



Modelado de datos para análisis conjunto de variables síntomas en turbinas a vapor 100 MW

Data modeling for joint analysis of symptom variables in 100 MW steam turbines

Francisco Antonio de la Torre Silva^I, Evelio Palomino Marín^{I,*}, Armando Díaz Concepción^I, Alexander Alfonso Álvarez^{II}, Joel Guillen García^{III}, Alejandra Elena García Toll^I

I. Universidad Tecnológica de la Habana José Antonio Echeverría, Centro de Estudios en Ingeniería de Mantenimiento, CEIM. La Habana, Cuba.

II. Universidad de la Serena. La Serena, Chile.

III. Universidad Técnica de Manabí. Puerto Viejo. Ecuador.

* Autor de correspondencia: evelio.palomino@gmail.com

Este documento posee una [licencia Creative Commons Reconocimiento-No Comercial 4.0 internacional](#)



Recibido: 28 de enero de 2024

Aceptado: 4 de abril de 2024

Resumen

El artículo presentó el modelado de una muestra de datos para el análisis conjunto de variables síntomas en turbinas a vapor de 100 MW. El objetivo principal fue obtener una muestra de datos patrón que represente la variabilidad normal del comportamiento mecánico dinámico funcional de la turbina. Se detalla el pre-procesamiento y procesamiento de la base de datos, aspecto que no había sido documentado previamente en la literatura para este tipo de turbina. Los resultados mostraron que la metodología aplicada permite identificar y caracterizar las condiciones operativas normales

y anormales de la turbina, proporcionando una base sólida para futuros estudios de monitoreo de condición y mantenimiento predictivo en sistemas de energía a vapor. Este enfoque mejoró la precisión y fiabilidad del diagnóstico de fallos y el análisis de desempeño.

Palabras claves: modelado de datos, control estadístico multivariable de proceso (MSPC), procesamiento estadístico de datos, patrón estadístico de datos, condición de turbinas a vapor, minería de datos.

Abstract

The paper presents the modeling of a data sample for the joint analysis of symptom variables in 100 MW steam turbines. The main objective is to obtain a pattern data sample representing the normal variability of the turbine's dynamic mechanical behavior. The preprocessing and processing of the database are detailed, which had not been previously documented in the literature for this type of turbine. The results show that the applied methodology allows for the identification and characterization of normal and abnormal

operating conditions of the turbine, providing a solid foundation for future condition monitoring and predictive maintenance studies in steam energy systems. This approach improves the accuracy and reliability of fault diagnosis and performance analysis.

Key words: modelling of data, multivariate statistical control of process (MSPC), statistical processing of data, statistical model of data, condition of steam turbines, data mining.

Cómo citar este artículo, norma Vancouver:

de la Torre Silva FA, Palomino Marín E, Díaz Concepción A, Alfonso Álvarez A, Guillen García J, García Toll AE. Modelado de datos para análisis conjunto de variables síntomas en turbinas a vapor 100 MW. Ingeniería Mecánica. 2024;27(2):e696.

1. Introducción

En las centrales termoeléctricas, la evaluación del comportamiento mecánico dinámico funcional de turbinas a vapor de 100 MW presenta desafíos, particularmente en el análisis integrado de parámetros de proceso y tecnológicos. Este trabajo aborda la integración de parámetros síntomas de la condición de la turbina mediante el modelado MSPC de una muestra de datos, analizando conjuntamente variables síntomas y realizando el pre-procesamiento y procesamiento de la base de datos.

Los modelos aleatorios de regresión son entrenados para predecir la calidad de superficies en rodamientos, logrando una exactitud del 94 % con el estadístico T^2 de Hotelling's [1, 2].

Por su parte la literatura refiere el desarrollo de un nuevo KPI para plantas hidroeléctricas, superando a los gráficos de control multivariantes convencionales, demostrando que el KPI propuesto funciona mejor que las gráficas multivariantes convencionales de control de proceso, como el índice Hotelling T^2 . [3, 4].

También se han publicado estudios de mantenimiento predictivo para un sistema de operación continua, donde se presenta un modelo de degradación para sistemas de operación continua [5].

Otra investigación sobre el desbalance del alojamiento del cojinete utiliza series temporales para investigar desbalances en cojinetes y predecir su vida útil remanente [6].

Así mismo, se han encontrado publicaciones para realizar pruebas de normalidad: Anderson-Darling, Ryan-Joiner, Shapiro-Wilk y Kolmogórov-Smirnov en datos de producción. Tales pruebas demuestran si la población de la cual se obtuvo la muestra sigue la distribución normal, esto resulta clave para la aplicación del procedimiento estadístico adecuado para el análisis de los datos en un determinado estudio [7].

Otra investigación se enfoca en la detección de fallos en turbinas a vapor mediante digitalización y seguimiento de valores de proceso [8].

Interesante resulta el análisis de las principales características del monitoreo de proceso usando una carta de control multivariante a partir de una carta de control chi-cuadrado con parámetros variables se puede ver en [9].

Otra publicación de mucho interés refiere la propuesta de un método para predecir la vida útil remanente de sistemas con fallos severos [10].

En la revisión bibliográfica también se ha encontrado información sobre el hecho de que, al usar métodos de control estadísticos de proceso, las gráficas de control Shewhart reportan muestras no correlacionadas entre sí por lo que, se proponen modificaciones que consisten en modelar la señal original en modelos estocásticos adecuados [11].

De especial importancia resulta la propuesta de aplicación de redes neuronales y reglas de asociación para predecir consumo de energía eléctrica. Los algoritmos usados fueron el perceptrón multicapa a nivel de redes neuronales y PART como regla de asociación, demostrando que las redes neuronales y reglas de asociación son alternativas viables para predecir los niveles de consumo y comprender los patrones de consumo de energía [12].

Por último, es necesario señalar que la bibliografía refiere también el desarrollo de modelos integrados para Mantenimiento Basado en Condición, *Condition Based Maintenance*, CBM y Control Estadístico de Procesos, Statistical Process Control, SPC [13].

Toda esta revisión bibliográfica denota la ausencia del tratamiento de una muestra de datos para el análisis conjunto de variables síntomas de turbinas a vapor de 100 MW empleando un modelo MSCP, *Multivariable Statistical Control Process*, así como del preprocesamiento y procesamiento de la base de datos observada para este tipo de turbina.

En este artículo se establecieron condiciones y criterios para seleccionar la muestra de datos de trabajo y la definición de la muestra patrón, así como su validación. Se emplearon técnicas de minería de datos para el pretratamiento de la muestra poblacional obtenida de sistemas de monitoreo en línea.

El objetivo principal fue obtener una muestra de datos patrón que represente la variabilidad normal del comportamiento mecánico dinámico funcional de la turbina. Se crea un modelo grupal de parámetros síntomas de la turbina utilizando datos del período de operación analizado.

Como resultado principal, se obtuvo una muestra de datos patrón representativa de la variabilidad normal del comportamiento mecánico dinámico funcional de la turbina y se realizó el preprocesamiento y procesamiento de la base de datos observada.

Esta investigación contribuyó al campo del monitoreo de condición de turbinas a vapor, proporcionando un enfoque integrado para el análisis de múltiples variables síntomas. El modelo desarrollado y la metodología de procesamiento de datos pueden mejorar las estrategias de mantenimiento predictivo y optimizar la operación de turbinas en centrales termoeléctricas.

2. Métodos y Materiales

Los materiales empleados en esta investigación fueron tomados de las bases de datos generadas por el sistema de monitoreo *on-line* instalado en las tres turbinas a vapor de 100 MW de una Central Termoeléctrica.

El método empleado fue el de crear un modelo MSPC a partir de condiciones y criterios para el establecimiento de la muestra de datos de trabajo y de la muestra de datos patrón que incluyen el preprocesamiento y limpieza de la muestra poblacional obtenida de las bases de datos generadas por los sistemas de monitoreo *on-line*, instalados en estas turbinas.

2.1. Muestra y población

Se generó un modelo grupal de la turbina analizada, en este caso la número 2 por ser la de comportamiento más estable, formado por los valores de 22 parámetros identificados como síntomas de comportamiento mecánico dinámico funcional de la turbina. Los parámetros síntomas analizados fueron:

- Vibración relativa del rotor respecto a su alojamiento (representados por la dirección de medición X y Y
- Expansión diferencial axial relativa del rotor respecto al cilindro, lado regulador D. Axial A y lado generador D. Axial B
- Expansión axial absoluta del conjunto turbo grupo Exp.T.
- expansión relativa del rotor respecto al cilindro de alta presión Exp.R CAP y el cilindro de media presión (Exp.R CMP)
- Expansión relativa del rotor respecto al cilindro de baja presión, lado regulador Exp.R CBP 1 y lado generador Exp.R CBP 2
- Revoluciones a las que rota el rotor de la turbina. El modelo grupal de los parámetros síntomas de la turbina se muestra en la figura 1.

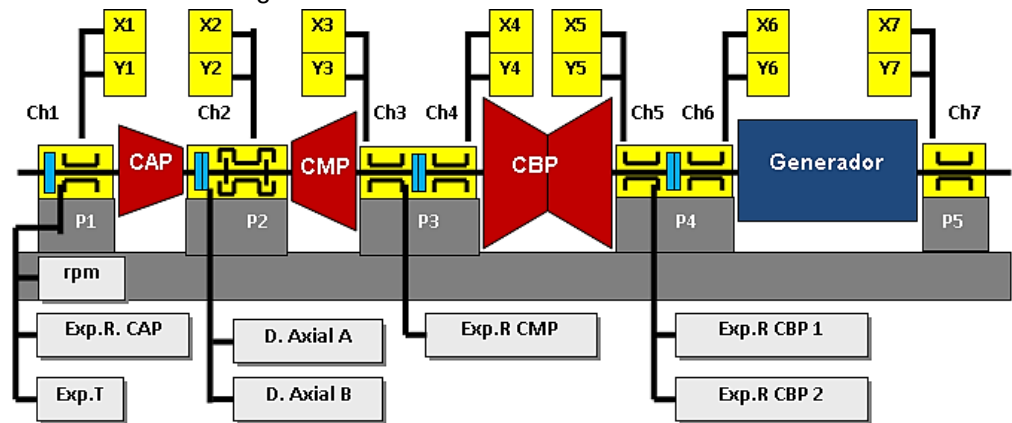


Fig. 1. Zonas de medición de los parámetros del modelo grupal turbina. Fuente: autores

2.1.1. Obtención de la muestra de datos de trabajo para el modelo MSCP

La forma genérica de actuación para obtener la muestra de datos de trabajo fue descargando la base de datos histórica en el período de operación analizado correspondiente en este caso al 23/06/20 hasta el 03/11/22 del sistema de monitoreado instalado en cada turbina. El acondicionamiento y limpieza de esta base de datos, permitió obtener una muestra de datos de trabajo semielaborada para el posterior procesamiento de tres turbinas de la misma denominación. En la muestra se tuvo en cuenta el cumplimiento de los requerimientos para la instalación, calibración, mantenimiento [2] y verificación de la instrumentación. Se plantea que el preprocesamiento y procesamiento de datos representa un 60 % del tiempo empleado en el desarrollo de proyectos basados en datos.

2.1.2. Acondicionamiento de la muestra de datos de trabajo para el modelo MSPC

Fue necesario acondicionar la muestra de forma organizada y estructurada convenientemente para las tres turbinas analizadas pertenecientes a los tres bloques de generación de esta central termoelectrica, en cada una de las hojas de cálculo correspondientes a 22 variables vibratorias y tecnológicas, cuyo esquema de configuración tecnológica se muestra en la figura 1. Las lecturas de cada una de estas 22 variables están realizadas con un intervalo de 10 minutos entre cada medición. Debido a una serie de factores de orden tecnológico y de la automática, la secuencia de captura automática de los valores de algunas de estas variables se ve afectado, lo cual conduce a fallos en la captura de estas lecturas. Esto conlleva en algunos casos, a que falten valores de las lecturas de algunas variables en la misma instancia de medición o que se pierdan tramos completos de medición para una o varias variables. Por lo tanto, para acondicionar la muestra de datos de trabajo se requiere cotejar el tiempo de lectura de las instancias de medición de cada variable analizada y posteriormente realizar un cotejo de fechas de captura, para estas instancias.

En la tabla 1 se presenta un resumen de los resultados del acondicionamiento de la muestra de datos de trabajo para la turbina 2. Estos requerimientos fueron valorados con los especialistas en diagnóstico, en mantenimiento, en automática, en operación y en inspección de turbinas, los cuales emitieron sus criterios sobre la factibilidad y uso de los datos de las variables generadas por el sistema de monitoreado instalado.

Tabla 1. Acondicionamiento de la muestra de datos de trabajo observada de la población. Fuente: autores

Parámetro	Turbina 2. 100 MW
Vibraciones relativas	3039 instancias de los atributos X1, Y1, X2, Y2, X3, Y3, X4, Y4, X5, Y5, X6, Y6, X7, Y7. (14 atributos en total).
Tecnológicos	2947 instancias de los atributos Exp Rel CAP, Exp Rel CMP, Exp Rel CBP 1, Exp Rel CBP 2, Axial A, Axial B, Exp Total. (7 atributos en total). <i>Nota:</i> fue desechado el atributo velocidad de rotación por su inconsistencia y poca cantidad de instancias de medición.
Proceso	n/d
Resumen cotejado	Cotejados: 2798 instancias de medición con 21 atributos en total

Nota: excluidos 92 instancias de los atributos vibraciones relativas, por no existir lecturas de medición. Fueron sesgadas 149 instancias de medición correspondientes a las interrupciones de funcionamiento y estadías de girador. Se evidencian problemas en grupos de días de instancias de medición. Excluido el atributo Vel Rot por su inconsistencia en las mediciones.

Todos los parámetros incluidos en la tabla 1 se pueden ver en la figura 1.

Del análisis de esta muestra de trabajo se pudo observar la poca regularidad en la cantidad de instancias por mes y la irregularidad en la cantidad de días con valores de medición por mes respectivamente, debido a interrupciones en la generación, fallos en la automática y de comunicación, entre otros factores. De haberse podido contar con la base de datos de otras turbinas, la muestra de datos de trabajo de una unidad experimental serían consideradas independientes de las otras, es decir, sin relación entre sí.

Basado en los resultados cotejados de la base de datos de cada turbina analizada, se tomó la decisión de trabajar con la base de datos de la turbina 2, por ser la que mejor consistencia presenta en cuanto a las instancias de medición de cada variable y la que mayor estabilidad en operación manifestó durante el período analizado, obteniéndose como resultado para esta turbina, una muestra de datos de trabajo de 2798 instancias de medición para 21 variables. La turbina 2 recibió un mantenimiento capital de sus conjuntos mecánicos principales y de gran parte de sus agregados, incluyendo los equipos auxiliares asociados a su funcionamiento, además se ejecutó la modernización de los sistemas automáticos de toda la unidad, este hecho favorece positivamente el tomar como caso de estudio a ésta turbina, ya que los trabajos de mantenimiento realizados presuponen la restauración del desempeño esperado de la turbina 2, además de estar asociado a un programa de mantenimiento predictivo.

2.1.3. Selección, limpieza y transformación de la muestra de datos de trabajo recopilada

El trabajo con la muestra de datos de trabajo observadas de la población no termina con su "limpieza", es necesario que estos datos tengan calidad (es decir, que estén íntegros, completos y consistentes) para poder ser trabajados con cualquier propósito (por ejemplo, extraer conocimiento o encontrar patrones). Si bien es cierto que una parte del proceso de selección, limpieza y transformación fue realizado durante el acondicionamiento de la muestra de datos de trabajo mostrados en la tabla 1, cuyo resumen se muestra en la tabla 2, generalmente, debido a las características propias de estas muestras de datos, es necesario profundizar la limpieza para obtener una materia prima que sea adecuada para el propósito concreto requerido.

Tabla 2. Resumen de las características del acondicionamiento de la muestra de datos de trabajo, para la turbina 2. Fuente: autores

Muestra cotejada de datos de trabajo	Fecha		Cantidad total		Atributos		
	Desde-hasta	Meses	Instancias	Atributos	Vibratorios	Tecnológicos	Proceso
Tur2_XYT_2798-21	23/06/20-03/11/22	30	2798	21	14	7	n/d
Tur2_P_45224-76	05/05/22-23/01/23	9	45224	76	n/d	n/d	76

Leyenda: XYT: atributos vibratorios y tecnológicos. P: atributos de proceso. n/d: atributos no disponibles

2.1.4. Análisis de valores perdidos o campos faltantes (*missing value*)

El análisis de los valores perdidos en la muestra de datos de trabajo denominada Tur2_XYT_2798-21, referenciado en la tabla 2, se realizó a través de un software especializado en procesamiento estadístico. Esta base de datos tuvo la característica de no tener campos faltantes debido al proceso de acondicionamiento recibido. El análisis de los valores perdidos en la muestra de datos de trabajo denominada Tur2_P_45224-76 referenciada en la tabla 2, se realizó con la ayuda de un software estadístico especializado. En total fueron detectados 140 campos (celdas) con valores faltantes (0,3 % del total). El análisis se realizó por grupos de atributos debido a la falta de memoria de cómputo. Las causas de los valores faltantes, se deben a interrupciones en la comunicación de la señal, sensor dañado o desconectado y a errores en el proceso de captación almacenamiento del valor por parte del sistema de monitoreado instalado.

2.1.5. Análisis de valores erróneos, duplicados, inconsistentes e irrelevantes

En la muestra Tur2_XYT_2798-21 no se detectaron valores duplicados sin embargo, en la muestra Tur2_P_45224-76 se detectaron tres variables con un gran número de lecturas duplicadas agrupadas en conjuntos diferentes de instancias. No se ejecutó acción alguna para su transformación.

Las implicaciones de cantidades grandes de valores duplicados perjudican la detección de la variabilidad de la variable, influyendo en los estadísticos media, desviación estándar. En cuanto a los valores erróneos, es importante destacar que un valor erróneo y uno anómalo no es lo mismo. Un valor puede considerarse erróneo si no se ajusta al formato establecido para el campo o que el sistema de monitoreado automáticamente ubique en el campo el código ERR o *** para significar que ha ocurrido fallo en la lectura del atributo o que no se ha efectuado la captura de la misma. Los campos con código ERR o *** fueron detectados mediante opciones de búsqueda en un tabulador electrónico. Para la muestra Tur2_XYT_2798-21, no rebasó el 0,13 % y para la muestra Tur2_P_45224-76, no rebasó el 0,028 %. En ambos casos, se tomó la decisión de sustituirlos por el

valor medio entre el valor precedente y postrero para los campos individuales y sustituirlos con el valor de la media para el caso de varios campos agrupados. En ningún caso los campos agrupados sobrepasaron las 12 instancias.

2.1.6. Análisis de valores nulos

Los campos con valores nulos para el caso de la muestra Tur2_XYT_2798-21, no rebasaron el 0,25 % y se tomó la decisión de sustituirlos por el valor medio entre el valor precedente y posterior para los campos individuales y sustituirlos con el valor de la media, para el caso de varios campos agrupados. Para el caso de la muestra Tur2_P_45224-76, no rebasó el 0,022 % y se tomó la decisión de sustituirlos por el valor que preserva la desviación estándar.

2.1.7. Análisis de valores anómalos, atípicos o extremos (outliers)

La detección de los valores anómalos no será realizada para las muestras de datos de trabajo Tur2_XYT_2798-21 y Tur2_P_45224-76, referenciadas en la tabla 2, ya que es conveniente conservar en éstas, aquellos valores que pueden causar estados de comportamiento anormal de la turbina o de incidencia en la variabilidad (dispersión) de la variable analizada respectivamente. Estos valores serán analizados y procesados durante la selección de las instancias de medición para la obtención de la muestra de datos patrón y la muestra de datos de referencia.

2.2. Definición de la muestra de datos patrón para el modelo MSPC. Fuente autores

- Muestra de datos patrón representativa de la condición normal de la turbina: grupo de instancias de medición para todas las variables, medidas en el mismo instante de tiempo, seleccionadas de una misma fuente, durante el período de tiempo de operación normal entre mantenimientos, cuya variabilidad de los valores de las variables síntomas del comportamiento mecánico dinámico funcional de la turbina, se corresponda con el rango de cambios aleatorios normales inherentes a la operación normal o estado funcional estable estacionario de la turbina, donde no este evidenciado algún problema de origen mecánico, tecnológico, dinámico o de operación que modifique este estado.
- Muestra de datos patrón entrenada: grupo de instancias de medición para todas las variables representativa de las posibles condiciones normales de variabilidad de los valores de las variables síntomas de la turbina.
- Muestra de datos patrón simulada: grupo de instancias de medición para todas las variables estadísticamente simuladas, cuya variabilidad (definida por la media y la desviación estándar) de los valores de las variables síntomas, simulen las características del comportamiento mecánico dinámico funcional, durante la operación normal o estado funcional estable estacionario de la turbina.
- Entrenamiento de la muestra de datos patrón: adición de instancias de medición (de todas las variables) a la muestra de datos patrón, que sean representativas de condición normal de la turbina.
- Actualización del modelo MSPC: durante el período de operación considerado, volver a calcular para una nueva muestra de datos patrón la media, covarianza (**LCS - límite de control superior**) para el estadístico T^2 de *Hotelling's*, así como la verificación de los requerimientos de datos.
- Actualización de la muestra de datos patrón: obtener una muestra de datos patrón, para otro grupo de instancias diferente al ya evaluado, dentro del período de tiempo objeto de estudio.
- Rango de cambios aleatorios normales inherentes a la operación normal de la turbina: rango definido por los valores máximos y mínimos de operación normal que puede alcanzar la variable cuantificada.

2.2.1. Selección de las instancias de medición para la muestra de datos patrón

Para la selección del grupo de instancias de medición que conformarán la muestra de datos patrón se tuvieron en cuenta las normas de operación de la turbina y los estándares de severidad de la vibración: Norma ISO 10816-2:1996. Mechanical vibration - Evaluation of machine vibration by measurements on non-rotating parts - Part 2, Norma ISO 7919-2: 1996. Mechanical vibration - Evaluation of machine vibration by measurements on rotating shafts - Part 2 así como las normas de monitoreo de la condición: Norma ISO 13380: 2002. Condition monitoring and diagnostics of machines General guidelines on using performance parameters.

Para la muestra de datos de trabajo Tur2_XYT_2798-21 referenciada en la tabla 2, se procedió a identificar visualmente para un primer análisis, los tiempos de operación con posibilidades de comportamiento normal. En la figura 2A y 2B se muestra la gráfica en el tiempo calendario del parámetro vibración relativa en la dirección de medición X y Y para todas las chumaceras, donde se encuentran señalados con paréntesis, tres tramos reducidos de tiempo calendario (R1: 02/09/20 al 23/01/21 con 222 instancias, R2: 24/01/21 al 25/04/21 con 300 instancias, R3: 24/05/21 al 28/09/21 con 282 instancias) que abarcan un total de 804 (28,73 % del total) instancias de medición. Igual análisis se realizó con los parámetros tecnológicos.

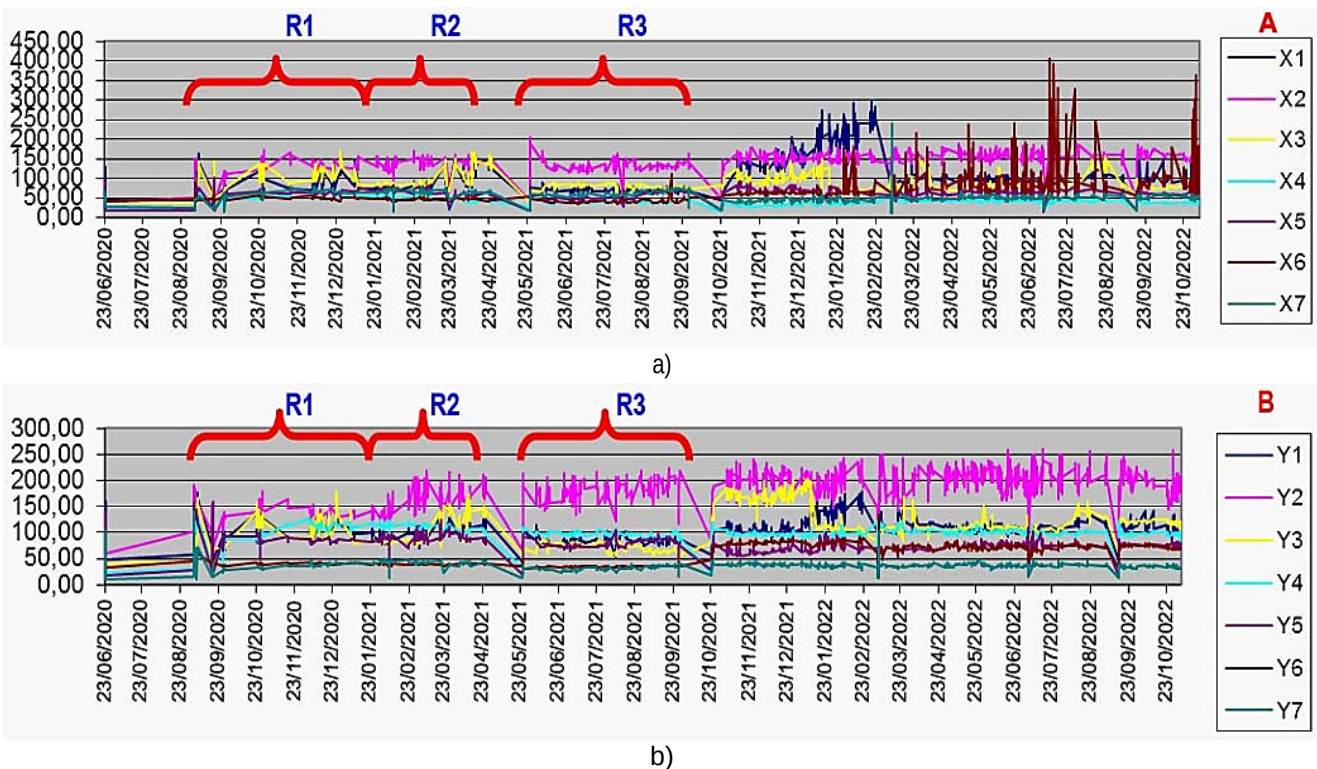


Fig. 2. Comportamiento en el tiempo del parámetro vibración relativa, para la muestra de datos de trabajo Tur2_XYT_ 2798-21, a) en la dirección de medición X, b) en la dirección de medición Y. Fuente: autores

Además, se tuvo en cuenta la historia de las incidencias ocurridas (basado en las interrupciones). Por lo tanto, se desecharon 49 instancias del tramo R1. En el tiempo de operación que abarca desde el 24/01/21 hasta el 28/09/21 (correspondiente a los tramos R2 y R3) se produjeron 41 interrupciones de funcionamiento de la unidad, siete de ellas debido a turbina, las cuales no requirieron de intervención que modificara la condición de la turbina, por lo que se consideró mantener las instancias seleccionadas para estos tramos.

Aunque no se tiene como referencia para cada instancia de medición de la muestra de datos de trabajo, el registro del dato de los valores de carga (punto de operación) a la que estuvo sometida la turbina, sí se tuvo en consideración (avalado por el conocimiento de los especialistas de operación, mantenimiento y diagnóstico) que en estos tres tramos, la operación de la turbina estuvo bajo un régimen de carga de 50, 60, 65, 70, 75, 80, 85, 90, 95 y 100 MW.

Se realizó el análisis de los valores anómalos a las 755 instancias (26,98 % del total) correspondientes a los tres tramos seleccionados, mediante el empleo de un software estadístico especializado, imponiendo como límites para la anomalía de cada atributo los límites de operación segura y protección tecnológica de la turbina 2. Se aplicó el criterio del sesgo de instancias para los casos en que los atributos tuvieran valores considerados anómalos por éstos límites, obteniéndose una muestra de datos patrón como resultado del análisis de los tramos R1, R2 y R3, con 442 instancias (15,79 % del total) representativa del comportamiento de la variabilidad normal natural de los valores de las variables, dentro del período de funcionamiento normal de la turbina entre mantenimientos, corroborado por la evidencia. La muestra de datos patrón abarca un período de 13 meses (durante los cuales no hubo registrado mantenimiento alguno, además de que no se evidenciara cambios que modificaran la condición de la turbina). En la tabla 3 se muestra el resumen de las muestras reducidas de datos que fueron utilizadas para acondicionar la muestra de datos patrón de comportamiento normal de la turbina 2 tomada como estudio, en base a la muestra de datos de trabajo Tur2_XYT_ 2798-21.

Tabla 3. Resumen de las características de la muestra de datos de trabajo patrón. Fuente: autores

Muestra de datos	Fecha		Cantidad total			Atributos
	Desde-hasta	Meses	Instancias	% del Total	Atributos	
Tur2_XYT_ 2798-21	23/06/20-03/11/22	30	2798	100	21	Vibratorios: 14 Tecnológicos: 7 Proceso: n/d
Patrón_XYT_ 442-21	25/09/20-27/09/21	13	442	15.79	21	
R1_XYT_ 173-21	25/09/20-23/01/21	5	173	6.18	21	
R2_XYT_ 210-21	24/01/21-25/04/21	4	210	7.50	21	
R3_XYT_ 59-21	04/06/21-27/09/21	4	59	2.10	21	

Legenda: XYT: atributos vibratorios y tecnológicos, n/d: atributos no disponibles para procesar

3. Resultados y Discusión

Normalidad en la distribución de los atributos de la muestra de datos patrón

En la tabla 4 se muestra un resumen de la prueba de normalidad realizada para cada atributo correspondiente a las instancias que conforman el tramo R1 de la muestra de datos patrón, mediante el empleo de las herramientas de un software especializado para tratamiento estadístico. En la prueba de Kolmogórov-Smirnov [7] se empleó el criterio de que si el estadístico D es significativo (valor alto resaltado en color rojo) y el valor $K-S$ (p -value) es significativo ($p \leq 0,01$ resaltado en color rojo), entonces la hipótesis de que la respectiva distribución sea normal, debería ser rechazada. En la prueba de Lilliefors se empleó el criterio de que si el estadístico Lilliefors (p -value) es significativo ($p \leq 0,01$ resaltado en color rojo), entonces la hipótesis de que la respectiva distribución sea normal, debería ser rechazada. Además se empleó el criterio de que, si el estadístico Shapiro-Wilk (W value) es significativo (valor pequeño de “W” y “p” resaltado en color rojo), entonces la hipótesis de que la respectiva distribución sea normal, debería ser rechazada. Puede observarse según los resultados, que no se evidencia normalidad en la distribución para las variables X2, Y2, Y5, Axial a, Axial b y Exp Total.

Tabla 4. Resumen de la prueba de normalidad de Kolmogórov-Smirnov, Lilliefors y Shapiro-Wilk para los atributos de la muestra patrón correspondiente al tramo R1. Fuente: autores

Variable	Tests of Normality (Spreadsheet3_21 by 173 Normal patrón tramo R1)					
	N	max D	K-S p	Lilliefors p	W	p
Var1_x1	173	0,051475	p > .20	p > .20	0,991281	0,377039
Var2_y1	173	0,054601	p > .20	p > .20	0,990749	0,326896
Var3_x2	173	0,106860	p < ,05	p < ,01	0,941834	0,000002
Var4_y2	173	0,132113	p < ,01	p < ,01	0,924406	0,000000
Var5_x3	173	0,037881	p > .20	p > .20	0,986149	0,085292
Var6_y3	173	0,048007	p > .20	p > .20	0,992132	0,468686
Var7_x4	173	0,050055	p > .20	p > .20	0,994590	0,781385
Var8_y4	173	0,048944	p > .20	p > .20	0,991571	0,406747
Var9_x5	173	0,046658	p > .20	p > .20	0,993526	0,643113
Var10_y5	173	0,092207	p < ,15	p < ,01	0,886514	0,000000
Var11_x6	173	0,054273	p > .20	p > .20	0,987580	0,131073
Var12_y6	173	0,045802	p > .20	p > .20	0,994910	0,820111
Var13_x7	173	0,040872	p > .20	p > .20	0,990693	0,321946
Var14_y7	173	0,040618	p > .20	p > .20	0,994535	0,774444
Var15_AxialA	173	0,091990	p < ,15	p < ,01	0,979091	0,010418
Var16_AxialB	173	0,087580	p < ,15	p < ,01	0,984059	0,045388
Var17_ExpRelCAP	173	0,055501	p > .20	p > .20	0,993523	0,642706
Var18_ExpRelCMP	173	0,055518	p > .20	p > .20	0,984396	0,050233
Var19_ExpRelCBP1	173	0,041473	p > .20	p > .20	0,986273	0,088528
Var20_ExpRelCBP2	173	0,050125	p > .20	p > .20	0,989457	0,227494
Var21_ExpTotal	173	0,169559	p < ,01	p < ,01	0,909749	0,000000

Para las 15 variables (X1, Y1, X3, Y3, X4, Y4, X5, X6, Y6, X7, y7, Exp Rel CAP, Exp Rel CMP, Exp Rel CBP1 y Exp Rel CBP2) en las que se evidencia normalidad (dado por los resultados de los estadísticos $K-S$ p -value, Lilliefors p -value y W value), se comprobó además el ajuste a una distribución normal mediante la prueba de bondad de ajuste chi-cuadrado, a partir de histogramas de frecuencia y gráficos $Normal$ p -plot. Se empleó el criterio de que, si p -value es significativo en el nivel 0,05, es decir $p < 0,026$, entonces la hipótesis de que la respectiva distribución sea normal, debería ser rechazada ya que la distribución se desvía significativamente de una distribución normal típica. En los histogramas de frecuencias, las siglas n.s. para los estadísticos $K-S$ p -value y Lilliefors p -value, representa no significativo. En la tabla 5 están reflejados los valores de la media y la desviación estándar correspondientes a la variabilidad natural del comportamiento funcional normal obtenida de las 15 variables y sus correspondientes 173 instancias que definieron la distribución normal (pertenecientes al tramo R1 de la muestra de datos patrón).

Tabla 5. Resumen de los estadísticos obtenidos para los atributos de la muestra patrón (tramo R1) que cumplieron la condición de normalidad en la distribución de los datos. Fuente: autores

Variable	Descriptive Statistics (Spreadsheet3_21 by 173 Normal patrón tramo R1)					
	Valid N	Mean	Minimum	Maximum	Variance	Std.Dev.
Var1_x1	173	85,9836	53,30000	120,1000	204,3148	14,29387
Var2_y1	173	109,7183	68,15000	153,7000	318,8010	17,85500
Var5_x3	173	106,1265	68,15000	145,1900	331,3471	18,20294
Var6_y3	173	107,7164	51,70000	158,3000	447,7299	21,15963
Var7_x4	173	51,7255	40,45000	64,5100	20,7367	4,55376
Var8_y4	173	100,1580	70,27000	126,3000	141,2215	11,88366
Var9_x5	173	64,5940	50,86000	79,9800	29,3292	5,41565
Var11_x6	173	51,6513	41,36000	59,1500	11,9736	3,46029
Var12_y6	173	39,8609	34,28000	44,6800	4,1447	2,03585
Var13_x7	173	64,8702	42,47000	78,3600	38,2329	6,18328
Var14_y7	173	35,5298	24,82000	46,4100	18,6227	4,31540
Var17_ExpRelCAP	173	0,1427	-0,44210	0,7250	0,0420	0,20503
Var18_ExpRelCMP	173	-0,1702	-1,31100	0,7783	0,2167	0,46555
Var19_ExpRelCBP1	173	-0,8637	-1,68800	-0,1040	0,1239	0,35200
Var20_ExpRelCBP2	173	2,8271	1,74700	4,1890	0,2349	0,48466

Siguiendo el mismo esquema de análisis e obtuvieron los valores de la media y la desviación estándar de los atributos del tramo R2 y R3, se obtuvo un resumen de la prueba de normalidad de Kolmogórov-Smirnov, Lilliefors y Shapiro-Wilk, así como los resultados de la prueba de bondad de ajuste chi-cuadrado, que evidenciaron normalidad. Finalmente quedan conformados los tramos patrón R1, R2 y R3 de la siguiente forma R1_XYT_173-15, R2_XYT_137-11 y R3_XYT_95-7.

Ajuste del modelo MSCP turbina

El ajuste del modelo turbina o su validación será realizado teniendo en cuenta el resultado obtenido en la verificación de la normalidad en la distribución de los atributos y en la verificación del volumen de datos requerido por la muestra de datos patrón. Ésta muestra de datos patrón es representativa de la población de datos colectados y representa la condición normal estándar de variabilidad del proceso que caracterizan el comportamiento mecánico dinámico funcional en turbinas a vapor de 100 MW.

Los gráficos de control son capaces de detectar anomalías en el comportamiento mecánico dinámico funcional. La evaluación del comportamiento mecánico dinámico funcional en turbinas a vapor de 100 MW puede ser realizado, a partir de un modelo MSCP [9], donde se emplea el estadístico T^2 de Hotelling's [1] como parámetro síntoma para la detección de la condición normal del funcionamiento de estas turbinas y el gráfico de control estadístico de proceso tipo *shewhart*, para la monitorización del estadístico T^2 de Hotelling's [3], o a partir de la aplicación del análisis de componentes principales [4]. La determinación del umbral (LCS) para el estadístico T^2 de Hotelling's, parte de considerar una distribución normal multivariada en los atributos de la muestra, por lo que se requiere de la comprobación de la normalidad en la distribución de los atributos de la muestra de datos patrón.

Finalmente, se puede afirmar que el resultado fundamental de esta investigación lo constituye la obtención de una muestra de datos patrón representativa de la variabilidad normal del comportamiento mecánico dinámico funcional de una turbina a vapor de 100 MW, presentando el preprocesamiento y procesamiento de la base de datos observada, lo cual no aparece documentado en la literatura consultada para este tipo de turbina.

Conclusiones

La metodología aplicada para el modelado de una muestra de datos con vista al análisis integrado de parámetros síntoma, proporciona una base sólida para la identificación y caracterización de las condiciones operativas de turbinas a vapor de 100 MW, mejorando así la precisión y confiabilidad del monitoreo de condición de este tipo de máquina. Este enfoque integrado no solo incrementa la eficacia en la detección temprana de fallos potenciales, sino que también facilita la toma de decisiones contribuyendo significativamente a la eficiencia y seguridad de las plantas termoeléctricas.

Referencias

- Ahmer M, Sandin F, Marklund P, Berglund K. Using Multivariate Quality Statistic for Maintenance Decision Support in a Bearing Ring Grinder. *Journals Machines. Section Machines Testing and Maintenance*. 2022;10(9):794-804 <https://doi.org/10.3390/machines10090794>
- Barriento VF, Achcar JA. Statistical analysis of equipment maintenance time in the food industry: a case study to identify sources of impact on performance. *Ingeniare. Revista chilena de ingeniería*. 2019;27(1):151-163 <http://dx.doi.org/10.4067/S0718-33052019000100151>
- Betti A, Crisostomi E, Paolinelli G, Piazzini A, Ruffini F. Condition monitoring and predictive maintenance methodologies for hydropower plants equipment. *Renewable Energy*.

- 2021;171(6):246-253
<https://doi.org/10.1016/j.renene.2021.02.102>
4. Braga J, Andrade AR. Multivariate statistical aggregation and dimensionality reduction techniques to improve monitoring and maintenance in railways: The wheel set component. *Reliability Engineering & System Safety*. 2021;216(12):107932-107942
<https://doi.org/10.1016/j.ress.2021.107932>
 5. Duc Hanh D, Phuc D, Tao QB, Si Hung NH. Degradation modelling and opportunistic maintenance for two-component systems with an intermittent operation component. *Computers & Industrial Engineering*. 2023;185(11):109698-109708
<https://doi.org/10.1016/j.cie.2023.109698>
 6. Falekas G; Karlis A; Antonino-Daviu JA. Industrial Steam Turbine Generator Predictive Maintenance Based on Measurement Availability. In: IEEE Energy Conversion Congress and Exposition (ECCE). Nashville, EU: IEEE; 2023
<https://doi.org/10.1109/ECCE53617.2023.10362627>
 7. Flores Tapia CE, Flores Cevallos KL. Pruebas para comprobar la normalidad de datos en procesos productivos: Anderson-Darling, Ryan-Joiner, Shapiro-Wilk y Kolmogórov-Smirnov. *Societas. Revista de Ciencias Sociales y Humanísticas. Universidad de Panamá*. 2021;23(2):83-106.
 8. Helin, M. Fault detection in steam turbines using data-driven methods. AEE - Master's Programme in Automation and Electrical Engineering. ELEC3025. 2020. Aalto University, Finlandia.
 9. Jalal T, Hiwa F, Hasan R. An integrated Markov chain model for the economic-statistical design of adaptive multivariate control charts and maintenance planning. *European Journal of Industrial Engineering*. 2022;17(1):1-28.
<https://doi.org/10.1504/EJIE.2023.10044732>
 10. Man J, Zhou, Q. Prediction of hard failures with stochastic degradation signals using Wiener process and proportional hazards model. *Computers & Industrial Engineering*. 2018;125(11):480-489. <https://doi.org/10.1016/j.cie.2018.09.015>
 11. Montgomery DC, Friedman DJ. Statistical process control in a computer-integrated manufacturing environment. Statistical process control in automated manufacturing. Florida, USA: CRC Press; 2020.
 12. Pincay Ponce JI, Angulo Murillo NG., Herrera Tapia JS, Delgado Muentes WR. Techniques of data mining as support for the management of a system of electric energy commercialization. *Mikarimin, Revista Científica Multidisciplinaria*. 2020;6(2):19-34.
 13. Rasay H, Fallahnezhad MS, ZareMehrdadi Y. Application of multivariate control charts for condition-based maintenance. *International Journal of Engineering*. 2018;31(4):597-604
<https://doi.org/10.5829/ije.2018.31.04a.11>

Conflicto de intereses

Los autores declaran que no existen conflictos de intereses

Contribución de los autores

Francisco Antonio de la Torre Silva. <http://orcid.org/0000-0002-1040-2367>

Participó en el diseño de la investigación, recolección de los datos, procesamiento de los datos recogidos para el estudio, análisis de los resultados e interpretación de los datos, investigación bibliográfica y en la revisión crítica de su contenido así como en la redacción y aprobación del informe final.

Evelio Palomino Marín. <http://orcid.org/0000-0001-7417-6189>

Trabajó en el diseño de la investigación, recolección de los datos, procesamiento de los datos recogidos para el estudio, análisis de los resultados e interpretación de los datos, revisión del estado del arte y en la revisión crítica de su contenido así como en la redacción y aprobación del informe final.

Armando Díaz Concepción. <http://orcid.org/0000-0001-9849-0826>

Participó en el diseño de la investigación, recolección de los datos, procesamiento de los datos recogidos para el estudio, análisis de los resultados e interpretación de los datos y en la revisión crítica de su contenido así como en la redacción y aprobación del informe final.

Alexander Alfonso Álvarez. <http://orcid.org/0000-0001-5129-7164>

Trabajó en la recolección de los datos, procesamiento de los datos recogidos para el estudio y en la revisión crítica de su contenido así como en la redacción y aprobación del informe final.

Joel Guillen García. <http://orcid.org/0000-0002-5714-2400>

Trabajó en la recolección de los datos, procesamiento de los datos recogidos para el estudio y en la revisión crítica de su contenido así como en la redacción y aprobación del informe final.

Alejandra Elena García Toll. <http://orcid.org/0000-0002-9208-5413>

Participó en la recolección de los datos, procesamiento de los datos recogidos para el estudio y en la revisión crítica de su contenido así como en la redacción y aprobación del informe final.