

## Introducción

La empresa XY se dedica a la comercialización de vehículos y cuenta con un equipo multidisciplinario que incluye ingenieros de diseño, pruebas y manufactura. En el proceso de desarrollo del producto, cada etapa presenta desafíos únicos que deben abordarse de manera ágil para garantizar que el producto llegue al mercado antes que los competidores. El equipo de pruebas tiene la responsabilidad de medir vehículos nuevos y existentes para proporcionar datos cruciales al equipo de diseño. Estos datos son fundamentales para tomar decisiones y realizar cálculos que aseguren la integridad estructural del vehículo. La entrega oportuna y precisa de estos datos es esencial, ya que influye directamente en la arquitectura del vehículo desde las primeras etapas del proyecto.

Ignorar o malinterpretar los datos de pruebas debido a la falta de tiempo para obtener insights podría tener consecuencias adversas. Por un lado, sobrediseñar el vehículo resultaría en costos innecesarios y un aumento de peso, lo que reduciría la eficiencia de combustible. Por otro lado, diseñar por debajo de las especificaciones podría provocar fallas estructurales en el vehículo, comprometiendo la seguridad del usuario y dañando la reputación de la marca. Es esencial evitar estos escenarios asegurando la precisión y prontitud en el análisis de los datos de pruebas.

El equipo de pruebas se enfoca principalmente en colocar sensores en los vehículos, replicar pruebas que imiten el uso del usuario final y realizar cálculos para determinar si el vehículo probado cumple con las especificaciones de estándares o leyes aplicables. En este proyecto, se utilizan datos reales capturados en campo para una prueba de especial interés. A pesar de que se omiten los detalles del contexto de la prueba y del vehículo por razones de privacidad, el desconocimiento de estos detalles no afecta la realización de un análisis estadístico y descriptivo de la información.

Los datos recopilados para este estudio provienen de una prueba realizada en un solo vehículo, dividida en 41 subconjuntos de operaciones. Cada operación implica la recopilación de información de 70 sensores. Estos datos son almacenados como series de tiempo con duraciones variables, oscilando entre 20 segundos y un máximo de 20 minutos. Cada uno de los 70 sensores produce 200 mediciones por segundo (200Hz). Además, los sensores están distribuidos en diversas partes del vehículo, lo que significa que, a pesar de tratarse de la misma prueba, los sensores pueden mostrar magnitudes, valles, picos, máximos y mínimos diferentes, así como cambios de polaridad (positivos o negativos). Esta variabilidad en los datos es esencial para comprender el comportamiento del vehículo durante diferentes operaciones y situaciones, lo que proporciona una perspectiva detallada y completa para el análisis del estudio.

El objetivo principal es salvaguardar la integridad estructural del vehículo. Por ende, el presente estudio se enfoca exclusivamente en el post procesamiento de información proveniente de una prueba específica. Esta prueba se considera la más importante debido a su duración prolongada (20 minutos) comparada con los 2 a 3 minutos de otras pruebas y a que el equipo de calidad determina que el cliente pasa el 90% del tiempo en esta operación. Además, es en esta fase donde los sensores detectan el mayor daño estructural. Para este análisis, se seleccionaron los sensores 14 y 53, previamente identificados como los más críticos dado que son los que tienen registros de deformación más altos. Para poder obtener la información necesaria para tomar decisión de diseño los objetivos son los siguientes:

- Transcribir datos del programa especializado de pruebas Y a un formato CSV limpio y sin información no necesaria para iniciar el postprocesamiento.
- Para la prueba más crítica encontrar el tiempo en que se minimice la suma de diferencias de un punto respecto a su media para todos los sensores.

- Identificar el ciclo de la serie de tiempo con el máximo daño estructural por fatiga

## Metodología

La limpieza de datos para comenzar con el estudio es sencilla dado que solo involucra transformar el archivo del programa de pruebas Y (.rst) a algo post procesable por Python (csv.). En primera instancia se eliminan toda la metadata de la prueba (columnas y renglones no necesarios) para solo quedarnos con las columnas que contienen la información de series de tiempo. Después se eliminan los registros que en alguna serie de tiempo les haga falta su valor, con la intención de que todas las series de tiempo tengan la misma cantidad de observaciones. Por último los valores que enumeran las observaciones de la serie de tiempo se dividieron entre 200 para transformar de observación a segundos dado que las lecturas de los sensores están medidas en Hertz con una frecuencia de 200. una vez realizado estos pasos el archivo está listo para los siguientes pasos de la metodología.

Las lecturas de los sensores de interés (XX y YY) miden deformación estructural; ambos sensores están por encima de la deformación permitida para cumplir con la vida estructural requerida por estándares internos de la compañía. Los ingenieros de pruebas realizan una estimación de vida, con teoría de falla por fatiga que se encuentra fuera de alcance de este poster, para después sugerir un porcentaje de reducción de deformación para ambos sensores. Dado la deformación no es un valor constante si no fluctuante, el ingeniero de diseño tiene que encontrar el punto de la serie de tiempo que resulte ser el más representativo de toda la serie para ahí tomar el valor base sobre el cual se intentará rediseñar para obtener la reducción previamente sugerida por el equipo de pruebas. Este valor es la media de la serie de tiempo, pero debido a que cada serie de tiempo tiene valores, picos y valles diferentes lo que se necesite es un valor representativo global, dicho valor se obtiene encontrando aquel valor de x (segundo) donde se minimice la diferencia de la media respecto a ese valor. Una vez seleccionado el segundo mas representativo global se puede utilizar las deformación de ese punto para rediseñar. Para encontrar dicho punto el siguiente algoritmo se utilizó.

1. Calcular el RMS (RMS\_Serie) de n series de tiempo. Cada serie de tiempo tendrá un valor de RMS.

$$RMS\_Serie = \sqrt{\frac{1}{2} * (x_1^2 + x_1^2 + ..... x_n^2)}$$

2. Restar el valor absoluto de cada observación (Obs) de la serie tiempo con el respectivo valor de RMS de su serie de tiempo del paso 1 para obtener DifObs. Si la serie de tiempo tiene m observaciones, se obtendrán m cantidad de diferencias para n series de tiempo. Se transforma al valor absoluto dado que los resultados del paso 1 solo son positivos y la series de tiempo pueden tomar valores positivos o negativos..

$$DifObs = abs(Obs) - RMS\_Serie$$

3. Sumar n cantidad de diferencias de DifObs que tengan mismo valor de x (segundos) para obtener por cada valor de m una diferencia global (DifGlobalObs). Si se tienen x cantidad de segundos se ontendran x cantidad de DifGlobalObs

$$DifGlobalObs = \sum_{1}^n DifObs$$

4. Seleccionar aquel valor de x (segundos) donde la DifGlobalObs sea la mínima.
- 5.

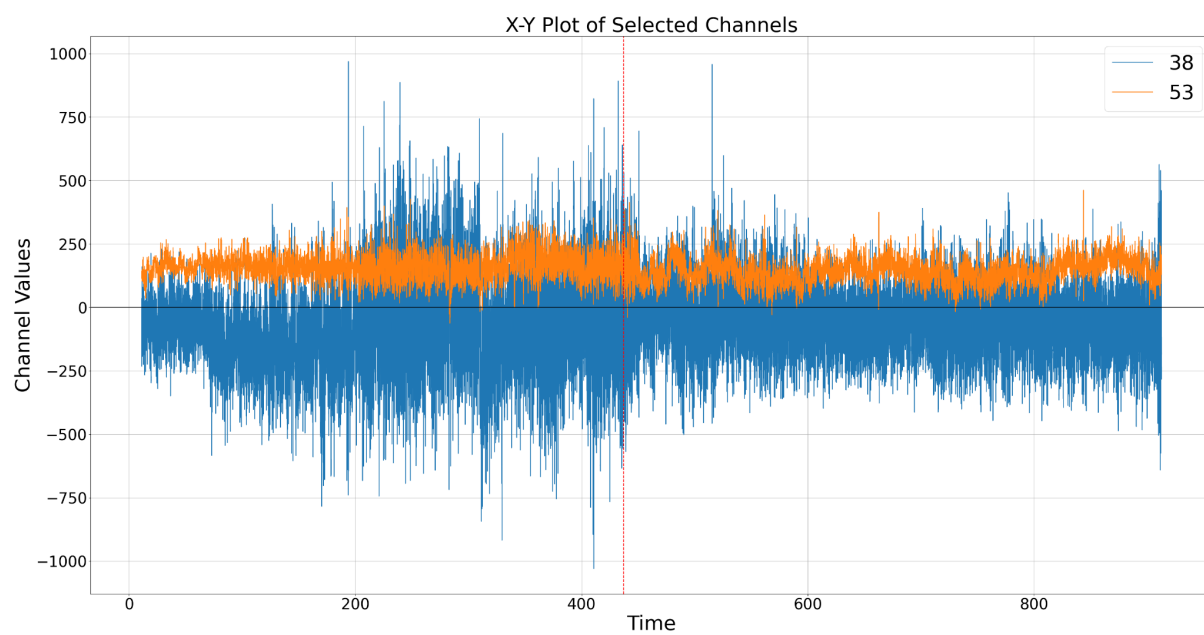
$$Segundo\ Seleccionado = Min(DifGlobalObs)$$

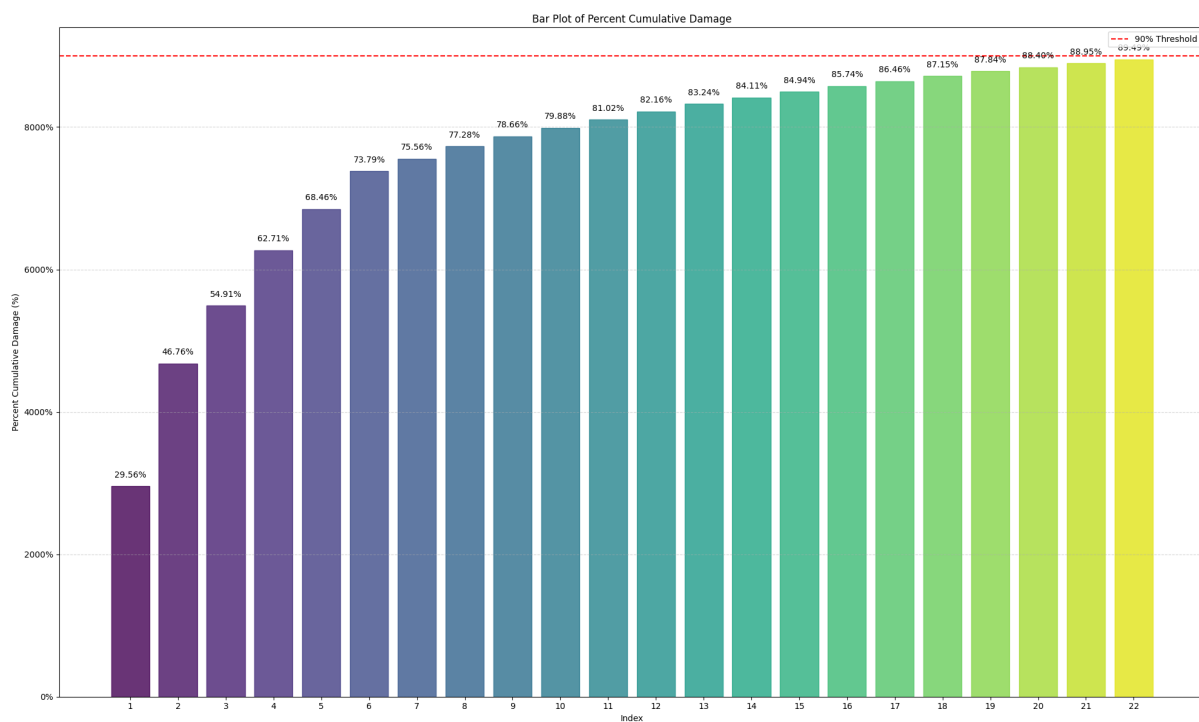
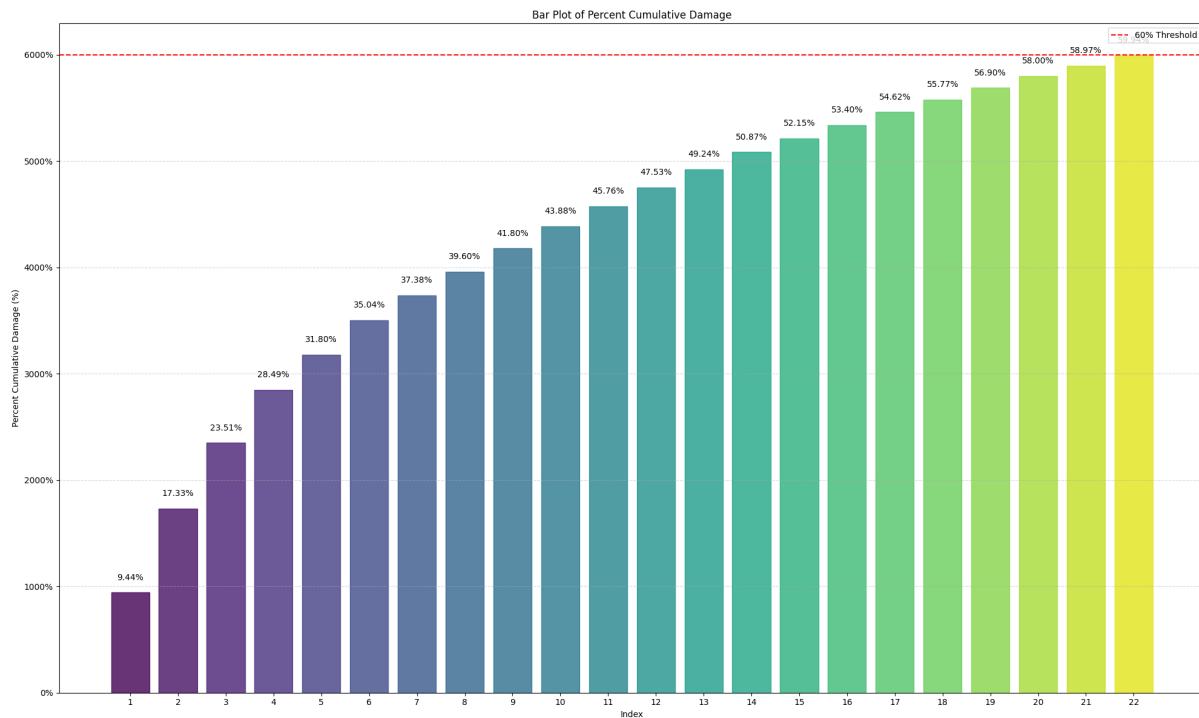
Para identificar el ciclo de la serie de tiempo con el máximo daño estructural por fatiga se utilizará un diagrama de pareto para mostrar cuántos ciclados son necesarios para cubrir el 90% del daño equivalente. La explicación de la teoría de daño equivalente se encuentra fuera del alcance de este poster. Lo esencial a entender con el diagrama de pareto es identificar cuáles y cuántos ciclos son los más representativos de la serie de tiempo. Entender cuáles son los ciclos con más daño estructural ayuda a centrarse en diseñar para ellos en lugar de arreglar ciclo por ciclo lo cual es prácticamente imposible. Para obtener los ciclos más relevante se siguen los siguientes pasos:

1. Filtrar de mayor a menor el daños estructural capturado por cada ciclo.
2. Del daño total calculado, añadir una columna que calcula el porcentaje del daño del ciclo que se está evaluando.
3. Añadir una columna del porcentaje acumulado.
4. Crear un pareto con la tabla anterior.

## Resultados

Sensor	Promedio	Valor 437.125 segundos
1	26.34	20.26
2	450.55	463.20
3	870.36	900.90
4	222.57	189.60
27	327.33	277.30
28	193.00	197.40
29	-91.33	-90.09
30	-65.33	-43.67
34	-760.14	-806.10
38	38.81	46.57
39	233.03	233.30
40	163.30	168.10
45	338.15	359.30
46	422.28	447.20
48	151.72	183.50
49	-366.84	-404.00
50	268.64	270.90
53	190.2	209.22
54	289.19	296.70
55	-522.90	-542.70
56	418.97	415.40
59	630.71	633.10
70	138.80	182.10





## Conclusiones

Con los dos algoritmos utilizados se logró el objetivo de identificar exactamente cuáles son los ciclos que más daño aportan a la estructura así como encontrar el segundo más representativo de la serie de tiempo. Teniendo estos dos datos el ingeniero de diseño ahora tiene la posibilidad de enfocarse solo en un solo tiempo del caso de carga más dañino en lugar de los miles de ciclos disponibles. Con este primer dato es posible comenzar a hacer iteraciones por medio de análisis de elemento finito para lograr la reducción propuesta por el equipo de pruebas.

También ahora se comprende cuáles son los ciclos más dañinos y su aportación porcentual así como el cúmulo porcentual hasta llegar al 90 o 60% según sea el interés del usuario. Esta información será clave para investigar, especialmente los primeros tres ciclos de daño que fue lo que lo causó, ¿Un impacto? ¿Una vuelta cerrada? ¿Un bache? Si bien el diagrama de Pareto no explica qué fue lo que sucedió es el inicio de la investigación.

Como siguientes pasos se desea obtener una metodología para eliminar eficientemente los datos que puedan ser considerados atípicos dado que para esta asignatura se realizaron limpiezas con zscore, zcore modificada y isolation forest pero se detectó que removía datos que no deberían ser considerados atípicos. Para ello se requiere un entendimiento más profundo de las aseveraciones de detección de atípicos en series de tiempo. La relevancia de datos atípicos para cálculo de vida estructural es muy importante puesto que entre mayor sea la diferencia en magnitud de deformación en un ciclo mayor daño se captura. Esta investigación continuará en los siguientes meses.