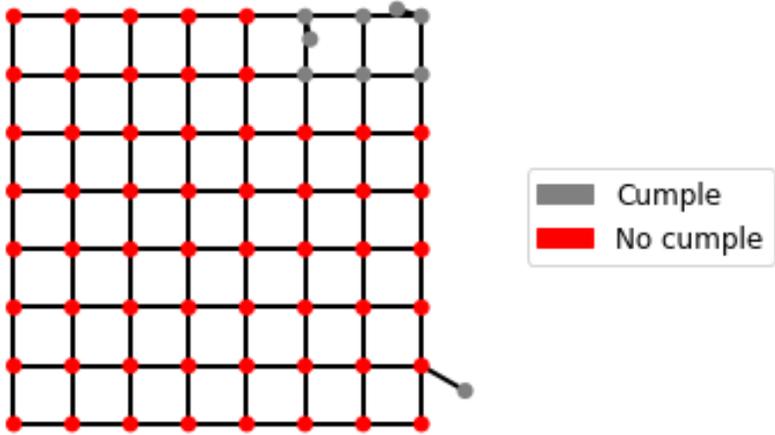


Figura 4.79: Distribución espacial nodos que no cumplen presión requerida escenario futuro año 5, red experimental, escenario sísmico 2.

Situación a diez años:

Se muestran los resultados de la simulación hidráulica para el escenario actual, considerando deterioro y el escenario sísmico 2, según las fallas de la Figura 61 (c) y Figura 4.58 (b):

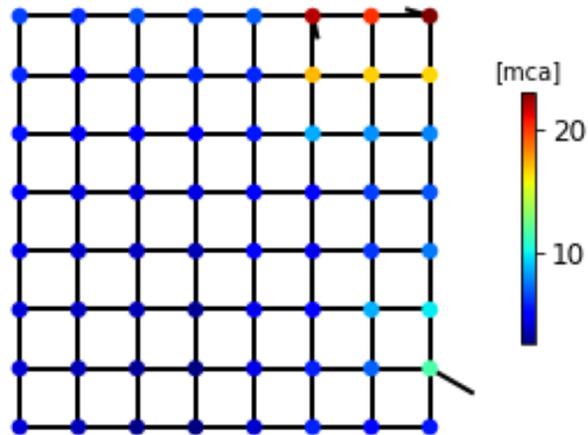


Figura 4.80: Presión promedio escenario futuro año 10, red experimental, escenario sísmico 2.

A continuación se muestran los resultados de las métricas hidráulicas. En la Figura 4.81 se muestra la evolución de la disponibilidad de agua durante el tiempo de estudio, la cual se encuentra en general cerca del 40% y, cuando inician los efectos sísmicos, la cobertura disminuye casi a cero.

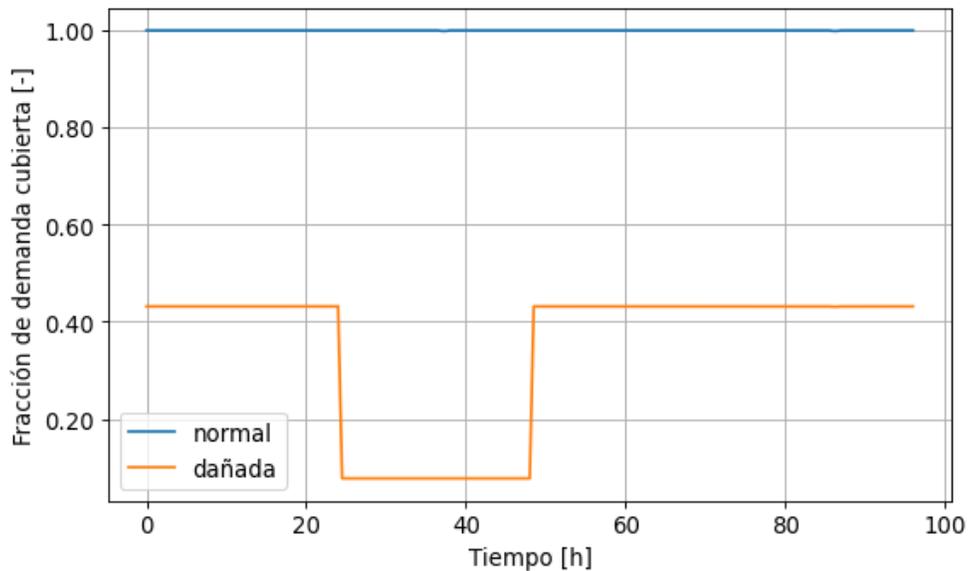


Figura 4.81: Fracción de demanda cubierta escenario futuro año 10, red experimental, escenario sísmico 2.

En la Figura 4.82 se muestra el porcentaje de la red que cumple con la presión requerida, observándose que en el momento en que inician los efectos del sismo, al igual que con la demanda cubierta, el cumplimiento de presiones cae casi a cero.

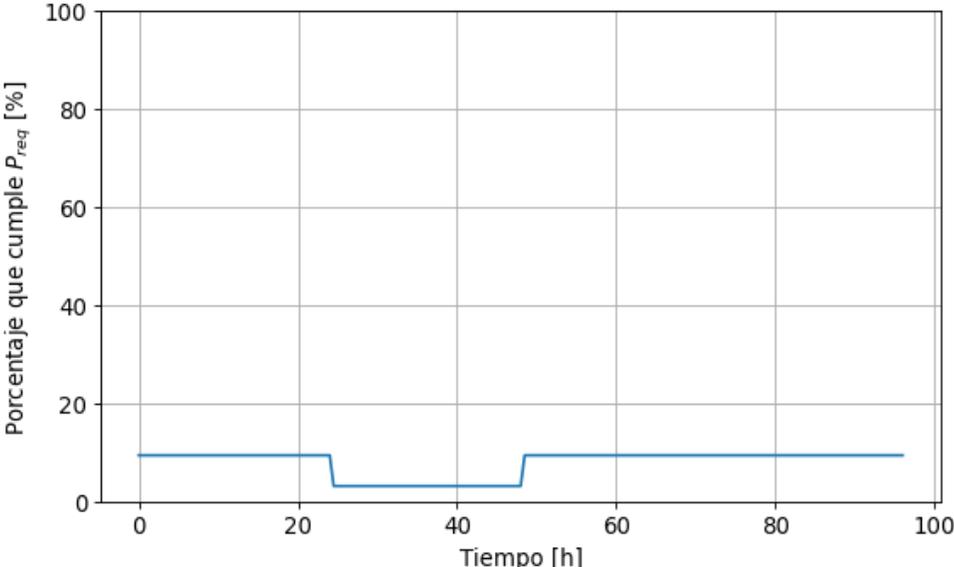


Figura 4.82: Porcentaje de cumplimiento de presiones escenario futuro año 10, red experimental, escenario sísmico 2.

Finalmente, en la Figura 4.83 se muestra la distribución espacial de los nodos que no cumplen con la presión mínima:

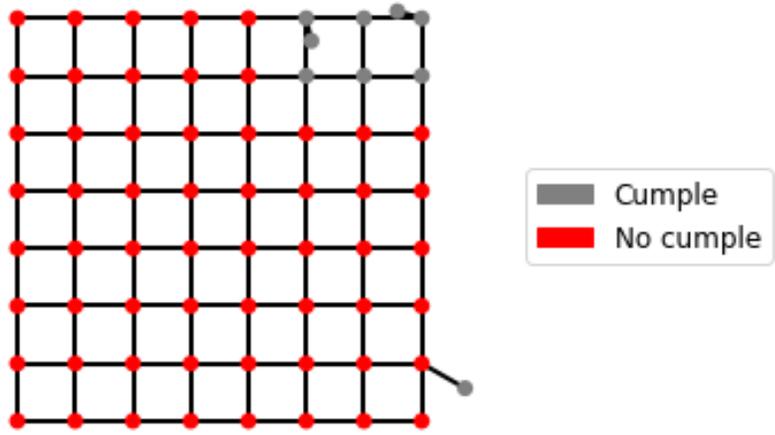


Figura 4.83: Distribución espacial nodos que no cumplen presión requerida escenario futuro año 10, red experimental, escenario sísmico 2.

En general, se observa que los efectos sísmicos de un evento con tan baja probabilidad de excedencia

5 Análisis de resultados

Se realiza el análisis de los resultados para los modelos de inteligencia artificial aplicados al caso de estudio (sección 4.1). En los resultados de los tres algoritmos se observa, para las configuraciones sin remuestreo, que los modelos prácticamente predicen que no hay fallas (ver Figura 4.5, Figura 4.7, Figura 4.17, Figura 4.19) debido al gran desbalanceo de los datos.

Por otro lado, para los casos con remuestreo, se observa una mejora considerable en la predicción de fallas, a la vez que hay una disminución en la precisión de la predicción de tuberías en buen estado (ver Figura 4.9, Figura 4.11, Figura 4.21, Figura 4.23, Figura 4.29, Figura 4.31). Cabe señalar que, para este caso, la curva ROC parece no ser una buena medida del desempeño de los modelos, con valores cercanos a 0,9 en la mayoría de los casos analizados (recordar que mientras más cercano a uno, mejor es la capacidad del modelo).

Al analizar los resultados de los algoritmos para el análisis de la comuna de La Florida, destaca el mal desempeño de la SVM, el que es particularmente apreciable en la Figura 4.34 y Figura 4.36, donde se observa que el modelo prácticamente sigue la línea de predicción aleatoria.

Sobre la métrica a usar para el proceso de búsqueda por grilla, se observa que no hay diferencias considerables en los resultados al optimizar según *Accuracy* o *Balanced Accuracy*. Sin embargo, se destaca la importancia de determinar la bondad del modelo según la métrica apropiada dado que, para los casos sin remuestreo, los valores de *Accuracy* están por sobre el 97%, mientras que los de *Balanced Accuracy* están en general por debajo del 60%.

Para el caso de sectorización por comuna no hay una diferencia apreciable en la predicción de fallas al usar *Balanced Accuracy* o *F1 Score*, pero se aprecia una mayor diferencia en la predicción de tuberías en buen estado (Figura 4.25 y Figura 4.27).

En cuanto al desempeño que consiguen en general los tres algoritmos utilizados, se observa que todos presentan una baja capacidad predictiva, con precisiones que van entre un 3% y un 71%. Esto se contradice con los resultados obtenidos por Bubiña et al. (2011), Giraldo-González & Rodríguez (2020), Kutyłowska (2015) entre otros que usaron metodologías y algoritmos similares a los presentados en este trabajo.

La causa principal del bajo desempeño de los algoritmos se relaciona con la calidad de los datos entregados al modelo. Como se señala en la sección 3.1, la base de datos incluye la red en condiciones actuales, es decir, supone que al momento de la modelación la red está 100% operativa, implicando una adición de alrededor de 290.000 datos de tuberías sin fallas. Dado que el registro es desde el año 2017 hasta mediados de 2021, la muestra queda muy desbalanceada. Si bien este problema se puede palear con las metodologías

de remuestreo usadas en los casos 3 y 4, el procedimiento no consigue llevar los resultados a niveles como los observados en la bibliografía.

En la misma línea de la calidad de los datos entregados al modelo, los datos de la red proporcionados por la sanitaria a la SISS en el PR012 parecen imprecisos si se estudian los años de instalación de las tuberías (Figura 5.1), observándose máximos en la instalación de tuberías de PVC en el año 1957 y la instalación de tuberías de asbesto cemento posterior al 2001, año en que se prohibió el uso de cualquier material que contuviese asbesto (decreto N°656 del MINSAL). Esto, sumado a los posibles errores en la asignación de las fallas a las tuberías (Figura 3.1), añade mucha incertidumbre a los datos entregados al modelo y puede ser la causa del bajo rendimiento.

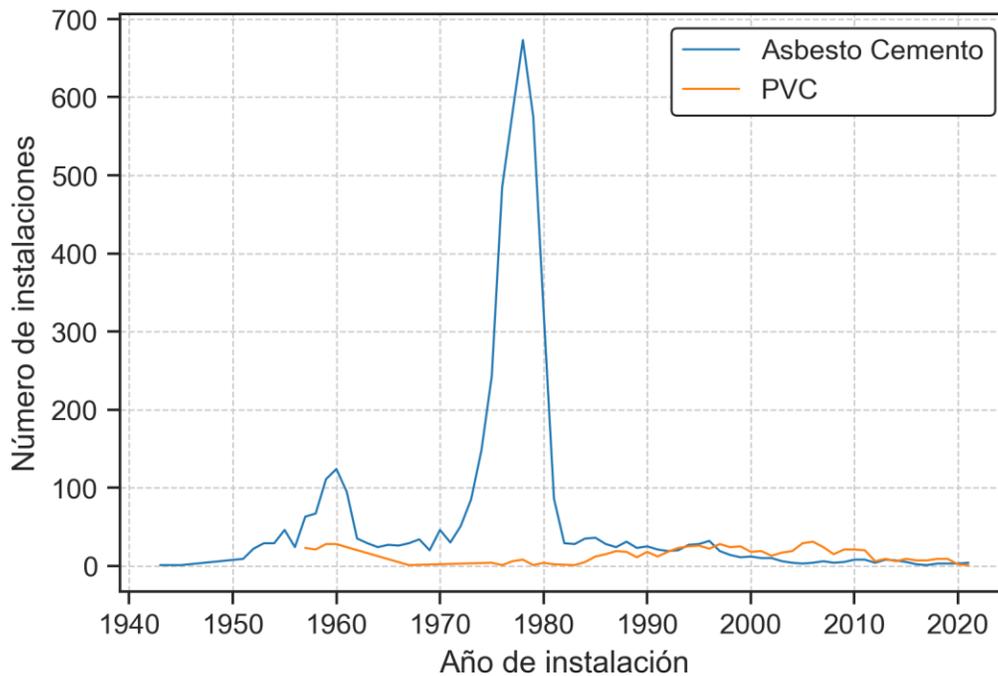


Figura 5.1: Evolución de la instalación de tuberías de asbesto cemento y PVC en la red estudiada.

Una de las opciones que no fue evaluada en este trabajo y que podría mejorar el desempeño en algunos casos es la de realizar un modelo de deterioro para cada material de la red. Este procedimiento es descrito en Giraldo-González & Rodríguez (2020) donde separan la red en tuberías de PVC y de asbesto cemento, haciendo un modelo de deterioro para cada uno de ellos debido a que presentan patrones de falla distintos. Sin embargo, con dicho procedimiento se pierde el impacto del reemplazo de tuberías con distinto material además de generar un número muy grande de modelos (la red de Santiago cuenta con nueve materiales distintos registrados), de los cuales no todos necesariamente contarán con registros suficientemente largos. De tomar esta solución se deberán estudiar otras maneras de construir la base de datos a modo de incorporar los efectos del reemplazo de tuberías.

Pese a los resultados, los modelos cuentan con un gran número de opciones de mejora además de las previamente señaladas en torno a la base de datos. Tal como sugiere Kutylowska (2015), una opción es agregar más parámetros de entrada a los algoritmos. Debido a la escasa información con la que se contaba, el número de variables explicativas es bastante bajo y se limita sólo a las características físicas de las tuberías. Agregar parámetros como tipo de suelo, tipo de tubería, caudal transportado, presión hidráulica o pH, podría aumentar la capacidad predictiva de los modelos. Sin embargo, un estudio de estas características está limitado porque las sanitarias son empresas privadas que no publican este tipo de información y la información que puede ser obtenida por el servicio de transparencia no es suficiente para la construcción de los modelos. Otra opción interesante sería agregar características climáticas como las señaladas en Hu & Hubble (2007) o Giraldo-González & Rodríguez (2020). Además, se podrían usar criterios de sectorización a modo de captar los patrones de falla de cada lugar específico, pese a que la sectorización por comunas no supuso mejoras considerables en el caso de La Florida (sectorizar según cuarteles, por ejemplo).

Otra de las opciones es cambiar el espacio de búsqueda de hiperparámetros. En este caso la metodología empleada consiste en una búsqueda por grilla con validación cruzada, proceso que podría ser cambiado por una búsqueda aleatoria con validación cruzada o usar métodos con *halving*, de modo de explorar un mayor número de combinaciones minimizando los gastos computacionales. Explorar un espacio más grande podría significar encontrar combinaciones que entreguen mejores resultados.

En cuanto a la comparación de los distintos algoritmos, se observa que todos presentan un comportamiento similar, teniendo un desempeño levemente superior el GBT al evaluar métricas como la TPR, la FNR o la ROC AUC.

Sin embargo, estas ventajas comparativas leves que tiene el GBT para este caso de estudio se reducen con su alto tiempo de cómputo, llegando a más de 33 horas en los casos 1 y 2, y a más de 1,3 horas en los casos 3 y 4. El MLP presenta métricas similares demorando hasta cinco veces menos, dependiendo de la configuración utilizada. Por otra parte, la SVM presenta tiempos excesivamente altos (ejecuciones de más de una semana sin obtener resultados) como para ser considerada en el análisis.

De las distintas configuraciones probadas, se observa que aquellas que consideran el remuestreo de datos obtienen mejoras sustanciales en sus resultados pese a la posible fuga de información por la reducción de los datos de entrenamiento. Además de la mejora en las métricas, este procedimiento supone una gran reducción del tiempo de cómputo de los algoritmos, pero sigue siendo necesaria una infraestructura de supercómputo como la que posee el NLHPC para la ejecución de los algoritmos.

Cabe recalcar que está ampliamente demostrada la eficiencia de los algoritmos de inteligencia artificial en la predicción de fallas y que el resultado del modelo está principalmente vinculado a la calidad de los datos entregados y a las variables explicativas utilizadas. Esto se hace más evidente al observar los resultados de la Tabla 4.3, donde las

métricas de evaluación para la predicción de fallas en redes de juguete, usando un historial artificial en base a un modelo estadístico extraído de la literatura, son mejores que las del caso de estudio.

Sobre la predicción de escenarios futuros aplicada en redes experimentales, (por ejemplo los presentados en la Figura 4.39) se observa una leve tendencia al alza en la probabilidad de falla de la red, que queda de manifiesto en la Tabla 4.4, con un aumento para las situaciones a cinco y diez años. Sin embargo, al observar en detalle la Figura 4.39 se destaca que hay tuberías que presentan una reducción de la probabilidad de falla con el tiempo, lo que en una primera instancia no tiene sentido. Este fenómeno puede ser explicado con el hecho de que, al hacer predicciones con muchos años de distancia, el modelo podría estar moviéndose muy lejos del rango de observaciones de entrenamiento, pudiendo provocar un comportamiento predictivo irregular (Winkler et al., 2018).

En cuanto a la aplicación en un modelo hidráulico, se destaca la facilidad de integración de los algoritmos definidos dada la naturaleza de sus resultados en la modelación hidráulica con paquetes como WNTR en Python o con análisis multiamenazas como el de Abarca (2022). Yendo a los resultados en particular de la sección 4.2.1, no se encuentra una relación apreciable entre la probabilidad de falla calculada y alguna de las características físicas del modelo (Figura 4.38 y Figura 4.40), por lo que el modelo no estaría capturando la dependencia de estas variables presentada en la sección 3.2.2, situación que se repite en los modelos de la sección 4.2.2.

Por otro lado, los efectos del deterioro se plasman en un cambio del régimen de presiones que se observa al comparar la Figura 4.41 y Figura 4.42 para las condiciones actuales y la Figura 4.41, Figura 4.46, Figura 4.50 para las condiciones futuras. Los efectos de las fugas se manifiestan, además, en una disminución de la demanda cubierta, estando en todos los casos por debajo del 80% (Figura 4.43, Figura 4.47, Figura 4.51).

En la situación con sismo, se analizaron los casos extremos de la curva de amenaza, observándose que, para el caso con menor probabilidad de excedencia, la red no sufre ningún daño, mientras que para el caso con mayor probabilidad de excedencia la mayor parte sufre daño mayor o daño moderado.

Para el caso con menor probabilidad de excedencia y por lo señalado anteriormente, la red no sufre alteraciones en su funcionamiento más allá de los efectos del deterioro. Por otro lado, en el escenario sísmico 2 y dados los efectos de la Figura 4.58, el funcionamiento de la red se reduce casi a cero para todos los escenarios de deterioro mientras duran los efectos del sismo (Figura 4.81, Figura 4.82, Figura 4.83).

Una de las limitaciones que aparece en la incorporación del modelo de deterioro en el modelo hidráulico es la determinación del tamaño de las fugas producidas por las fallas. Como se señala en la sección 3.2.3.1, a cada fuga se le asocia un tamaño proporcional al área de la sección transversal de la tubería, el cual no es determinado por el modelo de deterioro. Para tener una cuantificación más apropiada de la magnitud de las pérdidas en

redes de agua potable se deben desarrollar modelos de predicción del tamaño de fugas. Dentro de las posibles metodologías para este trabajo se encuentra la planteada por Islam (2018), donde se usan modelos con inteligencia artificial para predecir distintos tamaños de fuga.

Por otro lado, la variación de los coeficientes de pérdidas podría derivarse a partir de modelos construidos con datos nacionales, de este modo se podría tener mejores estimaciones de la variación temporal de los coeficientes de pérdida para el tipo de tubería y el tipo de agua usado en el país. La metodología Abdelmonem et al. (2020) parece una manera apropiada de hacerlo dados los estándares de calidad del agua transportada en Chile dictados según la NCh 409.

Para poder proyectar los análisis desarrollados en este trabajo a una red real, se necesita desarrollar estudios sobre, por ejemplo, el desarrollo de curvas de fragilidad o sobre la evolución temporal de los coeficientes de pérdida de carga para todos los materiales de la red que se desea modelar.

6 Conclusiones

En el presente trabajo se desarrollaron modelos de deterioro haciendo uso de algoritmos de inteligencia artificial, los cuales se combinaron con un análisis probabilístico de riesgo sísmico y se incorporaron dentro de un modelo hidráulico.

Los algoritmos del modelo de deterioro fueron aplicados a un caso de estudio, del que se desprende que el *Gradient Boosted Tree* es el que presenta un mejor desempeño. Sin embargo, todos los algoritmos presentaron un bajo desempeño en general, con precisiones que van entre un 3% y un 71%, hecho que se asocia principalmente a la calidad de los datos utilizados en la modelación y a la escasez de estos.

Este último punto resalta la importancia de una correcta recolección de información. En el desarrollo de cualquier tipo de modelo, pero en particular en el desarrollo de modelos con algoritmos de inteligencia artificial, la ausencia de datos acompañada de una alta incertidumbre como se describe en la sección 3.1.6 llevan a una reducción considerable de la capacidad de estos. En el contexto de escasez hídrica que vive el país, en particular la zona de Chile Central (Garreaud et al., 2017), los modelos predictivos de pérdidas en redes, así como los modelos de localización de fugas son fundamentales, y para que estos funcionen se deben tomar datos correctamente.

Se concluye que, si bien los modelos pueden ser útiles, se requiere de una gran cantidad de datos para poder realizarlos, lo que limita su aplicación a redes de mediana o gran edad. Una red joven no cuenta con un registro de fallas suficientemente extenso como para que los algoritmos de inteligencia artificial puedan aprender de los datos proporcionados. Por el contrario, una red de gran edad debería contar con una extensión de datos suficiente como para poder realizar el aprendizaje. Este criterio también aplica a tuberías de materiales que no tienen un registro extenso de fallas.

Como alternativa a los modelos conducidos por datos (inteligencia artificial o modelos estadísticos) está el uso de modelos físicos que estimen el tiempo de vida según la combinación de cargas a la que están sometidas las redes. Sin embargo, estos modelos requieren de una gran cantidad de información que es difícil o costosa de conseguir.

Se recomienda el uso de los modelos siempre y cuando se cuente con una extensión de registro suficiente para entrenarlos. En caso contrario, habría que buscar alternativas que no requieran registros de falla extensos (tales como los modelos físicos).

En cuanto a la incorporación de los modelos de deterioro en el análisis hidráulico y sísmico de redes, se destaca la facilidad con que estos pueden ser integrados. Sin embargo, para un análisis completo en este modelo integrado hacen falta las curvas de fragilidad para todos los materiales que componen la red, así como un modelo apropiado para los cambios en la rugosidad en la red y para la predicción y estimación de fugas.

7 Bibliografía

- Abarca, D. (2022). *Marco de análisis de riesgo multi-amenaza en redes de agua potable*. (Tesis de Magíster) Universidad de Chile (En Proceso).
- Abdelmonem, Y., Ead, S. A., & Shabayek, S. A. (2005). Effect of time on pipe roughness. *17th Canadian Hydrotechnical Conference; Edmonton, Alberta, Canada*.
- Arnejo, H. (2017). *Métodos para la mejora de predicciones en clases desbalanceadas en el estudio de bajas de clientes (CHURN)*. (Tesis de Magister) Universidad de Santiago de Compostela.
- Bubtiena, A. M., Elshafie, A. H., & Jafaar, O. (2011). Application of artificial neural networks in modeling water networks. *Proceedings - 2011 IEEE 7th International Colloquium on Signal Processing and Its Applications, CSPA 2011*, 50–57. <https://doi.org/10.1109/CSPA.2011.5759841>
- Dawood, T., Elwakil, E., Novoa, H. M., & Delgado, J. F. G. (2020). Artificial intelligence for the modeling of water pipes deterioration mechanisms. *Automation in Construction*, 120(January), 103398. <https://doi.org/10.1016/j.autcon.2020.103398>
- Dawood, T., Elwakil, E., Novoa, H. M., & Gárate Delgado, J. F. (2020). Water pipe failure prediction and risk models: state-of-the-art review. *Canadian Journal of Civil Engineering*, 47(10), 1117–1127. <https://doi.org/10.1139/cjce-2019-0481>
- Debón, A., Carrión, A., Cabrera, E., & Solano, H. (2010). Comparing risk of failure models in water supply networks using ROC curves. *Reliability Engineering & System Safety*, 95(1), 43–48. <https://doi.org/10.1016/j.ress.2009.07.004>
- Echávez, G. (1997). Increase in Losses Coefficient with Age for Small Diameter Pipes. *Journal of Hydraulic Engineering*, 123(2), 157–159. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1997\)123:2\(157\)](https://doi.org/10.1061/(ASCE)0733-9429(1997)123:2(157))
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367–378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Garreaud, R. D., Alvarez-Garreton, C., Barichovich, J., Pablo Boisier, J., Christie, D., Galleguillos, M., LeQuesne, C., McPhee, J., & Zambrano-Bigiarini, M. (2017). The 2010-2015 megadrought in central Chile: Impacts on regional hydroclimate and vegetation. *Hydrology and Earth System Sciences*, 21(12), 6307–6327. <https://doi.org/10.5194/hess-21-6307-2017>
- Garvs, I. (2020). *Desarrollo De Curvas De Fragilidad Para Sistemas De Distribución De Agua En Chile*. (Tesis de Pregrado) Universidad de Chile.
- Giraldo-González, M. M., & Rodríguez, J. P. (2020). Comparison of Statistical and Machine Learning Models for Pipe Failure Modeling in Water Distribution Networks. *Water*, 12(4), 1153. <https://doi.org/10.3390/w12041153>
- Guzmán, A. (2011). *Estudio de la interacción suelo-tubería en las redes de agua potable y aguas servidas del Gran Concepción*. (Tesis de Pregrado) Universidad Católica de la Santísima Concepción.
- Guzmán, A., Bravo, R., Villalobos, F., & Montalva, G. (2012). *Estudio preliminar de roturas de tuberías de agua potable en el Gran Concepción*.

- Hu, Y., & Hubble, D. W. (2007). Factors contributing to the failure of asbestos cement water mains. *Canadian Journal of Civil Engineering*, 34(5), 608–621. <https://doi.org/10.1139/L06-162>
- Islam, M. T. (2018). *Leak Detection, Localization and Size Prediction in Water Pipeline Systems*. (Tesis de Magister) Texas State University.
- Kettler, A. J., & Goulter, I. C. (1985). Analysis of Pipe Breakage in Urban Water Distribution Networks. *Canadian Journal of Civil Engineering*, 12(2), 286–293. <https://doi.org/10.1139/l85-030>
- Kleiner, Y., & Rajani, B. (2001). Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water*, 3(3), 131–150. [https://doi.org/10.1016/S1462-0758\(01\)00033-4](https://doi.org/10.1016/S1462-0758(01)00033-4)
- Kok, J. N., Boers, E. J. W., Kosters, W. A., Putten, P. Van Der, & Poel, M. (2010). Artificial Intelligence: Definition, Trends, Techniques and Cases. *Encyclopedia of Life Support Systems (EOLSS)*, 1096–1097. <https://www.eolss.net/Sample-Chapters/C15/E6-44.pdf>
- Kutyłowska, M. (2015). Neural network approach for failure rate prediction. *Engineering Failure Analysis*, 47(PA), 41–48. <https://doi.org/10.1016/j.engfailanal.2014.10.007>
- Lee, W.-M. (2019). *Python Machine Learning*. John Wiley & Sons, Inc.
- Miravet, B. A. (2021). *Mejora de las predicciones en muestras desbalanceadas*. (Tesis de Licenciatura) Universidad Autónoma de Madrid.
- Nefedov, A. (2016). *Support Vector Machines: A Simple Tutorial*. https://sustech-cs-courses.github.io/IDA/materials/Classification/SVM_tutorial
- Nishiyama, M., & Fillion, Y. (2013). *Review of Statistical Water Main Break Prediction Models*. 1–37. <https://doi.org/10.1139/cjce-2012-0424>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Rajani, B., & Kleiner, Y. (2001). Comprehensive review of structural deterioration of water mains: Physically based models. *Urban Water*, 3(3), 151–164. [https://doi.org/10.1016/S1462-0758\(01\)00032-2](https://doi.org/10.1016/S1462-0758(01)00032-2)
- Ress, E., & Roberson, J. A. (2016). *The Financial and Policy Implications of Water Loss*. *February*, 77–86. <https://doi.org/10.5942/jawwa.2016.108.0026>
- Rodrigo, J. A. (n.d.-a). *Gradient Boosting con Python*. https://www.cienciadedatos.net/documentos/py09_gradient_boosting_python.html
- Rodrigo, J. A. (n.d.-b). *Redes neuronales con Python*. <https://www.cienciadedatos.net/documentos/py35-redes-neuronales-python.html>
- Shahzad, A., & James, W. (2002). Loss in Carrying Capacity of Water Mains due to Encrustation and Biofouling, and Application to Walkerton, Ontario. *Journal of Water Management Modeling*, 6062. <https://doi.org/10.14796/JWMM.R208-19>
- Sharp, W. W., & Walski, T. M. (1988). Predicting Internal Roughness in Water Mains. *Journal*

- *American Water Works Association*, 80(11), 34–40. <https://doi.org/10.1002/j.1551-8833.1988.tb03132.x>

St. Clair, A. M., & Sinha, S. (2012). State-of-the-technology review on water pipe condition, deterioration and failure rate prediction models! *Urban Water Journal*, 9(2), 85–112. <https://doi.org/10.1080/1573062X.2011.644566>

Superintendencia de Servicios Sanitarios. (2020). *Informe de Gestión del Sector Sanitario, 2020*.

Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W., & Tscheikner-Gratl, F. (2018). Pipe failure modelling for water distribution networks using boosted decision trees. *Structure and Infrastructure Engineering*, 14(10), 1402–1411. <https://doi.org/10.1080/15732479.2018.1443145>

Zangenehmadar, Z., & Moselhi, O. (2016). Prioritizing deterioration factors of water pipelines using Delphi method. *Measurement: Journal of the International Measurement Confederation*, 90, 491–499. <https://doi.org/10.1016/j.measurement.2016.05.001>