ANÁLISIS DE ANOMALÍAS POR CORROSIÓN Y TIPO GRIETA EN DUCTOS DE TRANSPORTE DE HIDROCARBUROS

JOSÉ ANDRÉS SÁNCHEZ ARÉVALO[1]

Resumen

Este estudio presenta una metodología integral para la evaluación de la integridad estructural en ductos de transporte de hidrocarburos, mediante la aplicación de técnicas de aprendizaje automático. Se integraron datos provenientes de inspecciones en línea (ILI) y validaciones mediante inspecciones directas en campo, con el fin de modelar y predecir condiciones críticas asociadas a los principales mecanismos de daño: pérdida de metal (corrosión) y agrietamiento. Se implementaron modelos de regresión y algoritmos de ensamble, como *Random Forest* y *Gradient Boosting*, precedidos de una etapa de reducción de dimensionalidad mediante Análisis de Componentes Principales (PCA), lo que permitió seleccionar variables significativas relacionadas con geometría, localización espacial y temporal del daño. Los modelos predictivos para corrosión alcanzaron un coeficiente de determinación de hasta $R^2 = 0.989$ con errores mínimos (RMSE < 0.005), mientras que para anomalías tipo grieta, el modelo *Random Forest* obtuvo un $R^2 = 0.812$, superando a otros enfoques estadísticos y de *machine learning*. Asimismo, se aplicaron algoritmos de clusterización no supervisada (DBSCAN, HDBSCAN y OPTICS) para identificar agrupaciones espaciales de daño, facilitando el análisis evolutivo y la priorización de zonas críticas. Se desarrolló un análisis detallado de la vida remanente y tasa de deterioro, considerando variables geométricas y geográficas. Los hallazgos obtenidos permiten optimizar la asignación de recursos de mantenimiento, reducir la incertidumbre en la gestión de riesgos y fortalecer la confiabilidad operativa. Esta metodología representa una herramienta robusta y escalable para la toma de decisiones proactivas en programas de integridad de ductos.

Abstract

This study presents a comprehensive methodology for assessing the structural integrity of hydrocarbon transportation pipelines through the application of machine learning techniques. Data from in-line inspections (ILI) and direct field validation inspections were integrated to model and predict critical conditions associated with the main damage mechanisms: metal loss (corrosion) and cracking. Regression models and ensemble algorithms, such as Random Forest and Gradient Boosting, were implemented following a dimensionality reduction stage using Principal Component Analysis (PCA), which enabled the selection of significant variables related to the geometry and the spatiotemporal location of the damage. Predictive models for corrosion achieved a coefficient of determination of up to $R^2 = 0.989$ with minimal errors (RMSE < 0.005), while for crack-type anomalies, the Random Forest model reached $R^2 = 0.812$, outperforming other statistical and machine learning approaches. In addition, unsupervised clustering algorithms (DBSCAN, HDBSCAN, and OPTICS) were applied to identify spatial groupings of damage, supporting the evolutionary analysis and prioritization of critical zones. A detailed analysis of remaining life and degradation rate was conducted, incorporating geometric and geographic variables. The results enable

[1] Universidad El Bosque_Maestria en Estadística Aplicada y Ciencia de datos_ Contacto: jsanchezar@unbosque.edu.co

optimization of maintenance resource allocation, reduction of uncertainty in risk management, and enhancement of operational reliability. This methodology represents a robust and scalable tool for proactive decision-making in pipeline integrity management programs.

1. Introducción

La integridad estructural de los sistemas de transporte de hidrocarburos es un factor crítico para garantizar una operación segura y eficiente, al tiempo que se minimizan los riesgos asociados a fallas catastróficas, pérdidas económicas y afectaciones ambientales. En este contexto, la inspección en línea (ILI, por sus siglas en inglés) se ha consolidado como una técnica no destructiva clave para el diagnóstico de condiciones internas en ductos, permitiendo la detección de pérdidas de metal, grietas, deformaciones geométricas, geo amenazas y otras anomalías sin necesidad de interrumpir la operación continua. Las herramientas ILI, comúnmente denominadas "*pigs* inteligentes", emplean tecnologías avanzadas como ultrasonido (UT), fuga de campo magnético (MFL) y unidades de medición inercial calibradas (CAL-IMU), cuya selección depende del tipo de daño esperado y de las condiciones operativas del ducto (API_1163, 2021).

A pesar del alto nivel de automatización y precisión de estas tecnologías, los resultados de las inspecciones ILI deben ser validados mediante comparaciones con datos reales, obtenidos por inspecciones directas en zanja o mediante correlaciones históricas. Este proceso de validación, que tradicionalmente se realiza de forma manual o semiautomatizada, representa un desafío en términos de tiempo, recursos técnicos y consistencia de resultados. En este sentido, la incorporación de modelos basados en aprendizaje automático (*Machine Learning*, ML) representa una alternativa para automatizar los procesos de validación, es esperado que al implementar modelos de predicción haya reducción de costos y de carga operativa de los equipos de integridad.

El presente trabajo propone una metodología para la aplicación de técnicas de ML en el proceso de validación de datos obtenidos mediante ILI e inspecciones directas de validación, con el objetivo de identificar patrones, correlaciones y discrepancias entre ambas fuentes de información, con el fin de implementar un modelo de predicción que permita anticipar intervenciones preventivas antes que se produzcan perdidas de contención. Mediante el entrenamiento de modelos supervisados, se desarrolla una metodología que facilita la toma de decisiones orientadas a incrementar la confiabilidad del sistema de gestión de integridad de activos, optimizando a su vez los recursos destinados a mantenimiento e intervención.

2. Revisión de literatura

La gestión de la integridad de ductos en la industria del petróleo y gas ha evolucionado considerablemente, pasando de modelos tradicionales a sistemas inteligentes apoyados en técnicas de inteligencia artificial y aprendizaje automático. Tradicionalmente, la evaluación de la integridad residual de tuberías se ha basado en modelos empíricos como ASME B31G y DNV-RP-F101, que utilizan ecuaciones derivadas de la mecánica de materiales y datos experimentales simplificados (Sebastián et al., n.d.). No obstante, estos enfoques presentan limitaciones al enfrentar condiciones complejas y defectos irregulares. En respuesta a estas deficiencias, los modelos han ganado protagonismo, ofreciendo una mayor capacidad para manejar grandes volúmenes de información, modelar relaciones no lineales y mejorar significativamente la precisión de las

predicciones (Zhuang, Liu y Chen 2024). Entre las técnicas más utilizadas en esta transición destacan las redes neuronales artificiales (ANN) y las máquinas vectoriales (SVM), que han demostrado eficacia en la detección y predicción de defectos en ductos, así como en el análisis de crecimiento de la corrosión. Por ejemplo, el modelo desarrollado por (Mohd Zain & Md Noor, 2015)emplea ANN entrenadas con datos de inspección en línea (ILI), alcanzando un coeficiente de correlación de $R^2 = 0.9997$, lo cual evidencia su alta capacidad predictiva. También se destacan otras técnicas, como el análisis de Monte Carlo y algoritmos como *Random Forest*, los cuales permiten simular la progresión de la corrosión, clasificar defectos y optimizar estrategias de mantenimiento (Martinez Brandon Barrena, 2019) (Zhuang et al., 2024). Asimismo, modelos híbridos que combinan algoritmos tradicionales con técnicas de *machine learning* se han empleado para predecir parámetros críticos como el factor estimado de reparación (ERF). En (Martinez Brandon Barrena, 2019), se utilizó Segmentación jerárquica y *Random Forest* para priorizar áreas de alto riesgo, permitiendo una toma de decisiones más efectiva en cuanto a inspección y reparación. Otros avances incluyen el uso de *Graph Convolutional Networks* (GCN), que han mejorado la detección de fugas en redes de tuberías, superando a métodos tradicionales de análisis (Şahin & Yüce, 2023). Por otro lado, el enfoque híbrido ANFIS-PCA propuesto por (Phan & Duong, 2021) para la predicción de presión de estallido en tuberías con defectos ha mostrado resultados superiores a modelos clásicos. La integración del Análisis de Componentes Principales permite una reducción del ruido y dimensionalidad de los datos, mientras que ANFIS mejora la interpretación y predicción. Este modelo alcanzó un R^2 de 0.9919 y RMSE de 0.9883 MPa. Más recientemente, las investigaciones han explorado el uso de algoritmos evolutivos como el *Sparrow Search Algorithm* (SSA) para optimizar modelos neuronales, mejorando la precisión y estabilidad de las predicciones de corrosión (Zhuang et al., 2024). Además, se ha documentado un creciente interés en el uso de redes neuronales profundas (DNN) y técnicas de aprendizaje profundo (*deep learning*), que ofrecen ventajas considerables en términos de precisión, capacidad de generalización y procesamiento de grandes volúmenes de datos (Hussain et al., 2024). Un área crítica que también ha sido abordada con estas técnicas es la predicción de corrosión bajo tensión (SCC). Aunque persisten desafíos relacionados con la disponibilidad y calidad de los datos, las redes neuronales profundas han mostrado resultados prometedores en la predicción temprana de este tipo de fallas (Muhammad Hussain et al., 2024).

En conjunto, estas innovaciones están transformando la manera en que se evalúa y mantiene la infraestructura de transporte de hidrocarburos. El uso de ensayos no destructivos (END) integrados con *machine learning* ha permitido una evaluación más precisa del estado estructural de las tuberías, optimizando recursos, mejorando la seguridad operativa y enfrentando desafíos en términos de costos y capacidad de procesamiento. Así, se refuerza la necesidad de enfoques híbridos avanzados como pilares para una gestión eficiente y segura de activos críticos.

3. Metodología

La metodología adoptada en este estudio combina técnicas de análisis exploratorio, reducción de dimensionalidad y modelado predictivo mediante aprendizaje automático. Inicialmente, se recopilaron datos provenientes de múltiples campañas de inspección en línea (ILI) y de inspecciones directas de validación en zanja, los cuales fueron integrados y depurados para construir una base de datos consolidada de anomalías por perdida de metal (corrosión) y agrietamiento.

Después de integrar datos de inspección en línea y directas se llevó a cabo un análisis de clusterización con el objetivo de agrupar las zonas afectadas por pérdida de metal (corrosión) y agrietamiento que presentan coincidencias espaciales. Esta agrupación permite identificar regiones con comportamientos de daño similares a lo largo del tiempo, facilitando así su comparación y la estimación de sus velocidades de propagación. Para ello, se emplearon algoritmos de clustering no supervisado: DBSCAN, OPTICS Y HDBSCAN.

El análisis de Componentes Principales (PCA) como técnica de reducción de dimensionalidad, es implementado en este trabajo con el fin de identificar combinaciones de variables que concentran la mayor varianza explicada del sistema, sin redundancia ni colinealidad. Esta selección fue complementada con criterios técnicos basados en el conocimiento del fenómeno de corrosión y propagación de grietas, resultando en un conjunto final de variables relevantes para el entrenamiento del modelo predictivo, incluyendo características geométricas, espaciales y temporales de las anomalías.

Para analizar las anomalías y predecir sus valores, se implementan modelos de regresión, agrupados en dos categorías: modelos estadísticos (regresión lineal y regresión logarítmica) y modelos de machine learning (*Multilayer Perceptron, Support Vector Machines, Gradient Boosting y Random Forest*). La evaluación del desempeño se realizó mediante validación cruzada, utilizando como métricas principales el coeficiente de determinación (R^2), el error absoluto medio (MAE) y la raíz del error cuadrático medio (RMSE).

4. Análisis y discusión

En este estudio se analizan dos líneas de transporte de hidrocarburos en Colombia. La Línea A, construida en 1976 y ubicada en el norte del país, tiene una longitud de 103,65 km y una presión máxima de operación de 900 psi, y en ella se evalúan anomalías por pérdida de metal (corrosión). La Línea B, construida en 1985 en el centro occidente del país, tiene una longitud de 238,6 km, opera a una presión máxima de 2239 psi y presenta principalmente anomalías tipo grieta. Ambos tipos de defectos constituyen los mecanismos de falla más frecuentes y de mayor impacto sobre la integridad de la infraestructura de transporte de hidrocarburos en Colombia.

Para cada línea se recolectaron datos provenientes de cuatro inspecciones ILI: realizadas entre 2012 y 2024 en la Línea A, y entre 2017 y 2024 en la Línea B. Como parte fundamental del proceso de validación y mejora en la confiabilidad de los datos obtenidos mediante ILI, se analizan también los resultados de inspecciones directas en zanja realizadas entre 2021 y 2025. Estas inspecciones permiten verificar las dimensiones reales en campo, características geométricas y severidad de cada tipo de anomalía reportada.

4.1. Análisis descriptivo

4.1.1. Línea A de transporte de hidrocarburos

Entre 2012 y 2024 se llevaron a cabo cuatro inspecciones ILI, todas orientadas a la detección y caracterización de anomalías por pérdida de metal (corrosión). Estas inspecciones utilizaron tecnologías de geometría de alta resolución (XGP), mapeo inercial XYZ, ultrasonido de alta resolución RoCorr y pérdida de flujo magnético axial (MFL-A), con herramientas progresivamente más avanzadas en cuanto a resolución y

capacidad de mapeo. La Figura 1 muestra la cantidad de anomalías de pérdida de metal (corrosión) detectadas en las cuatro corridas ILI, según el tipo: anomalía individual o agrupación por su cercanía espacial en el ducto, y su representación en porcentaje

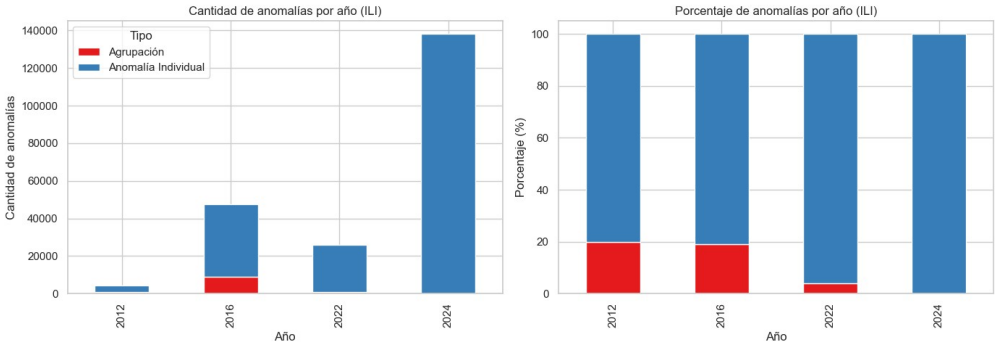

Figura 1. Anomalías detectadas en inspección ILI y su distribución porcentual por tipo.

A lo largo de estas campañas se observó un incremento en la cantidad de anomalías detectadas, lo cual evidencia tanto un aumento en el deterioro de la línea como una mejora en la sensibilidad de las tecnologías empleadas para detectarlas, permitiendo una discriminación más precisa entre anomalías individuales y agrupaciones. De 2022 a 2024 se observa un aumento del 500% de puntos de corrosión, registrándose en total 138.402 anomalías de perdida de metal (corrosión). Por otra parte, el 100% de las anomalías detectadas en 2024 fueron clasificadas como individuales, evidenciando la mejora en la resolución de la herramienta de inspección.

En la Figura 2 se presenta la distribución de las anomalías detectadas por pérdida de metal (corrosión) en la línea A, agrupadas cada 10 km en cada corrida.

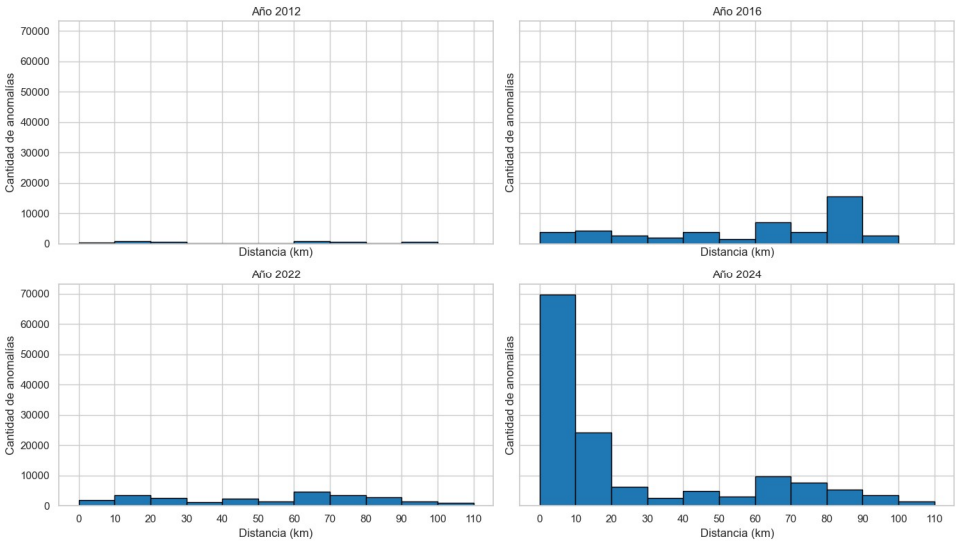

Figura 2. Línea A: Distribución de anomalías de corrosión cada 10 km por corrida

Entre 2012 y 2024, las inspecciones mediante ILI evidenciaron cambios en la cantidad y distribución, influenciadas por el avance en las tecnologías de inspección, condiciones externas y operativas del ducto. En 2012 se identificaron 4.506 anomalías, concentradas

entre el PK (Punto kilométrico) 0 y el PK 20, y entre el PK 60 y el PK 90. Para 2016, se registró un expresivo aumento, con un total de 47.507 anomalías, representando, un incremento del 954% respecto al valor base de 2012, con una intensificación notable entre el PK 60 y el PK 90, lo que sugiere un proceso acelerado de deterioro.

En 2022, la cantidad de anomalías disminuyó a 26.104 anomalías, representando una reducción del 45% respecto a 2016, y evidenciando un desplazamiento en la distribución hacia el PK 10 al PK 20, mientras que el tramo PK80 al PK90 mostró una disminución del 38% respecto a su máximo anterior. Finalmente, en 2024, con la introducción de una tecnología de inspección de mayor resolución, se detectaron 138.110 anomalías, cuadruplicando el registro de 2022 y triplicando el máximo histórico de 2016, lo cual, más allá de indicar un deterioro real de la tubería, refleja también una mejora en la sensibilidad de detección. La densidad de anomalías supera las 4.000 por cada 10 km entre el PK 60 y el PK 90, consolidándose como un segmento crítico.

En la Tabla 1 se presentan los valores máximos de pérdida de espesor y su desviación estándar registrados en la Línea A, divididos en intervalos de 10 km. Cabe señalar que los datos correspondientes al tramo entre el PK 100 y el PK 110 en la corrida de 2016 no están disponibles debido a un reemplazo de tubería, el cual redujo la longitud efectivamente inspeccionada.

Segmento (km)	**Profundidad Media (Desv. Est.)**				**% Perdida Máx**			
	2012	**2016**	**2021**	**2024**	**2012**	**2016**	**2021**	**2024**
0-10	12,54 (5,98)	12,73 (4,18)	12,22 (3,04)	13,77 (3,37)	81	67	40	61
10-20	14,49 (5,39)	15,7 (5,49)	15,51 (6,55)	14,84 (4,57)	42	49	72	68
20-30	15,31 (6,43)	17,15 (7,11)	16,35 (7,16)	16,09 (5,75)	52	51	64	85
30-40	16,5 (8,30)	19,6 (10,37)	17,24 (8,82)	16,82 (6,64)	69	69	73	66
40-50	13,6 (4,65)	16,01 (6,65)	16,75 (7,57)	16,85 (6,01)	34	71	64	69
50-60	14,2 (4,83)	15,27 (6,21)	16,35 (7,00)	16,36 (5,51)	35	51	66	64
60-70	14,76 (6,19)	17,05 (7,25)	16,21 (6,86)	16,69 (5,61)	54	72	70	66
70-80	15,72 (6,46)	18,42 (7,87)	17,69 (7,58)	17,13 (6,59)	48	57	69	72
80-90	13,8 (4,17)	20,63 (9,81)	16,45 (6,95)	17,51 (5,85)	33	70	82	66
90-100	13,99 (4,78)	19,09 (9,06)	15,64 (6,43)	16,94 (5,44)	36	68	72	65
100-110	14,22 (4,78)	-	22,02 (13,76)	16,44 (5,36)	32	-	85	48

Tabla 1. Características de profundidad de anomalías.

Se observa un incremento general en la profundidad media de pérdida de espesor a lo largo del tiempo en varios segmentos del ducto, aunque con algunas variaciones puntuales. En el segmento comprendido entre el PK 30 y el PK 40, la profundidad media aumentó de 16,5 % en 2012 a 19,6 % en 2016, para luego descender a 16,82 % en 2024. Por su parte, el segmento entre el PK 100 y el PK 110 presenta un valor atípico en 2021, con una profundidad media de 22,02 % y una desviación estándar elevada, lo que sugiere la presencia de anomalías de alta severidad o agrupaciones concentradas. En términos de pérdida máxima, se identifican incrementos expresivos en algunos tramos. En el segmento PK 20 a PK 30 el porcentaje máximo pasó de 52 % en 2012 a 85 % en 2024. De manera similar, en el tramo entre el PK 80 y el PK 90, la pérdida máxima se incrementa del 33 % en 2012 al 82 % en 2021, seguida de una reducción al 66 % en 2024, lo que podría indicar la intervención o reparación de la zona crítica, aunque persiste una velocidad de deterioro acelerado.

Se nota un deterioro progresivo en ciertos segmentos del ducto, con profundidades críticas que superan el 80 %, lo que subraya la necesidad de priorizar el monitoreo y la intervención en estas zonas para preservar la integridad estructural de la línea.

4.1.2. Línea B de transporte de hidrocarburos

Entre 2017 y 2024 se realizaron cuatro inspecciones en línea (ILI), orientadas a la detección y caracterización de anomalías tipo grieta. Las inspecciones emplearon tecnologías basadas en ultrasonido de haz angular, permitiendo una evaluación detallada del estado estructural del ducto. En la Figura 3 se muestra la cantidad total de anomalías detectadas en cada corrida ILI, según su clasificación como grieta individual, campo de grietas o entallas mecánicas, y su distribución por tipo. Se observa que predominan las grietas individuales y las colonias por corrosión bajo esfuerzos combinados.

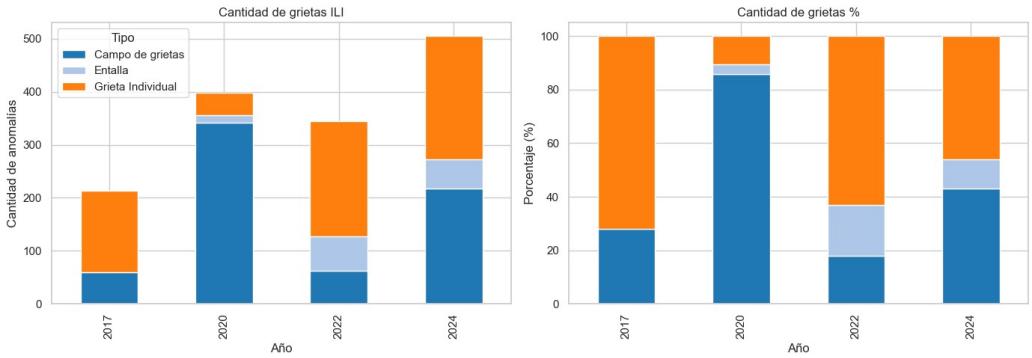

Figura 3. Cantidad de tipo de grietas por cada corrida para línea B.

Se identifica una tendencia creciente en la detección de anomalías tipo grieta en la Línea B. La cantidad total de grietas reportadas muestra un incremento acumulado del orden del 150% en el periodo analizado. Este aumento sostenido sugiere un deterioro progresivo del sistema, posiblemente asociado a condiciones operativas, factores de servicio o al envejecimiento inherente de los materiales constructivos. En 2020 se observa un cambio en la distribución tipológica de las anomalías: las grietas individuales, predominantes en 2017, fueron ampliamente superadas por los campos de grietas, los cuales representaron más del 80% del total registrado. Esta transformación es indicativa de una evolución en los mecanismos de daño por corrosión bajo esfuerzos combinados. En los años 2022 y 2024 se evidencia una diversificación en los tipos de defectos, destacándose un incremento relativo en la detección de entallas (aproximadamente entre 20% y 25%), lo cual puede estar vinculado a procesos de iniciación de grietas inducidos por discontinuidades superficiales. La variabilidad interanual observada refuerza la necesidad de realizar análisis complementarios que exploren la relación entre estas tendencias y variables operacionales, condiciones ambientales y propiedades metalúrgicas de la tubería.

La distribución de anomalías por grieta detectadas en la Línea B, agrupadas cada 10 km en cada una de las corridas, se presenta en la Figura 4.

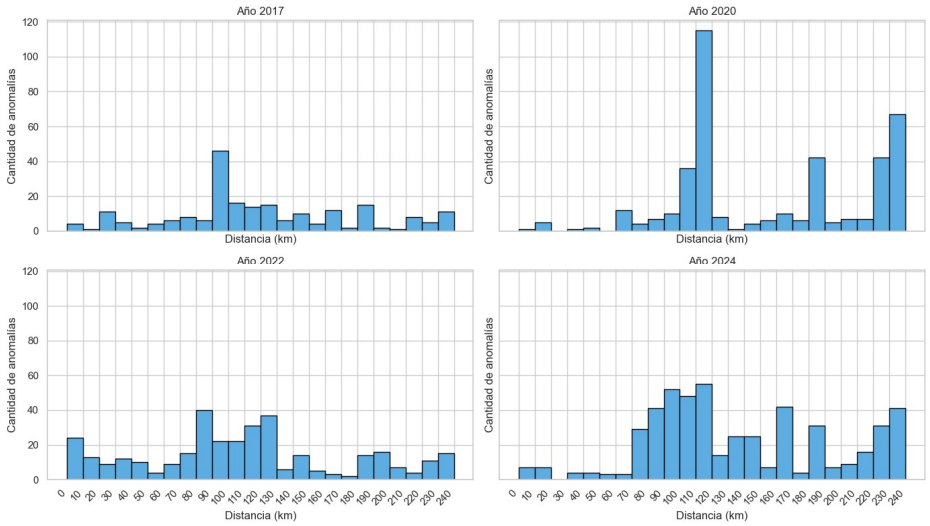

Figura 4. Distribución cantidad de anomalías por cada corrida por cada 10 km línea B.

Se observa una fuerte asimetría positiva en la distribución de frecuencia de anomalías por grietas en los años 2020 y 2024, con picos de ocurrencia entre los segmentos del PK100 al PK120 y del PK160 al PK180. El cambio observado en 2022, respecto a 2020, se atribuye a intervenciones de mantenimiento y sustitución de secciones afectadas por la severidad de los hallazgos en las dos primeras inspecciones. La dispersión de anomalías muestra un incremento progresivo, lo que sugiere una expansión de las zonas afectadas y un posible desplazamiento del daño hacia nuevas regiones. Además, la naturaleza multimodal de la distribución —especialmente en las inspecciones más recientes— indica la coexistencia de múltiples zonas críticas con alta concentración de daño. Este patrón refuerza la necesidad de aplicar metodologías de análisis espacial para identificar formalmente las regiones más vulnerables y optimizar tanto los planes de monitoreo como las estrategias de mantenimiento preventivo.

En la Tabla 2 se presentan los valores de profundidad media y su desviación estándar, profundidad mínima y máxima registrados en la Línea B, en intervalos de 10 km.

Intervalo km	**Profundidad Media (Desv. Estandar)**				**Profundidad mínima e máxima (em mm)**			
	2017	**2020**	**2022**	**2024**	**2017**	**2020**	**2022**	**2024**
0-20	1,12 (0,12)	1,35 (0,28)	0,55 (0,35)	0,65 (0,2)	1,05 - 1,20	1,00 - 1,50	0,10 - 1,85	0,50 - 1,05
20-40	1,10 (0,14)	0,55 (0,10)	0,48 (0,27)	0,47 (0,16)	1,00 - 1,35	1,00 - 0,50	0,10 - 1,30	0,55 - 1,60
40-60	1,13 (0,13)	0,6 0(0,12)	0,49 (0,23)	0,77 (0,43)	1,00 - 1,25	1,00 - 0,60	0,10 - 0,90	0,55 - 0,75
60-80	1,25 (0,20)	1,17 (0,32)	0,52 (0,25)	0,59 (0,15)	1,05 - 1,60	1,00 - 1,80	0,15 - 1,00	0,50 - 0,95
80-100	1,26 (0,24)	1,20 (0,30)	0,61 (0,45)	0,59 (0,12)	1,00 - 1,80	1,00 - 1,75	0,10 - 2,30	0,50 - 1,10
100-120	1,33 (0,38)	1,26 (0,53)	0,53 (0,35)	0,60 (0,13)	1,00 - 2,40	1,00 - 3,45	0,10 - 1,75	0,50 - 1,60
120-140	1,21 (0,27)	1,15 (0,27)	0,46 (0,32)	0,68 (0,32)	1,00 - 1,80	1,00 - 1,50	0,10 - 1,30	0,50 - 1,50
140-160	1,08 (0,16)	1,39 (0,44)	0,48 (0,21)	0,68 (0,20)	1,00 - 1,50	1,00 - 2,10	0,15 - 0,90	0,50 - 0,45
160-180	1,08 (0,01)	1,28 (0,51)	0,51 (0,22)	0,72 (0,14)	1,00 - 1,25	1,00 - 2,40	0,10 - 0,85	0,60 - 1,10
180-200	1,14 (0,14)	1,25 (0,39)	0,42 (0,25)	0,81 (0,28)	1,00 - 1,45	1,00 - 2,10	0,10 - 1,10	0,50 - 1,30
200-220	1,59 (0,12)	1,57 (0,84)	0,39 (0,15)	0,75 (0,24)	1,50 - 1,80	1,00 - 3,15	0,15 - 0,60	0,50 - 1,35
220-240	1,35 (0,38)	1,24 (0,4)	0,38 (0,20)	0,77 (0,21)	1,05 - 2,00	1,00 - 2,75	0,10 - 0,70	0,50 - 1,25

Tabla 2. Distribución de profundidades línea B.

Los resultados evidencian variaciones en la severidad y dispersión de las grietas en el tiempo y a lo largo del ducto. En varios segmentos se observa una disminución en la profundidad media entre 2020 y 2022, posiblemente asociada al reemplazo de tramos o a mejoras en la resolución de inspección. En 2024 se detecta un nuevo incremento en la profundidad media en ciertos tramos, como entre el PK 180 y el PK 220, lo que sugiere una reactivación del daño, aparición de nuevas grietas o avances en la tecnología de inspección. Los tramos entre el PK 200 y el PK 240 presentan condiciones críticas, con profundidades máximas superiores a 2,5 mm (40% del espesor de pared) y desviaciones estándar elevadas, indicando la presencia de grietas de distinta severidad, algunas potencialmente agrupadas. Estos valores representan una amenaza significativa para la integridad del ducto, especialmente bajo esfuerzos fluctuantes. Esta información es clave para priorizar evaluaciones, monitoreo continuo y planificación de intervenciones.

4.1.3. Unión de datos de ILI con validación en campo

La validación de los datos obtenidos mediante ILI a través de inspecciones directas en zanja constituye un componente crítico para verificar la precisión dimensional y la localización de las anomalías detectadas. Esta actividad se lleva a cabo bajo los lineamientos establecidos en los estándares internacionales API 1163, ASME B31.8S y NACE SP0502. En este contexto, los resultados de las inspecciones directas realizadas en campo entre los años 2021 y 2025, fueron integrados con los datos provenientes de las corridas ILI, manteniendo la misma estructura de información, que incluye distancia de registro, posición horaria, longitud, ancho y profundidad del defecto y todos los relacionados con la ubicación geográfica de cada indicación. Esta integración permite mejorar la consistencia de los datos, favoreciendo tanto la eficiencia operativa como la seguridad en el transporte de hidrocarburos. Como resultado, la base de datos consolidada presenta un total de 217.612 anomalías por pérdida de metal (corrosión) para la Línea A y 1.911 anomalías tipo grieta para la Línea B.

Como resultado de la integración de datos para la línea A, en la Figura 5, se representa la distribución final de las anomalías en intervalos de 10 km y el comportamiento de la perdida de metal en cada segmento.

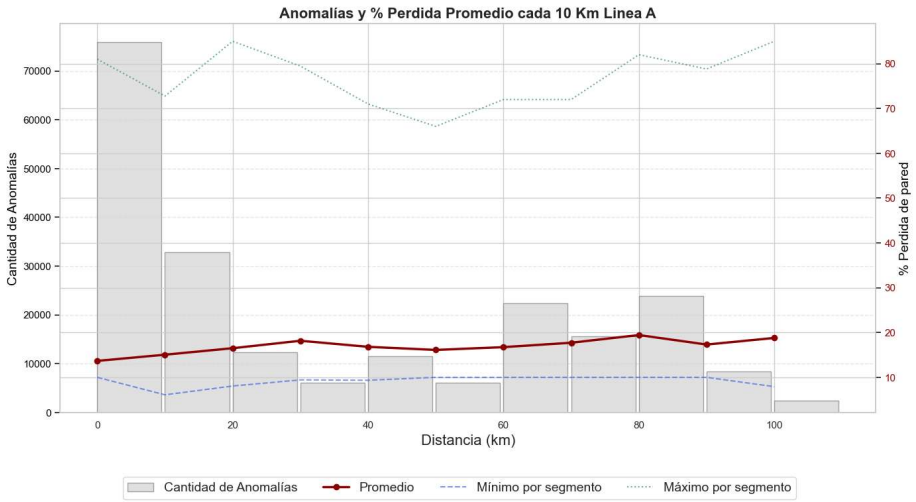

Figura 5. Distribución de anomalías cada 10Km y comportamiento del % perdida de metal para la línea A.

Una marcada concentración de anomalías se registra entre el PK 0 y el PK 10, con más de 70.000 pérdidas de metal. En este tramo, el porcentaje de pérdida de espesor presenta una media del 13%, alcanzando valores máximos de 81%. Adicionalmente, entre los PK 60 y PK 90 se identifica otra zona crítica con alta densidad de anomalías, donde la pérdida de metal varía entre el 10% y el 82%. A lo largo de toda la línea, la media del porcentaje de pérdida de pared se mantiene con poca variación, oscilando entre el 17% y el 20%. Esta estabilidad contrasta con la alta variabilidad observada en la cantidad de anomalías, lo que sugiere una débil o nula correlación lineal entre ambas variables. El reconocimiento de estos patrones es fundamental para la priorización de inspecciones detalladas y la planificación de mantenimiento predictivo, particularmente en segmentos con alta concentración de anomalías, aun cuando no presenten porcentajes elevados de pérdida de pared.

En la Figura 6 se muestra la distribución de anomalías con la profundidad de las grietas, agrupadas en segmentos de 10 Km a lo largo de la línea B.

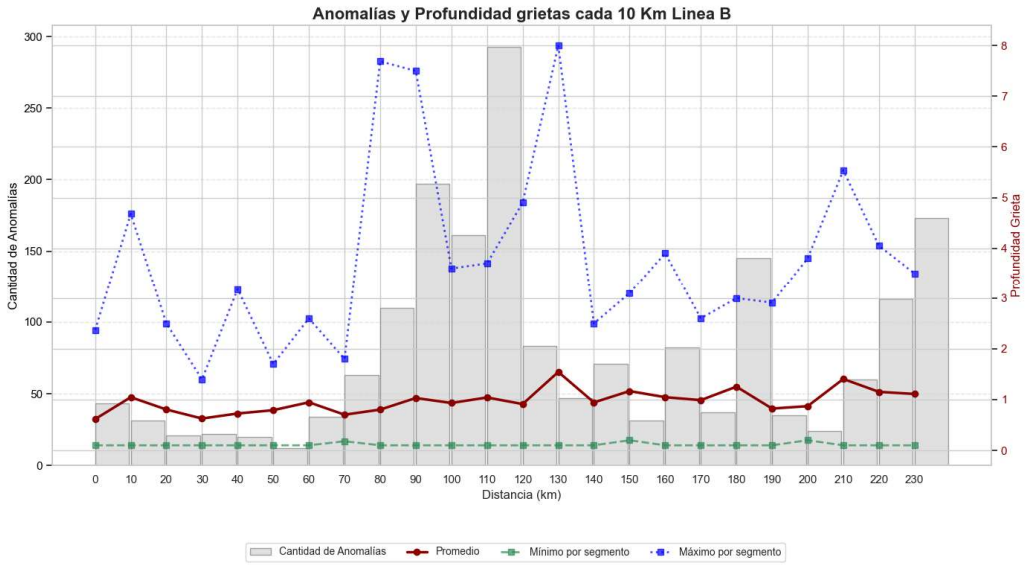

Figura 6. Distribución de anomalías cada 10Km y profundidad de las grietas para la línea B

El número medio de anomalías observadas es de 104 grietas, con una desviación estándar de 76 evidenciando una alta dispersión y presencia de valores atípicos entre el PK90 y PK130, donde se registran picos superiores a 200 y hasta 295 grietas (máximo en el PK110). En contraste, la profundidad promedio de las grietas permanece relativamente estable, con un valor medio de 0,98 mm y una desviación estándar de 0,21, lo cual sugiere una distribución homogénea. No obstante, la profundidad máxima de grietas alcanza valores de hasta 8 mm (zonas con pérdida de contención), con una media de 4,1 y una desviación estándar de 1,72, lo que refleja la presencia de eventos puntuales de severidad considerable, principalmente en los PK70, PK120 y PK130. El análisis conjunto indica una posible relación entre concentración de grietas y los valores extremos de profundidad máxima, aunque no necesariamente con la profundidad promedio. Los tramos entre el PK90 y PK130 representan zonas críticas tanto por frecuencia como por severidad potencial, y deberían ser objeto de monitoreo intensivo y priorización en estrategias de mantenimiento.

4.1.4. Análisis por Tubos

Las anomalías han sido representadas de forma continua a lo largo de los ductos; sin embargo, su aplicación práctica en estudios de integridad estructural requiere la identificación precisa de los segmentos individuales de tubería o tramos completos. Esta identificación se logra mediante la comparación de las coordenadas de las soldaduras detectadas en cada corrida de inspección, lo que permite delimitar zonas específicas con concentración de anomalías, analizar su evolución temporal, estimar su impacto sobre la integridad del activo y planear actividades de intervención.

Para este propósito, se consideró una longitud promedio de 12 metros por tubo, conforme a las especificaciones del tubo convencional comercial. No obstante, se identificaron secciones con segmentos de menor longitud, las cuales fueron tratadas como casos especiales. La comparación de coordenadas de soldadura entre corridas se realizó aplicando una tolerancia de ±300 mm, lo que permitió emparejar un total de 17.626 soldaduras correspondientes a la Línea A y 20.891 a la Línea B. Dado que el trazado de cada línea ha experimentado modificaciones a lo largo del tiempo, se segmentaron ambas líneas por tramos de tubería. Como resultado, se identificaron 12,821 segmentos emparejados para la Línea A y 20,700 para la Línea B, los cuales se utilizaron en la evaluación detallada de anomalías.

La metodología desarrollada en este trabajo presenta alta adaptabilidad a cualquier activo, consolidándose como un recurso clave para el análisis temporal de defectos en distintos ductos y campañas de inspección. Una vez establecidos los segmentos de tubería mediante la alineación espacial de soldaduras, se llevó a cabo la clusterización de las anomalías de cada línea.

4.2.Clusterización

En el proceso de clusterización de anomalías por pérdida de metal (corrosión) y por agrietamiento, tres algoritmos de *clustering* no supervisado son implementados: DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), OPTICS (*Ordering Points To Identify the Clustering Structure*) y HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*). La elección de estos métodos se fundamenta en su capacidad para identificar estructuras de agrupamiento de forma no paramétrica, sin requerir una especificación previa del número de clústeres, y en su robustez frente al ruido y a la variabilidad en la densidad de los datos, características comunes evidenciadas en el análisis descriptivo.

En la Línea A, el análisis de anomalías por pérdida de metal (corrosión) se basa en un proceso de clusterización que agrupa zonas de corrosión cercanas, siempre que la distancia entre ellas no exceda tres veces el espesor de pared, tanto longitudinal como circunferencialmente. Este enfoque asume una posible interacción entre áreas adyacentes con el mismo mecanismo de daño. Además, se agrupan zonas coincidentes en diferentes inspecciones para analizar su evolución temporal y estimar tasas de crecimiento. La Figura 7 ejemplifica este procedimiento, mostrando la formación de dos zonas de pérdida de metal bajo dicho criterio.

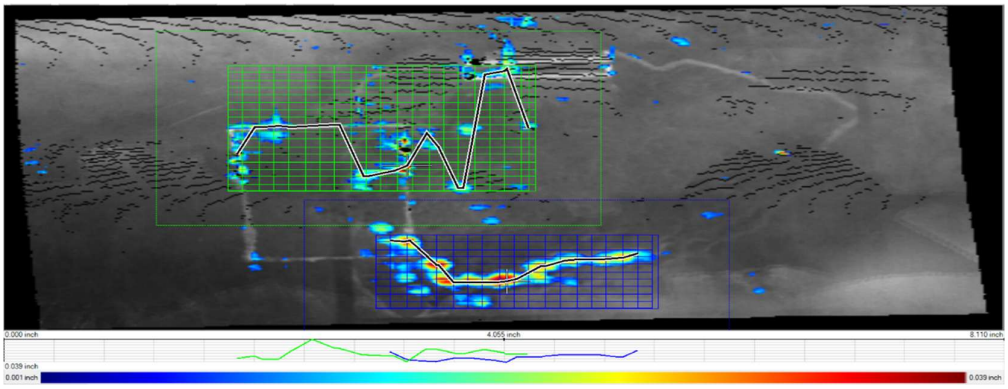

Figura 7. Agrupaciones de pequeñas zonas de corrosión en dos áreas, criterio 3t.

En la línea B, debido a que se trata de un modo de falla distinto, los criterios de clusterización se basan en aspectos espaciales: 100 mm en el sentido longitudinal y 3.5 segundos, equivalentes a 50 mm en el sentido circunferencial. De este modo, se agrupan todas las zonas con agrietamiento coincidentes en ubicación a través del tiempo, lo que permite realizar comparaciones coherentes. Las anomalías clasificadas como ruido se consideran eventos independientes cuya aparición se atribuye al periodo prolongado de tiempo transcurrido entre la última inspección y la fecha de fabricación del ducto. En la Figura 8 se presenta un ejemplo de clúster de agrietamiento identificado en campo mediante ensayo no destructivo.

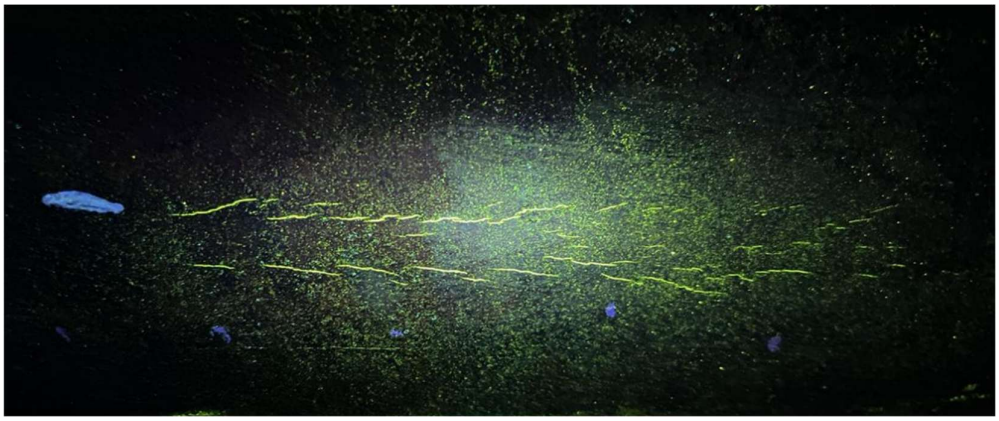

Figura 8. Clúster de grietas identificado mediante ensayo no destructivo.

Los métodos de clusterización son aplicados usando como variables, la fecha de corrida, como variable categórica que permite identificar la evolución de una misma zona con anomalías en diferentes inspecciones; la superficie de ubicación (interna o externa), también categórica, que caracteriza el ambiente de exposición; la distancia de registro y la posición horaria, que describen la localización espacial de cada anomalía en los ejes longitudinal y circunferencial del ducto, respectivamente; el segmento o tramo del ducto donde se detectó la anomalía; y las dimensiones geométricas del defecto, tales como longitud, ancho y profundidad. La combinación de estas variables permite capturar patrones espaciales y morfológicos significativos para identificar grupos de perdida de metal (corrosión), contribuyendo a una mejor comprensión de la corrosión a lo largo del tiempo y el espacio. La evaluación de los resultados obtenidos con cada algoritmo se

realizó mediante el coeficiente de silueta (*silhouette score*), utilizado como métrica de coherencia interna de los agrupamientos generados.

Con el fin de mantener criterios comparables entre los métodos de agrupamiento aplicados, se estableció un parámetro común relacionado con el tamaño mínimo de clúster. Para ello, se configuraron los parámetros min_samples o min_cluster_size con un valor igual a 2. Esta decisión se fundamenta en la naturaleza del mecanismo de daño, que busca identificar zonas de corrosión o agrietamiento recurrente o persistente a lo largo del tiempo. Dado que el objetivo es detectar agrupaciones de anomalías que representen la misma área afectada observada en al menos dos inspecciones distintas, un clúster válido debe contener como mínimo dos registros asociados a diferentes momentos temporales.

De este modo, se garantiza que cada grupo refleje un patrón evolutivo del modo de daño, permitiendo así un análisis más robusto de su progresión espacial y temporal. Además, se observó que valores mayores para este parámetro tienden a excluir clústeres pequeños pero relevantes, limitando la capacidad del modelo para detectar zonas emergentes de corrosión. Por el contrario, valores inferiores favorecen la formación de agrupamientos basados en ruido. En la Tabla 3, se presentan los agrupamientos obtenidos de la aplicación de estos algoritmos para la línea A y B.

Modelo	**Total, clúster**		**Total, puntos ruido**		**Promedio por clúster**		**Ruido por grupo**		**Promedio índice Silhouette**	
	A	**B**	**A**	**B**	**A**	**B**	**A**	**B**	**A**	**B**
DBSCAN	18.961	72	179.455	1.221	2,5	2,25	23,69	88,29	0,96	-0,706
HDSCAN	32.876	233	41.614	181	4,5	5,16	5,49	13,09	0,64	0,4589
OPTICS	68.444	74	26.925	1.221	9,03	2,19	3,55	88,29	0,62	-0,707

Tabla 3. Resultados algoritmos de clusterización para línea A y B.

En la línea A, se observa que el algoritmo DBSCAN alcanza el valor más alto del índice de silueta (0.96), atribuible a la clara separación entre un número reducido de grupos bien definidos. No obstante, este método genera 18.961 clústeres y clasifica una gran proporción de los puntos como ruido (179.455), lo que indica que numerosas anomalías potencialmente relevantes no son agrupadas. Esta limitación puede conducir a una subestimación del número real de agrupamientos, particularmente en entornos complejos influenciados por condiciones ambientales o de operación. Por su parte, HDBSCAN ofrece un equilibrio más adecuado, incrementando la cantidad de clústeres y reduciendo de forma significativa el número de puntos considerados como ruido. Aunque su índice de silueta (0,64) es inferior al de DBSCAN, sigue siendo aceptable, reflejando una cohesión interna moderada y cierta dificultad para capturar por completo la estructura espacial de los datos. En contraste, OPTICS, a pesar de presentar el índice de silueta más bajo (0,62), genera más del doble de clústeres que HDBSCAN y más del triple que DBSCAN, con el menor nivel de ruido tanto absoluto como relativo. Su capacidad para identificar estructuras jerárquicas y clústeres con densidades variables lo posiciona como una herramienta especialmente adecuada para el análisis de anomalías de perdida de metal (corrosión) distribuidos de forma heterogénea en tuberías de transporte. Adicionalmente, presenta el promedio más alto de elementos por clúster (9,03), lo que sugiere una mayor resolución espacial y una mejor granularidad en el análisis.

En la línea B, el análisis comparativo de los modelos basados en densidad, como DBSCAN y OPTICS, generaron un número similar de clústeres (72 y 74,

respectivamente), acompañados de una alta cantidad de puntos clasificados como ruido (1.221 en ambos casos) y un bajo índice de *Silhouette* (−0,7062 y −0,7076), lo que sugiere una pobre cohesión y separación entre grupos. HDBSCAN, aunque también es un método de densidad, logró identificar un mayor número de clústeres (233) y reducir considerablemente la cantidad de puntos de ruido (181), presentando un índice de Silhouette positivo (0,4589), lo que indica una mejor calidad en la segmentación. En conjunto, estos resultados muestran que HDBSCAN ofrecen un balance favorable entre segmentación y calidad del agrupamiento para el caso analizado.

4.2.1. Análisis de los clústeres para cada línea

En la línea A, la distribución correspondiente a los 20 segmentos más críticos, clasificados según la mayor cantidad de anomalías detectadas y el número de clústeres asociados a cada uno se presenta en la Figura 9.

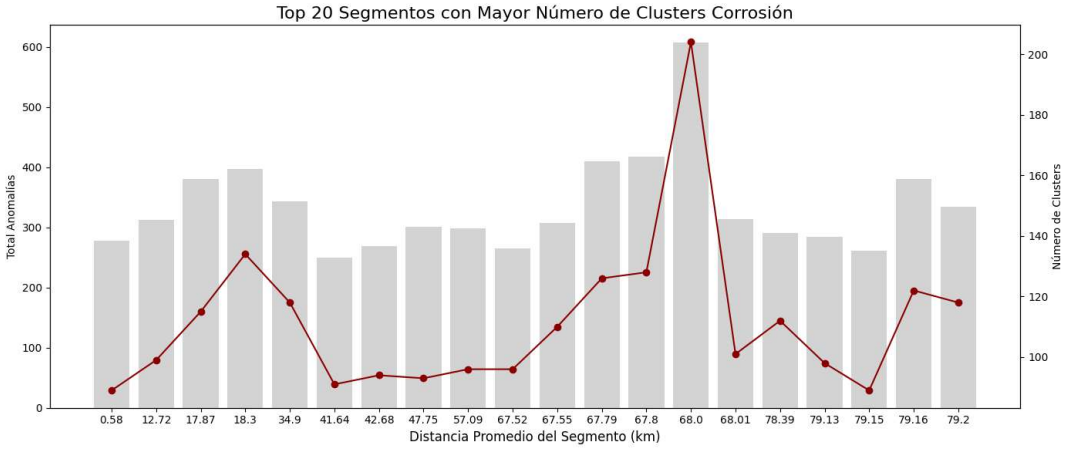

Figura 9. 20 segmentos más críticos por cantidad de anomalías y número de clúster según la distancia de registro de cada segmento o tubo.

El segmento identificado con el mayor nivel de concentración de corrosión interna corresponde al número 427 (está a 5,07km del inicio), en el cual se detectaron 151 clústeres distribuidos a lo largo de toda su longitud. La concentración del daño se presenta en los primeros 10 Km, evidencia de la necesidad de atención y monitoreo en esta sección. Entre los segmentos de mayor criticidad destacan los ubicados en los (PK) 5,07, PK 10,67 y PK7,73. El segmento en el PK 5,07 presenta los valores más altos tanto en número de anomalías como en cantidad de clústeres, lo que evidencia una región con alta densidad de daño y una distribución espacial compleja. Los tramos en PK 10,67 y PK 7,73 también registran métricas elevadas en ambas dimensiones, por lo que deben considerarse prioritarios para inspecciones detalladas o intervenciones correctivas. Este análisis permite establecer criterios objetivos para la priorización de segmentos críticos en sistemas de transporte de fluidos, al integrar no solo la magnitud del daño (número de anomalías), sino también su dispersión (cantidad de clústeres). La combinación de estas métricas ofrece una visión más robusta de la severidad y complejidad del deterioro, fundamental para el diseño de estrategias de mantenimiento basadas en riesgo.

Realizando la discriminación entre corrosión interna y externa, esta última influenciada por condiciones de operación y del ambiente que rodea la sección de tubería. En la Figura 10, se presentan los segmentos con mayor concentración de corrosión externa se presenta

el análisis correspondiente a las veinte zonas con mayor criticidad asociada a pérdida de metal por corrosión externa.

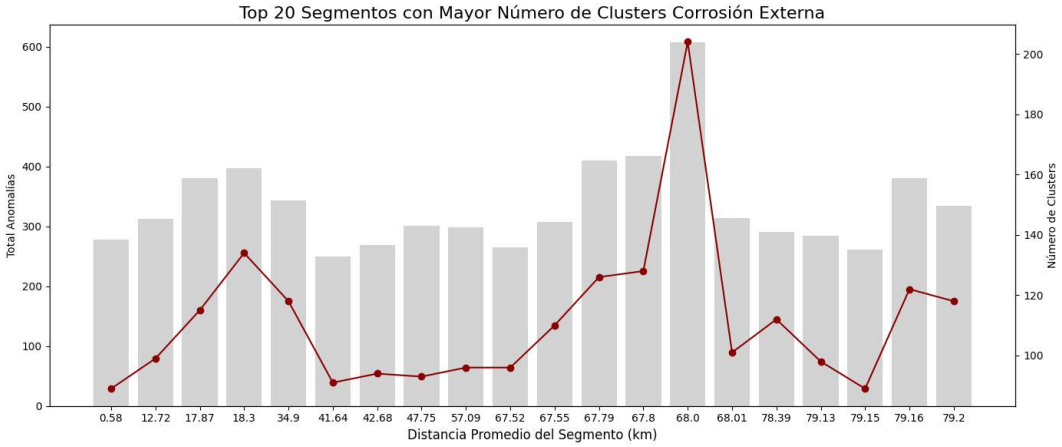

Figura 10. Top 20 de segmentos críticos para corrosión externa según la distancia de registro de cada segmento o tubo.

Se observa la mayor cantidad de anomalías en el segmento en el PK68.0, el cual presenta tanto el mayor número de clústeres (más de 200) como la mayor cantidad de clúster, evidenciando una zona crítica con alta densidad y dispersión de defectos superficiales. Este patrón sugiere un posible deterioro acelerado localizado, probablemente asociado a factores externos como condiciones ambientales, fallas en recubrimientos o interferencias externas. Se identifican otros segmentos relevantes entre el PK17.87 y PK18.3 así como entre los PK67.79 y PK67.8, que también presentan picos significativos tanto en el número de clústeres como en el volumen de anomalías, lo que podría indicar acumulación de defectos en zonas específicas. En términos más amplios, se distinguen dos zonas críticas a lo largo del trazado: la primera entre el PK0 y el PK35, y la segunda entre el PK41 y el PK79. Estas áreas deben considerarse prioritarias para inspección detallada y acciones de mitigación.

En la Figura 11, presenta los 20 segmentos con mayor número de clústeres asociados a corrosión interna a lo largo del ducto. Se muestra la cantidad total de anomalías detectadas por segmento (barras verdes, eje izquierdo) y el número de clústeres formados en cada uno (línea roja, eje derecho). Esta visualización permite identificar regiones críticas con alta densidad de defectos agrupados espacialmente, lo cual es clave para la evaluación de la severidad y extensión del deterioro interno del ducto.

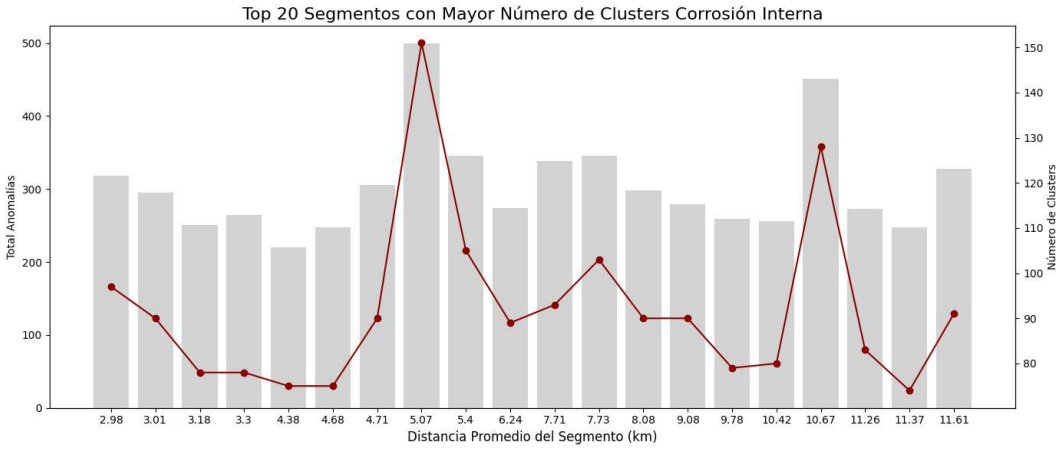

Figura 11. Top 20 de segmentos críticos para corrosión interna según la distancia de registro de cada segmento o tubo.

La severidad observada en la gráfica corresponde a fenómenos de corrosión interna, cuya mayor concentración se presenta en los primeros 10 km del ducto. Esta distribución sugiere la existencia de condiciones operativas particulares en dicho tramo que favorecen la degradación interna, como presencia de sedimentos o inadecuada dispersión de inhibidores. La identificación de esta zona crítica resalta la necesidad de implementar acciones correctivas y preventivas, tales como tratamientos químicos del hidrocarburo transportado, optimización de las condiciones de operación o la aplicación de técnicas específicas de mitigación derivadas de un análisis detallado del mecanismo de corrosión presente. Estas medidas permitirán una gestión proactiva de la integridad del sistema, orientada a reducir el riesgo de falla y extender la vida útil del ducto.

4.3.Análisis de aspectos de integridad mecánica

4.3.1. Análisis de perdida de metal (Corrosión)

4.3.1.1.Tasa de corrosión

La tasa de corrosión mide la velocidad de degradación del material y permite prever su vida útil y planificar el mantenimiento. Con el fin de identificar zonas críticas según la progresión, se analizó el valor máximo de la tasa de corrosión (TC) agrupado por intervalos de 5 Km según la distancia de registro. La tasa de corrosión es determinada dentro de cada clúster determinando la distancia en tiempo desde la fecha de fabricación y la inspección más reciente y la combinación entre fecha de inspección, tomando el escenario más crítico como valor predominante para cada clúster, las anomalías identificadas como ruido son analizadas de forma independiente ya que son anomalías individuales que no tiene comparación en el tiempo y son de elevada importancia para el cálculo de la vida remanente.

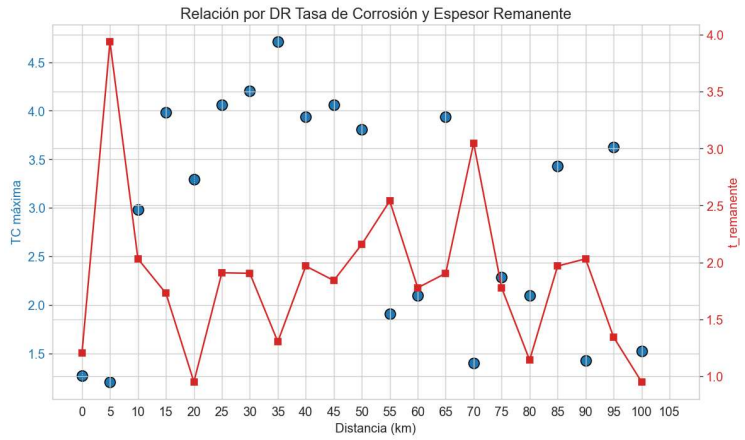

Figura 12. Relación distancia de registro, tasa de corrosión y espesor remanente.

En la Figura 12 se observa que la TC máxima presenta una dispersión considerable que oscila entre 1.3 y 4.7 mm por año, los mayores valores se centran en corrosión externa la cual se presenta a partir del PK10. Por su parte, el espesor remanente presenta un comportamiento más oscilatorio, pero con una tendencia general a la disminución en tramos específicos, especialmente hacia el extremo derecho entre el PK 90 y PK 105, donde se detectan valores mínimos cercanos a 1.0 mm. Este comportamiento sugiere que en varias ubicaciones donde la TC máxima es elevada, el espesor remanente tiende a ser bajo, indicando zonas potenciales de mayor deterioro. Esta relación inversa es consistente con los mecanismos típicos de pérdida de material por corrosión en sistemas metálicos, donde una mayor tasa de ataque acelera la reducción de espesor de pared.

4.3.1.2.Vida remanente (RL)

Con la tasa de corrosión ya determinada para cada clúster y anomalía individual, se dispone de una estimación de la velocidad de deterioro, lo que permite proyectar el tiempo estimado para una posible pérdida de contención. Este análisis, conocido como vida remanente (RL), es fundamental para evaluar el estado del sistema y priorizar las acciones de mantenimiento en el tiempo.

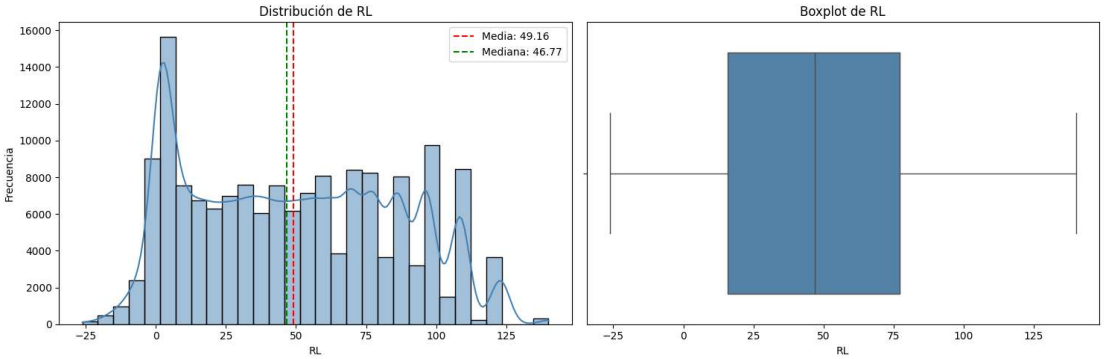

Figura 13. Distribución de la vida remanente y análisis de valores atípicos.

En la Figura 13 se presenta la distribución de la vida remanente en intervalos de 5 km, lo cual permite identificar tramos con vida útil comprometida. Se observa que la

distribución de vida remanente (RL) presenta asimetría positiva leve, con una concentración significativa de valores entre 0 y 50, lo que indica predominancia de tramos con vida útil baja a moderada. La media (49.16) es ligeramente superior a la mediana (46.77), reflejando la presencia de valores altos que elevan el promedio. Se identifican valores negativos de RL, que corresponden a zonas críticas fuera de especificación y deben ser priorizadas. La alta dispersión y presencia de *outliers* revelan una condición no homogénea a lo largo del sistema. Como se puede observar, estas dos variables están relacionadas de forma inversa a medida que la tasa de corrosión (TC) aumenta, la vida remanente (RL) tiende a disminuir.

En la Figura 14 se presentan los resultados segmentados cada 5 km a lo largo del ducto, lo que permite visualizar los tramos donde esta relación es más crítica. Esta representación facilita la identificación de zonas con mayor riesgo estructural, orientando la priorización de intervenciones y estrategias de mantenimiento basadas en el deterioro acumulado.

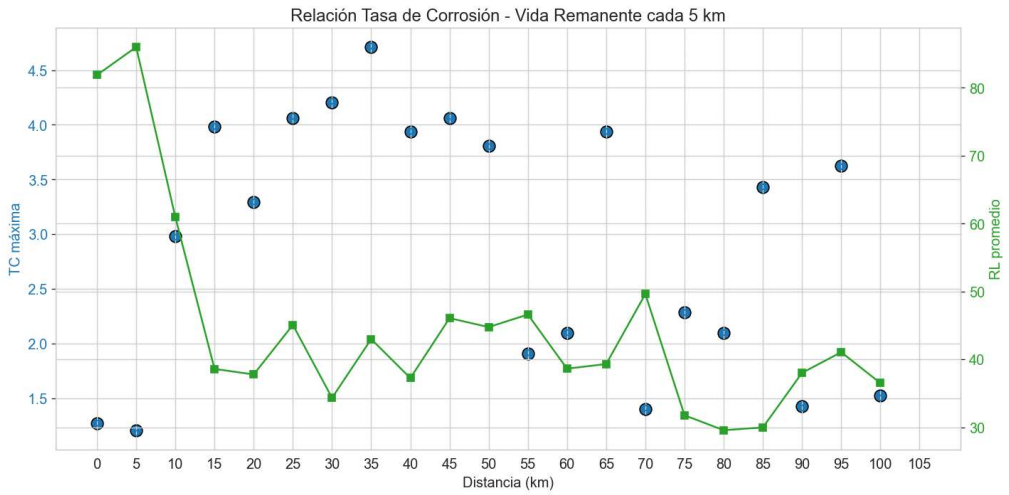

Figura 14. Relación entre la tasa de corrosión y vida remanente.

La tendencia inversa presentada en la Figura 14, es coherente con los principios de deterioro por corrosión. En los primeros 10 km se registran las TC más altas (4.5–4.7 mm/año) y las menores RL (menores a 40 años). Entre los km 15 y 70, la TC se estabiliza en valores moderados (1.5–2.2 mm/año), con una recuperación parcial de la RL (35–50 años), aunque con variaciones no lineales. A partir del km 70 reaparecen picos de TC, con RL descendiendo hasta los 30 años, señalando tramos críticos que requieren asignación de actividades de intervención.

4.3.2. Análisis de agrietamiento

4.3.2.1. Tasa de crecimiento de grietas

La estimación de la tasa de crecimiento de grietas se calculó basado en mecanismos de propagación por fatiga y estadísticas aplicada a la información disponible. Este análisis consideró tanto componentes dependientes del tiempo como del ciclo de carga, permitiendo simular el avance de defectos a partir de condiciones iniciales estimadas y parámetros de carga operacional. Se incorporaron criterios de tolerancia y factores de

incertidumbre en los datos de entrada, lo que permitió definir umbrales críticos de falla y establecer intervalos óptimos para la reevaluación estructural. En la Figura 15 se presenta una evaluación del perfil de espesor remanente (representado mediante barras amarillas) y la tasa de crecimiento del daño por corrosión (línea azul), con segmentos cada 10 km.

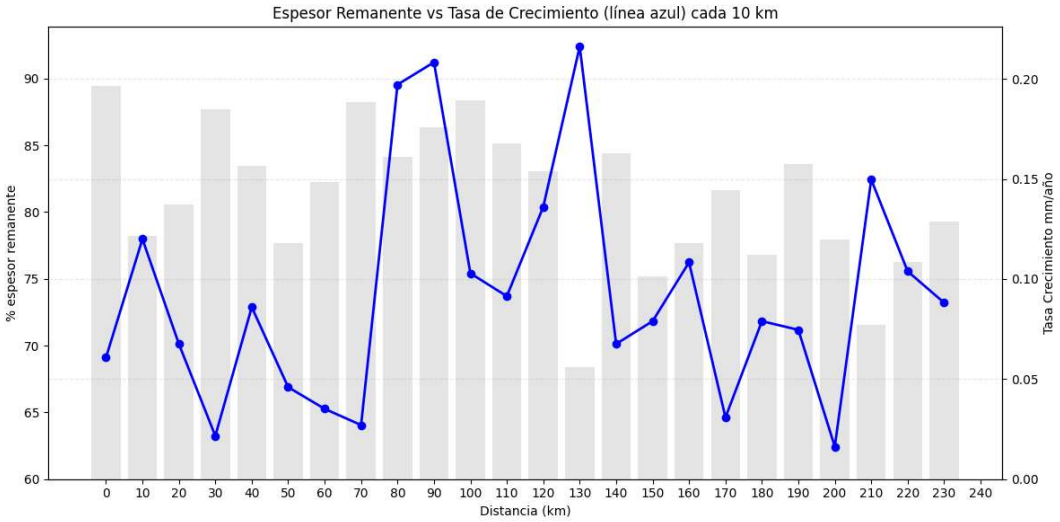

Figura 15. Relación entre tasa de crecimiento grietas Vs % espesor remanente.

Se observa una variabilidad significativa en ambas variables, donde el espesor remanente se mantiene mayoritariamente por encima del 75%, aunque existen sectores puntuales con valores cercanos al 65%. La tasa de crecimiento presenta picos importantes, destacando el registrado en el kilómetro 130, que supera los 0.1 mm/año, lo que sugiere un proceso acelerado de propagación de anomalías tipo grieta.

4.3.2.2.Relación presión de falla y vida remanente

La relación entre la presión de falla (FPR) y la vida remanente constituye un elemento crítico en la evaluación de integridad estructural de ductos con defectos tipo grieta. La FPR, definida como la relación entre la presión estimada de ruptura (ERP) y la presión de operación (MOP/MAOP o transitoria), permite cuantificar la cercanía de una anomalía a condiciones críticas de falla. Simultáneamente, la vida remanente se determina a partir del tiempo requerido para que una grieta inicial crezca hasta alcanzar un tamaño crítico, bajo la influencia de cargas cíclicas o procesos dependientes del tiempo, como la corrosión bajo tensión. Una FPR baja implica un margen de seguridad reducido, acelerando la necesidad de intervención, mientras que la estimación precisa de la vida

remanente guía los intervalos de reevaluación y define las prioridades de mitigación. En la Figura 16, se integran las dos variables en el análisis de defectos tipo grieta.

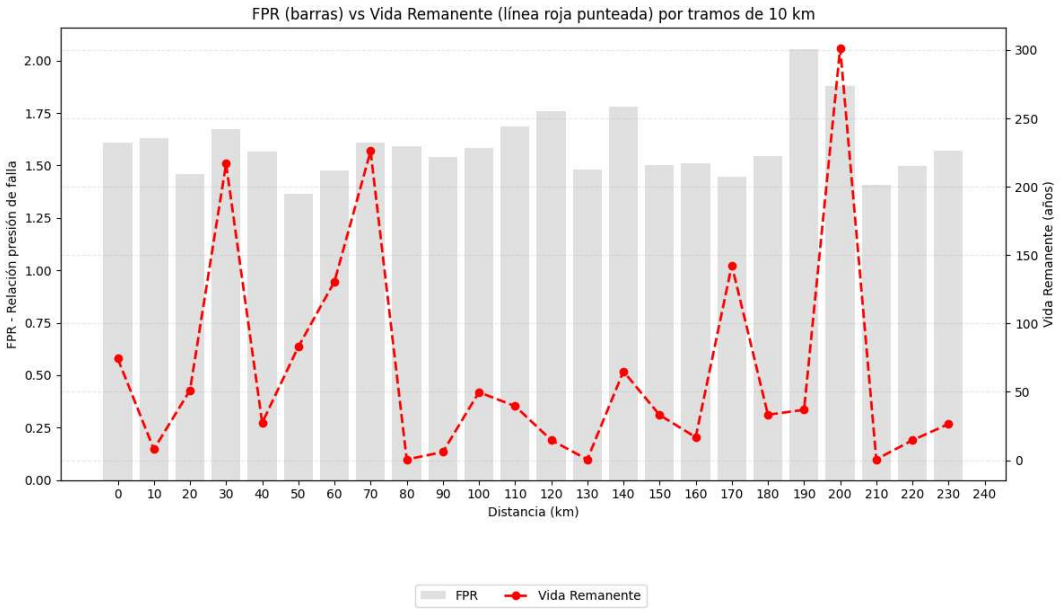

Figura 16. Relación entre FPR y vida remanente.

Se observa que la mayoría de los tramos mantiene valores de FPR superiores a 1.1, lo que indica que las anomalías presentes aún conservan un margen estructural aceptable frente a las condiciones operativas. No obstante, la vida remanente muestra una dispersión significativa, con varios segmentos registrando valores cercanos a cero, lo cual sugiere que, pese a la adecuada capacidad estructural actual, ciertos defectos están sujetos a mecanismos de crecimiento activo que podrían reducir rápidamente la seguridad del sistema si no se gestionan oportunamente. En contraste, algunos tramos alcanzan vidas remanentes superiores a 150 años, lo que podría asociarse a menores niveles de solicitación cíclica o defectos menos críticos.

4.4.Modelos predictivos

4.4.1. Modelos para anomalías de pérdidas de metal (Corrosión)

La gestión de la integridad estructural en sistemas afectados por corrosión es un desafío clave para industrias como la energética, petroquímica y de transporte de fluidos, donde las fallas imprevistas pueden tener consecuencias técnicas y económicas significativas. En este contexto, la tasa de corrosión es modelada considerando inicialmente 14 variables independientes: X_1 (distancia de registro, DR), X_2 (longitud de la anomalía), X_3 (ancho de la anomalía), X_4 (profundidad de la anomalía), X_5 (segmento o tramo del tubo), X_6 (posición horaria en segundos), X_7 (ubicación de la anomalía: superficie externa o interna), X_8 (número de clúster), X_9 (espesor nominal), X_{10} (tiempo de exposición a largo plazo), X_{11} (tiempo de exposición a corto plazo), X_{12} (espesor remanente en el clúster), X_{13} (espesor requerido por el código de diseño) y X_{14} (velocidad de corrosión a fecha de construcción). Estas variables representan una combinación de

características geométricas, geoespaciales, operativas y de integridad estructural, lo que permite un enfoque integral para la predicción de la tasa de corrosión.

El modelado predictivo incluye regresiones lineales y logarítmicas, así como algoritmos de aprendizaje automático. Entre estos se consideran el perceptrón multicapa (*Multilayer Perceptron*, MLP), adecuado para capturar relaciones no lineales complejas; las máquinas de vectores de soporte (*Support Vector Machines*, SVM), eficaces en contextos con ruido o datos limitados; y los bosques aleatorios (*Random Forest*, RF), reconocidos por su robustez y capacidad de interpretación.

Considerando la cantidad de variables y la posible existencia de multicolinealidad entre ellas, se aplicó un Análisis de Componentes Principales (PCA) como etapa preliminar de procesamiento. En la Tabla 3, se presenta la interpretación de los resultados obtenidos mediante la correlación entre los componentes principales y las variables originales.

Variable	PC1	PC2	PC3
DR[m]	**0,659**	0,659	-0,320
Longitud	0,394	**-0,590**	-0,423
Ancho	**0,591**	-0,536	-0,421
Profundidad	**0,804**	-0,164	0,504
segmento	**0,659**	0,659	-0,320
PH_Seg	0,013	0,003	-0,146
Sup_Externa_Num	0,000E+00	8,818E-18	3,303E-17
Cluster_OPTICS	0,095	0,128	0,146
t_nom	-2,536E-23	2,133E-30	1,100E-29
Dif_LT	**-0,528**	0,115	0,461
Dif_ST	-0,257	-0,121	-0,273
TC_LT_mmpy	**0,863**	-0,197	0,381
t_rem_cluster	**-0,743**	0,000	-0,355
treq_mm	0,000	0,000	0,000

Tabla 3. Correlación componentes principales y variables originales para corrosión.

En la Tabla 3, se consideran de alta importancia aquellas variables que presentan una correlación superior a 0.5, las cuales se destacan en negrita. Los dos primeros componentes principales explican el 61 % de la varianza total de los datos. El primer componente principal muestra una fuerte correlación con siete de las variables originales, aumentando con X_{1p} (distancia de registro), X_{2p} (ancho), X_{3p} (profundidad), X_{4p} (segmento o tubo), y X_{5p} (velocidad de corrosión a largo plazo), mientras que disminuye con X_{6p} (antigüedad respecto a la fecha de fabricación). Este componente se asocia principalmente con características geográficas y temporales del proceso de corrosión. Por su parte, el segundo componente principal exhibe una correlación negativa con X_{7p} (longitud del defecto), indicando que a medida que esta variable aumenta, la velocidad de corrosión tiende a disminuir.

Con las 8 variables seleccionadas mediante el Análisis de Componentes Principales, se procede a la comparación de los seis modelos utilizando métricas el coeficiente de determinación (R^2), que cuantifica la proporción de la variabilidad total explicada por el modelo; el error absoluto medio (MAE), que representa el promedio de las magnitudes de los errores sin considerar su signo; y la raíz del error cuadrático medio (RMSE), que penaliza de forma más severa los errores de mayor magnitud, proporcionando una medida sensible a grandes desviaciones. En la tabla 4 se presentan los resultados de las métricas para los 6 modelos propuestos.

Modelo	**R^2**	**MAE**	**RMSE**
Regresión lineal	0,680	0,063	0,118
Reg. Lineal Logaritmica	0,702	0,051	0,114
Gradient Boosting	0.985	0,012	0,050
Multilayer Perceptron	0,943	0,016	0,050
Support Vector Machines	0,895	0,016	0,050
Random Forest	0.989	0,011	0,050

Tabla 4. Resultados modelos de predicción para anomalías de perdida de metal (corrosión)

Los modelos de regresión lineal y logarítmica mostraron un desempeño moderado, con coeficientes de determinación de 0,680 y 0,702, respectivamente, y errores absolutos medios y cuadráticos, más altos que los modelos de aprendizaje automático. En contraste, los modelos basados en técnicas de *ensemble learning*, particularmente *Random Forest* y *Gradient Boosting*, presentaron un desempeño superior, alcanzando valores de R^2 mayores a 0.9, junto con errores MAE y RMSE muy bajos, lo que indica buena capacidad de generalización y precisión.

Para evaluar la posible presencia de sobreajuste, se compararon las métricas obtenidas en los conjuntos de entrenamiento y prueba. En el caso del modelo *Gradient Boosting*, se obtuvo un R^2 de 0.985, con valores de RMSE de 0.050 para entrenamiento y 0,050 para prueba. Por su parte, el modelo *Random Forest* también presentó un R^2 de 0.978 tanto en entrenamiento como en prueba, con valores de RMSE de 0,040 y 0,030, respectivamente, lo cual indica que el modelo explica el 97 % de la variabilidad en la variable la tasa de corrosión. Adicional se realizó una validación cruzada, obteniéndo un R^2 promedio de 0.985 para *Gradient Boosting* y de 0,978 para *Random Forest*. Esto implica que ambos modelos explican el 98.5 % y el 97,8 % de la varianza de la tasa de corrosión,

respectivamente. Los errores promedio de RMSE fueron de 0.045 ± 0.014 para *Gradient Boosting* y de 0.0026 ± 0.0016 para *Random Forest*.

4.4.2. Modelos para anomalías tipo grieta

La relación de presión es modelada considerando inicialmente 12 variables independientes: X_1 (distancia de registro, DR), X_2 (número de clúster), X_3 (año de corrida o inspección), X_4 (Espesor en zona sana), X_5 (espesor remanente en el clúster), X_6 (posición horaria en segundo), X_7 (longitud de la anomalía), X_8 (profundidad de la anomalía), X_9 (MAOP presión máxima permisible), X_{10} (Presión de falla), X_{11} (tasa de crecimiento en mm por año), X_{12} (vida remanente en años). Estas variables representan una combinación de características geométricas, geoespaciales, operativas y de integridad estructural, lo que permite un enfoque integral para la predicción de la relación de presión de falla.

Considerando la cantidad de variables y la posible existencia de multicolinealidad entre ellas, se aplicó un Análisis de Componentes Principales (PCA) como etapa preliminar de procesamiento. En la Tabla 5, se presenta la interpretación de los resultados obtenidos mediante la correlación entre los componentes principales y las variables originales.

Variable	PC1	PC2	PC3
DR_m	**-0,535**	-0,191	0,238
Cluster_DR_PH	**0,544**	0,474	0,207
Año	-0,045	0,149	**0,796**
Espesor	**0,771**	0,556	-0,076
t_rem	**0,715**	-0,561	0,179
PH_seg	0,065	0,052	0,340
Longitud	0,011	-0,029	-0,267
Profundidad	**-0,556**	0,713	-0,218
MAOP	**0,703**	0,275	-0,324
P_burst	**0,584**	0,559	0,231
Tasa_Crecimiento_mm_año	**-0,519**	0,646	-0,001
Vida_Remanente_aprox	0,428	-0,401	-0,155

Tabla 5. Correlación componentes principales y variables originales para grietas.

En la Tabla 5, se consideran de alta importancia aquellas variables que presentan una correlación superior a 0.5, las cuales se destacan en negrita. Los dos primeros componentes principales explican el 58 % de la varianza total de los datos. El primer componente principal muestra una fuerte correlación con siete de las variables originales, aumentando con el clúster asignado, espesor en zona sana, X_{1p} (espesor remanente), X_{2p} (máxima presión de operación) y X_{3p} (presión de falla), mientras que disminuye con la X_{4p} (distancia de registro), X_{5p} (profundidad de la grieta) y X_{6p} (tasa de crecimiento en mm por año). Este componente se asocia principalmente con características geográficas y dimensiones de la grieta. Por su parte, el segundo componente principal no aporta variables. El tercer componente exhibe una correlación positiva con X_{7p} (año de la inspección).

Las métricas empleadas incluyen: el coeficiente de determinación (R^2), que cuantifica la proporción de la variabilidad total explicada por el modelo; el error absoluto medio (MAE), que representa el promedio de las magnitudes de los errores sin considerar su signo; y la raíz del error cuadrático medio (RMSE), que penaliza de forma más severa los errores de mayor magnitud, proporcionando una medida sensible a grandes desviaciones. En la tabla 6 se presentan los resultados de las métricas para los 6 modelos propuestos.

Modelo	**R^2**	**MAE**	**RMSE**
Regresión lineal	0.555	0.252	0.346
Reg. Lineal Logarítmica	0.656	0.206	0.304
Gradient Boosting	0.766	0.127	0.251
Multilayer Perceptron	0.646	0.234	0.309
Support Vector Machines	0.679	0.183	0.294
Random Forest	0.812	0.107	0.225

Tabla 6. Resultados modelos de predicción para anomalías tipo grieta

El modelo *Random Forest* alcanzó el mejor rendimiento global con un R^2 de 0.812, MAE de 0.107 y RMSE de 0.225, lo que evidencia una alta capacidad de generalización y precisión. El modelo *Gradient Boosting* también mostró un desempeño competitivo (R^2 = 0.766), destacándose entre los métodos evaluados. En contraste, el perceptrón multicapa (MLP) obtuvo los peores resultados, con un R^2 de 0.646 y los errores más altos (MAE = 0.234; RMSE = 0.309), lo que sugiere una menor capacidad para capturar la complejidad del fenómeno.

5. Conclusiones

Los modelos *Random Forest* y *Gradient Boosting* demostraron un desempeño superior en la predicción de variables críticas asociadas tanto a corrosión como a agrietamiento. En el caso de la tasa de corrosión, ambos modelos alcanzaron un coeficiente de determinación R^2=0.9 y errores mínimos (RMSE < 0.02), mientras que, para la predicción de la relación de presión de falla en defectos tipo grieta, Random Forest alcanzó un R^2=0.810, destacándose sobre otros enfoques evaluados.

La aplicación del PCA permitió reducir la dimensionalidad del problema sin pérdida significativa de información, lo que facilitó la identificación de las variables más relevantes para el modelado. Esta estrategia mejoró la estabilidad y la interpretabilidad de los modelos, contribuyendo a una mayor eficiencia computacional y precisión en la predicción.

Los algoritmos DBSCAN, HDBSCAN y OPTICS permitieron identificar agrupaciones espaciales de anomalías con comportamiento evolutivo común, revelando zonas críticas recurrentes en el tiempo. Esta segmentación espacial y temporal es fundamental para priorizar intervenciones en tramos con alta densidad y severidad de daño.

La metodología propuesta en el presente documento constituye una herramienta para la toma de decisiones en programas de integridad estructural. Al integrar datos históricos y actuales con técnicas avanzadas de análisis, se logra una evaluación proactiva de la condición del ducto, optimizando los recursos de mantenimiento y fortaleciendo la confiabilidad operativa del sistema de transporte de hidrocarburos.

Como extensión del presente estudio, se proponen diversas líneas de investigación que permitan profundizar y optimizar la metodología desarrollada. En primer lugar, se plantea la incorporación de variables operacionales en tiempo real, como presión, temperatura y caudal, a fin de enriquecer los modelos predictivos con información dinámica que refleje mejor las condiciones a las que están sometidos los ductos.

Con base en los resultados obtenidos, se destaca la necesidad de que el sector de hidrocarburos avance hacia la implementación de sistemas integrados de apoyo a la toma de decisiones, fundamentados en los modelos predictivos desarrollados. Estos sistemas permitirán automatizar la priorización de intervenciones y optimizar los planes de mantenimiento de manera dinámica, considerando la evolución temporal del daño, la severidad estructural y criterios de costo-beneficio. La adopción de este enfoque facilitará una gestión proactiva de la integridad, maximizando la eficiencia operativa y reduciendo los riesgos asociados a las pérdidas de contención.

REFERENCIAS BIBLIOGRÁFICAS

API_1163. (2021). *In-line Inspection Systems Qualification*.

Hussain, M., Zhang, T., Chaudhry, M., Jamil, I., Kausar, S., & Hussain, I. (2024). Review of Prediction of Stress Corrosion Cracking in Gas Pipelines Using Machine Learning. In *Machines* (Vol. 12, Issue 1). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/machines12010042

Martinez Brandon Barrena. (2019). *Metodología para la evaluación de la integridad de tuberías corroídas mediante la predicción del factor estimado de reparación a través de la aplicación de algoritmos de Machine Learning*.

Mohd Zain, A., & Md Noor, N. (2015). *An artificial neural network modeling for pipeline corrosion growth prediction*. https://www.researchgate.net/publication/272353772

Muhammad Hussain, Tieling Zhang, Ishrat Jamil, Afzal Ahmed Soomro, & Intizar Hussain. (2024). ApplicationofMachineLearningApproachestoPredictionofCorrosionDefectsinEnergyPipelines. *Advances in Corrosion Modelling, Engineering Materials*, 1–40.

Phan, H. C., & Duong, H. T. (2021). Predicting burst pressure of defected pipeline with Principal Component Analysis and adaptive Neuro Fuzzy Inference System. *International Journal of Pressure Vessels and Piping*, *189*. https://doi.org/10.1016/j.ijpvp.2020.104274

Şahin, E., & Yüce, H. (2023). Prediction of Water Leakage in Pipeline Networks Using Graph Convolutional Network Method. *Applied Sciences (Switzerland)*, *13*(13). https://doi.org/10.3390/app13137427

Sebastián, J., Moyano -201423841, M., Amaya Gómez, R., Ríos, N., Camilo, R., Franco De Los Ríos, A., & Barrera Martínez, B. (n.d.). *Analysis of Source Models and Clustering in the Characterization of Corrosion-Caused Leaks*.

Zhuang, Q., Liu, D., & Chen, Z. (2024). Prediction on Failure Pressure of Pipeline Containing Corrosion Defects Based on ISSA-BPNN Model. *Energy Engineering: Journal of the Association of Energy Engineering*, *121*(3), 821–834. https://doi.org/10.32604/ee.2023.044054