CMC MÉXICO 2025

Sesión TOOLBOX

1

CONGRESO DE MANTENIMIENTO & CONFIABILIDAD MÉXICO 18ª EDICIÓN

TOOLBOX

Implementación de Gemelos Digitales en el Mantenimiento Predictivo – PdM

Adolfo Crespo Márquez

Presidente de INGEMAN

Prof. ETSI Universidad de Sevilla

España

2

Resumen

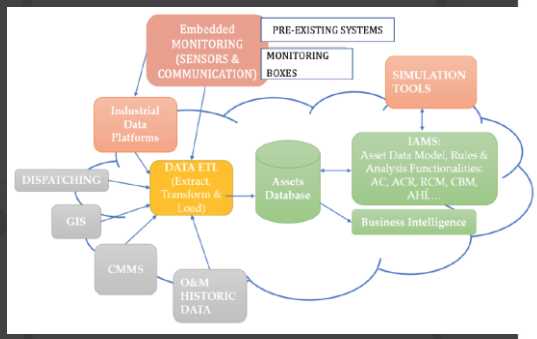
- Esta sesión presenta un **marco genérico para la descripción funcional de Gemelos Digitales (DT)** orientados al mantenimiento inteligente.
- El asistente finaliza la sesión **aprendiendo a describir la arquitectura del GD, y cómo interactúa los los responsables de mantenimiento** para la resolución del problema de PdM.
- Aprenderemos a hacerlo describiendo un sistema de PdM de la empresa TALGO, utilizando el marco de referencia que proponemos en el curso.

3

Contenido

- Resumen – utilización del marco de referencia para GDM
- El nuevo contexto industrial – El problema / Necesidad Marco
- Marco de gestión digital del mantenimiento (GDM)
- Introducción al caso de GD PdM de TALGO
- Requisitos a cumplir fijados por los técnicos/empresas para los GD
- Estudio con detalle del GD PdM de TALGO: Analítica utilizada: detección, diagnosis y prognosis.
- La Interacción de los técnicos con el GD: DIAGRAMA DE EVENTOS-ACCIONES-ESTADOS
- Cumplimiento de los requisitos de gemelo

4

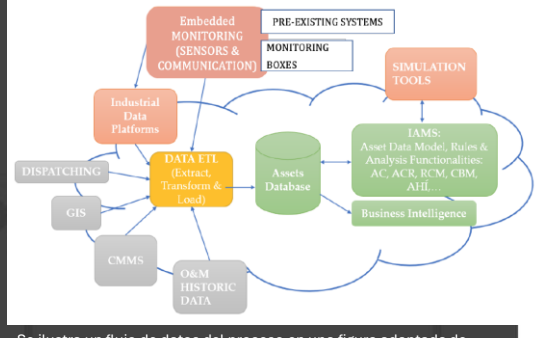


Se ilustra un flujo de datos del proceso en una figura adaptada de Crespo et al. (2020).

El nuevo contexto

- Avances rápidos en tecnologías digitales, analítica de datos e inteligencia artificial están revolucionando el mantenimiento.
- Estos enfoques permiten comprender mejor el funcionamiento de sistemas industriales complejos y gestionarlos de forma más eficaz.
- Los datos son clave en la mejora de la gestión del mantenimiento, facilitando decisiones basadas en información precisa.
- Se utilizan procesos automáticos de extracción, transformación y carga (ETL) de datos para apoyar decisiones de mantenimiento.
- Las aplicaciones inteligentes de gestión de activos (IAMS Apps) organizan y analizan los datos, respaldando la toma de decisiones.

5



Se ilustra un flujo de datos del proceso en una figura adaptada de Crespo et al. (2020).

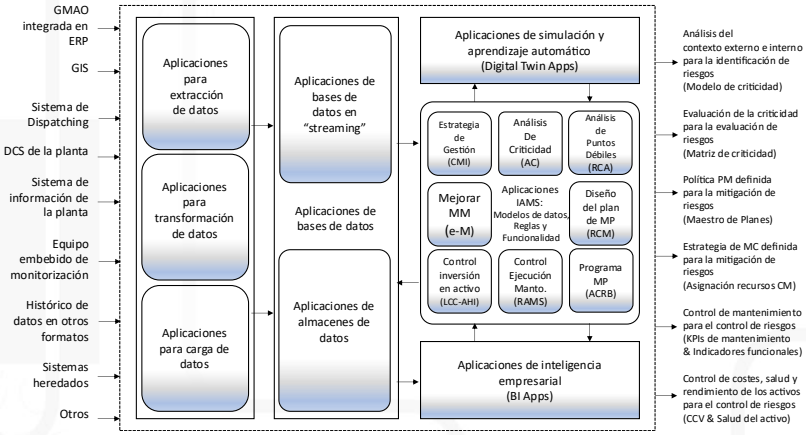
El nuevo contexto – La problemática asociada

- Las aplicaciones inteligentes de gestión de activos son complejas pues acaban integrándose con herramientas adicionales como simuladores, ofreciendo servicios analíticos complementarios.
- Estas herramientas acaban enriqueciendo las bases de datos con sus resultados y el flujo de información se complica.
- La información abastece también nuevas aplicaciones de descubrimiento, creación y almacenamiento de información sobre activos, además de inteligencia empresarial (BI Apps).
- La arquitectura del sistema completo es difícil muchas veces de explicar, el ingeniero de mantenimiento experimenta entonces una sensación de frustración ante la difícil interacción con los ingenieros de sistemas y los proveedores de software.

Este es el primer punto a resolver en el curso

6

Un marco de referencia para la Gestión Digital del Mantenimiento



#CMCMéxico2025

7

Interacción "Operarios- GD". Requisito a cumplir por el GD

- Aplicable a más equipos o fallos.
- Combinable con diferentes modelos y datos.
- Extensible con el desarrollo de nuevos modelos.
- Seguimiento preciso del comportamiento y estado del sistema.
- Toma de decisiones e interacción humana.
- Integración general de procesos y arquitectura del sistema.

#CMCMéxico2025

8

Dificultades clave en la interacción de los operarios con un Gemelo Digital (GDM)

1. Nivel conceptual: Barreras de comprensión

- **Desconexión entre el modelo digital y el modelo mental del operario**:
  El GDM se basa en analítica avanzada (IA, modelos predictivos, algoritmos de RUL), mientras que el operario tiene una visión práctica basada en experiencia y observación directa
- **Falta de visualización intuitiva**:
  Muchos GDM no presentan sus resultados de forma accesible (gráficos comprensibles, alertas claras, explicaciones del "por qué", acciones prácticas inmediatas).
- **Sospecha o desconfianza hacia "cajas negras"**:
  El operario puede dudar de las recomendaciones automáticas si no entiende cómo se han generado, la supervisión de cada acción es necesaria.

📌 *Ejemplos*:
¿Cómo puede el GD apreciar el problema en estas condiciones tan cambiantes de operación?
¿Por qué si puede el GD y no puede el TCMS "Train Control Monitoring System"?
¿Son los mismos datos los que reciben ambos sistemas?

Dificultades clave en la interacción de los operarios con un Gemelo Digital (GDM)

2. Nivel operativo: Fricción en el uso diario

- **Retrasos en la actualización de datos** o *latencia* que afecta decisiones en tiempo real.
- **Volumen de alertas o información irrelevante**: "Ruido digital" que satura al usuario y hace que pierda confianza en el sistema.
- **No integración con las tareas reales**: Si el GDM detecta una anomalía, pero no lanza una orden de trabajo o no se alinea con el proceso del GMAO, no aporta valor inmediato al operario.

Dificultades clave en la interacción de los operarios con un Gemelo Digital (GDM)

3. Nivel organizativo: Cultura, formación y roles

- **Falta de formación adecuada**: El operario no ha sido entrenado para interpretar modelos o indicadores avanzados como RUL o probabilidad de fallo.
- **Cambio de rol del operario no gestionado**: Pasan de "ejecutores" a "supervisores inteligentes" sin que se les haya explicado o valorado su nuevo rol.
- **Falta de participación en el diseño del GDM**: Si no se ha involucrado a los técnicos desde el principio, el GDM difícilmente se ajustará a sus necesidades reales.
- **Temor al reemplazo**: Puede percibirse el GDM como una amenaza a su puesto o a su experiencia tradicional.

📌 *Ejemplos*:
¿Pero quién habrá diseñado este proceso de locos?

#CMCMéxico2025

11

¿Cómo hacer para que la interacción del GD con los operarios de mantenimiento sea fácil?

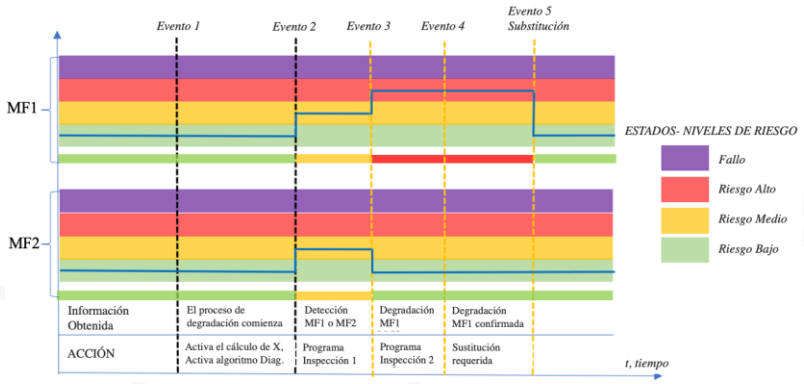
Segundo punto a resolver en la sesión

12

Diagramas DIO-GD

Un **diagrama de interacción operativa del gemelo digital (DIO-GD)** es una herramienta visual que representa cómo los eventos técnicos detectados por el sistema (eventos), las respuestas propuestas o ejecutadas (acciones) y el estado del sistema o del componente (estado) se interrelacionan a lo largo del tiempo para apoyar decisiones de mantenimiento.

Frente a un diagrama de estados tradicional de ingeniería de software, **DIO-GD se posiciona como una herramienta híbrida entre la ingeniería de sistemas, el mantenimiento y la gestión del cambio.**



#CMCMéxico2025

13

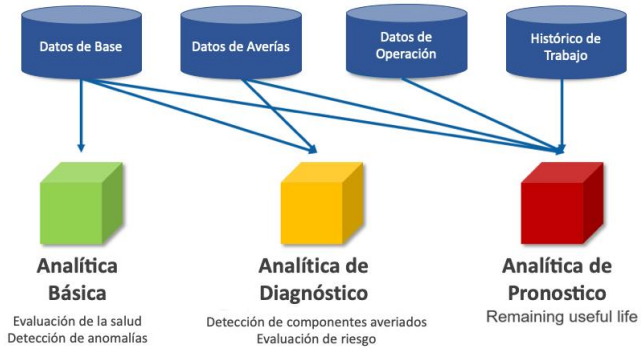
🛠 GD para PdM en TALGO. Caso a utilizar para aprender

- **Desafío**: Detectar y predecir fallos en cojinetes de ejes de trenes de alta velocidad con base únicamente en la temperatura del rodamiento.
- **Limitaciones previas**:
  - Modelos físicos complejos y de difícil implementación en condiciones reales.
  - Dificultad para distinguir entre fallos internos (degradación) y externos (problemas de guiado).
  - Volumen excesivo de datos poco útil para la toma de decisiones.
- **Solución implementada**:
  - Desarrollo de un **Gemelo Digital** orientado al mantenimiento basado en condición (CBM).
  - Sustitución del modelo físico por uno **basado en datos** (machine learning + reglas de negocio).
  - Clasificación automática de modos de fallo y estimación de la **vida útil remanente (RUL)**.
- **Impacto**:
  - Mejora de la **eficacia diagnóstica**, reducción del volumen de datos necesarios, y planificación proactiva del mantenimiento con alta fiabilidad.

#CMCMéxico2025

14

🛠 El Caso: GD de CBM de Rodamientos de Tren



Alcance específico de este caso:

- Detección de fallos en rodamientos del eje de tren.
- Clasificación de fallos en rodamientos del eje de tren.
- Predicción de fallos en rodamientos del eje de tren.

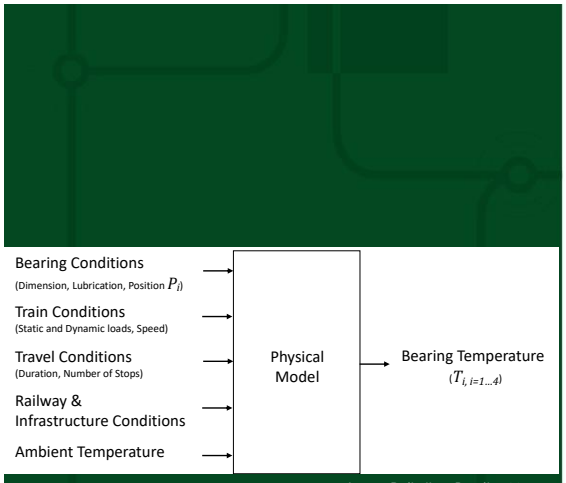
Nota importante : Monitorización limitada exclusivamente a la temperatura de cada rodamiento.

15

Veamos cómo se implementó cada elemento de la analítica utilizada

#CMCMéxico2025

16



Lunys, Dailydka y Bureika, 2015, Mironov AA, 2008.

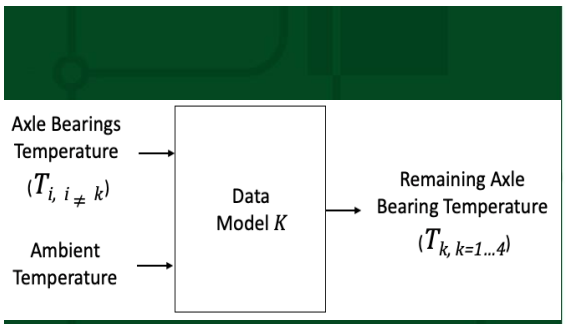
🛠 Detección de anomalías

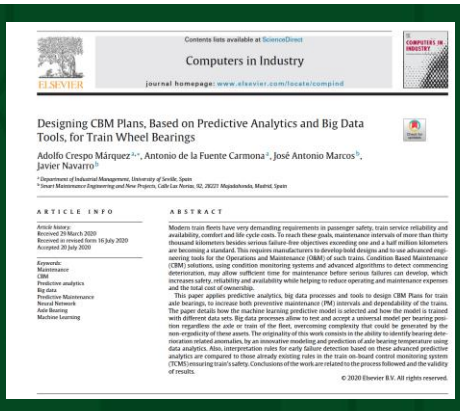
La temperatura del rodamiento del eje de un tren en régimen de circulación continua depende de los siguientes factores:

- Tipo y dimensiones del rodamiento.
- Propiedades antifricción e hidrodinámicas del lubricante.
- Holguras entre rodillos y anillos del rodamiento.
- Cargas estáticas y dinámicas sobre el rodamiento.
- Velocidad de circulación del tren.
- Duración del recorrido sin paradas.
- Temperatura ambiente.
- Curvas del trazado ferroviario.

Lunys, Dailydka y Bureika, 2015,
Mironov AA, 2008.

17





Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Designing CBM Plans, Based on Predictive Analytics and Big Data Tools, for Train Wheel Bearings

Adolfo Crespo Márquez[a,*], Antonio de la Fuente Carmona[a], José Antonio Marcos[b], Javier Navarro[b]

[a] Department of Industrial Management, University of Seville, Spain
[b] Smart Maintenance Engineering and New Projects, Calle Las Norias, 92, 28227 Majadahonda, Madrid, Spain

ARTICLE INFO

ABSTRACT

Modern train fleets have very demanding requirements in passenger safety, train service reliability and availability, comfort and life-cycle costs. To reach these goals, maintenance intervals of more than thirty thousand kilometers besides serious failure-free objectives exceeding one and a half million kilometers are becoming a standard. This requires manufacturers to develop bold designs and to use advanced engineering tools for the Operations and Maintenance (O&M) of such trains. Condition Based Maintenance (CBM) solutions, using condition monitoring systems and advanced algorithms to detect commencing deterioration, may allow sufficient time for maintenance before serious failures can develop, which increases safety, reliability and availability while helping to reduce operating and maintenance expenses and the total cost of ownership.

This paper applies predictive analytics, big data processes and tools to design CBM Plans for train axle bearings, to increase both preventive maintenance (PM) intervals and dependability of the trains. The paper details how the machine learning predictive model is selected and how the model is trained with different data sets. Big data processes allow to test and accept a universal model per bearing position regardless the axle or train of the fleet, overcoming complexity that could be generated by the non-ergodicity of these assets. The originality of this work consists in the ability to identify bearing deterioration related anomalies, by an innovative modeling and prediction of axle bearing temperature using data analytics. Also, interpretation rules for early failure detection based on these advanced predictive analytics are compared to those already existing rules in the train on-board control monitoring system (TCMS) ensuring train's safety. Conclusions of the work are related to the process followed and the validity of results.

© 2020 Elsevier B.V. All rights reserved.

🛠 Detección de anomalías

- Este Gemelo Digital sustituye el modelo físico teórico por un modelo basado en datos, (Crespo Márquez, de la Fuente Carmona, Marcos, & Navarro, 2020).
- Las entradas y salidas del modelo basado en datos se presentan en la Figura.
- Para estimar la temperatura de un rodamiento de eje, las únicas entradas consideradas son:
  - La temperatura de los demás rodamientos del eje.
  - La temperatura ambiente.
- Este es el principio fundamental, y un enfoque muy innovador, para construir todas las capacidades de analítica predictiva requeridas por el Gemelo Digital.

18

Resultado del algoritmo de predicción

A. Crespo Márquez, A. de la Fuente Carmona, J.A. Marcos et al. / Computers in Industry 122 (2020) 103292

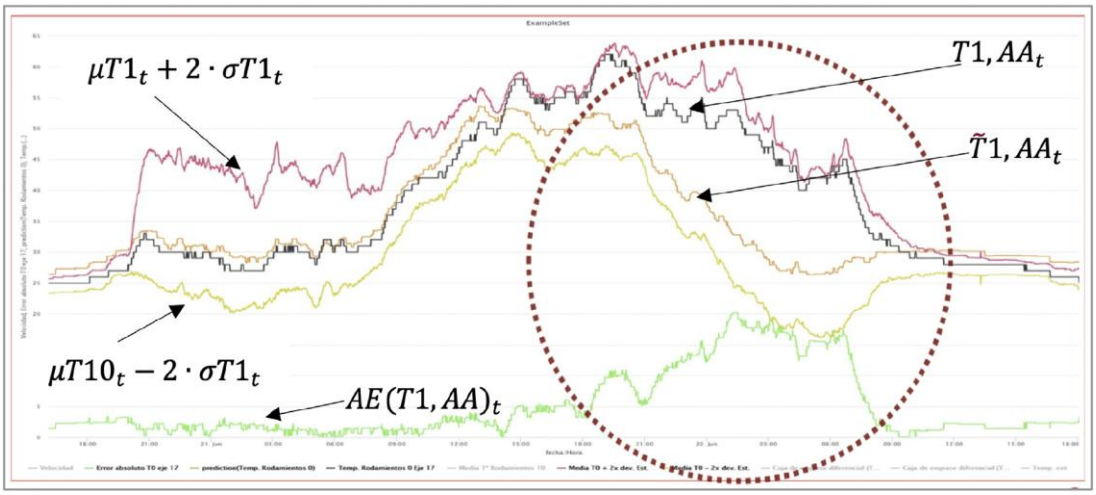


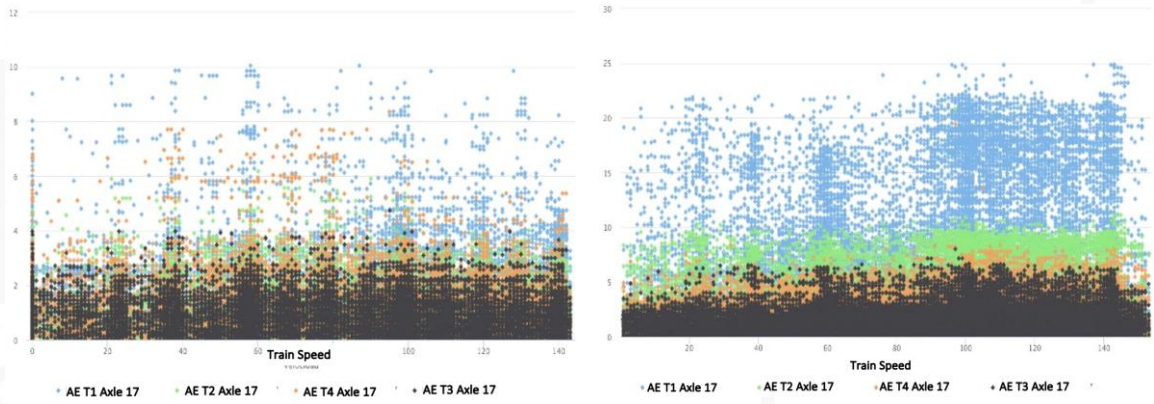
Fig. 6. Variable determining C rules in Axle Bearing (1, AA), train TRA, with $TS_{TH1} = 50\,km/h$.

#CMCMéxico2025

19

Resultado del algoritmo de predicción

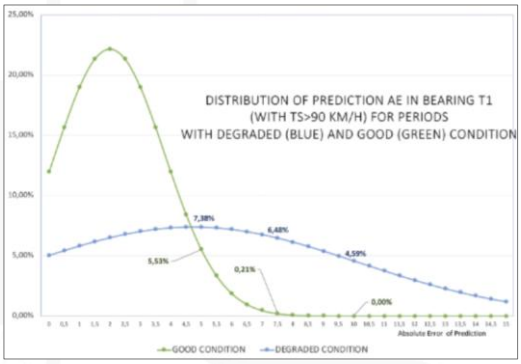
Gráficos con (izquierda) y con anomalía en T1 (derecha). errores absolutos de predicción (AE) en cuatro rodamientos de dos semiejes de un rodal
de tren sin anomalía



#CMCMéxico2025

20

Estableciendo las reglas de detección



- La regla de detección de anomalías diseñada puede identificar rodamientos dañados con un **100 % de precisión**, a cualquier velocidad del tren, utilizando un umbral de **10 °C de Error Absoluto (AE)** en la temperatura predicha del rodamiento.
- Se introdujo un **umbral de velocidad del tren** únicamente para **reducir los conjuntos de datos utilizados** y, con ello, se esperaba una mejora en la precisión de la regla. Sin embargo, se observó que la mejora en la precisión **no fue significativa en todos los casos**.
- Para ilustrar la diferencia en la **distribución del AE** cuando el rodamiento está en **buen estado** frente a cuando está **degradado**, esta figura muestra la distribución del error absoluto de predicción de temperatura:
  - **Verde**: períodos de buen estado.
  - **Azul**: períodos en estado degradado.

•Condición de análisis: velocidades del tren **$TS_t \geq 90$ km/h**.

#CMCMéxico2025

21

Resumen del Proceso seguido para la Detección de Anomalías

Modelo basado en datos:
Se reemplaza el modelo físico por uno de aprendizaje automático que predice la temperatura del rodamiento usando:

- Temperatura de los otros rodamientos del eje.
- Temperatura ambiente.

Criterio de detección de anomalías (positivos):
Se define un "positivo" cuando:

- Error Absoluto (AE) ≥ 10 °C.
- Velocidad del tren ≥ 90 km/h.
- Condición mantenida por más de 1 minuto.

Evaluación del modelo de predicción:

- Precisión del 100 % en la detección de rodamientos dañados.
- La velocidad del tren se usó para filtrar datos, pero no mejoró significativamente la precisión.

22

🛠 Clasificación o diagnóstico de Fallos

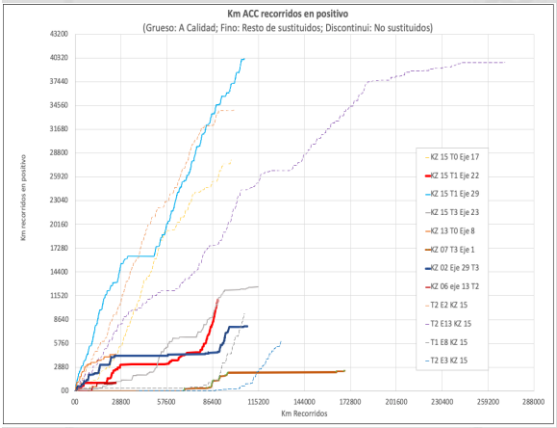
El modelo de clasificación de modos de fallo (FM) de los rodamientos del eje del tren es el **segundo modelo incluido en el Gemelo Digital (GD)** de la aplicación CBM de este estudio.

Este esfuerzo de modelado, orientado a **identificar un modo de fallo específico** del rodamiento, requirió procesos ETL adicionales y distintas herramientas de modelado.

- El principal desafío fue la decisión (tomada por el Departamento de Mantenimiento Inteligente junto con el Departamento de Ingeniería de Mantenimiento de la empresa) de **modelar ciclos de temperatura en lugar de puntos de temperatura**. Este es un **método popular** (Healey et al., 2021) para analizar datos de fatiga en componentes mecánicos.
- En estos casos, es común **reducir un espectro de esfuerzos variables** a un conjunto más simple y equivalente de esfuerzos.
- Se utilizan métodos que **extraen ciclos sucesivamente más pequeños** de una secuencia para **simplificar el cálculo de la vida a fatiga** de un componente a partir de esos ciclos más simples (Healey et al., 2021).

#CMCMéxico2025

23



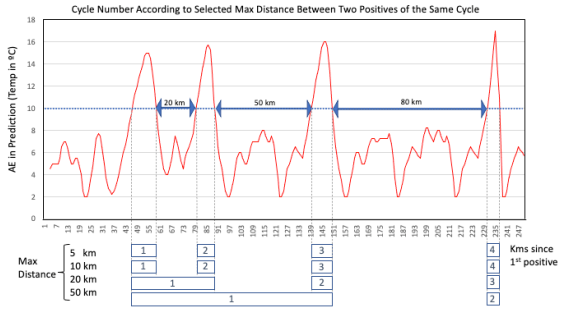
Pasos seguidos

Determinación inicial de nuevas variables, calculadas a partir de las ya extraídas (ver Figura):

- **Error absoluto acumulado (Acc AE):** Suma del error absoluto desde que se registra el primer valor positivo.
- **Kilómetros acumulados desde el primer positivo:** Total de kilómetros recorridos por el tren desde que se detectó el primer positivo.
- **Kilómetros acumulados en positivo:** Total de kilómetros recorridos en condición positiva desde el primer positivo.

24

Pasos seguidos: Transformando variables



La figura muestra cómo varía el número de ciclos detectados al modificar la **distancia máxima permitida entre dos positivos consecutivos** dentro de un mismo ciclo. En el gráfico:

- El eje vertical representa el **Error Absoluto (AE)** en la predicción de temperatura (en °C), y el eje horizontal los **kilómetros recorridos desde el primer positivo**.
- La línea roja indica la evolución del AE a lo largo del recorrido.
- Se aplica un umbral de **10 °C** para identificar los positivos.
- Las **flechas azules horizontales** indican distintas configuraciones de distancia máxima entre positivos (20 km, 50 km, 80 km).
- En la parte inferior, se muestra cómo varía el número de ciclos detectados (etiquetados como 1, 2, 3, 4) según se aumenta dicha distancia máxima:
  - Con distancias máximas más cortas (por ejemplo, 5 km), se detectan más ciclos.
  - Con distancias más largas (por ejemplo, 50 u 80 km), los positivos se agrupan dentro de menos ciclos.
- Esto permite ajustar la sensibilidad del análisis de ciclos de fallo en función del comportamiento del rodamiento y de los parámetros del modelo de predicción.

25

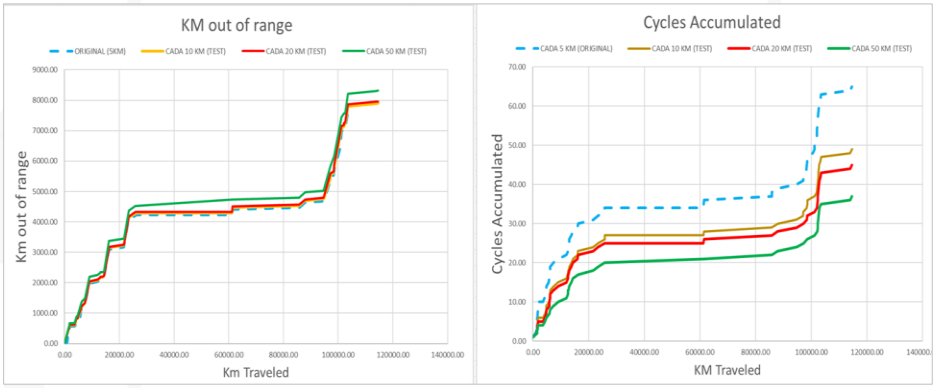
Pasos seguidos: Transformando variables

Obtención de nuevas variables de interés según el análisis de ciclos realizado:

- **Kilómetros al inicio del ciclo:** Desde el primer positivo hasta el inicio de un nuevo ciclo.
- **Kilómetros al final del ciclo:** Desde el primer positivo hasta el fin del ciclo actual.
- **Kms por ciclo:** Distancia recorrida dentro de un ciclo (el ciclo termina si el siguiente positivo está más lejos que el límite definido).
- **Kms entre ciclos:** Distancia entre el fin de un ciclo y el comienzo del siguiente.
- **Ciclos acumulados:** Número total de ciclos desde el primer positivo.
- **Porcentaje de kms en ciclo activo:** Proporción de kilómetros recorridos dentro de ciclos respecto al total desde el primer ciclo.
- **Total de kms en ciclo activo:** Suma de kilómetros recorridos dentro de ciclos.
- **Kms acumulados entre ciclos:** Suma de kilómetros recorridos entre ciclos.
- **Media de kms entre ciclos:** Promedio de los kilómetros recorridos entre ciclos, calculado progresivamente.

#CMCMéxico2025

26



- **Gráfico izquierdo: "KM fuera de rango" vs "KM recorridos"**
- Muestra cuántos kilómetros **quedan fuera** de los ciclos detectados.
- A medida que se **aumenta la distancia máxima permitida** entre positivos (de 5 a 50 km), se **reducen** los km fuera de rango.
- Esto significa que se agrupan más positivos en menos ciclos, abarcando más km dentro de los ciclos.

- **Gráfico derecho: "Ciclos acumulados" vs "KM recorridos"**
- Muestra cuántos **ciclos acumulados** se detectan al recorrer más kilómetros.
- Con una distancia máxima más **estricta (5 km)** se detectan **más ciclos**.
- Con una distancia más **amplia (50 km)** se detectan **menos ciclos**, pero más largos.

#CMCMéxico2025

27

Pasos seguidos: Transformando variables

Si bien el principal objetivo del proceso de transformación es aproximar el modelo físico de degradación de forma más simple para un determinado modo de fallo, se observó que la **cantidad de datos a considerar y almacenar** para el análisis de los rodamientos también se **reducía significativamente**.

Bearing Samples	REDUCTION of DATA POINTS for a Maximum distance between positives of a cycle of				
	1km	5km	10km	20km	50km
KZ02 T3 AXLE 29	89.346	65	55	51	42
KZ15 T2 AXLE 1	78.318	416	281	207	152

Tabla: Reducción lograda en los datos a almacenar por cada rodamiento analizado se presenta para dos ejemplos en la **Tabla 1**.

✅ **Conclusión:** El uso del algoritmo de ciclos, ajustando la distancia entre positivos, **reduce significativamente los datos necesarios**, lo cual:

- Mejora la eficiencia del almacenamiento.
- Simplifica el análisis.
- Facilita el despliegue en sistemas con recursos limitados.

#CMCMéxico2025

28

Pasos seguidos: Transformando variables

	Bearing	Replaced	Damaged	Kms end of cycle	Kms between cycles	Cycle Kms	Cumulative Cycles	Kms in act cycle	%Kms in act cycle	Acc Kms betw Cycles	Avg Kms betw cycles
2	KZ15 T0 EJE 17	1	1	345,6	2619,9	342,6	1,0	342,6	0,3%	2619,9	2619,9
62	KZ15 T0 EJE 17	1	1	66704,7	58,5	31,7	61,0	10667,0	6,4%	56093,2	919,6
63	KZ15 T1 EJE 22	1	1	25,1	8,0	22,0	1,0	22,0	0,0%	8,0	8,0
201	KZ15 T1 EJE 22	1	1	92086,1	41,1	154,1	139,0	9694,0	5,0%	82430,0	593,0
202	KZ15 T1 EJE 29	1	1	1211,8	13,4	1209,9	1,0	1209,9	1,2%	13,4	13,4
459	KZ15 T1 EJE 29	1	1	105226,1	870,3	18,3	258,0	38379,0	18,7%	67715,5	262,5
460	KZ13 T3 EJE 23	1	1	60,9	23,0	33,8	1,0	33,8	0,0%	23,0	23,0
743	KZ13 T3 EJE 23	1	1	110897,4	3865,3	0,8	284,0	12251,3	5,8%	102484,3	360,9
744	KZ07 T3 EJE 1	1	1	79561,5	72,7	1,8	1,0	1,8	0,0%	72,7	72,7
811	KZ07 T3 EJE 1	1	1	197818,7	130,8	3,3	58,0	2342,3	0,8%	116047,5	2000,8
812	KZ02 T3 EJE 29	1	1	5,3	979,8	5,3	1,0	5,3	0,0%	979,8	979,8
883	KZ02 T3 EJE 29	1	1	114651,1	2383,2	10,2	65,0	7770,8	3,6%	109263,4	1681,0
884	KZ15 T2 EJE 1	0	0	82,7	8199,1	59,8	1,0	59,8	0,1%	8199,1	8199,1
1299	KZ15 T2 EJE 1	0	0	203641,3	7,7	3,6	416,0	27284,0	9,0%	176342,1	423,9
1300	KZ15 T2 EJE 2	0	0	4,2	18,6	3,4	1,0	3,4	0,0%	18,6	18,6
1804	KZ15 T2 EJE 2	0	0	207253,8	508,7	0,0	505,0	32954,6	10,7%	174807,1	346,2
1805	KZ15 T2 EJE 3	0	0	186233,4	2958,9	66,2	1,0	66,2	0,0%	189126,1	189126,1
1949	KZ15 T2 EJE 3	0	0	245440,2	25,8	12,4	116,0	5735,5	2,3%	239730,5	2066,6
1950	KZ15 T0 EJE 8	1	1	46,8	14,2	46,8	1,0	46,8	0,0%	14,2	14,2
1982	KZ15 T0 EJE 8	1	1	20361,6	7,9	4,7	33,0	2510,0	2,1%	17859,4	541,2
1983	KZ15 T1 EJE 8	0	0	18,5	5,5	18,5	1,0	18,5	0,0%	5,5	5,5
2059	KZ15 T1 EJE 8	0	0	184943,8	81,2	3,3	77,0	9383,9	3,3%	175641,1	2281,1
2060	KZ15 T2 EJE 13	0	0	13,7	89,8	12,6	1,0	12,6	0,0%	89,8	89,8
2937	KZ15 T2 EJE 13	0	0	270318,8	51,9	8,0	878,0	38209,0	10,3%	232160,5	264,4
2938	KZ11 T1 EJE 15	1	0	9,8	10,1	9,8	1,0	9,8	0,0%	10,1	10,1
3072	KZ11 T1 EJE 15	1	0	123716,9	235,0	6,5	116,0	13641,2	6,1%	110310,6	951,0
3073	KZ15 T1 EJE 19	1	0	32,8	23543,6	32,8	1,0	32,8	0,0%	23543,6	23543,6
3117	KZ15 T1 EJE 19	1	0	232583,5	44,1	6,3	36,0	1812,1	0,5%	230815,4	6411,5
3118	KZ02 T2 EJE 27	1	0	41,6	46,0	41,5	1,0	41,5	0,0%	46,0	46,0
3400	KZ02 T2 EJE 27	1	0	255001,7	6,4	197,9	248,0	13459,6	3,8%	241548,5	974,0
3401	KZ02 T3 EJE 27	1	0	7,3	35,9	7,3	1,0	7,3	0,0%	35,9	35,9
3463	KZ02 T3 EJE 27	1	0	319807,6	637,8	201,6	54,0	2890,3	0,7%	317554,9	5880,6
3464	KZ16 T1 EJE 29	1	0	3,0	4619,3	3,0	1,0	3,0	0,0%	4619,3	4619,3
3660	KZ16 T1 EJE 29	1	0	253375,3	8,0	1,6	171,0	7604,0	2,2%	245779,2	1437,3

Ejemplo de un extracto de tablas finales a utilizar en el algoritmo de clasificación, con datos de varios rodamientos, mostrando el número de líneas de datos por rodamiento (asumiendo 5 km como distancia máxima entre positivos de un ciclo).

#CMCMéxico2025

29

Pasos seguidos: Selección de algoritmo

- Una vez preparada la base de datos, el proceso continúa con el diseño, prueba y validación del algoritmo.
- El objetivo del algoritmo es **distinguir rodamientos con deterioro interno** de aquellos con **sobretemperatura por causas externas**, principalmente por fallos en el guiado del eje del tren.
  Para ello, es clave conocer el **diagnóstico final** de los rodamientos con ciclos de sobretemperatura: si fueron reemplazados o no, y si se detectó deterioro interno tras su análisis.
- Los rodamientos **no reemplazados**, pero con ciclos de sobretemperatura, se clasifican como **“no deteriorados”**, ya que recuperaron condiciones normales al resolverse los problemas del guiado.
- El algoritmo de clasificación se elige según criterios como **curva ROC**, **error**, **ganancia**, **tiempo de entrenamiento y ejecución**, etc.
- En este caso, se ha utilizado un modelo de **Deep Learning**.

#CMCMéxico2025

30

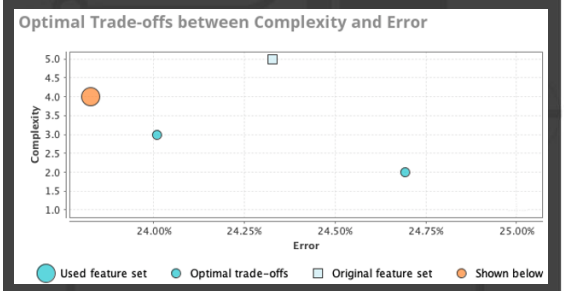


Figura A. Compromisos entre la dimensionalidad (complejidad) del modelo y el error, incluyendo la selección final de variables y sus pesos (RapidMiner ®).

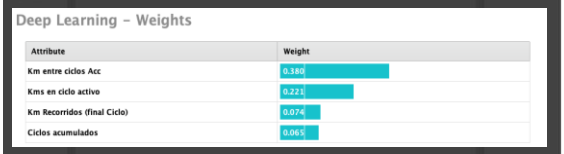


Figura B: **pesos de importancia** asignados a cada variable de entrada por el modelo de **Deep Learning**, utilizados en el proceso de clasificación de rodamientos.

Pasos seguidos: Selección de variables

En relación con la **selección final de variables** para el modelo, se observa que una configuración de **4 variables** alcanzó una **tasa de error más baja** que el conjunto original de 5 (que incluía también la **duración del ciclo** como una de las variables).

Los puntos representan diferentes combinaciones de variables:

- **Cuadro azul (Original feature set)**: modelo original con 5 variables.
- **Círculos azul claro (Optimal trade-offs)**: otras combinaciones evaluadas.
- **Círculo naranja grande (Shown below)**: modelo **final elegido** : 4 variables y menor error (~24%).

Por tanto, el modelo es **menos complejo** y aún así **más preciso** que el espacio de variables original (representado como un **cuadrado azul en la Figura A**).

Las **4 variables seleccionadas** (eliminada la variable Duración de Ciclo) en el modelo final con menor error y menor complejidad (círculo naranja) son las de **mayor peso** en la tabla de la Figura B (obtenida con RapidMiner®).

Usar menos variables también implica que los **modelos pueden entrenarse más rápido**. A continuación, se presenta el **conjunto de variables** utilizado para construir el modelo final

31

Criterion	Value	STD
Accuracy	76.3%	± 0.2%
Classification error	23.7%	± 0.2%
AUC	91.6%	± 0.2%
Precision	100%	± 0.0%
Recall	5%	± 1.0%
F Measure	9.5%	± 1.9%
Sensitivity	5%	± 1.0%
Specificity	100%	± 0.0%

Tabla A. Métricas de rendimiento del algoritmo de clasificación

	True range 1	True range 2	Class Precision
Predicted range 1	785	248	75.99%
Predicted range 2	0	13	100.00%
Class Recall	100.00%	4.98%	

Tabla B. Matriz de confusión del algoritmo de clasificación (rango 1: fallo por guiado; rango 2: fallo interno).

Medida del rendimiento del algoritmo

🎯 Estrategia de uso del modelo actual

- Usar **como apoyo** para detectar fallos por guiado (FM1), donde tiene alta precisión.
- **No confiar en él para fallos internos (FM2)**: bajo recall (5 %) = alto riesgo de no detección. Requiere supervisión por parte del personal de mantenimiento u otros modelos complementarios (por ejemplo, basados en reglas o estadísticos) para detectar estos fallos.
- Complementar con **otras alertas o inspecciones** hasta mejorar el modelo.
- Comienzo como **modelo en pruebas controladas**, no aún para producción automática.

32

🔧 Resumen del Proceso seguido para la Detección de Anomalías

1.Adquisición de datos brutos

1. Temperatura de rodamientos.
2. Velocidad del tren.
3. Kilometraje.

2.Preprocesado y transformación

1. Aplicación del umbral de temperatura.
2. Detección de “positivos” (AE > 10 °C).
3. Agrupación en ciclos según distancia máxima entre positivos.

3.Reducción de datos

1. Cálculo de variables por ciclo (km dentro/fuera, acumulados...).
2. Disminución del volumen de datos según el umbral (1 km, 5 km, etc.).

4.Generación de base de datos final

1. Estructura por ciclo.
2. Etiquetado con diagnóstico final (deteriorado / no deteriorado).

5.Entrenamiento del modelo

1. Diseño, validación y prueba del algoritmo de clasificación.
2. Selección del mejor modelo (Deep Learning en este caso).

6.Clasificación de rodamientos

1. Separación entre fallos internos y sobretemperaturas externas.
2. Apoyo a la toma de decisiones de mantenimiento.

33

🛠 Pronóstico de Fallos: Análisis de datos

- Según **ISO 13381-1:2004**, la **prognosis de fallos** es la **estimación del tiempo hasta el fallo (ETTF)** y del riesgo de aparición de uno o más modos de fallo.
- En la mayoría de la literatura especializada se utiliza el término **RUL (Remaining Useful Life – Vida Útil Remanente)** en lugar de ETTF.
- El concepto de **RUL** es ampliamente utilizado en investigación operativa, fiabilidad y estadística, con aplicaciones relevantes en ciencia de materiales, bioestadística y econometría.
- La definición de vida útil remanente depende del **contexto y de las condiciones operativas**.
- Existen **tres enfoques principales** para estimar el RUL:
  - **Estadísticos**
  - **Basados en inteligencia artificial (IA)**
  - **Basados en modelos físicos**

#CMCMéxico2025

34

🛠 Pronóstico de Fallos: Análisis de datos

- En este caso de estudio se siguió un **enfoque estadístico para estimar la RUL** (vida útil remanente) de un rodamiento del tren, una vez que aparece un "positivo" o anomalía.
- Un **positivo** se define como un **error absoluto (AE) ≥ 10 °C** entre la temperatura real del rodamiento y la predicha por la red neuronal (ANN), **cuando el tren circula a más de 90 km/h y durante más de 1 minuto**.
- La **RUL se modela como una variable aleatoria**, estimada desde la aparición del primer positivo, y permite prever el reemplazo del rodamiento por **sobretemperatura o ruido**.
- El reemplazo se realiza tras:
  - La activación de la **alarma de seguridad del sistema de control del tren (TCMS)**, o
  - Una **inspección programada**, probablemente durante el mantenimiento semanal.
- La **alarma se dispara** cuando la **diferencia de temperatura entre los cuatro rodamientos del mismo eje supera los 25 °C** durante más de 1 minuto.
- 🎯 **Objetivo de la empresa:** anticipar el momento óptimo de **reemplazo del rodamiento tras el primer positivo**, sin necesidad de inspección previa, utilizando estimaciones estadísticas.

#CMCMéxico2025

35

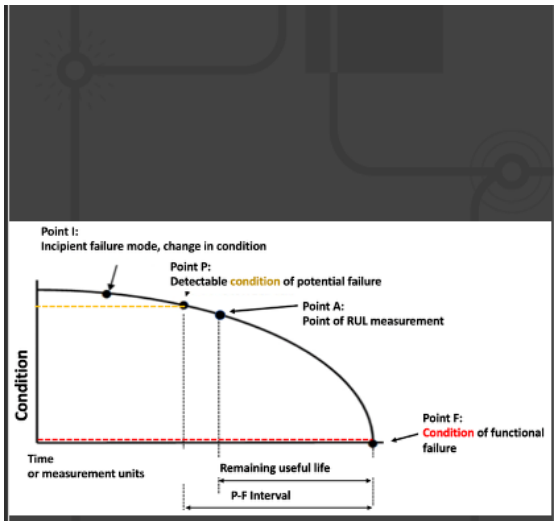


Figura. Curva P-F e intervalo de tiempo P-F. Tiempo estimado hasta el fallo (RUL).

Criterio para la RUL

- Para calcular la **RUL** en el punto A, es necesario modelar la variable aleatoria **"intervalo PF"**, es decir, el tiempo (o km) que transcurre entre:
  - El primer **positivo** (punto P, definido por AE ≥ 10 °C y velocidad TS ≥ 90 km/h), y
  - El posible **reemplazo** del rodamiento por **sobretemperatura o ruido**.
- El **punto F** se considera generalmente cuando se activa la **alerta de seguridad del sistema TCMS**. No implica una **falla funcional**, pero sí una condición de **bajo nivel de seguridad**.

Para este caso de estudio se definió:

$$RUL = Intervalo\ AF = Intervalo\ PF - Intervalo\ PA$$

- La **estimación de la RUL** se obtiene modelando la **distribución del intervalo PF**, mediante técnicas estadísticas como el **análisis de Weibull**.

36

Resumen del Proceso seguido para el Pronóstico de la RUL

1.Definición del positivo (inicio del pronóstico):

- Se detecta un positivo cuando:
  AE ≥ 10 °C y velocidad ≥ 90 km/h durante más de 1 minuto.

2.Identificación del punto de fallo (F):

- Se considera cuando se activa la **alarma del sistema TCMS** (por diferencia de temperatura entre rodamientos ≥ 25 °C durante más de 1 minuto), o cuando el rodamiento es reemplazado tras una **inspección**.

3.Modelado del intervalo PF:

- Se mide el tiempo o los km entre el **primer positivo (P)** y el **fallo (F)**.
- Este intervalo se modela como una **variable aleatoria**.

4.Cálculo de la RUL en un punto intermedio (A):

- Se calcula como:

$$RUL = Intervalo\ AF = Intervalo\ PF - Intervalo\ PA$$

- Es decir, lo que queda desde A hasta F.

Técnica empleada:

- Se utiliza **análisis estadístico**, concretamente el **modelo de Weibull**, para estimar la distribución del intervalo PF y predecir la RUL.

Objetivo:

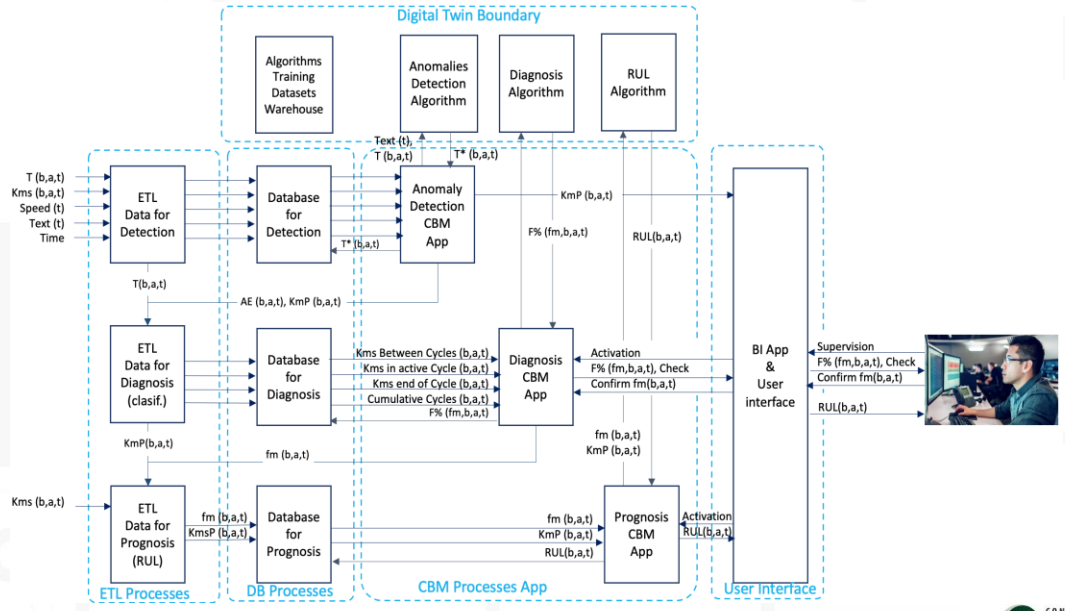
- **Anticipar el reemplazo del rodamiento** tras el primer positivo, **sin necesidad de inspección previa**, solo con estimaciones estadísticas.

37

ARQUITECTURA DEL GD
-
APLICANDO EL MARCO de GDM al GD-PdM de Talgo

#CMCMéxico2025

38



#CMCMéxico2025

39

🧩 La interacción de los Técnicos con el GD 🔄

Concepto	Tipos
Evento: Tiempo en el que debe reanalizarse el nivel de riesgo de los modos de fallo afectados.	• **Evento de monitorización:** Eventos provocados por la App CBM (y sus algoritmos de gemelo digital). Pueden ser de detección, diagnóstico o pronóstico. • **Eventos de mantenimiento preventivo:** Mantenimiento programado o no. Pueden ser inspecciones u otras actividades de mantenimiento preventivo.
Estado: Nivel cualitativo de riesgo en un momento determinado. Cada evento puede causar un cambio en ese nivel.	• **Avería:** Estado tras la ocurrencia del fallo. Se requiere reparación o sustitución inmediata. • **Riesgo alto:** Estado próximo al fallo. Se planifican acciones a corto plazo para reducir el riesgo. • **Riesgo medio:** Anomalía detectada, pero se puede operar con cierta seguridad. Se planifican acciones a medio plazo. • **Riesgo bajo**: Estado operativo normal del ítem.
Modos de fallo gestionados por el CBM.	• Modo de fallo **primario** (MFP) • Modo de fallo s**ecundario** (MFS): Iniciado por un MFP.

Tabla . Conceptos clave en la interacción del Gemelo Digital (DT) con las tecnologías de mantenimiento. Adaptado de (Martínez-Galán Fernández, Guillén López, Márquez, Gómez Fernández y Marcos, 2022)

#CMCMéxico2025

40

La interacción con el GD en la práctica

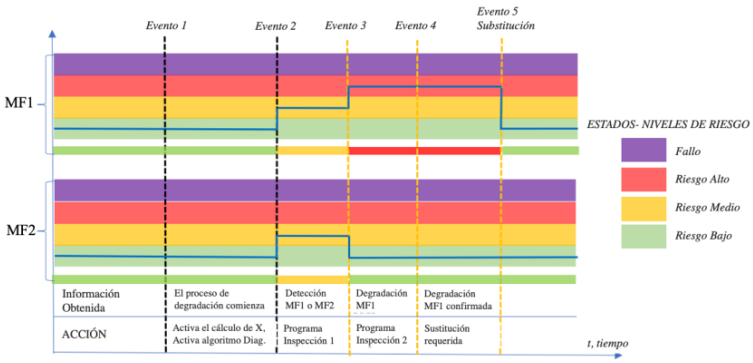


Figura. Representación gráfica de la interacción de la APP CBM con Gemelo Digital (DT) y los técnicos de mantenimiento. Adaptado de (Martínez-Galán et al., 2022).

- Evolución temporal de los modos de fallo **FM1 (degradación interna)** y **FM2 (fallo externo de guiado)**.
- Los **eventos de monitorización** y de **mantenimiento preventivo** modifican el nivel de riesgo del MF.
- **Niveles de riesgo** en colores:

Bajo | Medio | Alto | Fallo.

- Conforme se obtiene información (p.ej. detección de una anomalía o confirmación de fallo), se ejecutan **acciones**: cálculos, inspecciones y sustituciones.
- Una correcta planificación puede evitar o retrasar la aparición del fallo definitivo.

41

Conclusiones

✅ Se ha demostrado cómo utilizar el marco DMM para definir funcionalmente un gemelo digital (DT) que dé soporte a una aplicación CBM basada en analítica predictiva.

📊 A tal efecto se presentaron los modelos de datos específicos para cada algoritmo predictivo implementado en el caso de PdM de los rodamientos de trenes.

👤 Se analizó la interacción entre el usuario final y la App, destacando cómo esta se produce y cómo influye en el riesgo. Esto se facilita con el diagrama DIO-GD

📊 A tal efecto se diseño el diagrama de interacción (eventos-acciones-estados) para el GD de los rodamientos de los trenes, una vez decidida su puesta en operación.

✅ Son casos reales de aplicación con soluciones implantadas

📑 Se propone documentar herramientas de este tipo usando un esquema como el DMM y controlarlas conforme a los requisitos definidos.

🔧 Este enfoque puede aplicarse no solo a soluciones operativas como el CBM, sino también a herramientas estratégicas como análisis de criticidad o salud de activos a largo plazo.

#CMCMéxico2025

42

CONGRESO DE MANTENIMIENTO & CONFIABILIDAD MÉXICO
18ª EDICIÓN
¡Gracias!
Adolfo Crespo Márquez
adolfo@us.es
ingeman
ASOCIACIÓN PARA EL DESARROLLO DE LA INGENIERÍA DE MANTENIMIENTO

43