



TESIS DOCTORAL

CONTROL ESTADÍSTICO APLICADO EN MOTORES DIÉSEL MARINOS

David Boullosa Falces

Directores:

Miguel Ángel Gómez Solaetxe

Juan Luis Larrabe Barrena

Bilbao 2017

“Todos tenemos sueños. Pero para que se hagan realidad, se necesita una gran determinación, dedicación, autodisciplina y esfuerzo”

Jesse Owens.

Agradecimientos

En primer lugar quiero agradecer, al Departamento de Ciencias y Técnicas de la Navegación, Maquinas y Construcciones Navales de la UPV/EHU, por darme la oportunidad de desarrollar esta tesis doctoral en el programa de doctorado, Ingeniería Náutica, Marina y Radioelectrónica Naval.

A Miguel Ángel Gómez Solaetxe (UPV/EHU), director de la presente tesis, quien desde un primer momento entendió mi situación, y con quien pase tardes y tardes, discutiendo uno u otro enfoque de la tesis, sacrificando parte de su tiempo. Por su labor de revisión, crítica y conducción de esta tesis, por su apoyo incondicional, por su ayuda, su paciencia y sabios consejos. Gracias, Miguel Ángel.

A Juan Luis Larrabe Barrena (UPV/EHU), también director de la presente tesis doctoral, que cuyas labores de crítica han servido, sin duda alguna, para que este trabajo tenga una mayor consistencia. Y que con sus muestras de ánimo y superación me ha ayudado a seguir adelante.

A Belén Río Calonge, Manuel Alfredo Girón Portilla, Luis Manuel Vega Antolín y Carlos Ángel Pérez Labajos, por permitirme realizar parte de la tesis doctoral dentro del marco de una estancia de investigación en el Departamento de Ciencias y Técnicas de la Navegación y de la Construcción Naval de la universidad de Cantabria.

A Jaime Menendez, y por extensión a la Naviera Ibaizabal, y sus tripulaciones, por su inestimable colaboración y total disponibilidad, siempre que ha sido requerida, facilitando medios materiales y humanos, de manera altruista, con los que sin ellos, no hubiese sido posible realizar este trabajo de investigación.

A Fernando Marcos (M.A.N.), quien con las aportaciones de información técnica, ha hecho que la ardua búsqueda de algunos datos técnicos haya sido mucho más sencilla.

A Alberto Lopez Arraiza (UPV/EHU), quien me ha dado ánimos constantemente y accedió a realizar la última revisión de la tesis.

Y por supuesto el agradecimiento más profundo y sentido es para mi familia.

A mis padres, quien con mucho cariño, sacrificio y esfuerzo, me han educado y formado siguiendo unos valores, de los que hoy, solo puedo sentirme orgulloso. Lo que tengo y lo que soy se lo debo a ellos. Gracias!!

A mi hermana, con quien he pasado muchos momentos en la vida, y siempre ha estado a mi lado para ayudarme, mostrándome la mejor cara ante la adversidad de los problemas.

A mi tía M^a Carmen, por su apoyo constante, pendiente de mis avances, animándome siempre a continuar, en los momentos difíciles de este trabajo.

A mis suegros y mis cuñados, que en todo momento han estado para lo que he necesitado, dándome ánimo para seguir luchando.

Finalmente, si hay una persona que haya padecido los efectos de este trabajo, esa es mi mujer, Begoña. Quien desde antes de comenzar los estudios universitarios ha estado a mi lado, y cuando las cosas se han puesto cuesta arriba me ha animado y me ha dado su apoyo incondicional para seguir adelante, “tranquilo, ya verás cómo al final lo consigues”. Pues aquí estoy.

Gracias Bego y también, gracias a nuestra hija Paula por vuestra paciencia, comprensión y apoyo moral, por el tiempo que me habéis concedido, un tiempo robado a la historia familiar. Sin vuestro apoyo este trabajo nunca hubiese sido posible y, por eso, este trabajo es también vuestro. Os quiero.

A todas las personas que, aunque no aparecen en estas líneas, me han apoyado a lo largo de esta etapa. Gracias a todos ellos, este trabajo ha podido llegar a buen puerto.

Resumen

Frecuentemente, puede ser complejo, identificar cuando un proceso tiene una desviación desde sus condiciones de trabajo normal, cuando las variables están monitorizadas individualmente. La monitorización univariante, no tiene en cuenta la interacción entre las variables, dando una diagnosis deficiente del origen de la alarma.

Los sistemas de monitorización actuales, implementados en motores marinos, monitorizan individualmente una gran cantidad de variables del proceso, siendo complejo para el usuario visualizar todas las variables al mismo tiempo. Estos sistemas de diagnosis, generan una alarma cuando su nivel supera un valor previamente establecido por el fabricante, sin tener en cuenta la condición del motor para una determinada operación del buque, pudiendo llevar a error en la interpretación del origen de la alarma.

Los gráficos de control multivariantes son una gran aportación al control de procesos, siendo el gráfico T^2 de Hotelling, la opción más utilizada.

La carta de control T^2 de Hotelling, presenta la evolución del proceso en un solo gráfico y es capaz de detectar cambios de más de 1,5 desviaciones típicas con respecto a una determinada condición.

Sim embargo, estos gráficos, presentan una serie de dificultades, en la identificación de la variable que origina la desviación en el proceso, cuando el número de variables a monitorizar es elevado. Además, pierden sensibilidad en la detección de cambios pequeños y progresivos.

Ante esta necesidad de mejora, en la presente tesis doctoral, se desarrolla una metodología, denominada SSDM, “Method for detection Small and Sudden Deviations”, para la detección de cambios pequeños y progresivos, además de bruscos y elevados. Con capacidad de reducir el número de variables a monitorizar, sin pérdida de información del proceso, incluso cuando las variables están pobremente correlacionadas.

Gracias a la inestimable colaboración de la Naviera Ibaizabal, contamos con los datos de funcionamiento del motor propulsor MAN B&W 6S70ME-C8 instalado en un buque tanque, pudiendo aplicar la metodología en dos de los procesos de funcionamiento del motor: proceso de lubricación de cilindros y proceso de combustible.

A través de un análisis de regresión multivariable y análisis de correlaciones, se seleccionaron las variables representativas de los procesos, asegurando un ajuste entre las variables y sus variables predictivas, con coeficientes de determinación R_{ajust}^2 , superiores a 0,8.

Las variables representativas del proceso, se monitorizaron completamente y de forma multivariable, para unas condiciones determinadas del buque. La carta T^2 de Hotelling detecto eficazmente desviaciones de gran magnitud fuera de control estadístico con respecto a las condiciones normales de operación del motor e identifico la variable que había originado el cambio en el proceso.

Las variables predictivas de la variable originaria del cambio en el proceso, se monitorizaron a través de gráficos univariantes Cusum y se identifico el origen de cambios pequeños y progresivos por debajo del umbral de alarma preestablecido por el fabricante.

La detección prematura de las desviaciones en el proceso, ayudo a los responsables del funcionamiento del motor a tomar las acciones preventivas correctas; sin esta identificación, hubiesen tenido que esperar a que una de las alarmas y de forma individual estuviese por encima del umbral marcado por el fabricante.

La metodología SSDM, se podría implementar en cualquier tipo de motor propulsor, ya que es capaz de detectar cualquier variación en el proceso, cualquiera que sea su magnitud y puede ser customizado de forma sencilla y económica, a requerimiento de las condiciones del motor y operación del buque.

Índice de contenidos

1. INTRODUCCIÓN	3
1.1 Enmarque de la tesis	3
1.2 Organización de la memoria de la tesis	5
2. ESTADO DEL ARTE	9
2.1 Gráficos de control univariantes	9
2.1.1 Introducción	9
2.1.2 Gráficos Shewhart	10
2.1.3 Gráficos de control Cusum.....	16
2.1.4 Gráficos EWMA	23
2.2 Gráficos de control multivariantes.....	27
2.2.1 Introducción	27
2.2.2 Gráfico MCusum.....	28
2.2.3 Gráfico MEWMA	30
2.2.4 Gráfico T ² de Hotelling.....	33
2.3 Motivación de la tesis	43
3. OBJETIVOS	47
4. MATERIALES Y MÉTODOS	51
4.1 Materiales utilizados	51
4.2 Metodología	52
4.2.1 Adquisición de datos	53
4.2.2 Datos perdidos.....	54
4.2.3 Elección del tamaño de la base de datos preliminar.....	56
4.2.4 Selección de variables	57
4.2.5 Proceso de purgado de outliers.....	60
4.2.6 Normalidad multivariante	63
4.2.7 Control del proceso	64

4.2.8 Identificación de la variable indicativa de fallo	65
4.2.9 Control de las variables predictivas	67
4.3 Aplicación práctica	69
4.3.1 Proceso de Lubricación de cilindros	69
4.3.1.1 Descripción del proceso.....	69
4.3.1.2 Selección de variables	71
4.3.1.3 Adquisición de datos	75
4.3.1.4 Base de datos preliminar	77
4.3.1.5 Proceso de purgado de outliers	78
4.3.1.6 Normalidad multivariante.....	82
4.3.1.7 Control del proceso.....	84
4.3.2 Proceso de combustible	85
4.3.2.1 Descripción del proceso.....	85
4.3.2.2 Adquisición de datos	87
4.3.2.3 Datos perdidos	88
4.3.2.4 Base de datos preliminar	94
4.3.2.5 Selección de variables	95
4.3.2.6 Proceso de purgado de outliers	98
4.3.2.7 Normalidad multivariante.....	107
4.3.2.8 Control del proceso.....	109
5. RESULTADOS Y DISCUSIÓN	113
5.1 Proceso de lubricación de cilindros	113
5.1.1 Aplicación estadístico T^2 de Hotelling.....	113
5.1.2 Descomposición de los valores T^2	114
5.2 Proceso de combustible	116
5.2.1 Aplicación estadístico T^2 de Hotelling.....	116
5.2.2 Descomposición de los valores T^2	117

5.2.3 Suma acumulada, Cusum	117
5.3 Validación de la metodología propuesta.....	122
6. CONCLUSIONES GENERALES	135
6.1 Conclusiones.....	135
6.2 Líneas futuras.....	138
6.3 Publicaciones y Congresos	138
7. BIBLIOGRAFÍA	143

Índice de figuras

Figura 2.1: Distribución normal.	11
Figura 2.2: Gráfico Cusum V-Mask	19
Figura 4.1: Sistema de lubricación de cilindros. [121].....	70
Figura 4.2: T^2 Control Chart – Step 1.....	78
Figura 4.3: T^2 Control Chart – Step 2.....	79
Figura 4.4: T^2 Control Chart – Step 3.....	79
Figura 4.5: T^2 Control Chart – Step 4.....	80
Figura 4.6: T^2 Control Chart – Step 5.....	80
Figura 4.7: T^2 Control Chart – Step 6.....	81
Figura 4.8: T^2 Control Chart – Step 7.....	81
Figura 4.9: Análisis de normalidad multivariable, Base de datos preliminar.....	83
Figura 4.10: Análisis de normalidad multivariable, HDS	83
Figura 4.11 - Sistema de combustible. [121].....	86
Figura 4.12: T^2 Control Chart – Step 1.....	99
Figura 4.13: T^2 Control Chart – Step 2.....	99
Figura 4.14: T^2 Control Chart – Step 3.....	100
Figura 4.15: T^2 Control Chart – Step 4.....	100
Figura 4.16: T^2 Control Chart – Step 5.....	100
Figura 4.17: T^2 Control Chart – Step 6.....	101
Figura 4.18: T^2 Control Chart – Step 7.....	101
Figura 4.19: T^2 Control Chart – Step 8.....	101
Figura 4.20: T^2 Control Chart – Step 9.....	102
Figura 4.21: T^2 Control Chart – Step 10.....	102
Figura 4.22: T^2 Control Chart – Step 11.....	102
Figura 4.23: T^2 Control Chart – Step 12.....	103
Figura 4.24: T^2 Control Chart – Step 13.....	103

Figura 4.25: T^2 Control Chart – Step 14.....	103
Figura 4.26: T^2 Control Chart – Step 15.....	104
Figura 4.27: T^2 Control Chart – Step 16.....	104
Figura 4.28: T^2 Control Chart – Step 17.....	104
Figura 4.29: T^2 Control Chart – Step 18.....	105
Figura 4.30: T^2 Control Chart – Step 19.....	105
Figura 4.31: T^2 Control Chart – Step 20.....	105
Figura 4.32: T^2 Control Chart – Step 21.....	106
Figura 4.33: T^2 Control Chart – Step 22.....	106
Figura 4.34: T^2 Control Chart – Step 23.....	106
Figura 4.35: Análisis de normalidad multivariable, Base de datos preliminar.....	108
Figura 4.36: Análisis de normalidad multivariable, HDS	108
Figura 5.1: Carta de control.....	113
Figura 5.2: Carta de control.....	116
Figura 5.3: Scavenge air cooler air inlet temperature ($ARL_1=1.23$).	118
Figura 5.4: P (scav) ($ARL_1=4.29$)	119
Figura 5.5: Estimate Effective Power ($ARL_1=0.79$)	119
Figura 5.6: Pcom ($ARL_1=3.37$)	120
Figura 5.7: SFOC ($ARL_1=0.87$)	120
Figura 5.8: Carta de control.....	126
Figura 5.9: Descomposición T^2 . Observación 1	127
Figura 5.10: Descomposición T^2 . Observación 2	127
Figura 5.11: Descomposición T^2 . Observación 3	127
Figura 5.12: Descomposición T^2 . Observación 4.....	128
Figura 5.13: Descomposición T^2 . Observación 5	128
Figura 5.14: Descomposición T^2 . Observación 6	128
Figura 5.15: Descomposición T^2 . Observación 7	129

Figura 5.16: Descomposición T^2 . Observación 8	129
Figura 5.17: Descomposición T^2 . Observación 9	129
Figura 5.18: Descomposición T^2 . Observación 10	130
Figura 5.19: Descomposición T^2 . Observación 11	130
Figura 5.20: Descomposición T^2 . Observación 12	130
Figura 5.21: Descomposición T^2 . Observación 13	131

Índice de Tablas

Tabla 2.1: Límites de control, grafico \bar{X} -R	12
Tabla 2.2: Límites de control, grafico \bar{X} -S	13
Tabla 2.3: Representación del ARL para el Cusum tabular con $K=0,5$ [30].....	21
Tabla 2.4: Valores de ARL para Cusum tabular y Cusum -Shewhart [31].....	22
Tabla 4.1: Datos característicos del motor principal.	52
Tabla 4.2: Datos característicos del buque.	52
Tabla 4.3: Variable dependiente: Engine Load	72
Tabla 4.4: Variable dependiente: Fuel Index	73
Tabla 4.5: Variable dependiente: Fuel Plunge Stroke	73
Tabla 4.6: Variable dependiente: Hydraulic oil Pressure after pump P1	73
Tabla 4.7: Variable dependiente: Total Oil	73
Tabla 4.8: Variable dependiente: Scavenge Air receiver Pressure.....	73
Tabla 4.9: Variable dependiente: Scavenge Air receiver Temperature.....	73
Tabla 4.10: Variable dependiente: Exhaust gas temperatura	74
Tabla 4.11: Variable dependiente: Exhaust gas pressure	74
Tabla 4.12: Variable dependiente: Maximum Pressure, Pmax	74
Tabla 4.13: Representación de las variables Dependientes / Explicativas	74
Tabla 4.14 Base de datos adquirida.	76
Tabla 4.15: Media, desviación típica, valores máximos y mínimos.	77
Tabla 4.16: Pasos hasta conseguir el HDS.	82
Tabla 4.17: Nueva entrada de datos.....	84
Tabla 4.18: Modelos generados, variable dependiente: Scavenge air cooler air inlet temperatura.	89
Tabla 4.19: Modelos generados, variable dependiente: Exhaust gas temperature at turbine inlet.....	90
Tabla 4.20: Modelos generados, variable dependiente: Turbocharger Speed.....	90
Tabla 4.21: Modelos generados, variable dependiente: SFOC	91

Tabla 4.22: Modelo seleccionado, variable dependiente: Scavenge air cooler air inlet Temperature.....	91
Tabla 4.23: Modelo seleccionado, variable dependiente: Exhaust gas temperature at turbine inlet.....	91
Tabla 4.24: Modelo seleccionado, variable dependiente: Turbocharger speed.....	92
Tabla 4.25: Modelo seleccionado, variable dependiente: SFOC.....	92
Tabla 4.26: Base de los datos adquiridos.	92
Tabla 4.27: Medias y desviaciones típicas y valores máximos y mínimos de las variables del proceso de combustible. Base de datos adquiridos.	94
Tabla 4.28: Medias y desviaciones típicas - valores máximos y mínimos de las variables del proceso de combustible. Base de datos preliminar.	95
Tabla 4.29: Correlación entre variables. Base de datos preliminar.	96
Tabla 4.30: Correlación entre variables. Base de datos adquiridos.....	96
Tabla 4.31: Modelos generados. Variables seleccionadas / Variables predictoras	98
Tabla 4.32: Pasos hasta conseguir el HDS.	107
Tabla 4.33: Nueva entrada de datos.....	109
Tabla 5.1: Cronología de observaciones.	114
Tabla 5.2: Descomposición de los valores T^2	115
Tabla 5.3: Descomposición de los valores T^2	117
Tabla 5.4: Medias y desviaciones típicas de las variables predictivas.	118
Tabla 5.5: Cronología de observaciones fuera de rango.	121
Tabla 5.6. Parte del HDS.....	124
Tabla 5.7: Nueva entrada de datos.....	125

Acrónimos

SPC	Statistical Process Control
Cusum Chart	Cumulative Sum Chart
EWMA Chart	Exponential Weighted Moving Average Chart
Mcusum Chart	Multivariate Cumulative Sum Chart
MEWMA Chart	Multivariate Exponential Weighted Moving Average Chart
MYT	Mason, Young and Tracy
RWE	Western Electric Rules
ARL	Average Run Legth.
LCS	Límite de Control Superior
LC	Límite de Control
LCI	Límite de Control Inferior
PCA	Principal Component Analysis
PLA	Partial Least Squares
HDS	Historical Data Set
SSDM	Method for Detection of Small and Sudden Deviations
CoCos EDS	Computer Controlled Surveillance – Engine Diagnostics System
IAS	Integrated Automation System
PMI	Pressure Monitoring Indicator
HCU	Hydraulic Cylinder Unit
CCU	Cylinder Control Unit
ACC	Alpha Adaptative Cylinder oil control
DO	Diesel Oil
FO	Fuel Oil
ECU	Engine Control Unit
SFOC	Specific Fuel Oil Consumption

Capítulo 1

Introducción

1. INTRODUCCIÓN

La presente tesis doctoral titulada: “Control estadístico aplicado en motores diésel marinos”, ha sido desarrollada en la Escuela de Ingeniería de Bilbao, dentro del programa de doctorado “Ingeniería Náutica, Marina y Radioelectrónica Naval”, del Departamento de Ciencias y Técnicas de la Navegación, Maquinas y Construcciones Navales, de la Euskal Herriko Unibertsitatea (UPV/EHU). Igualmente se ha contado con la inestimable colaboración de la empresa Ibaizabal S.L. y del Departamento de Ciencias y Técnicas de la Navegación, Maquinas y Construcciones Navales, de la Universidad de Cantabria.

1.1 Enmarque de la tesis

El control estadístico de procesos, SPC, por sus siglas en inglés, ha llegado a ser muy importante en el control de la calidad en los procesos industriales. Su objetivo, es controlar el rendimiento de un proceso en el tiempo, con el fin de verificar que el proceso se mantiene en un estado de control estadístico [1].

El SPC, nació a finales de los años 20 en los laboratorios Bell Telephone Laboratories, de la mano del Dr. W. A. Shewhart, a quien se considera como el padre del control estadístico [2, 3], quien en su libro “Economic Control of Quality of Manufactured Products (1931)”, marco las pautas de lo que hoy se conoce como control estadístico de procesos.

Los gráficos de control propuestos por Shewhart, son los más utilizados debido a su sencillez, y a que presentan un comportamiento adecuado en la detección de cambios de gran magnitud, más de dos desviaciones típicas en la media del proceso. Sin embargo, estos gráficos son poco sensibles a cambios pequeños.

Posteriormente, como solución a esta limitación, aparecieron dos nuevos métodos para detectar pequeñas variaciones en el proceso; uno propuesto por E. S. Page [4], denominado gráfico de sumas acumuladas (CUSUM) y el segundo, denominado gráfico de medias móviles exponencialmente ponderadas (EWMA) propuesto por S. W. Roberts [5]. Estos gráficos acumulan la información obtenida de las muestras anteriores, y de esta manera pueden detectar de forma más rápida cambios pequeños en el proceso.

Más tarde, ante la necesidad de monitorizar diversas variables al mismo tiempo, se dio impulso al control estadístico multivariable; el primer estudio original sobre el control de la calidad multivariado fue introducido por Harold Hotelling [6], quien propuso un gráfico de control multivariado basado en la distancia de Mahalanobis [7], adecuado, al igual que los gráficos Shewhart univariantes para cambios de elevada magnitud.

Posteriormente, ante la necesidad de detección de pequeñas variaciones en procesos multivariantes, surgieron aportaciones tales como la realizada por Crosier [8], que presentó una extensión multivariada del gráfico de control de sumas acumuladas \bar{M} cusum, y la de Lowry [9] a través de la extensión del gráfico de control EWMA para el caso multivariante MEWMA.

En la presente tesis, queremos enfocar la investigación de los gráficos de control en el sector naval, más concretamente en el motor propulsor donde la monitorización y diagnóstico se basa en la generación de alarmas una vez se ha superado un umbral preestablecido por el fabricante, sin tener en cuenta la condición del buque, para una determinada operación del mismo.

En el último cuarto del siglo XX, la monitorización de cualquier proceso en general y de los motores propulsores marinos en particular, ha alcanzado tal dimensión, en la cual las casas fabricantes han desarrollado potentes herramientas de adquisición de datos, que ha derivado en un crecimiento de la computarización y monitorización que permite el registro en tiempo real con frecuencias de hasta milisegundos de una gran cantidad de variables.

Monitorizar variables redundantes innecesariamente, además de incrementar los costes de medida, puede derivar en acciones erróneas por el usuario debido a la dificultad en la interpretación de una gran cantidad de variables monitorizadas al mismo tiempo, disminuyendo la eficiencia del proceso.

Cabe pensar, que en el caso del motor propulsor de un buque, conocer cuando alguno de sus procesos está comenzando a desviarse de su condición normal, sin tener que esperar a que una de las variables este por encima de un umbral preestablecido, tiene una importancia especial, ya que una parada producto de una avería, puede tener consecuencias catastróficas, no solo por el coste de la pérdida de la actividad, sino,

porque si ocurre cerca de la costa o en el paso de un canal o durante una maniobra puede suponer la pérdida del buque e incluso pérdidas humanas.

En este contexto la presente tesis doctoral, plantea la idea de diseñar una novedosa metodología basada en el gráfico T^2 de Hotelling, que además de monitorizar de manera completa y multivariable, y detectar, desviaciones bruscas de un proceso concreto del motor principal del buque, tenga también la capacidad de detectar desviaciones progresivas y de baja magnitud, con reducción del número de variables monitorizadas sin pérdida de información relevante del proceso.

1.2 Organización de la memoria de la tesis

El **capítulo 1**, introduce el tema de investigación, su propósito e interés tanto a nivel científico como de posterior aplicación industrial.

En el **capítulo 2**, se ha llevado a cabo una revisión bibliográfica del tema a estudio, donde se abordan los gráficos de control mayormente utilizados en la industria, y se analizan las investigaciones más recientes sobre este tema. Además, se realiza un análisis de la motivación que nos ha llevado al estudio sobre esta línea de investigación, donde se destacan las principales dificultades y carencias detectadas que han surgido para establecer los objetivos a abordar en la presente tesis doctoral.

En el **capítulo 3**, se indican los objetivos que se establecen en la presente tesis.

En el **capítulo 4**, se describen los materiales y métodos empleados en el desarrollo de la tesis doctoral, se aplica la metodología a una aplicación práctica y se validan los métodos propuestos.

En el **capítulo 5**, se presentan y discuten los resultados obtenidos en el capítulo anterior.

Las conclusiones y aportaciones más significativas de este trabajo, se resumen en el **capítulo 6** junto con las recomendaciones para futuros trabajos.

Capítulo 2

Estado del arte

2. ESTADO DEL ARTE

A continuación, se va a llevar a cabo una revisión bibliográfica del estado del arte actual de los métodos de control univariantes y multivariantes. Primero se analizarán los gráficos univariantes más utilizados en procesos industriales: Shewhart, Sumas acumuladas (Cusum) y medias exponencialmente ponderadas (EWMA), por su sencillez y fácil aplicación. Posteriormente, se abordarán los siguientes gráficos multivariantes: MCusum y MEWMA, extensiones multivariantes de los gráficos Cusum y EWMA respectivamente y los gráficos basados en distancias, T^2 de Hotelling.

Finalmente, se analizarán las soluciones existentes a los inconvenientes que presentan los gráficos multivariantes T^2 de Hotelling, y se presentarán nuevas soluciones, base fundamental en la que se basa la presente tesis.

2.1 Gráficos de control univariantes

2.1.1 Introducción

En este apartado, analizaremos los gráficos de control univariantes de mayor aceptación en la industria por su sencillez y efectividad práctica, como son los gráficos de control Shewhart, los de sumas acumuladas (Cusum) y los de medias móviles exponencialmente ponderadas (EWMA).

Los gráficos Shewhart son muy utilizados debido a su sencillez; estos gráficos presentan muy buen comportamiento en la detección de cambios grandes, cambios por encima de dos desviaciones típicas en la media del proceso. Y por otro lado, los gráficos Cusum y EWMA, son efectivos en la detección de cambios de pequeña magnitud, al considerar en cada momento analizado toda la información pasada.

Primero, analizaremos la estructura de los gráficos de control Shewhart y posteriormente abordaremos los gráficos Cusum y EWMA.

2.1.2 Gráficos Shewhart

El Dr. Walter A. Shewhart durante su estancia en la compañía Western Electric, ante la necesidad de sustituir los rudimentarios sistemas de calidad de la época, consistentes en el chequeo de productos acabados y sustitución de los defectuosos, se propuso disminuir la variabilidad del proceso, en torno a unos cánones de calidad establecidos.

Estableció la idea de que los problemas ocurridos durante un proceso, pueden ser por causas comunes o aleatorias y causas especiales o asignables. Las primeras, presentes en un estado bajo control estadístico, son fuentes de variación puramente aleatorias, no identificables e imposibles de evitar, cuyo resultado sobre el producto se pueden considerar habitualmente pequeños; Las segundas, especiales o asignables, se manifiestan de forma esporádica y de una forma repentina tal que resulta complicada su previsión, siendo responsables de llevar el proceso a un estado de fuera de control estadístico de forma imprevista [10].

Los gráficos de control, se basan en la premisa de que la distribución de los datos recogidos se pueden ajustar a una distribución normal cuando el proceso en si está bajo control, por lo que este, estará bajo control estadístico y la variabilidad únicamente será debida a causas comunes de variación.

La distribución normal, es probablemente la distribución más importante de aplicación en estadística, y es aquella distribución de probabilidad que depende de μ y σ (media y desviación típica). Presenta una forma de campana de Gauss y es simétrica respecto a la media, tal y como se representa en la figura 2.1, es característico que el 68,2 % de la población de los datos se encuentran delimitados por $\mu \pm 1\sigma$, el 94,4% se encuentran entre $\mu \pm 2\sigma$ y el 99,73 % entre $\mu \pm 3\sigma$.

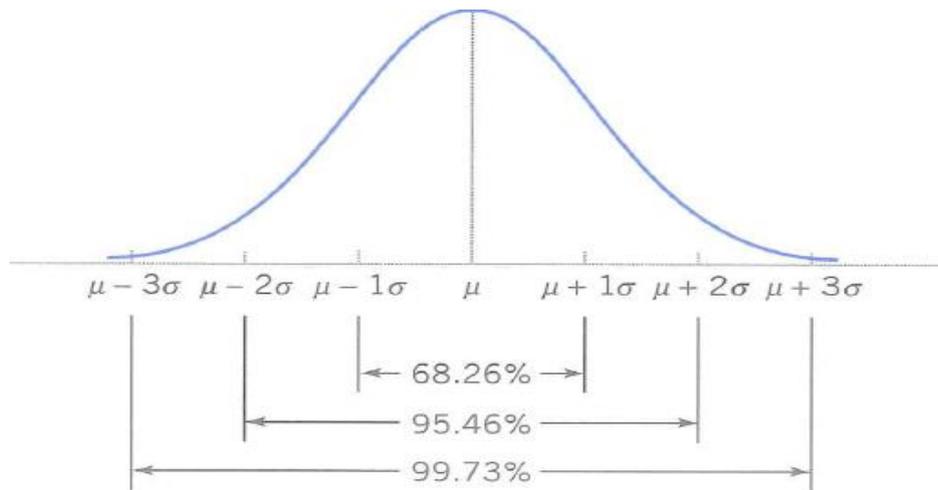


Figura 2.1: Distribución normal.

Por lo tanto, partiendo de la premisa que un proceso se puede explicar bajo una distribución $N(\mu, \sigma)$, se puede obtener una muestra n de datos de la población total, cuya media muestral puede encontrarse en el intervalo $\mu \pm 3\sigma/\sqrt{n}$, los cuales se denominan límites de control superior e inferior, respectivamente. Esto implica una probabilidad de 0,27% de estar fuera de los límites de control sin que se haya alterado la distribución estadística asignada, es decir, incurrir en un error tipo β . Estos límites los justificó indicando que, bajo el supuesto de normalidad del proceso, obtener observaciones situadas a más de tres desviaciones típicas con respecto a su media, era suficientemente infrecuente, por lo que se podría considerar que se había alterado el estado del proceso.

A partir de los límites de control, diseñó los gráficos para controlar la media y la variabilidad del proceso mediante dos gráficos de control, por un lado el gráfico de media – rango (\bar{X} -R) y por otro lado el gráfico de media – desviación típica (\bar{X} -S). La diferencia entre ambos gráficos reside en la medida de dispersión utilizada, si bien el más utilizado es el (\bar{X} -R), a continuación se estudiarán con más detalle cada uno de ellos.

- Gráfico \bar{X} -R

Este gráfico muestra posibles cambios en la media y como medida de dispersión utiliza el rango. Presenta la ventaja que se calcula muy fácilmente y para valores muestrales pequeños, menores a diez se comporta bien, mientras que para valores mayores, da una estimación sobrevalorada de la dispersión de la población.

Los límites de control para el gráfico \bar{X} -R se presentan en la tabla 2.1:

Tabla 2.1: Límites de control, gráfico \bar{X} -R.

<u>Gráfico \bar{X}:</u>	<u>Gráfico R:</u>
• $LCS = \bar{\bar{X}} + A_2 \cdot \bar{R}$	• $LCS = D_4 \cdot \bar{R}$
• Línea central: $\bar{\bar{X}}$	• Línea central: \bar{R}
• $LCI = \bar{\bar{X}} - A_2 \cdot \bar{R}$	• $LCI = D_3 \cdot \bar{R}$

Dónde:

$$\bar{\bar{X}} = \frac{1}{m} \sum_{i=1}^m \bar{x}_i$$

\bar{x}_i es la media muestral de la muestra i-ésima, calculada con los n valores de cada muestra y m es el número total de muestras) y:

$$\bar{R} = \frac{1}{m} \sum_{i=1}^m R_i$$

R_i es el rango de la muestra i-ésima.

Existe una relación entre las distribuciones normales de la desviación típica y el rango, llamado rango relativo $W=R/\sigma$, establecida a través de una constante d_2 , media de W , dependiente del tamaño de las muestras y tabulado a través de un estimador de σ , a través de la siguiente expresión:

$$\hat{\sigma} = \frac{\bar{R}}{d_2} \tag{2.1}$$

De igual manera, la desviación típica del rango σ_R se puede estimar a partir de la desviación típica de la variable corregida por una constante d_3 , a través de la expresión:

$$\hat{\sigma}_R = d_3 \frac{\bar{R}}{d_2} \tag{2.2}$$

Los valores (A_2, D_4 y D_3) que acompañan al rango medio en los límites de control, son valores tabulados que recogen el efecto de los factores de corrección de las estimaciones utilizadas (d_2 y d_3), dependientes del tamaño muestral, para situar los límites de control a tres desviaciones típicas del rango medio.

- Gráfico \bar{X} -S

En este caso, la medida de dispersión utilizada es la desviación típica y al contrario que en el gráfico anterior es adecuado para tamaños muestrales grandes por encima de diez.

Los límites de control para el gráfico \bar{X} -S se presentan en la tabla 2.2:

Tabla 2.2: Límites de control, grafico \bar{X} -S.

<u>Gráfico \bar{X}:</u>	<u>Gráfico S:</u>
• $LCS = \bar{\bar{X}} + A_3 \cdot \bar{S}$	• $LCS = B_4 \cdot \bar{S}$
• Línea central: $\bar{\bar{X}}$	• Línea central: \bar{S}
• $LCI = \bar{\bar{X}} - A_3 \cdot \bar{S}$	• $LCI = B_3 \cdot \bar{S}$

Dónde:

$$\bar{\bar{X}} = \frac{1}{m} \sum_{i=1}^m \bar{x}_i$$

\bar{x}_i es la media muestral de la muestra i -ésima, calculada con los n valores de cada muestra y m es el número total de muestras.

Los valores (A_3 , B_3 y B_4), son valores tabulados que dependen del tamaño de las muestras, de la distancia a la que están situados los límites de control de tres desviaciones típicas y de un factor C_4 , utilizado para obtener una estimación insesgada de la desviación típica.

Los valores tabulados que aparecen tanto en el gráfico \bar{X} -R, como en el gráfico \bar{X} -S pueden encontrarse en numerosos libros de control estadístico [11].

Una vez se han representado en el gráfico de control, las posiciones medias junto con los límites de control superior e inferior, se determina si el proceso se encuentra bajo control o no. Para ello se comprobaba si alguna de las muestras esta fuera de los límites de control. En caso afirmativo, se considera el proceso fuera de control y se procedería a buscar la causa asignable que pudiera explicar dicha situación. Si, por el contrario, todas las muestras están dentro de los límites de control, y no se observan patrones de comportamiento aleatorio, se considera que el proceso se encuentra en control.

Sin embargo, esto no siempre es así, y aun careciendo los gráficos de control de puntos fuera de los límites establecidos, pautas de comportamiento no aleatorio, son a veces difícilmente detectables solamente con la referencia de los límites de control.

En aras de detectar pautas de comportamiento no aleatorio en un gráfico de control, la compañía Western Electric, publico en el Western Electric Handbook [12], una serie de reglas para determinar este tipo de situaciones. Estas reglas, también denominadas RWE por sus siglas en inglés, son las siguientes:

- 1- Un punto cae fuera de los límites de control 3-sigma.
- 2- Dos de tres puntos consecutivos caen fuera de los límites 2-sigma.
- 3- Cuatro de cinco puntos consecutivos se encuentran fuera de los límites 1-sigma.
- 4- Ocho puntos consecutivos están en un mismo lado de la línea central.

Estas reglas, están basadas en situaciones poco probables, sabemos que para una distribución normal, el 68,2 % de la población de datos se encuentran en el rango de $\pm 1\sigma$, el 95,4 % se encuentra en $\pm 2\sigma$ y el 99,73% se encuentra en torno al $\pm 3\sigma$ sigma; por lo tanto, la probabilidad de que un punto esté predestinado a encontrarse fuera de los límites de control de tres desviaciones típicas, se antoja solo al 0,27 % de los datos. De igual forma, se identifican otra serie de situaciones poco probables como la probabilidad de observar dos puntos de tres consecutivos por encima de dos desviaciones típicas, la probabilidad de observar cuatro puntos de cinco consecutivos en un rango entre una desviación típica y ocho puntos consecutivos situados en un mismo lado de la línea central.

La aplicación de estas reglas, aumenta la sensibilidad de los gráficos, dándose falsas alarmas de manera más habitual. Este hecho es importante debido a que se pueden derrochar recursos económicos, y tiempo en la búsqueda del origen de una causa asignable cuando en realidad es una causa común. Por lo tanto, ante la aparición de un aviso de acuerdo con las reglas RWE, no se tratara como una alarma, si no únicamente como una advertencia para aumentar la vigilancia del proceso.

El uso de los gráficos de control de Shewhart está muy implantado en el control de procesos industriales, debido tanto a su sencillez a la hora de construirlos como para interpretar los resultados; estos son muy adecuados para detectar sesgos muy significativos o un incremento en las fluctuaciones aleatorias. Sin embargo, no todo son ventajas, entre los inconvenientes que presenta la utilización de este tipo de gráficos es que no son eficientes en la detección de pequeños sesgos, por debajo de dos desviaciones típicas, o derivas lentas en el proceso. Esto es debido, a que solo representan la información contenida en una muestra, no teniendo en cuenta la información contenida en las observaciones pasadas. Este problema se puede ver mitigado con el uso de las reglas RWE, desarrolladas por la Western Electric, aunque como se muestra en [13], esto conllevaría a una reducción considerable del ARL_0 , con respecto al del gráfico Shewhart sin usar estas reglas.

Para solucionar este problema, se desarrollaron otra serie de gráficos que tendrían en cuenta la información pasada, entre los que cabe destacar el de sumas acumuladas (Cusum) y el de medias móviles exponencialmente ponderadas (EWMA), que se desarrollaran en los apartados siguientes.

2.1.3 Gráficos de control Cusum

Los gráficos de sumas acumuladas (Cusum), propuestos por primera vez por Page [4], surgieron ante la necesidad de cubrir las limitaciones existentes en los gráficos de Shewhart, por su característica de falta de memoria considerando solo la última muestra obtenida, perdiendo la información provista por las muestras anteriores, y no siendo muy efectivos en la detección de cambios pequeños en el proceso; el enfoque principal de estos nuevos gráficos, fue considerar información histórica para detectar cambios pequeños en el proceso [14]. Desde que Page los propusiera por primera vez en 1954, con posterioridad, muchos otros autores los han estudiado [15-20].

Posteriormente, muchos han sido los estudios comparando la efectividad de los gráficos de Shewhart y los propuestos inicialmente por Page. Así, en [21], en la simulación del proceso de taladrado de una pieza, se analiza la respuesta de ambos gráficos frente a una pequeña desviación de 0,5 desviaciones típicas en el proceso. Mientras que el gráfico XR, permanece en control ante la perturbación provocada, el gráfico Cusum si fue capaz de detectar el pequeño cambio provocado.

También se han aplicado con efectividad en la detección de posibles defectos en el cojinete principal de un aerogenerador [22]; y que por su capacidad de memoria ofreció una estimación fehaciente en el desarrollo del desgaste del cojinete en función del tiempo de operación.

Como ya se ha mencionado anteriormente, los gráficos Cusum, incorporan toda la información histórica de los valores de las muestras, mediante la representación de las sumas acumuladas de las desviaciones de los valores de la muestra desde un valor objetivo.

El estadístico de sumas acumuladas adopta la siguiente forma:

$$C_i = \sum_{j=1}^i (\bar{X}_j - \mu_0) = X_i - \mu_0 + \sum_{j=1}^{i-1} (\bar{X}_j - \mu_0) = (X_i - \mu_0) + C_{i-1} \quad (2.3)$$

Donde i es el número de muestras consideradas, \bar{X}_j es el promedio de la j -ésima muestra y μ_0 es el valor objetivo, normalmente la media global de la variable monitorizada cuando el proceso está bajo control.

Se observa que si el proceso permanece bajo control para el valor objetivo μ_0 , el estadístico definido en la ecuación (2.3), debería variar aleatoriamente alrededor de cero. Sin embargo, si la media varía un valor $\mu_1 > \mu_0$, se observaría un desplazamiento positivo, es decir una tendencia creciente, en la suma acumulada C_i . Al contrario, si la media varía un valor $\mu_1 < \mu_0$, se observaría un desplazamiento negativo, es decir una tendencia decreciente, en la suma acumulada C_i . Por lo tanto, si se observa una tendencia significativa en el estadístico, tanto positiva como negativa, deberíamos considerarlo como una evidencia de que la media del proceso ha cambiado y se debería investigar alguna causa asignable responsable del cambio en el proceso con respecto a las condiciones de control.

Existen dos planteamientos diferentes para establecer los límites de control en este tipo de gráficos: Cusum tabular o algorítmico y máscara en V o V-Mask. De los dos planteamientos, el Cusum tabular es el más recomendado; las desventajas de la representación mediante el V-Mask serán expuestas más adelante.

En el primero de los casos, normalmente denominado Cusum tabular, calcula las desviaciones acumuladas desde el valor objetivo μ_0 , distinguiendo entre desviaciones positivas C^+ y negativas C^- . Los estadísticos C^+ y C^- , tienen la forma:

$$C_i^+ = \max[0, X_i - (\mu_0 + K) + C_{i-1}^+] \quad (2.4)$$

$$C_i^- = \max[0, (\mu_0 - K) - X_i + C_{i-1}^-] \quad (2.5)$$

Donde los valores iniciales son $C^+ = C^- = 0$, μ_0 es el valor medio de la característica de calidad cuando el proceso se encuentra bajo control, X_i es el valor de la i -ésima observación y K es el valor de referencia.

Normalmente, el valor de tolerancia K , es la mitad entre el valor objetivo μ_0 y el valor fuera de rango μ_1 que se está interesado en detectar o si la desviación es expresada en desviaciones típicas como $\delta^* = (\mu_1 - \mu_0) / \sigma$.

$$K = \frac{|\mu_1 - \mu_0|}{2} = \frac{\delta}{2} \sigma = k \cdot \sigma \quad (2.6)$$

Los límites de control o intervalos de decisión están dados por:

$$H = h \cdot \sigma \quad (2.7)$$

Donde, los parámetros k y h son seleccionados para aportar un buen comportamiento del ARL. Estudios desarrollados por numerosos autores, [20, 23, 24], han permitido establecer un valor de k igual a la mitad de la desviación que queremos detectar y valores de h iguales a 4 ó 5, para los que se proveerá un gráfico Cusum con buenas propiedades en el valor de longitud promedio de corrida (ARL) ante pequeñas desviaciones en la media del proceso.

Otra forma de analizar el gráfico Cusum, es a través del Cusum V-Mask propuesto por Barnard [25]. Para la representación gráfica de este tipo de Cusum se utiliza el siguiente estadístico:

$$C_i = \sum_{j=1}^i y_j = y_i + C_{i-1} \quad (2.8)$$

Donde y_i es la observación estandarizada $y_i = (x_i - \mu_0)/\sigma$

Una máscara en V , como se observa en la figura 2.2, es una V de costado que se superpone en el gráfico de sumas acumuladas. El punto de origen de la máscara de V se coloca sobre el último punto de suma acumulada y los puntos anteriores se examinan para ver si alguno cae por encima o por debajo de los lados de la V . Si todos los puntos anteriores se encuentran entre los lados de la V , el proceso se considera bajo control, por el contrario, como en el gráfico representado más abajo, si alguna suma acumulada cae fuera de la V , se sospecha que el proceso está fuera de control. Esta máscara debe aplicarse a cada nuevo punto que se trace en el gráfico Cusum, los brazos de la nueva máscara se consideran una extensión de los brazos determinados anteriormente.

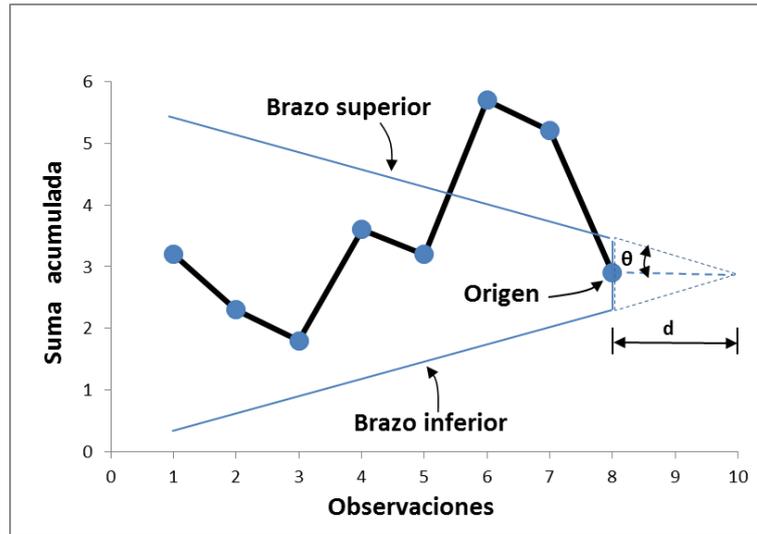


Figura 2.2: Gráfico Cusum V-Mask.

Está claro que el comportamiento de este gráfico está determinado por la distancia d y el ángulo θ . Johnson [26], propone una forma general para determinar ambos parámetros mediante las siguientes ecuaciones:

$$\theta = \tan^{-1} \left(\frac{\delta}{2A} \right) \tag{2.9}$$

$$d = \left(\frac{2}{\delta^2} \right) \ln \left(\frac{1 - \beta}{\alpha} \right) \tag{2.10}$$

Dónde:

2α ; es la probabilidad de una falsa alarma.

β ; la probabilidad de no detectar un cambio en la media del proceso, cuando realmente si ha ocurrido.

δ ; La cantidad del cambio en la media que queremos detectar, expresado como un múltiplo de la desviación típica.

A ; mide la distancia horizontal de la máscara V entre dos puntos sucesivos en términos de unidad de distancia al eje vertical.

Cuando el valor de β es muy pequeño, el valor de la distancia d se puede aproximar por:

$$d \approx -2 \frac{\ln(\alpha)}{\delta} \tag{2.11}$$

Para que el Cusum tabular y el Cusum de mascara en V tengan propiedades similares, los parámetros de ambos gráficos deben guardar la siguiente relación:

$$K = A \cdot \tan \theta; h = A \cdot d \cdot \tan \theta = d \cdot k \tag{2.12}$$

Sin embargo, esta forma de análisis del gráfico Cusum ha tenido sus detractores; Johnson [26] afirmó que la V-Mask da resultados inexactos e imprecisos además de obtener peores valores para el ARL bajo control y ser menos sensible a los cambios en la media del proceso. Montgomery [11], no recomienda el uso de este tipo de representación debido a que en ocasiones es difícil determinar hasta donde deben extenderse los brazos de la máscara V, lo que complica su interpretación. Por estas consideraciones, en la práctica es más utilizado el Cusum tabular que el V-Mask.

Aunque existen diferentes técnicas para calcular el ARL, Vance [27] propone un software bastante exacto, Hawkins [28] propone un cálculo, con una precisión entre 1% y el 3% del valor real del ARL, basado en una ecuación de aproximación donde se requiere una tabla de constantes. Mientras que Woodall y Adams [20] recomiendan el cálculo del ARL, basándose en la aproximación dada por Siegmund [29] por la simplicidad de los cálculos, como:

$$ARL = \frac{\exp(-2\Delta b) + 2\Delta b - 1}{2\Delta^2} \tag{2.13}$$

Donde, para C_i^+ , $\Delta = \delta^* - k$ y para C_i^- , $\Delta = -\delta^* - k$ y $b = h + 1.166$ y $\delta^* = (\mu_1 - \mu_0) / \sigma$.

El valor de δ^* representa el cambio en la media en función de σ , para el que el ARL está siendo calculado. Por lo tanto, si $\delta^* = 0$, el ARL_0 se calcula siguiendo la expresión (2.13), mientras si $\delta^* \neq 0$, se debe calcular el ARL_1 correspondiente al cambio en δ^* .

Para calcular el ARL teniendo en cuenta ambos lados del estadístico ARL^+ y ARL^- , se sigue la siguiente expresión:

$$\frac{1}{ARL} = \frac{1}{ARL^+} = \frac{1}{ARL^-} \tag{2.14}$$

Como se observa en la tabla 2.3 presentada en [30], los valores de ARL obtenidos mediante la aproximación de Siegmund están muy próximos a los valores exactos presentados por Lucas [17].

Tabla 2.3: Representación del ARL para el Cusum tabular con $K=0,5$ [30].

Cambio en la media (múltiplo de σ)	h=4		h=5		Shewhart Control chart
	valores exactos	Aproximación de Siegmund	valores exactos	Aproximación de Siegmund	
0	168	169,55	465	471,9	371
0,25	74,2	77,55	139	142	281,14
0,5	26,6	26,6	38	37,86	155,22
0,75	13,3	13,28	17	17	81,22
1	8,38	8,38	10,4	10,4	44
1,5	4,75	4,76	5,75	5,75	14,97
2	3,34	3,35	4,01	4	6,3
2,5	2,62	2,62	3,11	3,1	3,24
3	2,19	2,2	2,57	2,58	2
4	1,71	1,7	2,01	2,01	1,19

Como se puede observar, para detectar un cambio de una desviación típica es necesario examinar 8,38 o 10,4 muestras para valores de h de 4 o 5 respectivamente; si comparamos este resultado con las 44 muestras obtenidas para el gráfico de Shewhart, vemos que hemos obtenido una mejora sustancial. Mientras que para detectar un cambio de tres o más desviaciones típicas los gráficos Cusum requieren valores de ARL superiores a los de Shewhart.

Para el caso de una situación bajo control, el número de muestras examinadas hasta encontrar una falsa alarma es de 168 y 465 en función del valor elegido de h respectivamente; mientras que es necesario elegir un valor de h igual a 4,77, para obtener un valor de ARL_0 igual a 370, igual al obtenido en el caso de Shewhart con límites de control situados a tres desviaciones típicas.

Como se observa, el Cusum tabular es más eficiente en la detección de cambios pequeños en la media por debajo de 1,5 desviaciones típicas, pero presenta problemas para la detección de grandes cambios.

Una alternativa para solucionar este problema consiste en el uso de la combinación de gráficos Cusum junto con gráficos Shewhart. Es recomendable, en este caso, situar los límites de control de Shewhart aproximadamente a 3,5 desviaciones típicas del valor objetivo μ_0 . En [31], se dio una buena discusión de esta técnica, los resultados obtenidos se presentan en la tabla 2.4.

Tabla 2.4: Valores de ARL para Cusum tabular y Cusum -Shewhart [31].

Cambio en la media (múltiplo de σ)	Cusum con $k=0,5$ y $h=5$	Cusum -Shewhart (Límites de control de Shewhart $3,5\sigma$)
0	465	391
0,25	139	130,9
0,5	38	37,2
0,75	17	16,8
1	10,4	10,2
1,5	5,75	5,58
2	4,01	3,77
2,5	3,11	2,77
3	2,57	2,1
4	2,01	1,34

Se observa que mediante la combinación de estos dos gráficos de control se consigue una mejora del ARL para grandes cambios. Por ejemplo, cuando se intentan detectar desviaciones en el proceso de tres desviaciones típicas, es necesario examinar 2,57 muestras para el caso de los gráficos Cusum, mientras que con la combinación de ambos gráficos es necesario examinar 2,10 muestras para detectar un cambio, lo que demuestra una ligera mejora en el ARL, ya que serán capaces de detectar más rápidamente los cambios producidos en el proceso. Por lo tanto, la combinación de ambas técnicas mejora el comportamiento de los gráficos Cusum ante grandes cambios en el proceso.

Aunque, lo más habitual es encontrar estos gráficos desarrollados para el control de la media del proceso, se ha estudiado también el desarrollo de métodos Cusum para el

control de la variabilidad. Concretamente en [18], suponiendo que la distribución de la variable bajo control x_i es normal con media μ_0 y desviación típica σ , se sugiere desarrollar un gráfico de control para el estadístico v_i .

$$v_i = \frac{\sqrt{|y_i|} - 0,822}{0,349} \quad (2.15)$$

Donde $y_i = (x_i - \mu_0) / \sigma$ de x_i .

En este estudio, se sugirió que el valor v_i era más sensible a cambios en la varianza que a cambios en la media, por lo que lo usara en un gráfico de control denominado (Scale Cusum) para la variabilidad del proceso estableciendo lo siguientes estadísticos:

$$S_i^+ = \max[0, v_i - k + S_{i-1}^+] \quad (2.16)$$

$$S_i^- = \max[0, -k - v_i + S_{i-1}^-] \quad (2.17)$$

Donde los valores iniciales son $S_i^+ = S_i^- = 0$, siendo los valores k y h elegidos de manera similar al Cusum para el control de la media.

La interpretación del Scale Cusum, es similar a la interpretación del Cusum para el control de la media. Si la desviación típica del proceso aumenta, el valor de S_i^+ aumentara y excederá del valor h en el que están situados los límites de control, por el contrario si la desviación típica disminuye, el valor de S_i^- aumentara y excederá del valor h . Aunque se pueden representar separadamente ambos Cusum, para la media y para la desviación típica, Hawkins [24], sugiere representar en el mismo gráfico ambos Cusum y si en el Scale Cusum aparece una señal, deberíamos de sospechar de un cambio en la varianza, mientras que si en ambos Cusum aparece una señal, se debería de sospechar de un cambio en la media.

2.1.4 Gráficos EWMA

Los gráficos de medias móviles exponencialmente ponderadas (EWMA), son una buena alternativa, al igual que los gráficos de control de sumas acumuladas (Cusum), al gráfico Shewhart para detectar pequeñas desviaciones en el proceso. Estos gráficos

fueron propuestos por Roberts [5], manteniendo la filosofía de utilizar la información previa, aunque han sido estudiados por otros autores, [32-34]. El estadístico del EWMA se define como:

$$Z_i = \lambda \cdot x_i + (1 - \lambda) \cdot Z_{i-1} \quad (2.18)$$

Dónde:

λ ; es una constante entre cero y uno que determina el mayor o menor peso que tienen los datos anteriores. Siendo el valor inicial de λ igual que el valor objetivo μ_0 .

X_i ; es el promedio de la i -ésima muestra.

Z_{i-1} ; es el valor del estadístico EWMA correspondiente al periodo anterior al momento en el que estamos situados.

Si las observaciones x_i son variables aleatorias independientes con varianzas σ^2 , entonces definió la varianza de z_i como:

$$\sigma_{z_i}^2 = \sigma^2 \cdot \left(\frac{\lambda}{2 - \lambda} \right) \cdot [1 - (1 - \lambda)^{2i}] \quad (2.19)$$

Los límites de control están determinados por:

$$LCS = \mu_0 + L \cdot \sigma \sqrt{\frac{\lambda}{(2 - \lambda)} \cdot [1 - (1 - \lambda)^{2i}]} \quad (2.20)$$

$$LC = \mu_0 \quad (2.21)$$

$$LCI = \mu_0 - L \cdot \sigma \sqrt{\frac{\lambda}{(2 - \lambda)} \cdot [1 - (1 - \lambda)^{2i}]} \quad (2.22)$$

Donde L es la anchura de los límites de control.

Como se puede observar en las ecuaciones anteriores, estos límites no son constantes, sino que van a aumentar conforme aumenta el número de observaciones, tendiendo a estabilizarse, ya que el término $[1 - (1 - \lambda)^{2i}]$ converge a la unidad con grandes valores de i . Por lo que después de varios periodos, los límites de control se pueden definir como:

$$LCS = \mu_0 + L \cdot \sigma \sqrt{\frac{\lambda}{(2-\lambda)}} \quad (2.23)$$

$$LC = \mu_0 \quad (2.24)$$

$$LCI = \mu_0 - L \cdot \sigma \sqrt{\frac{\lambda}{(2-\lambda)}} \quad (2.25)$$

La utilización de las ecuaciones (2.20) y (2.22) para pequeños valores de i , según [11], mejorara el funcionamiento de la carta de control en la detección de señales inmediatamente después de haber comenzado con el estadístico.

El cálculo del ARL ha sido estudiado por numerosos autores desde distintos puntos de vista. En [35] evaluaron el ARL utilizando la expansión de Edgeworth. En [32, 33] muestra que puede ser descrita como una integral de Fredholm de segundo orden. Pero probablemente el trabajo más amplio en el estudio del ARL para los gráficos EWMA es el desarrollado en [34] en el que elaboran tablas que recogen el comportamiento para distintos valores de λ y L . En general, se ha comprobado que para valores $0,05 \leq \lambda \leq 0,25$ suelen tener un buen comportamiento en la mayoría de las situaciones.

Hunter [36] también estudio el gráfico EWMA, y sugirió asignar un valor de λ tal que el peso dado a las observaciones actuales y previas, fuese muy parecido al dado a estas observaciones mediante el gráfico de Shewhart con el uso de las reglas Western Electric.

Aunque este tipo de gráficos se suelen utilizar para el control de la media del proceso, al igual que ocurría con el gráfico Cusum, también se han estudiado para el control de la variabilidad. Suponiendo que la distribución de la variable bajo control es normal con media μ_0 y desviación típica σ , en [37], se sugiere desarrollar un gráfico de control, definiendo lo que se denomina error cuadrático medio exponencialmente ponderado (EWMSE) de sus siglas en inglés, como:

$$S_i^2 = \lambda \cdot (x_i - \mu)^2 + (1 - \lambda) \cdot S_{i-1}^2 \quad (2.26)$$

Se puede demostrar que $E(S_i^2) = \sigma^2$ para muestras de tamaño grande y además si las observaciones son independientes y están normalmente distribuidas, entonces S_i^2 / σ^2 tiene aproximadamente una distribución chi-cuadrado con $v = (2 - \lambda) / \lambda$ grados de libertad. Por lo tanto, si σ_0 representa el valor bajo control de la desviación típica del proceso, la raíz cuadrada del error cuadrático medio exponencialmente ponderado (EWMSE) $\sqrt{S_i^2}$, puede trazarse gráficamente junto a los límites de control dados por:

$$LCS = \sigma_0 \sqrt{\frac{X_{v, \alpha/2}^2}{v}} \quad (2.27)$$

$$LCI = \sigma_0 \sqrt{\frac{X_{v, 1-(\alpha/2)}^2}{v}} \quad (2.28)$$

También señalan que el estadístico EWMS puede ser sensible tanto a cambios en la media como en la desviación típica del proceso. Para solucionar este problema proponen sustituir $\hat{\mu}_i$ en cada punto del tiempo por la estimación de este valor mediante el estadístico EWMA (z_i), de tal forma que el valor a representar gráficamente, junto a los límites de control correspondientes se determina como:

$$S_i^2 = \lambda \cdot (x_i - z_i)^2 + (1 - \lambda) \cdot S_{i-1}^2 \quad (2.29)$$

Al igual que con los gráficos Shewhart y Cusum este tipo de gráfico también ha sido objeto de comparación. En este contexto, en el proceso de elaboración de un compuesto químico [38], fueron examinadas 40 muestras. A partir de la muestra 30, provocaron un cambio en el proceso de una desviación típica. Si no se tienen en cuenta las reglas RWE, el gráfico de Shewhart no fue sensible a la desviación provocada, mostrando que el proceso se encontraba en control estadístico. El gráfico EWMA, indicó que el proceso estaba en una situación de fuera de control a partir de la muestra 38, detectando además que a partir de la muestra 30 comenzó a desviarse de su condición normal.

Diferentes autores, también han realizado estudios comparando para un mismo proceso las técnicas univariantes estudiadas en la presente tesis; así, podemos encontrar el trabajo realizado en [14], donde en la industria automovilística, se compararon los gráficos Shewhart, Cusum y EWMA, para detectar pequeñas desviaciones en un proceso de fabricación de una pieza. Los resultados indicaron que ante una misma

magnitud en la desviación del proceso, los gráficos Cusum y EWMA, detectaron el cambio prácticamente en el mismo número de muestra, mientras que los gráficos Shewhart no alcanzaron a detectar el cambio.

En conclusión, se puede afirmar que los gráficos Shewhart son muy sencillos de utilizar y se comportan muy bien en la detección de cambios grandes, por encima de $1,5 \sigma$, mientras que, ante variaciones menores en el proceso, se hace necesario la utilización de gráficos como Cusum o EWMA. Si bien los gráficos Shewhart como se expone en [38], pueden trabajar razonablemente, bien incluso en situaciones en los que se viola de manera moderada el supuesto de normalidad, como fue señalado por [39], para subgrupos de tamaño menores de cuatro, la no normalidad puede generar serios problemas, en particular con la aparición de falsas alarmas. Como se denota en [40], la estructura de EWMA no es sensible a la normalidad, mientras que los gráficos Cusum si lo son.

Por otro lado, como se expone en [41], aunque ambos gráficos (Cusum y EWMA) tienen un buen comportamiento ante variaciones pequeñas, cuando el cambio comienza a aumentar, el gráfico EWMA tienen un mejor comportamiento, sin embargo, los cálculos del ARL para este gráfico se hacen muy tediosos y es necesario la utilización de un software especial.

Los gráficos univariantes, comentados en este apartado son eficientes cuando la monitorización, se realiza para una sola variable. En numerosos procesos, es necesario, analizar el comportamiento de dos o más variables al mismo tiempo; en este supuesto, por razones obvias, los gráficos univariantes no pueden ser utilizados y hay que hacer uso de técnicas multivariantes que se analizaran en el siguiente apartado.

2.2 Gráficos de control multivariantes

2.2.1 Introducción

Las técnicas de control univariantes, están diseñadas para controlar únicamente una característica de calidad. Sin embargo, en los procesos actuales hay muchas situaciones en la cual el control simultáneo de dos o más variables relacionadas con la calidad es necesario [42].

Una primera posibilidad, para solventar esta situación sería aplicar técnicas de control de calidad univariantes de forma independiente para cada una de las características. Sin embargo, no se podría controlar el efecto de las interacciones entre las diversas variables así como su variación a lo largo del tiempo. Para solucionar este problema, se dio impulso a técnicas de control multivariado, con las que es posible analizar varias características de la calidad al mismo tiempo [43]. Dentro de estas técnicas multivariantes, por su desarrollo y mayor uso cabe destacar las siguientes: técnicas basadas en distancias (T^2 de Hotelling), MEWMA y MCusum.

En este apartado, primero se analizarán los gráficos de control MCusum y MEWMA, extensión multivariante de los gráficos Cusum y EWMA, respectivamente. Posteriormente, se abordarán los gráficos de control T^2 de Hotelling, indicando cuáles son sus limitaciones, con respecto a un aumento desmesurado de las variables a analizar y la deficiencia que tienen en la detección de cambios pequeños en la media. Esta última parte es la base en la que se sustenta el desarrollo de la presente tesis, en la que se intentará dar solución a estos dos principales inconvenientes en este tipo de gráfico.

2.2.2 Gráfico MCusum

El gráfico MCusum es la extensión multivariante del ya comentado en el apartado anterior gráfico univariante Cusum. Existen dos aproximaciones distintas para extender el gráfico Cusum al caso multivariante; una es realizar una modificación del esquema Cusum univariante, obteniendo una aproximación para procesos multivariantes y elaborar un MCusum para las observaciones de las distintas variables y la otra, consiste en un análisis de forma simultánea de varios procedimientos Cusum univariantes, reduciendo las observaciones multivariantes a un escalar y elaborar un Cusum de escalares.

En [44], proponen la primera extensión multivariante del gráfico Cusum, a través de la elaboración simultánea de Cusum univariantes y de este modo evaluar el comportamiento del conjunto de gráficos. Este método, se verá afectado por la dirección del cambio en el vector de medias, aunque presenta un ARL en control inferior al del gráfico Cusum univariante.

Posteriormente, en [8] se propusieron, dos procedimientos para la elaboración de un gráfico de control MCusum: el primero consiste en la aplicación de un Cusum sobre la raíz cuadrada de los valores T^2 obtenidos de un gráfico de control de Hotelling y el segundo consiste en aplicar el gráfico Cusum directamente sobre los vectores observaciones.

Sin embargo, Pignatiello y Runger [45], probablemente elaboraron, los gráficos MCusum que mejor comportamiento tienen en términos de ARL para detectar cambios de forma rápida.

Ellos proponen dos métodos basados en la forma cuadrática del vector de medias. La diferencia entre los dos métodos MCusum reside en la forma en la que se lleva a cabo la acumulación. En el primer procedimiento, denominado Multivariate CUSUM # 1 o MC1, acumula el vector de observaciones x antes de producir la forma cuadrática, mientras que en el segundo procedimiento denominado Multivariate CUSUM # 2 o MC2, calcula la forma cuadrática para cada valor del vector de observaciones x y entonces se acumulan estas formas cuadráticas.

El primer procedimiento MC1, se basa en el siguiente estadístico:

$$MC1_i = \max[|C_i| - kn_i; 0] \tag{2.30}$$

Dónde:

$$|C_i| = \sqrt{C_i \Sigma^{-1} C_i}$$

$$C_i = \sum_{j=i-n_i+1}^i (\bar{X}_j - \mu_0)$$

$$n_i = \begin{cases} n_{i-1}, & n_i = MC1_{i-1} > 0 \\ 0, & \text{en otro caso} \end{cases}$$

Donde MC1 son los valores del estadístico que se representan en el gráfico de control, $|C_i|$ es el módulo de C_i estandarizado, C_i es la suma acumulada multivariante, n_i es el número de muestra correspondiente a cada muestra que vamos incorporando, y k es un valor de referencia dependiente de la magnitud de la distancia entre la media cuando el

proceso se encuentra bajo control y la media cuando el proceso está fuera de control, que al igual que en el caso univariante suele ser 0,5.

Al igual que en el caso univariante, cada valor calculado a través del estadístico se representa en un gráfico de control. Si alguno de estos valores excede del límite de control previamente establecido, a la mitad del tamaño del cambio en la media que queremos detectar, entonces se considera que el proceso está fuera de control.

El segundo gráfico de control propuesto en [45], MC2, tiene en cuenta el cuadrado de la distancia de cada media muestral desde el valor objetivo μ_0 y entonces acumula estas distancias cuadráticas. Para calcular la distancia, establece el siguiente estadístico:

$$D_t^2 = (X_t - \mu_0)^t \cdot \Sigma^{-1} \cdot (X_t - \mu_0) \quad (2.31)$$

Este estadístico tiene una distribución χ^2 con p grados de libertad cuando el proceso está bajo control y distribución χ^2 no central cuando el proceso está fuera de control. A partir de este estadístico se puede construir un gráfico de control Cusum como:

$$MC2_t = \max[0, MC2_{t-1} + D_t^2 - k] \quad (2.32)$$

Siendo para los valores iniciales $MC2_0=0$

En términos de ARL, el método propuesto por Crosier [8] y el MC1 propuesto en [45], tienen el mismo comportamiento, siempre que el cambio en la media del proceso afecte a una de las variables; sin embargo el MC1, se comporta mejor si el cambio afecta a más de una variable. Si las variables no presentan un comportamiento independiente, se ve aumentada la mejoría en el comportamiento.

2.2.3 Gráfico MEWMA

Los gráficos de Media móvil exponencialmente ponderada multivariable, MEWMA por sus siglas en inglés, son una extensión multivariante de los gráficos EWMA. Estos gráficos al igual que los MCusum, nos va a permitir considerar la información contenida en la serie, pero por el contrario, se pierde la independencia entre observaciones.

El ARL en este tipo de gráficos puede ser de dirección invariante o específica. El ARL de dirección invariante va a venir determinado por la distancia existente entre el valor de la media y el valor objetivo, valor de la media cuando estamos bajo control, mientras que en el de dirección específica el ARL es una función de dicha distancia, pero también tienen en cuenta la dirección.

El estadístico para el gráfico univariante venía dado por:

$$Z_i = \lambda \cdot x_i + (1 - \lambda) \cdot Z_{i-1} \quad \text{para } i=1, 2, \dots, m \quad (2.33)$$

Donde λ es una constante entre cero y uno, que determina el mayor o menor peso que tienen las observaciones anteriores, X_i es el promedio de la i -ésima muestra y Z_{i-1} es el i -ésimo valor EWMA.

Análogamente, la extensión multivariante de la ecuación anterior adopta la forma:

$$Z_i = \Lambda \cdot x_i + (1 - \Lambda) \cdot Z_{i-1} \quad (2.34)$$

Donde Λ es una matriz diagonal formada por los valores lambda para las distintas variables ($\lambda_1, \lambda_2, \dots, \lambda_p$), p es el número de variables, X_i es el promedio de la i -ésima muestra. Z_{i-1} es el i -ésimo valor MEWMA.

A partir de estos valores, habrá que calcular un estadístico que utilizaremos para analizar el estado del proceso, adoptando la siguiente expresión:

$$T_i^2 = Z_i^t \cdot \Sigma_{Z_i}^{-1} \cdot Z_i \quad (2.35)$$

Donde $\Sigma_{Z_i}^{-1}$ es la matriz de varianzas y covarianzas de los valores MEWMA.

Ha sido demostrado por Lowry et al. [9], que el elemento (k, l) de la matriz Σ_{Z_i} , es:

$$\Sigma_{Z_i}(k, l) = \lambda_k \lambda_l \frac{[1 - (1 - \lambda_k)^i (1 - \lambda_l)^i]}{[\lambda_k + \lambda_l - \lambda_k \lambda_l]} \sigma_{k,l} \quad (2.36)$$

Donde $\sigma_{k,l}$ es el elemento (k, l) de la matriz de varianzas y covarianzas de las variables analizadas.

Si el valor de todas las lambdas son iguales para todas las variables que estamos analizando, la expresión anterior se puede simplificar como:

$$\Sigma_{Z_i} = \frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2i}] \Sigma \quad (2.37)$$

Donde Σ es la matriz de varianzas y covarianzas de las variables.

Con grandes valores de i , la expresión $[1 - (1 - \lambda)^{2i}]$, se aproxima mucho a la unidad, por lo que, en algunos casos la matriz de varianzas y covarianzas puede definirse como:

$$\Sigma_{Z_i} = \frac{\lambda}{2 - \lambda} \Sigma \quad (2.38)$$

El gráfico MEWMA permite usar pesos específicos para cada variable medidos en términos del valor de λ_i de cada variable. Aquí es donde se diferencia un análisis direccional de un análisis sin dirección específica. Si se decide utilizar el mismo peso para todas las variables, el gráfico es de dirección invariante debido a que las causas de variación llevan igual peso en todas las variables. Sin embargo, si damos distinto peso a las variables, el gráfico se va a considerar de dirección específica y, en este caso, una señal fuera de control puede ser atribuida a los valores específicos de aquella variable a la que hemos otorgado un peso mayor.

Ha sido demostrado en Prabhu y Runger [46], mediante el análisis de los valores del ARL para distintos cambios en la media del proceso, que el gráfico MEWMA es más rápido en su respuesta ante pequeños cambios en la media del proceso, lo que representa una gran ventaja con respecto a otros modelos de control multivariantes. Si bien es cierto, que es menos efectivo que el gráfico T^2 de Hotelling en la detección de grandes cambios en el proceso.

En este trabajo, también se muestra una guía para la selección del límite de control superior, dependiente del cambio en la media que queramos detectar, del valor lambda seleccionado y del número de variables que estemos analizando. Estos valores del límite superior son determinados de tal forma que el $ARL_0=200$. Otros, como, [47], propusieron un software para calcular el comportamiento del ARL de un gráfico de este tipo.

2.2.4 Gráfico T^2 de Hotelling

El gráfico de control de T^2 de Hotelling se puede considerar como la extensión multivariante al gráfico univariante de Shewhart. Estos gráficos pueden aplicarse para tamaños de subgrupos mayores o iguales que uno. Geométricamente, el estadístico T^2 es proporcional a la distancia al cuadrado entre una observación multivariada y el vector de valores objetivo, donde puntos equidistantes forman elipsoides alrededor de dicho vector, siendo mayor la distancia entre la observación y el valor objetivo cuanto mayor es el valor de T^2 [48].

Lowry y Montgomery [49], sugirieron que el uso apropiado del estadístico podría estar dividido en cuatro categorías:

- 1- Fase I, trabajando con observaciones individuales ($n=1$).
- 2- Fase I, cuando se trabaja con subgrupos de observaciones ($n>1$).
- 3- Fase II, trabajando con observaciones individuales ($n=1$).
- 4- Fase II, cuando se trabaja con subgrupos de observaciones ($n>1$).

Tal y como fue discutido por diferentes autores [49-53], en la fase I se construye el gráfico de control para establecer si el proceso se encontraba en control cuando se obtuvieron las observaciones preliminares y se establece el control estadístico. En la fase II, se monitorizan las futuras observaciones, para comprobar si el proceso ha sufrido algún cambio, con respecto al estado inicial.

En la aplicación de este estadístico, se asume que las variables que se están analizando, siguen una distribución normal multivariable $N_p(\mu_0, \Sigma_0)$, donde la función de normalidad multivariable adopta la siguiente forma:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-1/2 (\mathbf{x}-\mu)^t \Sigma^{-1} (\mathbf{x}-\mu)} \quad (2.39)$$

Dónde: $-\infty < x_j < \infty; j = 1, 2, \dots, p$.

Cuando trabajamos con subgrupos de observaciones ($n>1$), donde tenemos p variables y m muestras de tamaño n , el estadístico sigue la siguiente expresión:

$$D_i^2 = n \cdot (\bar{X}_i - \vartheta_0)^t \cdot Z_0^{-1} \cdot (\bar{X}_i - \vartheta_0) \quad (2.40)$$

Donde \bar{X}_i es el vector de medias de los *-ith* subgrupos, ϑ_0 y Z_0 son el vector de medias y la matriz de covarianzas apropiadas respectivamente. El estadístico D_i^2 representa la distancia de Mahalanobis desde un punto hasta el valor objetivo ϑ_0 . Por lo tanto, si el valor de D_i^2 supera el límite de control superior, se considera que existen señales que hacen que el proceso este fuera de control. A diferencia de los gráficos de control univariantes de Shewhart, en estos no se representa ni el límite central (LC) ni el límite de control inferior (LCI), puesto que la distancia de Mahalanobis por su forma cuadrática nunca es negativa.

Si se conocen los verdaderos parámetros de la distribución, es decir, conocemos el vector de medias y la matriz de varianzas-covarianzas, entonces $\vartheta_0 = \mu_0$ y $Z_0 = \Sigma_0$ y el estadístico T^2 sigue una distribución chi-cuadrado con p grados de libertad.

En el caso bivalente, este estadístico tendría la siguiente forma:

$$X_0^2 = \frac{n}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} [\sigma_2^2 (\bar{X}_1 - \mu_1)^2 + \sigma_1^2 (\bar{X}_2 - \mu_2)^2 - 2\sigma_{12}(\bar{X}_1 - \mu_1)(\bar{X}_2 - \mu_2)] \quad (2.41)$$

Donde \bar{x}_1 y \bar{x}_2 son las medias muestrales de las observaciones, μ_1 y μ_2 son las medias poblacionales, σ_1 y σ_2 las desviaciones típicas y σ_{12} es la covarianza. Si se analizan más de dos variables, el estadístico chi-cuadrado tiene la siguiente forma:

$$X_0^2 = n \cdot (\bar{X}_i - \mu)^t \cdot \Sigma^{-1} \cdot (\bar{X}_i - \mu) \quad (2.42)$$

Donde μ es el vector de medias, Σ es la matriz de varianzas y covarianzas y n es el tamaño muestral. El límite de control superior se va a situar para un nivel de significación dado por $X^2_{(\alpha,p)}$, donde $X^2_{(\alpha,p)}$ es el α th percentil superior de X^2_p

Cuando el verdadero valor poblacional no es conocido, es necesaria su estimación. Si tenemos p variables y m muestras de tamaño n , la media y la cuasivarianza muestral se calculan como:

$$\bar{X}_{jk} = \frac{1}{n} \sum_{i=1}^n x_{ijk} \begin{cases} j = 1, 2, \dots, p \\ k = 1, 2, \dots, m \end{cases} \quad (2.43)$$

$$S_{jk}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ijk} - \bar{X}_{jk})^2 \begin{cases} j = 1, 2, \dots, p \\ k = 1, 2, \dots, m \end{cases} \quad (2.44)$$

Donde x_{ijk} es la i -ésima observación en la j -ésima característica en la muestra k .

La cuasicovarianza entre dos características de calidad j y h se calcula para la muestra k como:

$$S_{jhk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ijk} - \bar{X}_{jk}) (x_{ihk} - \bar{X}_{hk}) \begin{cases} k = 1, 2, \dots, m \\ j \neq h \end{cases} \quad (2.45)$$

Con las expresiones anteriores podemos determinar la media y la varianza para las m muestras y con estos valores estimar la media y la varianza, para cada una de las variables, promediando para las m muestras obtenidas. Estas medias y varianzas, así como el valor de la covarianza para cada una de las variables se estimaran como:

$$\bar{X} = \frac{1}{m} \sum_{k=1}^m \bar{X}_{jk} \quad j = 1, 2, \dots, p \quad (2.46)$$

$$\bar{S}_j^2 = \frac{1}{m} \sum_{k=1}^m S_{jk}^2 \quad j = 1, 2, \dots, p \quad (2.47)$$

$$\bar{S}_{jk} = \frac{1}{m} \sum_{k=1}^m S_{jhk} \quad j \neq h \quad (2.48)$$

Si usamos estos valores para calcular la distancia T^2 de Hotelling, vamos a tener que la expresión de esta distancia queda de la siguiente forma

$$T_i^2 = n \cdot (\bar{X}_i - \bar{X})^t \cdot S^{-1} \cdot (\bar{X}_i - \bar{X}) \quad (2.49)$$

Cuando se utilizan observaciones individuales, $n=1$, la distancia D_i^2 tiene la siguiente forma:

$$D_i^2 = (X_i - \vartheta_0)^t \cdot Z_0^{-1} \cdot (X_i - \vartheta_0) \quad (2.50)$$

Donde, X_i es la i th observación, $i = 1, 2, \dots, m$ siguiendo $N_p(\mu_0, \Sigma_0)$, ϑ_0 y Z_0 son el apropiado vector de medias y matriz de varianza-covarianza en cualquiera de las fases I y II.

Cuando los parámetros poblacionales no son conocidos, la forma de estimar la matriz de varianzas-covarianzas, anteriormente descrita, para el caso de subgrupos de observaciones, no va a ser adecuada, ya que obtendríamos una sobreestimación de dicha matriz.

Además de poder estimar la matriz directamente de la información muestral existen otras alternativas como la propuesta por Sullivan y Woodall [54], consistente en usar las diferencias entre pares sucesivos de observaciones. Vargas [55], utilizó esta propuesta como una alternativa a la estimación clásica de los parámetros del gráfico de control T^2 de Hotelling. Además, compara esta alternativa con distintas alternativas de estimación robusta.

Si, $\vartheta_0 = \bar{X}$ y $Z_0 = S$, entonces el estadístico T^2 de Hotelling adquiere la siguiente forma:

$$T^2 = (X_i - \bar{X})^t \cdot S^{-1} \cdot (X_i - \bar{X}) \quad (2.51)$$

Donde, X_i es la i th observación individual, \bar{X} es la media muestral y S es la matriz de varianzas-covarianzas.

Mientras que cuando los valores poblacionales son conocidos, $\vartheta_0 = \mu_0$, $Z_0 = \Sigma_0$ y X_i es la i th observación, entonces el estadístico T^2 sigue una distribución chi-cuadrado con p grados de libertad,[56].

Los límites de control para el gráfico T^2 de Hotelling, se determinan de tal forma, que cuando el proceso está bajo control, la probabilidad de aparición de una falsa alarma es de 0,0027, igual que en el caso univariante, a una distancia de tres desviaciones típicas [10].

Para subgrupos de observaciones $n > 1$, el estadístico T^2 de Hotelling cuando se conocen los parámetros poblacionales, sigue una distribución Chi-cuadrado con p grados de libertad. Por el contrario, cuando no se conocen los valores poblacionales y hay que estimarlos, la distribución que describe el estadístico T^2 viene dado por una distribución

F de Fisher y las ecuaciones de los límites de control para las fases I y II van a venir dadas como:

$$LCS = \left\{ \frac{p(m-1)(n-1)}{mn-m-p+1} \right\} F_{\{\alpha;p;mn-m-p+1\}} \quad (2.52)$$

Para la fase I donde $F_{\{\alpha;p;mn-m-p+1\}}$, es el α th percentil de $F_{\{p;mn-m-p+1\}}$

Y el límite de control para la fase II vienen dado por:

$$LCS = \left\{ \frac{p(m+1)(n-1)}{mn-m-p+1} \right\} F_{\{\alpha;p;mn-m-p+1\}} \quad (2.53)$$

Donde $F_{\{\alpha;p;mn-m-p+1\}}$, es el α th percentil de $F_{\{p;mn-m-p+1\}}$

En el caso del gráfico de control T^2 de Hotelling para observaciones individuales, si no se conocen los parámetros poblacionales, se distinguen dos situaciones: en la primera fase donde el vector de observaciones no es independiente de las estimaciones de los parámetros, es decir, toda la información que contienen el vector de observaciones es usada en la estimación de los parámetros, el límite superior de control vienen dado por:

$$LCS = \left\{ \frac{(m-1)^2}{m} \right\} \beta_{\{\alpha;\frac{p}{2};(m-p-1)/2\}} \quad (2.54)$$

Donde $\beta_{\{\alpha;\frac{p}{2};(m-p-1)/2\}}$, es el α th percentil de $\beta_{\{\frac{p}{2};(m-p-1)/2\}}$

O bien, en la segunda fase, donde el vector de observaciones no es utilizado en la estimación de los parámetros, el límite superior va a venir dado por:

$$LCS = \left\{ \frac{p(m+1)(m-1)}{m(m-p)} \right\} F_{\{\alpha;p;m-p\}} \quad (2.55)$$

Dónde: $F_{\{\alpha;p;m-p\}}$, es el α th percentil de $F_{\{p;m-p\}}$

Cuando los parámetros son conocidos, en la fase II, como se ha comentado anteriormente, el límite de control estará dado por $LCS = x^2(\alpha, p)$, donde, $x^2(\alpha, p)$ es el α th percentil de una distribución chi-cuadrado con p grados de libertad.

Lowry y Montgomery [49], indican que el límite de control basado en la distribución Chi-cuadrado solo es apropiado si los parámetros de la matriz de covarianza son conocidos y aun así, debe de ser utilizada con precaución. Si el número de variables es mayor o igual que diez, serán necesario al menos 250 muestras, para que la aproximación del límite de control basado en la distribución Chi-cuadrado tenga un valor correcto.

Una vez definidos los límites, se puede analizar fácilmente si el proceso se encuentra bajo control mediante la representación gráfica de cada uno de los valores T^2 junto con dicho límite de control. Si los valores T^2 para todas las muestras se encuentran por debajo del límite de control superior (LCS), se considera que el proceso está en control estadístico, por el contrario si algún valor se encuentra por encima del LCS, se considerará que el proceso ha sufrido alguna variación con respecto a su estado en control. Como se ha comentado anteriormente, en estos no se representa ni el límite central ni el límite de control inferior (LCI), puesto que el estadístico T^2 por su forma cuadrática nunca es negativo.

En los gráficos de control multivariados, al igual que en los gráficos de control univariantes, interpretar una situación fuera de control es una tarea relativamente sencilla, sin embargo, determinar las causas que han originado ese estado es más complicado, debido a que una situación fuera de rango puede haber sido originada por la existencia de una variable fuera de control o a un cambio en la relación existente entre dos o más variables [57].

Se han desarrollado algunas técnicas para ayudar en la interpretación de las señales fuera de control, como analizar gráficos de control univariantes para cada una de las variables e intentar identificar que variable es la que ha provocado la desviación en el proceso multivariable. Sin embargo, esta forma de proceder tienen varios inconvenientes, primero que cuando el proceso multivariable está compuesto por muchas variables, esta técnica tiende a ser tediosa, ya que hay muchos gráficos a analizar y la segunda, y más importante, es que normalmente una señal fuera de control no es causada por una sola variable, sino por la interacción de varias variables, por lo que en muchas ocasiones, los gráficos univariantes no van a mostrar señales fuera de control y por lo tanto no van a mostrar cual es la variable que está generando el estado fuera de control en el proceso multivariable [58].

Alt [59], desarrollo gráficos de control de medias individuales con límites de control tipo Bonferroni, indicando también que esto no implicaba que los gráficos univariantes deberían ser utilizados en lugar de los multivariantes. En [60], utilizaron un procedimiento de intervalos de confianza simultáneos tipo Bonferroni para cada una de las características de calidad. Estos intervalos de confianza son unos sustitutos de los gráficos de control de media individuales y suelen ser más efectivos en la identificación de la característica de calidad que causa la señal fuera de control en los gráficos multivariantes.

Hawkins [24, 61] y Wade y Woodall [62] separadamente, utilizaron ajustes de regresión en variables individuales para mejorar la potencia de diagnóstico del T^2 después de la detección de la señal. En [63] propusieron el uso de distancias métricas diferentes. En [59], proponen utilizar la región de control elíptica, aunque no puede ser aplicable cuando el número de características de calidad es mayor de dos. Murphy [64], recomienda un método que consiste en desarrollar un procedimiento basado en un análisis discriminatorio que permita agrupar las observaciones en grupos. Jackson [65], propone representar componentes principales en lugar de las variables originales, con el posible problema de interpretación de esos resultados cuando no sea fácil encontrar el significado de dichas componentes.

Aunque, la técnica que ha demostrado ser más eficiente en la interpretación de señales en el gráfico T^2 de Hotelling fue la presentada por Mason, Young y Tracy, [66, 67] en la que proponen una técnica que consiste en la descomposición del estadístico T^2 de forma que mida la influencia de cada una de las variables sobre el valor de T^2 . Esta descomposición aparece en [68], y consiste en dividir el estadístico T^2 en dos componentes: por un lado el término condicional, que recoge el valor del estadístico para cada una de las variables y por otro lado, el término incondicional que recoge el comportamiento del estadístico basado en los residuos de la regresión de cada variable sobre las demás, de forma que se recogen las variaciones que se producen en la estructura de correlación de las variables.

Esta técnica, denominada MYT (Mason, Young and Tracy), por las iniciales de sus autores, ha sido probada con resultados satisfactorios en numerosas aplicaciones; por ejemplo, en la industria alimenticia [69], en un proceso de extrusión de película plástica [70] o en la monitorización de un reducido número de variables pertenecientes al

proceso de lubricación de un motor marino [71], en el que el estadístico T^2 de Hotelling fue implementado en combinación con la técnica MYT, concluyendo que el control multivariable T^2 se mostró eficiente en la detección de posibles cambios en el proceso y la descomposición estadística, identificó las variables que generaron desviaciones en el proceso, facilitando la diagnosis de los cambios que se habían producido. También, en [72], se utilizó esta técnica, en la monitorización de la condición de lubricación del engranaje de la hélice de un buque.

Sin embargo, como se indica en [68], cuando el número de variables comienza a ser moderadamente grande, superior a diez, las posibles combinaciones entre ellas se incrementan exponencialmente y hace que sea confusa la interpretación de la variable que con mayor peso ha contribuido en el valor de T^2 .

En [73], indican que esta situación, ha llevado al desarrollo de software que puede producir rápidamente los componentes significativos de la descomposición para conjuntos de variables moderadamente grandes. Sin embargo, la cuestión de cómo estos algoritmos informáticos trabajaran cuando el número de señales a monitorizar sea de cientos de miles, todavía no se ha respondido.

Hay muchos factores que influyen la capacidad de este proceso, como tamaño y velocidad del ordenador, tamaño del conjunto de variables a analizar y la programación de un complejo algoritmo adecuado para esta aplicación.

Hay que tener en cuenta que la eficiente selección y reducción de variables a monitorizar, es un buen camino para optimizar el proceso, maximizando la eficiencia y además reduciendo los costes de medida [74].

Existen diferentes técnicas de reducción de variables, como la propuesta por Jackson [65], que consiste en el análisis de componentes principales (PCA), capaz de reducir la dimensionalidad de las variables en el espacio, generando componentes principales (PC), no correlacionados entre sí [75] o la contribución de [1, 73, 76] que tienen como objetivo reducir la dimensionalidad de los procesos con un gran número de variables a través de métodos de proyección como PLS (Partial Least Square).

Este tipo de técnicas han sido utilizadas en diferentes aplicaciones, como por ejemplo, en el sector marítimo, [77], propusieron reducir los datos de las variables monitorizadas de la maquinaria principal de un buque para la transmisión de los mismos vía satélite. El

algoritmo de compresión basado en PCA, redujo la gran cantidad de datos adquiridos a bordo, disminuyendo el tiempo y el coste de la comunicación con respecto a enviar todos los datos.

Yifei Wang [78], propuso un método basado en el análisis de componentes principales, con el objetivo de identificar un conjunto de variables desde una gran cantidad de variables medidas, y así reducir el número de sensores instalados para la monitorización de la condición. El algoritmo de selección redujo la dimensión de los datos mientras mantenía con bastante exactitud la información, asegurando la condición del sistema. Aparisi [79], estudio los gráficos de componentes principales y su optimización.

Gonzalez y Sanchez [80], presentan un esquema donde la dimensionalidad se reduce pero de forma permanente, esto crea un riesgo de indetectabilidad durante el proceso de ciertas alteraciones que podrían aparecer en las variables que dejaron de ser consideradas.

En [81], se compararon las técnicas de componentes principales (PCA) y Partial Least Squares (PLS) en una base de datos de 252 casos, 17 variables predictivas y 1 variable dependiente, con el objetivo de reducir la dimensionalidad.

Para seleccionar el mejor modelo de regresión, ellos usaron la predicción de suma de cuadrados residuales (PRESS), determinando que el mejor modelo para PCA era de 6 componentes y para la regresión PLS de 7 componentes; por razones de comparación, ambos modelos fueron estimados con 6 componentes, siendo los valores PRESS para cada uno de ellos 88.31 y 266.54 respectivamente. Estos resultados mostraron que la regresión PLS supero a la de PCA.

Estas técnicas de reducción de variables, se pueden combinar con los gráficos de control T^2 de Hotelling para disminuir la limitación que tienen estos cuando el número de variables es elevado; S.Joe Qin [82], analizo el uso de gráficos de control T^2 de Hotelling junto con PCA y otros métodos de detección, identificación y diagnosis de fallos; En [75], propusieron una metodología basada en la combinación de PCA y gráficos de control T^2 de Hotelling capaz de tratar procesos no estacionarios y con múltiples set points. La metodología propuesta la implementaron en una planta termoeléctrica, monitorizando en tiempo real y detectando cambios en las condiciones de operación de dos unidades críticas de la planta, caldera y turbina. En [83], se

utilizaron las técnicas PCA, PLS y T^2 en diferentes equipos de una planta de tratamiento de metanol, con el fin de mejorar la capacidad de decisión del mantenimiento basado en la condición, asegurando mayor confiabilidad y seguridad a la maquinaria.

Finalmente, en la industria del metal [84], se propusieron los gráficos de control T^2 de Hotelling y los gráficos de control de análisis de componentes principales (PCA), para la monitorización y control de un proceso normal multivariado.

Los gráficos de control T^2 de Hotelling, detectaron cuando el proceso tuvo una desviación con respecto a su condición de operación normal, pero no fue capaz de identificar las variables que estaban fuera de rango o la posible tendencia que podían tener las variables en el proceso, mientras que los gráficos de control mediante el uso de PCA, detectaron la desviación en el proceso y además también mostraron la posible tendencia que hacía que el proceso estuviese en una situación de fuera de rango con respecto a su condición normal.

Sin embargo, el principal inconveniente que presentan estas técnicas de reducción de variables, es que tienen dificultades en la detección de la variabilidad cuando la correlación entre las variables que definen el proceso es baja [85], y se hace complicado la identificación de la variable indicativa de fallo al estar las variables originales representadas por componentes principales.

Por lo tanto, cuando la correlación entre variables es baja, como ocurre en procesos térmicos, habrá que buscar otro tipo de técnicas que suplante a las técnicas de reducción de variables comúnmente utilizadas. En [86] propusieron una técnica basada en el análisis de regresión de variables que redujo el número de variables físicas del proceso de combustible de un motor marino, de manera satisfactoria y sin pérdida de información del proceso.

Otra de las características del gráfico de control de Hotelling, es que aunque resulta muy efectivo en la detección de cambios grandes en la media del proceso, al igual que ocurría en los gráficos Shewhart para el caso univariante, pierden sensibilidad ante pequeños cambios, por debajo de 1,5 desviaciones típicas, y progresivos en el vector de medias del proceso [87]. Este inconveniente, presenta una dificultad en la monitorización de máquinas rotativas de baja velocidad, ya que como se indica en [88], en este tipo de máquinas, las fallas pueden desarrollarse lentamente e incluso estar

latentes hasta un punto crítico de su desarrollo, cuando ya es demasiado tarde para actuar preventivamente.

Por lo tanto habrá que recurrir a otro tipo de técnicas que complementen la pobre detección ante este tipo de comportamientos, o al aumento de la potencia del gráfico mediante la inclusión de señales adicionales como se propone en [89].

2.3 Motivación de la tesis

Los sistemas de monitorización actuales implementados en motores marinos, monitorizan individualmente las variables del proceso y generan una alarma cuando su valor excede de un umbral establecido, sin tener en cuenta la condición del motor para cierta operación del buque; esto puede llevar a error en la interpretación del origen de la alarma y consecuentemente en las acciones para corregir la desviación en el proceso. Además, como se desprende de la revisión bibliográfica, la monitorización univariante no tiene en cuenta la interacción entre las diversas variables, dando una diagnosis deficiente del origen de la alarma.

En este contexto, valoramos la idea de implementar una técnica estadística multivariable en un buque Tanque de la Naviera Ibaizabal, con el objetivo de implementarlo como parte del protocolo de mantenimiento predictivo.

Por su probada aplicación en numerosos procesos industriales, se eligieron los gráficos de control T^2 de Hotelling como técnica multivariable. Sin embargo, aunque este tipo de gráficos tienen una gran efectividad ante desviaciones bruscas en el proceso, pierden sensibilidad de detección ante cambios pequeños y progresivos y además tienen dificultades en la identificación de la variable que dio origen al cambio en el proceso, cuando el número de variables a monitorizar es elevado.

Por otro lado, de la revisión bibliográfica, se desprende que los métodos de reducción de variables, como PCA, no pueden realizar la función para la que fueron diseñados, cuando la correlación entre variables es baja, y se hace una tarea difícil la identificación de la variable original, al utilizar componentes principales para representarlas.

Ante esta necesidad de mejora en este tipo de gráficos multivariables, en esta tesis doctoral, proponemos una novedosa metodología para la detección de desviaciones

pequeñas, y bruscas en el proceso, la cual denominaremos “Method for detection of Small and Sudden Deviations” (SSDM), por sus siglas en inglés, aplicable en procesos donde la correlación entre variables sea baja, como es característico en procesos de propulsión marina.

Esta metodología, se aplicara en dos procesos de un motor MAN B&M, instalado como propulsor en un buque tanque, donde primeramente, a través de un análisis de correlaciones, y un análisis de regresión lineal multiple, se reducirá el número de variables a monitorizar, poco correlacionadas entre sí, mejorando así, la limitación que tienen los gráficos T^2 de Hotelling, cuando el conjunto de variables a monitorizar es elevado.

Después, las variables seleccionadas, serán monitorizadas a través de una combinación entre gráficos T^2 de Hotelling y gráficos Cusum univariantes, para detectar las variables, que puedan originar desviaciones tanto pequeñas, como bruscas en el proceso, que no se hubiesen podido detectar con la única aplicación de los gráficos multivariados.

Esta nueva metodología, puede generar un nuevo campo de investigación dentro de la monitorización de equipos de propulsión en el sector marino.

Capítulo 3

Objetivos

3. OBJETIVOS

A partir de la revisión del estado del arte y de haber justificado el interés del presente estudio, se establecen los siguientes objetivos principales:

- Mejora del gráfico T^2 de Hotelling, en la detección de las variables que han provocado un cambio en el proceso, en situaciones cuando el número de variables a monitorizar es elevado, mayor de diez.
- Dotar al gráfico T^2 de Hotelling de capacidad para detectar desviaciones progresivas y de pequeña magnitud que se puedan producir en un proceso.

De estos objetivos generales, se desprenden los siguientes objetivos específicos:

- Elección de un proceso multivariable a monitorizar dentro del sector naval.
- Reducción del número de variables a monitorizar, sin pérdida de información del proceso.
- Desarrollo de una metodología, basada en el estadístico T^2 , capaz de detectar desviaciones tanto de elevada magnitud como de pequeña magnitud.
- Contrastación empírica de la metodología propuesta, mediante la aplicación en el proceso elegido.
- Validación de la metodología desarrollada, comparándola con otra técnica ampliamente testada (T^2 de Hotelling + MYT).
- Desarrollar un software, que realice los algoritmos del método desarrollado, de acuerdo a las necesidades del usuario, con una interfaz amigable y que provea la información que necesita al usuario para la correcta toma de decisiones.
- Implementación de la metodología desarrollada a bordo del buque, para que sirva como complemento de los sistemas de alarmas actualmente instalados o en las oficinas centrales como parte del protocolo de mantenimiento.

Capítulo 4

Materiales y métodos

4. MATERIALES Y MÉTODOS

En este capítulo, se especifica la metodología utilizada a lo largo de la tesis. Primero se detallan los materiales utilizados, después se explica el procedimiento seguido en la metodología, y finalmente se aplica la metodología desarrollada en dos procesos del motor principal; por un lado en el proceso de lubricación de cilindros y por otro lado en el proceso de combustible. Los resultados obtenidos se presentan en el siguiente capítulo.

4.1 Materiales utilizados

Uno de los grandes problemas a los que nos enfrentamos cuando comenzamos este estudio de investigación, fue la disponibilidad de un equipo desde donde obtener muestras reales de funcionamiento.

Ante esta tesitura, inicialmente, se modelizaron las condiciones de funcionamiento de un aerogenerador de General Electric de 1,5 MW, a partir de datos de viento obtenidos de una estación meteorológica.

Tras los primeros resultados parciales, se intentó implementar este algoritmo en un caso real, pero no se consiguió disponer de los datos de funcionamiento de un equipo en servicio, previa solicitud a los diferentes fabricantes de aerogeneradores, por lo que nos vimos obligados a cambiar la máquina objeto de la investigación.

A partir de ese momento orientamos nuestra aplicación al sector naval, gracias a que contamos con la inestimable colaboración de una naviera del entorno.

Con su ayuda, contamos con los datos de funcionamiento de un motor de 2T, que es frecuentemente instalado como motor propulsor en buques tanque y bulkcarriers.

La máquina que seleccionamos para realizar nuestro nuevo y estudio final, fue un motor Diesel MAN B&W 6S70ME-C8, instalado en un buque tanque. Los datos técnicos del motor se presentan en la tabla 4.1.

Tabla 4.1: Datos característicos del motor principal.

Fabricante	MAN B&W
Tipo	6S70ME-C8
Ciclo de trabajo	Low-speed 2 stroke
Velocidad nominal	91 r.p.m.
Potencia nominal	19.620 kW
Numero de cilindros	6
Carrera	2.800 mm
Diámetro	700 mm

El motor estaba instalado en un buque “Suez Max Crude Carrier”, cuyas características se presentan en la tabla 4.2.

Tabla 4.2: Datos característicos del buque.

Nombre del Buque	Confidencial
Astillero de Construcción	Confidencial
Año de construcción	2012
Tipo de buque	Crude Oil Tank
Clase	Suez Max
Eslora (Total)	274,20 m
Manga (de trazado)	48,04 m
Puntal (de trazado)	23,20 m
Calado	17 m
Tonelaje bruto	81,187
Tonelaje neto	51,148

Este buque, normalmente tiene una línea regular, entre el Oeste de África y el Norte de Europa, donde es descargado. Aunque a veces se modifica su rumbo, en función de la demanda del mercado.

4.2 Metodología

En este apartado, se desarrolla una metodología que será capaz de reducir el número de variables a monitorizar del motor MAN B&W 6S70ME-C8, sin pérdida de información del proceso.

Además, también será capaz de detectar desviaciones bruscas, pequeñas y progresivas e identificar las variables que han generado las desviaciones.

Esta metodología, denominada: “Method for detection of Small and Sudden Deviations” (SSDM), mejorará la dificultad que tienen los gráficos T^2 de Hotelling en la identificación de las variables fuera de rango, cuando aumenta el número de variables a monitorizar y mejorará la pobre sensibilidad de este tipo de gráficos en la detección de desviaciones pequeñas y progresivas.

4.2.1 Adquisición de datos

Normalmente, como ya se ha comentado, el buque a estudio, realiza una línea regular entre África y Europa de una duración aproximada de un mes; no obstante en función de la demanda del mercado, su rumbo puede ser cambiado.

La adquisición de datos, se realizó cuando el buque se encontraba en la condición de carga realizando la línea regular mencionada anteriormente. Se valoró, la posibilidad de realizar también el estudio para la condición de lastre, descartándola, debido a la elevada variabilidad que existía en la operatividad del buque y por consiguiente en las condiciones de funcionamiento del motor.

El motor principal dispone de tres equipos de monitorización y adquisición de datos, CoCos EDS, sistema de control de sobrevigilancia y diagnóstico, donde se recogen datos generales del funcionamiento del motor, PMI, Pressure Monitoring Indicator donde se recogen los datos pertenecientes al proceso de combustión, y por otro lado el IAS, Integrated Automation System, donde se recogen los datos pertenecientes al proceso termodinámico.

La monitorización del motor principal a través de los equipos ya mencionados, era permanente. Sin embargo, la adquisición de datos era manual, es decir era el responsable de la guardia el que tenía que dar la orden de adquisición para el almacenamiento de los datos. El hecho de que el almacenamiento de los datos fuera manual y no automático, pudo provocar la existencia de datos perdidos en alguna serie de muestras.

Al cierre de cada viaje, los datos recibidos fueron categorizados y analizados; primero, diferenciando entre condición de carga y condición de lastre del buque, y una vez realizada esta diferencia, se realizó una selección de los datos para unas condiciones de operación normales, analizando la velocidad de avance del buque, deslizamiento, temperatura ambiente, temperatura en sala de máquinas y temperatura del agua de mar.

4.2.2. Datos perdidos

Un problema frecuente en el análisis estadístico es la carencia de parte de los datos [90]. Estos pueden ocurrir por numerosas razones, el operador no recoge el dato, fallo del equipo en el momento de la toma de datos, son algunas de las razones por las que se pueden generar datos faltantes.[68, 91].

El impacto de los datos ausentes es perjudicial no solo por sus potenciales sesgos, sino también por su efecto en el tamaño de la muestra disponible para el análisis[92].

La eliminación o supresión de todos los casos que tengan un valor perdido es un procedimiento frecuentemente utilizado, sin embargo, tiene una serie de inconvenientes, como son, un evidente recorte en la potencia de la prueba con respecto a la potencia pretendida al fijar el tamaño de la muestra original y en segundo lugar se renuncia a observar toda la información disponible desperdiciando la información contenida en la variable que si se ha observado [93, 94].

Existen alternativas a la supresión, como la imputación, que consiste en sustituir los valores perdidos con otros valores estimados a partir de la información disponible. Los trabajos más clásicos plantean algunas soluciones sencillas como la sustitución del valor perdido por la media de la variable calculada con los casos disponibles [95, 96], este procedimiento es conservador, la media de la distribución no cambia; Sin embargo la varianza de la variable se reduce ya que la media está más próxima a si misma que a los valores perdidos que sustituye, lo que hará que las correlaciones con el resto de variables también disminuya [97].

La rapidez de procesamiento que proporciona la informática, ha permitido el desarrollo de otros modelos, que aun siendo menos parsimoniosos mejoran la estimación de los parámetros estadísticos [98]. Así, por ejemplo, el uso de la regresión multivariable [99,

100], donde la variable cuyos valores perdidos se quiere estimar actúa como variable dependiente, mientras que el resto de variables actúa como independientes. Los coeficientes de regresión, que han sido estimados, solo utilizando los datos en los que la variable dependiente no tomaba valores perdidos, se utiliza para estimar esos valores, puesto que las variables que han actuado como independientes sí que son casos completos.

En [68], se aplicó la técnica de regresión multivariable en un proceso químico que estaba siendo monitorizado por 6 variables. Consideraron 21 observaciones en control, simulando un dato perdido en la última observación de la variable X_5 ; Obtuvieron la ecuación de regresión de la variable X_5 en X_1 , X_2 , X_3 , X_4 y X_6 derivadas desde las 20 primeras observaciones y calcularon el valor estimado para la variable X_5 .

El valor estimado fue calculado como 98,58. Este estaba cerca con respecto al valor real de 98,41.

El valor de T^2 , con esta estimación del valor perdido, fue 5,94 también comparado con el valor real de 5,92. Además, sustituyendo el valor perdido tuvo una pequeña influencia en el estadístico T^2 . De manera similar, es insignificante el cambio en la media de X_5 . La media de la observación sin el valor perdido fue 98,21 contra una media de 98,23 cuando el valor estimado fue incluido.

Una comparación de las correlaciones entre X_5 y las otras cinco variables, también con y sin el valor predicho, mostro una pequeña diferencia entre estas.

Este procedimiento de imputación es mucho más razonable que la imputación simple de la media, ya que tienen en cuenta mucha más información [101]. Sin embargo, entre sus limitaciones están el hecho de que el valor estimado es tan bueno como lo sea la ecuación de regresión que lo generó, y el hecho de que las estimaciones pueden afectar a las estimaciones de la varianza [68].

Como se ha comentado en el apartado anterior, aunque los sistemas de monitorización que disponíamos a bordo del buque eran on-line, la adquisición no estaba programada por lo que necesitaba de una orden de grabado por parte del responsable de la guardia. No ocurría lo mismo con los instrumentos controlados por el IAS, donde era el propio operario el que realizaba la lectura y posterior transcripción de los datos manualmente, este procedimiento provoco datos perdidos en algunas de las variables.

Por otro lado, debido a las condiciones de operatividad que lleva intrínsecas un buque, y a los sistemas de adquisición que se disponían a bordo, el tamaño de la muestra no fue lo suficientemente grande como se esperaba.

Con el objetivo de no perder potencia en la prueba y no renunciar a explotar toda la información disponible, para el tratamiento de los datos perdidos se eligió la técnica de regresión.

El software utilizado para generar los modelos de regresión fue el SPSS; La variable cuyos valores se quiso estimar fue la variable dependiente, y las variables que se deseaban incluir en la ecuación de regresión como variables independientes. El método utilizado para generar los modelos fue la regresión por pasos, mediante este método el SPSS, intenta maximizar el ajuste del modelo de regresión utilizando el mínimo número posible de variables.

Para ello se basa en el criterio de significación, esto es, solo se incorporan al modelo de regresión aquellas variables que contribuyen de forma significativa al ajuste del modelo. La contribución individual de una variable al ajuste del modelo se establece contrastando, a partir del coeficiente de correlación parcial, la hipótesis de independencia entre esa variable y la variable dependiente. Para decidir si se mantienen o rechaza esa hipótesis de independencia, se utilizó la probabilidad de F , como criterio de significación.

Con este criterio, una variable pasa a formar parte del modelo de regresión si el nivel crítico asociado a su coeficiente de correlación parcial al contrastar la hipótesis de independencia es menor que 0,05 (probabilidad de entrada), y queda fuera del modelo de regresión si ese nivel crítico es mayor que 0,10 (probabilidad de salida) [102].

Con los modelos generados, se generaron los valores perdidos de cada una de las variables con las que se construyó la base de datos adquiridos.

4.2.3 Elección del tamaño de la base de datos preliminar

Dentro de un proceso multivariable, son muchas las variables que tienen que ser examinadas y muchos los parámetros que necesitan ser estimados. Estos requerimientos resultan en la necesidad de tamaños muestrales grandes.

En cualquier proceso, debido a las diferencias en las desviaciones estándar de cada una de las variables, cada variable individual puede requerir tamaños muestrales estimados diferentes para alcanzar una precisión dada. El objetivo entonces es la combinación de las diferentes estimaciones en un único valor de n .

El uso del estadístico T^2 requiere una muestra, en la que el número de observaciones n exceda el número de variables p . De lo contrario si $p < n$, la inversa de la matriz de covarianza Σ^{-1} o su estimación S^{-1} nunca podrá ser calculada.

Además, cuando el vector de medias y la matriz de covarianza son desconocidas, un gran número de parámetros debe ser estimado. Para proveer estimaciones estables y exactas de estos parámetros, n debe de ser lo suficientemente grande. Por lo tanto, el número mínimo de muestras tiene que cumplir la siguiente ecuación [103]:

$$\text{numero de muestras} = 2p + \frac{p(p-1)}{2} \quad (4.1)$$

4.2.4 Selección de variables

La selección de variables en un control estadístico del proceso, SPC, por sus siglas en inglés, siempre debe seguir unos criterios técnicos según la operación del mismo, por lo que ser conocedor del proceso nos ayudara en gran medida a realizar una selección de variables correcta y sin pérdida de información del proceso [80].

Sin embargo, las propiedades estadísticas de los datos, pueden ayudarnos en esta tarea. Como ya se ha comentado en el estado del arte, el método más popular de reducción de la dimensión en un conjunto de variables es el análisis de componentes principales (PCA). Este tipo de técnicas tienen una serie de desventajas como dificultad en la detección de la variabilidad cuando la correlación entre variables que definen el proceso es baja [85] y que al estar las variables originales representadas por componentes principales, existe el problema de interpretación de los resultados cuando no es fácil encontrar el significado de dichas componentes.

Nosotros proponemos un método para seleccionar un reducido número de variables en procesos multivariantes, basado en el análisis de correlación de Pearson y el análisis de regresión lineal multivariable, que complemente la información de los grandes

concedores del proceso como son el fabricante del motor, los inspectores de la naviera y el Jefe de Máquinas del buque.

El coeficiente de correlación de Pearson [104] es quizá el mejor coeficiente y el más utilizado para estudiar el grado de relación lineal existente entre dos variables cuantitativas, se suele representar por “ r ” y sigue la siguiente expresión:

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (4.2)$$

Donde σ_{xy} es la covarianza de (x,y), σ_x es la desviación típica de x y σ_y es la desviación típica de y.

Los valores que puede tomar el coeficiente de correlación “ r ” son $-1 < r < 1$. Si $r > 0$, la correlación es positiva, si sube el valor de una variable sube el de la otra. Por el contrario, si $r < 0$, la correlación es negativa y cuando el valor de una variable sube la otra baja. La correlación es tanto más fuerte cuanto más se aproxime a los valores extremos, -1 ó $+1$, no existiendo relación entre variables cuando $r = 0$.

En estadística el coeficiente de determinación R^2 , es un estadístico usado en el contexto de un modelo estadístico cuyo principal propósito es predecir futuros resultados. El coeficiente determina la calidad del modelo para replicar los resultados y la proporción de variación de los resultados que puedan explicarse por el modelo.

Hay varias definiciones diferentes para el coeficiente de determinación, que son algunas veces equivalentes, las más comunes se refieren a la regresión lineal. En este caso, el R^2 es el cuadrado del coeficiente de correlación de Pearson, lo cual es solo cierto para la regresión lineal simple.

Además, de estudiar la relación entre pares de variables, podemos estudiar la relación entre una variable y todas las demás.

La ecuación de regresión lineal simple, se convertirá entonces en una ecuación de regresión lineal múltiple, que nos permitirá hacer una predicción entre la variable dependiente y las variables explicativas.

La valoración de la capacidad predictiva de la ecuación de regresión múltiple se puede realizar a través del coeficiente de determinación R^2 o coeficiente de correlación múltiple al cuadrado, que es una medida descriptiva que sirve para evaluar la bondad de

ajuste del modelo, es decir, da la proporción de variación de la variable dependiente que es explicada por las variables explicativas o predictoras.

El coeficiente de determinación R^2 se define como el cociente entre la variabilidad explicada por la regresión y la variabilidad total, esto es:

$$R^2 = \frac{\text{VARIANZA EXPLICADA}}{\text{VARIANZA TOTAL}} = 1 - \frac{\text{VARIANZA NO EXPLICADA}}{\text{VARIANZA TOTAL}} \quad (4.3)$$

El valor de R^2 , puede tomar valores entre 0 y 1. Cuanto más se acerque a 1 mayor será la capacidad predictiva del modelo. Si llegara a ser igual a uno, entonces las variables predictoras explicarían toda la variación de la variable dependiente, mientras que si llegara a ser igual a cero, significaría que las variables predictoras tendrían capacidad nula de predecir la variación de la variable dependiente.

Como alternativa al uso de R^2 como medida de idoneidad de un modelo, a veces, se utiliza el coeficiente de determinación múltiple ajustado, denotado por R_{ajust}^2 este está dado por:

$$R_{ajust}^2 = R^2 - \left[\frac{p(1 - R^2)}{(n - p - 1)} \right] \quad (4.4)$$

Donde p , es el número de variables independientes y n , es el tamaño de la muestra.

Se observa que R_{ajust}^2 toma en cuenta tanto el tamaño de la muestra n como el número de parámetros del modelo. R_{ajust}^2 siempre es menor que R^2 y lo que es más importante, no puede forzarse hacia uno con solo agregar más y más variables independientes al modelo. [105].

Con frecuencia es importante medir la correlación entre una variable dependiente y una variable independiente en particular, cuando las demás variables involucradas se mantienen constantes, es decir, cuando se elimina el efecto de todas las otras variables.

Esto puede obtenerse a través del coeficiente de correlación parcial; por ejemplo, para el caso concreto de una variable dependiente X_1 y dos variables independientes X_2 y X_3 , si se denota $r_{12.3}$ el coeficiente de correlación parcial entre X_1 y X_2 , manteniendo X_3 constante, se encuentra que:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} \quad (4.5)$$

Mediante, la utilización de las técnicas expuestas en este apartado, se intentara reducir el número de variables monitorizadas en cada proceso. Como es lógico, el nivel de reducción al que se puede llegar mediante la técnica desarrollada, estará supeditado al nivel de complejidad intrínseco en cada uno de los procesos.

4.2.5 Proceso de purgado de outliers

Los outliers son aquellas observaciones para los que una, dos o múltiples variables toman valores extremos que los hace diferir del comportamiento del resto de la muestra y permiten al investigador sospechar que han sido generados por mecanismos distintos al resto [106]. Barnett y Lewis [107] indican que una observación atípica u outlier, es una observación que se desvía notablemente de otras observaciones de la muestra en la que se produce, de manera similar, Johnson [108], define un outlier como una observación en un conjunto de datos que parece ser incongruente con el resto de observaciones de ese conjunto de datos.

Las causas que generan la existencia de valores atípicos en un proceso pueden ser diversas, errores de medición o transcripción, cambios en el instrumento de medición o a heterogeneidad intrínseca de los elementos observados [85].

La inclusión de datos atípicos en un conjunto de datos distorsiona los resultados, oscureciendo el patrón de comportamiento del resto de observaciones, llegando a conclusiones que sin ellos serían completamente distintas. En definitiva, la razón por la que se busca detectar y eliminar las observaciones atípicas de la base de datos histórica, HDS, por sus siglas en inglés, es porque su inclusión puede producir estimaciones erróneas en el vector de medias y en la matriz de covarianza, dando un control del proceso impreciso [68].

Una vez detectado un outlier, hay que decidir qué hacer con él, eliminarlo o no. Cuando el valor atípico es un error evidente, nadie discrepa sobre su eliminación. Sin embargo el debate es mucho más intenso cuando se trata de qué hacer con un valor atípico legítimo [91]. Algunos autores, creen que la eliminación es el último recurso y que se puede intentar suavizar su influencia transformando las variables mediante raíces

cuadradas o logaritmos [109]. Otros, consideran que la mejor alternativa es su eliminación para asegurar que las estimaciones son correctas para la mayoría de la población [110].

Planteada la importancia que pueden tener los outliers para el control del proceso, toca elegir la técnica de detección. Esta detección se puede producir desde una perspectiva univariante, bivalente o multivalente.

Hay que tener en cuenta que un caso puede no tomar valores atípicos en dos variables consideradas individualmente, pero si hacerlo si se consideran conjuntamente. Un individuo de 12 años puede ser un elemento muestral lógico de nuestra investigación, un sujeto con una carrera universitaria también, pero un sujeto de 12 años con una carrera universitaria será, casi con toda seguridad, un caso atípico [91]. La detección de outliers desde una perspectiva univariante, ignora la contribución de las correlaciones entre variables.

La perspectiva univariante de identificación de casos atípicos examina la distribución de observaciones, seleccionando como casos atípicos aquellos casos cuyo valor estandarizado de la variable analizada caigan fuera de los rangos de la distribución. Por lo tanto, para establecer si un valor determinado es atípico es necesario determinar el umbral que debe superar, que normalmente se expresa en número de desviaciones estándar. En [92], sugieren para muestras pequeñas, menores a 80 observaciones, identificar como atípicos, aquellas observaciones con valores estándar de 2,5 ó superiores, y valores de 3 ó 4 desviaciones estándar para muestras mayores.

Un procedimiento también utilizado por algunos autores [111], para la detección univariante de observaciones atípicas es el test Grubbs [112, 113].

Además de la evaluación univariante, pueden evaluarse conjuntamente pares de variables mediante un gráfico de dispersión. Los casos que caigan manifiestamente fuera del rango del resto de las observaciones pueden identificarse como puntos aislados en el gráfico de dispersión [92].

No es tan problemático que un caso sea atípico respecto a una variable sino respecto al conjunto de las que se incorporan al análisis. Dado que la mayoría de los análisis multivariantes implican más de dos variables, se necesita un procedimiento que

contemple simultáneamente todas las variables para determinar si algún caso tiene un comportamiento anómalo [91].

Un procedimiento consiste en calcular la distancia de cada observación al centroide del conjunto de datos. Cuanto más lejos este una observación de la media, más probable es que sea un caso atípico. Harold Hotelling [6], propuso un método basado en la distancia de Mahalanobis, [7], utilizando el vector de medias y la matriz de covarianza de variables aleatorias con distribución normal multivariada, para realizar el gráfico de control T^2 de Hotelling.

El método está formado por dos fases, en la fase I, se define la base histórica de datos o distribución de referencia y en la fase II, se evalúa el comportamiento del proceso con respecto a la distribución de referencia.

Se considera dentro de la fase I, el proceso de purgado donde un único vector de la base de datos preliminar $X^t=(x_1, x_2, \dots, x_p)$ está siendo monitorizado para el control del proceso utilizando la carta de control T^2 de Hotelling, según la ecuación (4.6). Las observaciones de la base de datos preliminar siguen una distribución normal multivariable con un vector de medias μ y desviación típica σ estimadas para el proceso multivariable.

$$T^2 = (X_i - \bar{X})^t \cdot S^{-1} \cdot (X_i - \bar{X}) \tag{4.6}$$

Dónde:

$$X_i = (X_{i1}; X_{i2}; \dots; X_{ip})^t$$

\bar{X} = Vector de medias muestrales

S^{-1} = Inversa de la matriz de covarianza.

Dependiendo de las circunstancias, el estadístico T^2 puede ser descrito por tres funciones diferentes de probabilidad: distribución Beta, F y chi-cuadrado.

Cuando (μ, σ) son estimadas, se utiliza la distribución Beta en el proceso de purgado multivariable de la fase I, mientras que la distribución F , se utiliza en el desarrollo del proceso de control de la fase II. [68].

El proceso de purgado, comienza, realizando una primera detección de los puntos atípicos a través de la base de datos preliminar.

Para el cálculo del LCS (Limite de control Superior), se utilizó la distribución β de $\alpha=0,05$, para errores tipo I [114]. El nivel de α , puede ser variable, haciendo más o menos estricto el método. La elección del valor de α , se corresponde con el que se utiliza en los procesos industriales y depende de las condiciones de operación del buque.

Una vez elegido un valor de α para el proceso de purgado, existe la posibilidad de cometer un error tipo I, probabilidad de equivocarnos al rechazar la hipótesis nula [91], esto es, si una observación es declarada como señal cuando en realidad no lo es. Por lo que observaciones correctas serán excluidas del HDS.

Para el valor de α elegido, todos los vectores observación cuyos valores son menores o iguales al límite de control superior (LCS) permanecen en la base de datos preliminar. El límite de control superior se determina mediante la siguiente expresión:

$$LCS = \left\{ \frac{(n - 1)^2}{n} \right\} \beta_{\left\{ \alpha; \frac{p}{2}; (n-p-1)/2 \right\}} \quad (4.7)$$

Dónde: n , es el número de muestras, p , es el número de variables y $\beta_{\left\{ \alpha; p/2; (n - p - 1)/2 \right\}}$, es el ath , quantile de la distribución beta, $\beta_{\left\{ p/2; (n - p - 1)/2 \right\}}$

Si el valor de T^2 , que se monitorizo para una observación, excedía del LCS, la observación era eliminada de la base de datos preliminar.

Con las restantes observaciones, se calculan nuevas estimaciones del vector de medias y matriz de covarianza. Otra vez, se detectan y eliminan outliers, y se repite el proceso, hasta conseguir un set de observaciones homogéneo.

Finalmente, el set final de datos será el (HDS), del modo de operación normal del proceso.

4.2.6 Normalidad multivariante

Uno de los supuestos básicos del análisis multivariante es la normalidad de los datos, en referencia al perfil de la distribución de los datos para una única variable y su

correspondencia con una distribución normal, punto de referencia de los métodos estadísticos.

La normalidad multivariante implica que las variables individuales sean normales en un sentido univariante y que sus combinaciones también sean normales. Por lo tanto, si una variable es normal multivariante, es también univariante.

Sin embargo, dos o más variables normales univariantes no son necesariamente normal multivariante. Por tanto, una situación en la que todas las variables exhiben normalidad univariante ayudara a obtener normalidad multivariante pero no lo garantiza [115].

La técnica utilizada para analizar la normalidad multivariante de los datos en la presente tesis es a través de los gráficos Q-Q Plot, siguiendo la distribución beta [116].

Siguiendo la distribución β , se acepta la hipótesis de normalidad de los datos, siempre que los puntos observados, se encuentren dentro del valor Beta esperado, es decir, siempre que los puntos en el gráfico tengan un comportamiento suficientemente rectilíneo [117].

4.2.7 Control del proceso

En esta fase, se comprueba si una nueva entrada de datos, genera cambios bruscos en el proceso, con respecto a la base de datos de histórica (HDS), previamente calculada.

Primero se calculan los valores T^2 , para la nueva entrada de datos, siguiendo la siguiente expresión (4).

$$T^2 = (X_i - \bar{X})^t \cdot S^{-1} \cdot (X_i - \bar{X}) \quad (4.8)$$

Dónde:

\bar{X} y S^{-1} , son el vector de las medias muestrales y la inversa de la matriz de covarianza, obtenidas desde el HDS y X_i , la nueva entrada de datos.

$$X_i = (X_{i1}; X_{i2}; \dots; X_{iP})^t$$

Para el cálculo del LCS (Limite de control Superior), se utilizó la distribución F de $\alpha=0,05$, para errores tipo II [118]. El nivel de α elegido es el que se utiliza normalmente en procesos industriales y puede ser variable, siendo más o menos estricto el método.

El LCS, se determinó por la siguiente expresión (5):

$$LCS = \left\{ \frac{p(n + 1)(n - 1)}{n(n - p)} \right\} F_{\{\alpha;p;(n-p)\}} \tag{4.9}$$

Dónde p , es el número de variables, n , el tamaño del HDS y $F\{\alpha; p; (n - p)\}$, es el α th, quantile de $F\{p; (n - p)\}$.

Los valores de T^2 que excedieron del LCS, fueron declarados como señales y ello concluyo que la observación estaba fuera de rango con respecto al funcionamiento de operación normal del proceso.

4.2.8 Identificación de la variable indicativa de fallo

Una vez que el estadístico T^2 detecto muestras fuera de rango en el proceso desde las condiciones normales de operación, se utilizó la descomposición MYT [67, 119, 120], para identificar las variables con más peso, responsables del estado fuera de rango para cada muestra.

El estadístico T^2 para un vector de observaciones p -dimensional $(x)^t = (x_1, x_2, \dots, x_p)$ puede representarse como:

$$T^2 = (x - \bar{x})^t \cdot S^{-1} \cdot (x - \bar{x}) \tag{4.10}$$

Si x_p es una variable de interés particular, el vector centrado se puede particionar en la forma:

$$(x - \bar{x})^t = [(x^{(p-1)} - \bar{x}^{(p-1)}), (x_p - \bar{x}_p)] \tag{4.11}$$

Dónde, $x^{(p-1)} = (x_1, x_2, \dots, x_{p-1})$ representa el vector de las primeras $(p-1)$ variables y $\bar{x}^{(p-1)}$ el vector de medias correspondiente. Análogamente, se obtiene una partición similar en la matriz S de varianzas y covarianzas:

$$S = \begin{pmatrix} S_{XX} & S_{xX} \\ S'_{xX} & S_p^2 \end{pmatrix} \quad (4.12)$$

Siendo S_{XX} la matriz de covarianzas de las primeras $(p-1)$ variables, S_p^2 la varianza de la variable x_p y S_{xX} el vector de dimensión $(p-1)$ que contiene las covarianzas entre x_p y las $(p-1)$ variables. El estadístico T^2 puede descomponerse como un agregado de dos sumandos independientes:

$$T^2 = T_{p-1}^2 + T_{P,1,2,\dots,P-1}^2 \quad (4.13)$$

Dónde:

$T_{p-1}^2 = (\mathbf{X}^{(p-1)} - \bar{\mathbf{X}}^{(p-1)})^t S_{XX}^{-1} (\mathbf{X}^{(p-1)} - \bar{\mathbf{X}}^{(p-1)})$, es el estadístico T^2 definido para las primeras $(p-1)$ variables, y el segundo sumando es el cuadrado del valor de la variable x_p ajustado por los estimadores de la media y la desviación típica en la distribución condicional $x_p/(x_1 \dots x_{p-1})$. Es decir:

$$T_{P,1,2,\dots,P-1}^2 = \frac{(x_p - \bar{x}_{P,1,2,\dots,P-1})^2}{S_{P,1,2,\dots,P-1}^2} \quad (4.14)$$

Siendo:

$\bar{x}_{P,1,2,\dots,P-1} = \bar{x}_p + b_p'(\mathbf{X}^{(p-1)} - \bar{\mathbf{X}}^{(p-1)})$ y, $b_p = S_{XX}^{-1}S_{xX}$ es un vector de dimensión $(p-1)$ que contiene los coeficientes de la regresión de x_p en función de las $(p-1)$ variables restantes. La estimación de la varianza condicional por su parte queda:

$$S_{p,1,2,\dots,p-1}^2 = S_p^2 - S'_{xX}S_{XX}^{-1}S_{xX} \quad (4.15)$$

Como T_{p-1}^2 es también un estadístico del tipo T^2 , se puede descomponer a su vez en la forma:

$$T_{p-1}^2 = T_{p-2}^2 + T_{P-1,1,2,\dots,P-2}^2 \quad (4.16)$$

Continuando la partición y las iteraciones de esta manera, se obtiene descomposición general para “p” variables del estadístico T^2 de Hotelling:

$$T^2 = T_1^2 + T_{2.1}^2 + T_{3.1,2}^2 + T_{4.1,2,3}^2 + \dots + T_{P.1,\dots,P-1}^2 = T_1^2 \sum_{j=1}^{P-1} T_{j+1.1, \dots, j}^2 \quad (4.17)$$

El término T_1^2 , es el estadístico de Hotelling para la primera variable. Esto se reduce al cuadrado del estadístico t univariado para la variable inicial:

$$T_1^2 = \frac{(X_{i1} - \bar{X}_1)^2}{S_1^2} \quad (4.18)$$

4.2.9 Control de las variables predictivas

En esta última fase del método, queremos comprobar si existen cambios pequeños y progresivos en el proceso, que no hayan sido detectados por medio del estadístico T^2 de Hotelling. Para ello, monitorizaremos independientemente las variables predictivas de la variable indicativa de fallo.

Una de las mayores desventajas de los gráficos de Shewhart, para el análisis univariante, es que solo utiliza la información del proceso de las observaciones monitorizadas, e ignora, la información de la secuencia de observaciones [11]. Esta característica de falta de memoria, hace que no sean muy efectivos en la detección de cambios pequeños en el proceso, del orden de 1,5 desviaciones típicas.

Una alternativa, es el uso del gráfico de suma acumulada Cusum [4], el cual utilizamos para monitorizar cada una de las variables predictivas, de la variable indicativa de fallo y detectar posibles cambios pequeños y progresivos en el proceso.

Como se comentó en el capítulo dos, existen dos tipos de Cusum: Cusum tabular y V-Mask. Para la representación de los límites de control se eligió el Cusum tabular, por sus ventajas con respecto al V-Mask.

El Cusum tabular, acumula las desviaciones de cada valor sobre el valor objetivo μ_0 , con un estadístico C^+ y acumula las desviaciones de cada valor por debajo del valor objetivo μ_0 , con un estadístico C^- . Estos se definen de la siguiente forma:

$$C_i^+ = \max[0, X_i - (\mu_0 + K) + C_{i-1}^+] \quad (4.19)$$

$$C_i^- = \max[0, (\mu_0 - K) - X_i + C_{i-1}^-] \quad (4.20)$$

Dónde: $C_0^+ = C_0^- = 0$

El valor de K , llamado valor de referencia, suele ser la mitad entre el valor objetivo μ_0 y el valor medio μ_1 para el que queremos detectar que el proceso se ha desajustado, considerando para este un 15% sobre μ_0 , valor inferior al umbral establecido por el fabricante. Por lo tanto,

$$K = \frac{(\mu_1 - \mu_0)}{2} \tag{4.21}$$

Los valores de las sumas acumuladas C_i^+ y C_i^- , se comparan con el intervalo de decisión H , donde:

$$H = h \cdot \sigma \tag{4.22}$$

Dónde: σ , es la desviación típica del proceso en control y $h = 5$, valor utilizado normalmente en los procesos industriales. Si C_i^+ y C_i^- , excedían del intervalo de decisión, se consideró que la variable estaba fuera de rango.

Para el cálculo del ARL, utilizamos la aproximación dada por Siegmund [29]. Para cada uno de los estadísticos, C_i^+ o C_i^- con parámetros h y k , la aproximación de Siegmund es:

$$ARL = \frac{(-2\Delta b) + 2\Delta b - 1}{2\Delta^2} \tag{4.23}$$

Para $\Delta \neq 0$, donde $\Delta = \delta^* - k$ para el lado positivo C_i^+ , $\Delta = -\delta^* - k$ para el lado negativo C_i^- , $b = h + 1.166$, y $\delta^* = (\mu_1 - \mu_0) / \sigma$

La cantidad de δ^* , representa el cambio en la media, en unidades de σ , para el que el ARL se calcula. Por lo tanto, si $\delta^* = 0$, se considera un ARL_0 de 48, correspondientes a las muestras obtenidas durante 1 día. Mientras que si $\delta^* \neq 0$, se calculara el ARL_1 , correspondiente al cambio de tamaño δ^* .

Para obtener el ARL de los dos estadísticos del Cusum, ARL^+ y ARL^- , se utilizó la siguiente expresión:

$$\frac{1}{ARL} = \frac{1}{ARL^+} = \frac{1}{ARL^-} \tag{4.24}$$

4.3 Aplicación práctica

Este trabajo quedaría incompleto si no se llevara a cabo una aplicación práctica donde quedara de manifiesto la utilidad de la metodología desarrollada.

Aunque todos los procesos involucrados en el funcionamiento de un motor diésel son de vital importancia, la metodología propuesta la vamos a aplicar en dos procesos del motor principal: lubricación de cilindros y combustible.

En ambos procesos, se hace necesario un enfoque multivariable, por la cantidad de variables a monitorizar de las que dependen y la baja correlación que existe entre algunas de ellas, lo hacen idóneo para constatar las propiedades de la metodología desarrollada.

4.3.1 Proceso de Lubricación de cilindros

4.3.1.1 Descripción del proceso

El sistema de lubricación de cilindros denominado MAN B&M Alpha Cylinder Lubrication System, está diseñado, para inyectar el aceite de cilindros intermitentemente con control electrónico del avance y dosificación en una posición definida.

Este sistema es independiente de los demás circuitos de lubricación del motor (Cigüeñal, árbol de levas, turbocompresor...)[121].

El aceite de lubricación de cilindros es bombeado desde el tanque almacén hasta el tanque de servicio. El aceite de lubricación de cilindros alimenta al “Alpha Cylinder Lubrication System” por gravedad a una altura de unos 3 metros sobre el motor principal. El aceite de alimentación de los inyectores responsables para la inyección del aceite dentro de los cilindros, es presurizado por medio del Alpha Lubricator instalado en el HCU, (Hydraulic Cylinder Unit) y equipado con pequeñas bombas multi pistón; estas bombas son movidas por aceite de potencia a una presión de 300 Bar; este sistema se muestra en la Figura 4.1.

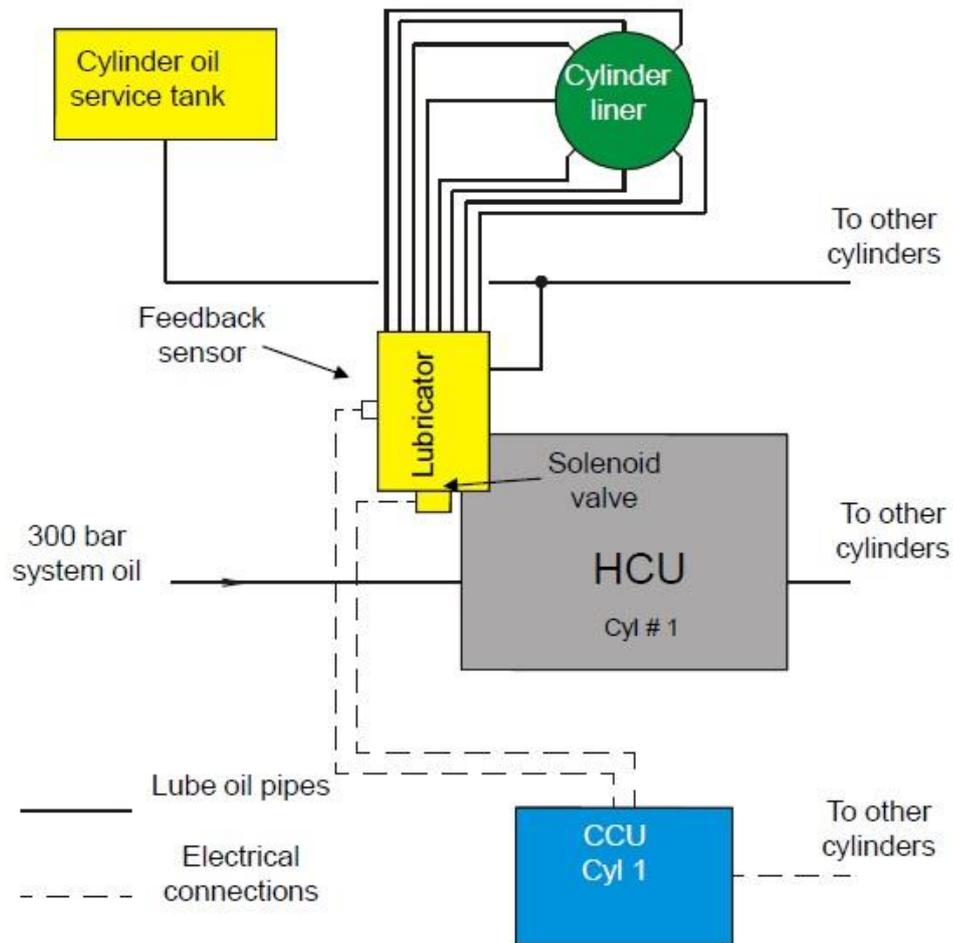


Figura 4.1: Sistema de lubricación de cilindros. [121]

El sistema completo es controlado por el Cylinder Control Unit (CCU), el cual controla la frecuencia de inyección de aceite en función de la velocidad del motor y la carga del motor (fuel index).

Las necesidades de aceite en los cilindros varían con las condiciones de operación, tales como la carga y la calidad del fuel. Consecuentemente, para mejorar el rendimiento de la óptima lubricación, tanto desde el punto de vista económico como técnico, la dosificación de aceite de lubricación de cilindros, debería seguir por lo tanto estas variaciones.

La dosificación óptima de aceite inyectada en los cilindros, es controlada por el ACC (Alpha Adaptive Cylinder Oil Control) [122, 123], en función de las necesidades del sistema, siguiendo los siguientes criterios:

- La dosificación de aceite, debe ser proporcional a la cantidad de azufre contenida en el combustible.
- La dosificación de aceite debe ser proporcional a la carga del motor.

La aplicación de estos dos criterios, permite una dosificación de aceite de cilindros óptima, proporcional a la cantidad de azufre existente en los cilindros.

Este principio se basa en que el principal desgaste en las camisas de los cilindros, provienen de naturaleza corrosiva, por lo que la cantidad de componentes alcalinos necesarios en el cilindro deberían ser proporcionales a la cantidad de azufre dentro de los cilindros.

Esta dosificación además de contrarrestar los efectos del azufre, también satisface los otros requerimientos básicos de la lubricación, como, proveer una adecuada película de aceite y propiedades detergentes.

4.3.1.2 Selección de variables

Como se ha comentado en el estado del arte, uno de los principales inconvenientes que tienen los gráficos de T^2 Hotelling en la monitorización de procesos, es que su rendimiento disminuye a medida que aumenta el número de variables monitorizadas [124].

Para mejorar esta limitación, se monitorizo el proceso de lubricación de cilindros con el mínimo número de variables, sin pérdida de información de las variables no seleccionadas.

Inicialmente el proceso de lubricación de cilindros estaba constituido por 10 variables, que se definen a continuación:

- 1- Engine Load (%), medida en el control local.
- 2- Fuel Index (%), medida en el control local.
- 3- Fuel Plunge Stroke C1 (mA), se refiere a la apertura del inyector.
- 4- Hydraulic oil after pump 1 (Bar), perteneciente al hydraulic power pack y accionado a través del motor principal.

- 5- Total oil (Kg), aceite de lubricación de cilindros consumido por el motor diariamente.
- 6- Scavenge air received pressure (Bar).
- 7- Scavenge air received temperature (°C), medidas en el intercambiador de aire de carga.
- 8- Exhaust gas temperature turbine inlet (°C), Temperatura a la entrada del turbo compresor.
- 9- Exhaust gas receiver pressure (Bar), medida en el colector principal de los gases de escape.
- 10- Maximum Pressure, Pmax (Bar) medida en la cámara de combustión.

Sin embargo, el objetivo fue monitorizar únicamente las variables representativas del proceso. Para ello, con muestras suministradas por la Naviera, adquiridas siguiendo unas condiciones de operación normal, velocidad de avance del buque de entre 14 y 15 nudos con un deslizamiento inferior al 17%, temperatura media del agua de mar de 20 °C, temperatura media ambiente de 30 °C y temperatura media de la sala de máquinas de 37 °C, se llevó a cabo un análisis de regresión multivariable entre las variables que constituían el proceso de lubricación de cilindros. Con este análisis, se consiguió comprobar, que variables dependían de quien y así poder seleccionar las más representativas del proceso.

De los modelos generados tras realizar la regresión entre variables, se eligieron aquellos que tuvieran un coeficiente de determinación $R^2 > 0,8$, que asegurara un buen ajuste del modelo y se descartaron aquellos modelos que aun teniendo un ajuste superior al modelo seleccionado, las variables explicativas añadidas no contribuían de forma importante a explicar la proporción de variación de la variable dependiente.

Los modelos seleccionados, después de realizar el análisis de regresión, se presentan en las tablas 4.3 a 4.12:

Tabla 4.3: Variable dependiente: Engine Load.

Modelo 1	Coefficientes	R ²	R ² ajustado	Correlación Pearson	Correlación Parcial
Cte/. modelo	-63,045	0,975	0,974		
Maximum Pressure, Pmax	0,851			0,987	0,987

Tabla 4.4: Variable dependiente: Fuel Index.

Modelo 1	Coefficientes	R ²	R ² ajustado	Correlación Pearson	Correlación Parcial
Cte/. modelo	21,91	0,97	0,968		
Engine Load	0,725			0,985	0,985

Tabla 4.5: Variable dependiente: Fuel Plunge Stroke.

Modelo 1	Coefficientes	R ²	R ² ajustado	Correlación Pearson	Correlación Parcial
Cte/. modelo	1,165	0,971	0,968		
Engine Load	0,024			0,979	0,983
Scavenge air T ^a	0,007			0,376	0,546

Tabla 4.6: Variable dependiente: Hydraulic oil Pressure after pump P1.

Modelo 1	Coefficientes	R ²	R ² ajustado	Correlación Pearson	Correlación Parcial
Cte/. modelo	200,34	0,821	0,812		
Maximum Pressure, Pmax	0,292			0,906	0,906

Tabla 4.7: Variable dependiente: Total Oil.

Modelo 1	Coefficientes	R ²	R ² ajustado	Correlación Pearson	Correlación Parcial
Cte/. modelo	95,313	0,473	0,446		
Exhaust gas Pressure	43,108			0,687	0,687

Tabla 4.8: Variable dependiente: Scavenge Air receiver Pressure.

Modelo 1	Coefficientes	R ²	R ² ajustado	Correlación Pearson	Correlación Parcial
Cte/. modelo	-2,276	0,971	0,969		
Maximum Pressure, Pmax	0,027			0,985	0,985

Tabla 4.9: Variable dependiente: Scavenge Air receiver Temperature.

Modelo 1	Coefficientes	R ²	R ² ajustado	Correlación Pearson	Correlación Parcial
Cte/. modelo	-44,033	0,418	0,357		
Exhaust gas T ^a	0,241			0,494	0,646
Total Oil	-0,038			-0,027	-0,479

Tabla 4.10: Variable dependiente: Exhaust gas temperatura.

Modelo 1	Coefficientes	R ²	R ² ajustado	Correlación Pearson	Correlación Parcial
Cte/. modelo	246,528	0,67	0,653		
FuelPlunge Stroke C1	48,842			0,818	0,818

Tabla 4.11: Variable dependiente: Exhaust gas pressure.

Modelo 1	Coefficientes	R ²	R ² ajustado	Correlación Pearson	Correlación Parcial
Cte/. modelo	-2,106	0,93	0,926		
Engine Load	0,095			0,964	0,964

Tabla 4.12: Variable dependiente: Maximum Pressure, Pmax.

Modelo 1	Coefficientes	R ²	R ² ajustado	Correlación Pearson	Correlación Parcial
Cte/. modelo	75,033	0,975	0,974		
Engine Load	1,145			0,987	0,987

Posteriormente, se representó de manera más amigable, las variables explicativas junto con sus correspondientes variables dependientes para cada modelo seleccionado, representado en la tabla 4.13.

Tabla 4.13: Representación de las variables Dependientes / Explicativas.

		Variables explicativas												
		Número de identificación de variables												
		1	2	3	4	5	6	7	8	9	10			
Variables dependientes	Número de identificación de variables	1												
		2												
		3												
		4												
		5												
		6												
		7												
		8												
		9												
		10												

Después, se eligió el mínimo número de variables, teniendo en cuenta que las variables no seleccionadas, tenían que tener una correlación igual o superior a 0,49, con alguna de las variables seleccionadas. Las opiniones críticas de la casa fabricante, inspectores de

la naviera y jefe de máquinas del buque, también se tuvieron en cuenta para la selección de la variables.

Finalmente, el proceso quedo monitorizado, por las variables cuya numeración esta resaltada en el cuadro superior, que corresponden a las siguientes 7 variables representativas del proceso:

- 1- Engine Load (%).
- 2- Fuel Index (%).
- 3- Fuel Plunge Stroke C1 (mA).
- 4- Hydraulic oil after pump 1 (Bar).
- 5- Total oil (Kg).
- 6- Exhaust gas receiver pressure (Bar).
- 7- Maximum Pressure, Pmax (Bar).

A partir de aquí, se utilizara la nueva numeración dada a las variables seleccionadas.

4.3.1.3 Adquisición de datos

Una vez, se conocieron las variables que fueron tomadas en cuenta en el proceso, la adquisición de datos, se llevó a cabo en la condición de carga, siguiendo las siguientes condiciones de operación normal del buque mencionadas anteriormente.

Bajo estas condiciones de operación, se adquirió 1 muestra al día, de las siete variables seleccionadas, durante dos viajes en condición de carga.

El primer viaje tuvo lugar desde Cabo Lopez (Gabón) el 03 de noviembre de 2015, donde el buque fue cargado, y finalizo, en el puerto de Rotterdam (Países bajos) el 17 de noviembre de 2015, donde fue descargado.

El segundo viaje se realizó desde la terminal de Odudu (Nigeria), el 08 de diciembre de 2015 a Rotterdam (Países bajos) el 19 de diciembre de 2015. Obteniendo un total de 21 muestras normales, válidas según los criterios mencionados anteriormente, representadas en la tabla 4.14.

Tabla 4.14: Base de datos adquirida.

		Número de identificación de variables						
		1	2	3	4	5	6	7
Número de muestra	1	56	64,7	2,9	251,9	240,3	1,75	140,3
	2	56	63,7	2,9	249,6	249,2	1,75	140,1
	3	55	63,8	2,9	251,9	240,3	1,7	138,6
	4	59	67,3	2,9	250,5	249,2	1,7	139,2
	5	51	59,6	2,7	250,9	240,3	1,7	139,1
	6	60	70,1	3	253,4	489,5	1,7	139,6
	7	57	63,6	2,9	253,7	249,2	1,7	140,1
	8	59	67,5	3	253,5	258,1	1,7	139,2
	9	56	62,7	2,9	251,5	249,2	1,7	140
	10	54	63,4	3	252,3	267	1,7	140,7
	11	63	69,4	3,1	253	231,4	1,7	137,9
	12	56	64,3	2,9	251,1	240,3	1,7	143
	13	57	64,5	2,9	250,7	240,3	1,7	141
	14	59	65	2,9	249,4	249,2	1,65	141
	15	57	63,9	3	256	249,2	1,7	142
	16	56	63,7	2,9	250,6	258,1	1,7	141
	17	58	61,7	3,1	254,2	249,2	1,65	144
	18	55	63,2	2,8	252,4	258,1	1,68	141
	19	62	68,5	2,9	253,4	249,2	1,68	142
	20	56	62,9	2,8	251,1	507,3	1,68	141
	21	55	65,3	2,7	250,3	249,2	1,68	137

Después de analizar las muestras adquiridas, se realizó un análisis estadístico univariante para detectar posibles outliers originados durante las mediciones a través del test Grubbs [107].

Mediante este análisis, en la variable Total Oil, se detectaron y eliminaron dos outliers, correspondientes a las muestras (6) y (20), representadas en la tabla 4.14. Después de la eliminación de los outliers correspondientes, la variable Total Oil, quedó finalmente representada por un total de 19 muestras válidas.

La media, desviación estándar y los valores máximos y mínimos para cada variable, se muestran en la tabla 4.15.

Tabla 4.15: Media, desviación típica, valores máximos y mínimos.

N.º	Variables	Unidades	Media(μ)	Desviación típica (σ)	Valor Min.	Valor Max.
1	Engine Load	%	57	2,702	51	63
2	Fuel Index	%	64,705	2,567	59,6	70,1
3	Fuel plunger stroke C1	mA	2,91	0,1044	2,7	3,1
4	Hydraulic oil pressure After Pump 1	Bar	251,971	16,665	249,4	256
5	Total Oil	kg.	248,263	83,356	231,4	267
6	Exhaust Gas received Presure	Bar	1,6962	0,02376	1,65	1,75
7	Maximum Pressure, P. máx.	Bar	140,371	16,328	137	144

4.3.1.4 Base de datos preliminar

Como ya se ha comentado en el apartado de materiales y métodos, el periodo de muestreo disponible a bordo fue demasiado lento, por lo que para poder aplicar el método, se hizo necesaria la utilización de una técnica de interpolación para conseguir un número mínimo de muestras que cumpliera la ecuación (4.1)

$$\text{numero de muestras} = 2p + \frac{p(p-1)}{2} \tag{4.1}$$

Considerando que la duración del viaje era de 15 días y la adquisición era de una muestra cada hora, se generaron 360 muestras de cada variable, siguiendo una distribución normal de media μ y desviación estándar σ , $N_p(\mu, \sigma)$.

La distribución aleatoria normal univariante [125], tiene la siguiente forma,

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-1/2(x-\mu)^2/\sigma^2} \tag{4.25}$$

Dónde: $-\infty < X < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$

La variabilidad de la aleatoriedad de la muestras obedeció a las diferentes condiciones de operación del barco durante su navegación desde el Golfo de Guinea al Mar del Norte.

Con esto se consiguió el número de muestras necesario para validar el estudio.

4.3.1.5 Proceso de purgado de outliers

En esta fase, siguiendo el método, desarrollado en el apartado 4.2.5, se monitorizaron las 360 muestras de la base de datos preliminar de media μ y desviación estándar σ , siguiendo una distribución normal, $N_p(\mu, \sigma)$, estimadas para el proceso multivariable, a través de la carta de control T^2 de Hotelling, siguiendo la ecuación (4.6):

$$T^2 = (X_i - \bar{X})^t \cdot S^{-1} \cdot (X_i - \bar{X}) \tag{4.6}$$

$X_i = (X_{i1}; X_{i2}; \dots; X_{ip})^t$, el vector de medias, $\bar{X} = (56,93 \ 64,42 \ 2,90 \ 252,05 \ 248,26 \ 1,69 \ 140,45)$ y la inversa de la matriz de covarianza S^{-1} es:

$$S^{-1} = \begin{pmatrix} 0,1509 & -0,0070 & 0,2917 & -0,0024 & -0,0011 & 0,0497 & 0,0160 \\ -0,0070 & 0,1621 & 0,0900 & 0,0038 & 0,0015 & 0,3959 & -0,0117 \\ 0,2917 & 0,0900 & 94,8547 & -0,0734 & 0,0553 & 35,0473 & -0,0065 \\ -0,0024 & 0,0038 & -0,0734 & 0,3229 & -0,0002 & -0,2194 & -0,0239 \\ -0,0011 & 0,0015 & 0,0553 & -0,0002 & 0,0142 & 0,0320 & -0,0009 \\ 0,0497 & 0,3959 & 35,0473 & -0,2194 & 0,0320 & 1899,1 & 0,5663 \\ 0,0160 & -0,0117 & -0,0065 & -0,0239 & -0,0009 & 0,5663 & 0,4362 \end{pmatrix}$$

A continuación se presentan en las figuras 4.2 a 4.8 , los gráficos de control hasta conseguir el HDS del funcionamiento óptimo del proceso.

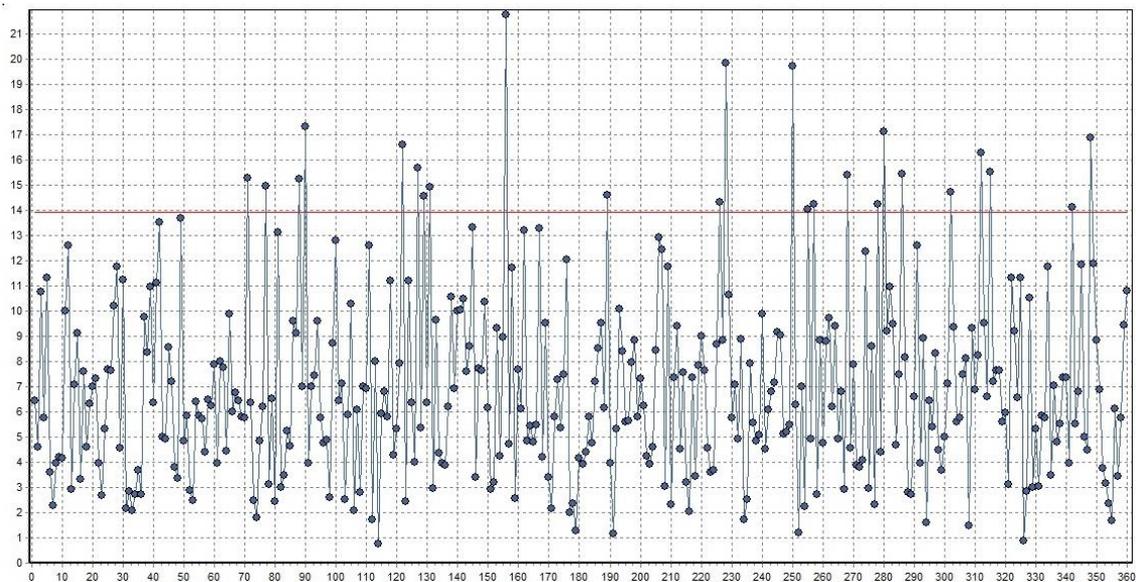


Figura 4.2: T^2 Control Chart – Step 1.

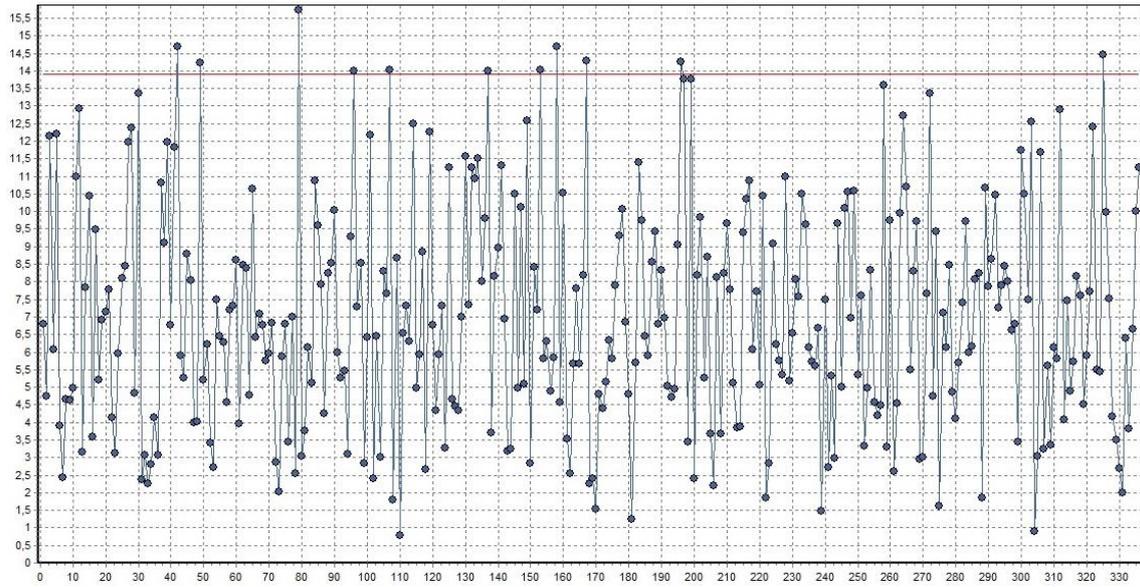


Figura 4.3: T^2 Control Chart – Step 2.

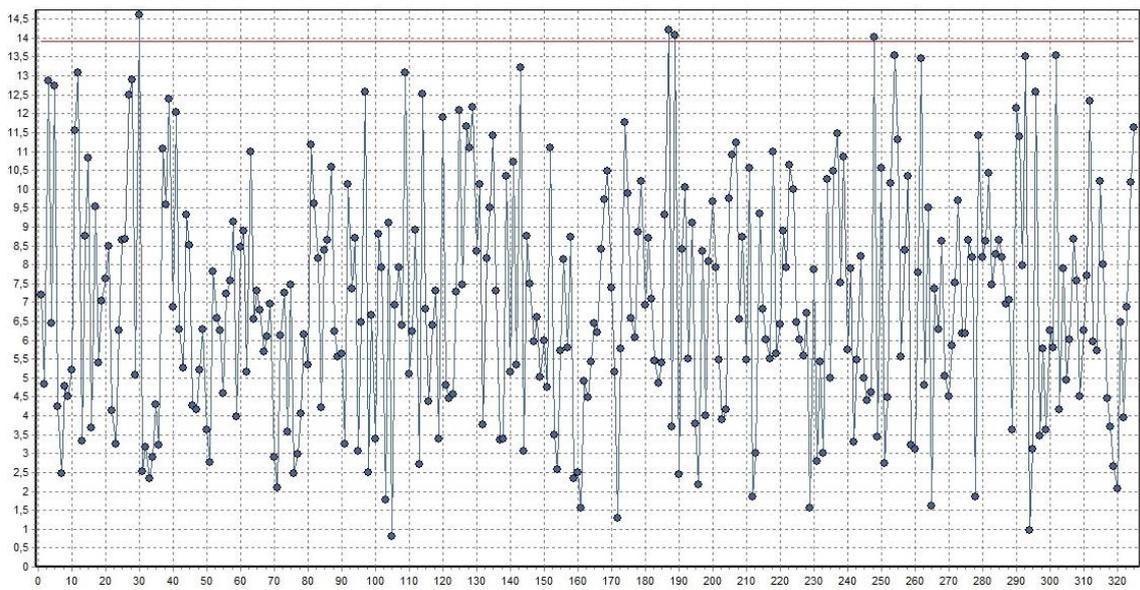


Figura 4.4: T^2 Control Chart – Step 3.

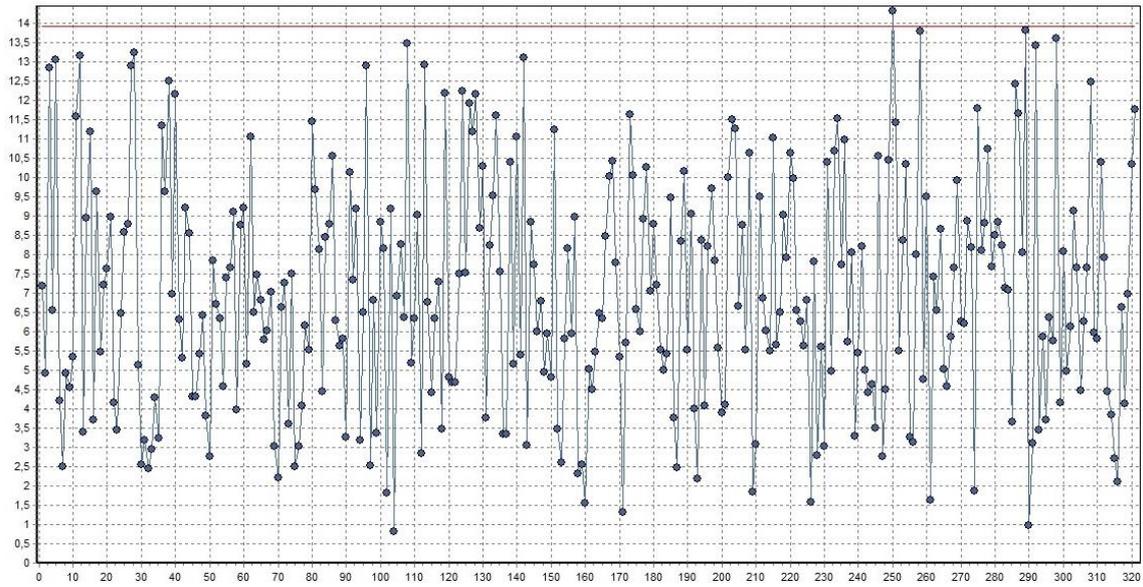


Figura 4.5: T² Control Chart – Step 4.

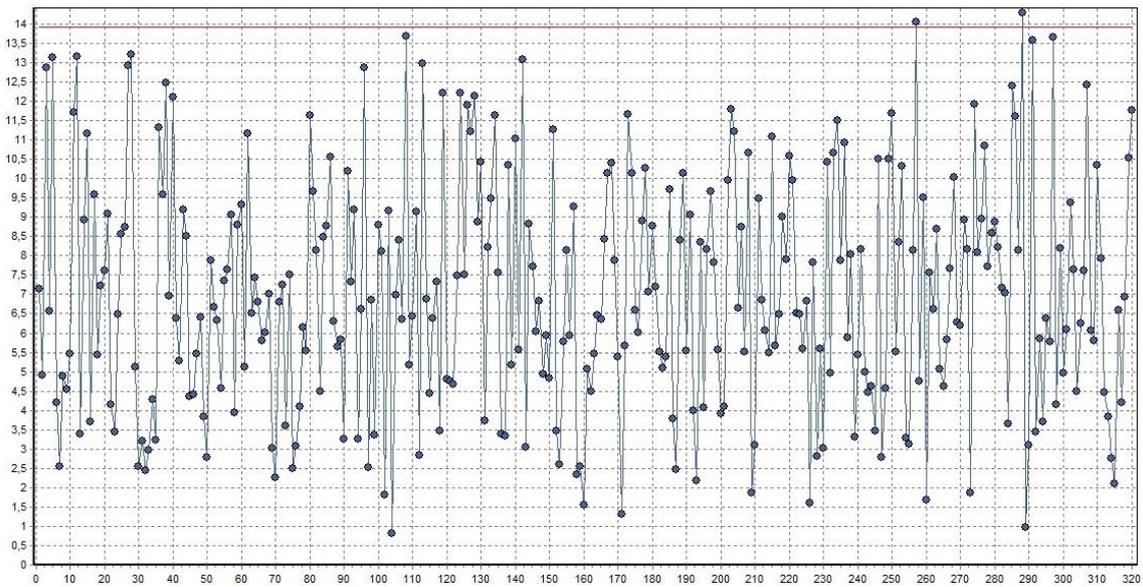


Figura 4.6: T² Control Chart – Step 5.

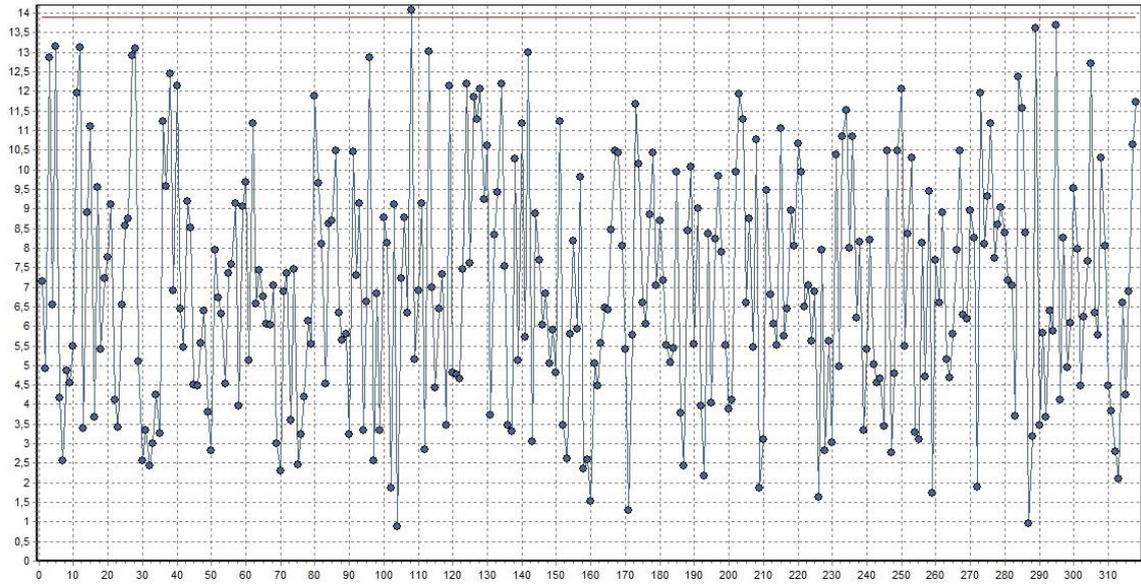


Figura 4.7: T^2 Control Chart – Step 6.

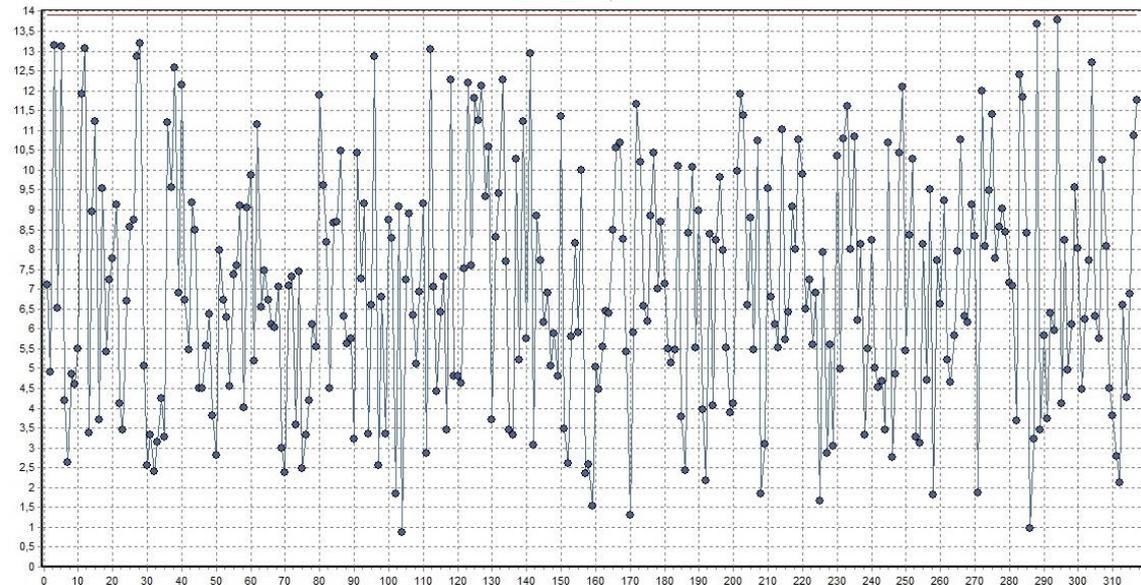


Figura 4.8: T^2 Control Chart – Step 7.

En la tabla 4.16, se representa el LCS calculado para cada paso y el número de outliers detectados hasta conseguir el HDS.

Tabla 4.16: Pasos hasta conseguir el HDS.

N.º de observaciones	LCS	N.º Outliers detectados
360	13,92	24
336	13,91	11
325	13,91	4
321	13,91	1
320	13,91	2
318	13,91	1
317	13,91	0

De la tabla se desprende un elevado número de outliers detectados, la generación de datos desde los datos adquiridos, ha podido ser la causa. Si el periodo de muestreo hubiese sido mayor, consideramos que lo ideal hubiesen sido dos muestras cada hora, hubiésemos tenido suficientes muestras disponibles para aplicar el método y el número de outliers hubiese sido menor.

4.3.1.6 Normalidad multivariante

La normalidad multivariante se comprobó mediante gráficos Q-Q, siguiendo una distribución beta.

En las figuras 4.9 y 4.10, se representa la normalidad multivariable de la base de datos preliminar, y después del proceso de purgado, HDS, respectivamente.

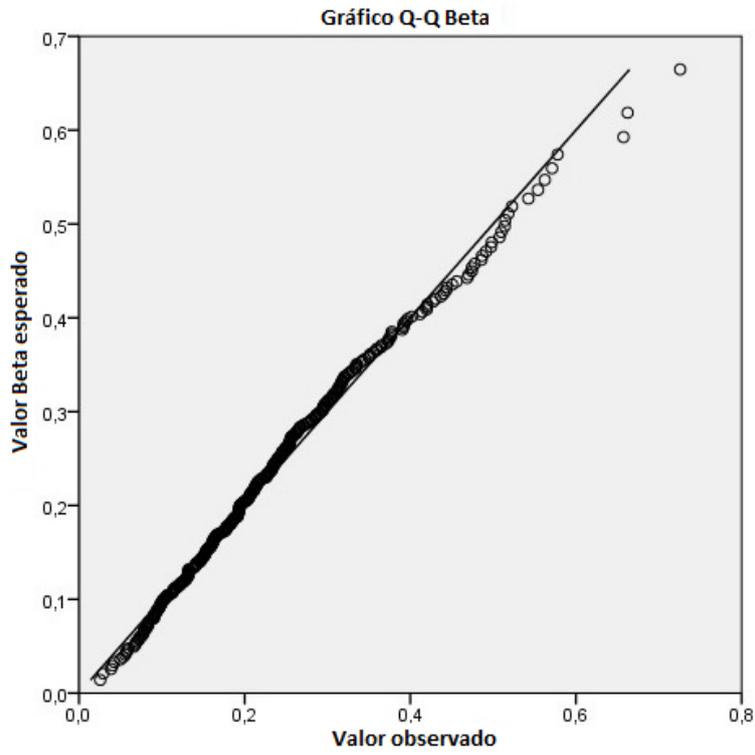


Figura 4.9: Análisis de normalidad multivariable, Base de datos preliminar.

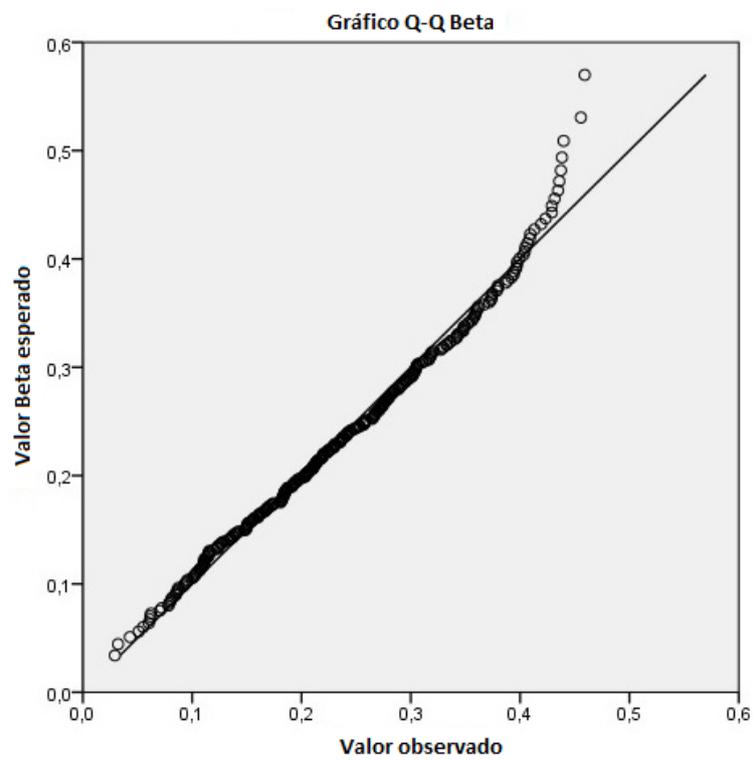


Figura 4.10: Análisis de normalidad multivariable, HDS.

4.3.1.7 Control del proceso

Una vez obtuvimos el HDS de las condiciones óptimas de funcionamiento en la Fase I, se comprobó si una nueva entrada de datos generaba alguna señal con respecto a las condiciones óptimas de funcionamiento.

La nueva entrada de datos, correspondió a muestras adquiridas durante dos viajes en condición de carga, obteniendo un total de 23 muestras validas, representadas en la tabla 4.17, después de analizarlas según el criterio de las condiciones normales de operación, expuestas en el apartado 4.3.1.2.

Tabla 4.17: Nueva entrada de datos.

		Número de identificación de variables						
		1	2	3	4	5	6	7
Número de muestra	1	56,00	62,80	2,90	248,40	249,2	1,75	140,00
	2	56,00	63,00	2,90	251,40	249,2	1,75	140,00
	3	55,00	62,80	2,90	248,50	249,2	1,70	140,00
	4	55,00	63,30	2,90	252,50	249,2	1,70	141,00
	5	58,00	65,80	2,90	249,40	240,3	1,70	142,00
	6	56,00	63,90	2,90	251,40	258,1	1,70	142,00
	7	57,00	65,60	2,90	254,70	258,1	1,70	144,00
	8	58,00	64,00	3,00	257,10	258,1	1,70	145,00
	9	55,00	64,20	2,90	248,30	249,2	1,70	140,00
	10	54,00	63,20	2,80	245,60	240,3	1,70	139,00
	11	52,00	60,30	2,70	241,40	231,4	1,70	135,00
	12	53,00	61,10	2,80	244,50	222,5	1,70	138,00
	13	55,00	63,90	2,90	250,00	249,2	1,70	137,00
	14	57,00	63,90	2,90	251,40	204,7	1,75	143,00
	15	55,00	63,50	2,90	250,20	258,1	1,70	142,00
	16	62,00	67,60	3,00	253,20	213,6	1,70	143,00
	17	58,00	64,60	2,90	252,20	240,3	1,70	143,00
	18	60,00	69,00	3,00	255,80	240,3	1,70	143,00
	19	60,00	68,80	3,00	259,30	258,1	1,70	148,00
	20	59,00	67,30	2,90	256,00	222,5	1,70	143,00
	21	55,00	64,40	2,90	249,40	231,4	1,70	141,00
	22	56,00	62,10	2,90	252,40	231,4	1,70	140,00
	23	50,00	60,70	2,70	248,30	231,4	1,70	138,00

Los resultados obtenidos ante esta nueva entrada de datos, se muestran en el siguiente capítulo.

4.3.2 Proceso de combustible

4.3.2.1 Descripción del proceso

El sistema de combustible del motor principal, está diseñado para consumir diésel oil (DO) y Heavy fuel oil (FO); el DO se utiliza durante las maniobras y durante el arranque del motor; mientras que el FO se utiliza durante la navegación.

Nuestro estudio se centró en la condición de navegación en carga, por lo que solo se tuvo en cuenta el consumo de FO.

El sistema de combustible del motor principal [121], cuenta con un tanque de servicio desde donde el fuel oil es bombeado a través de las bombas de suministro, accionadas eléctricamente, desde donde una presión de 4 bar puede ser mantenida en la parte de baja presión del sistema de circulación del combustible, evitando la gasificación del fuel en la venting box en los rangos de temperatura aplicados.

El venting box, está conectado con el tanque de servicio, por medio de una válvula de desaireación automática, que permite liberar los gases que se producen en el combustible.

Desde la parte de baja presión el sistema de combustible se alimenta a través de las bombas de circulación, accionadas eléctricamente, las cuales bombean el combustible a través de los calentadores y los filtros, situados inmediatamente antes de la entrada del motor. Este sistema se muestra en la figura 4.11.

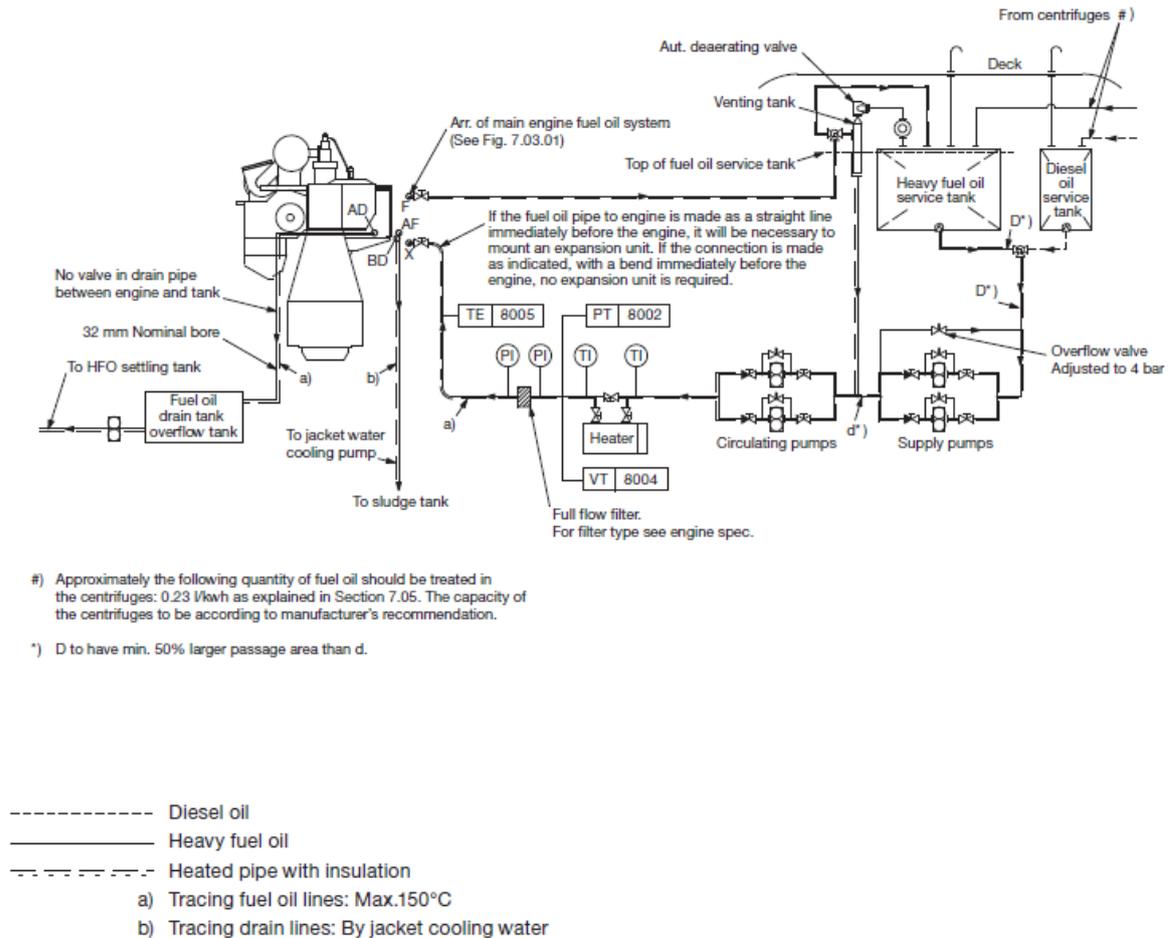


Figura 4.11: Sistema de combustible. [121]

A continuación pasamos a describir el funcionamiento del sistema de combustible:

Cuando el motor está parado, las bombas de circulación continúan bombeando combustible a través del sistema de fuel oil, de esta manera se mantienen las bombas de fuel a la temperatura de servicio.

La inyección de fuel se realiza a través del “electronically controlled pressure booster”, localizado en la Hydraulic Cylinder Unit (HCU), una por cada cilindro, la cual también contiene el actuador para la activación de la válvula de escape.

Esta, a su vez recibe señal desde el Cylinder control Unit (CCU), que calcula el timing de la inyección y la activación de la válvula de escape (FIVA) y las válvulas de aire de arranque (SAV), de acuerdo con los datos recibidos desde la Engine Control Unit (ECUs).

El sistema de control se compone de dos Engine Control Unit (ECU), operando en paralelo y realizando la misma tarea, uno en stand-by y el otro en funcionamiento. Si uno de ellos falla, el otro toma el control sin ningún tipo de interrupción.

Las funciones que desempeñan los ECUs en el motor principal son:

Funciones de velocidad de gobierno, secuencias de arranque y paro, tiempo de la inyección de fuel, tiempo de activación de la válvula de escape, etc.

Para asegurar el nivel de combustible del HCU, la capacidad de las bombas de circulación es mayor que la cantidad de fuel consumido por el motor, recirculando el excedente desde el motor al “venting box”.

4.3.2.2 Adquisición de datos

La adquisición de datos se realizó, durante la travesía que tuvo lugar desde la terminal offshore de Bonny (Nigeria) el 12 de abril de 2016 donde el buque fue cargado y finalizo en el puerto de Fos (Francia) el 24 de abril de 2016 donde fue descargado.

Para la elección de las variables que definieron el proceso de combustible, se contó con la colaboración de los ingenieros del buque, y con datos del fabricante; el proceso quedo definido por 11 variables, que se describen a continuación:

- 1- Engine Load (%), medida en el control local.
- 2- Fuel Index (%), medida en el control local.
- 3- Fuel Plunge Stroke (mA), corresponde a la media del valor del conjunto total de inyectores.
- 4- Scavenge air cooler air inlet temperature (°C), medida a la entrada del intercooler.
- 5- Exhaust gas temperature at turbine inlet (°C), medida a la entrada del turbo.
- 6- Turbocharger speed Rpm (r.p.m.), medida en el control local.
- 7- P (scav) (Bar), presión del aire a la entrada de la cámara de combustión.
- 8- Estimate Effective Power (kW), medida al eje.
- 9- Pcom (Presión de compresión) (Bar), corresponde a la media de todos los cilindros.

10- Pmax (Presión máxima) (Bar), corresponde a la media de todos los cilindros, medida en la cámara de combustión.

11- SFOC (g/kWh), corresponde al combustible consumido por el motor.

Se tomaron 4 muestras al día, de todas las variables, durante 1 viaje en condición de carga, obteniendo un total de 47 muestras normales, válidas según los siguientes criterios de operación normal: velocidad de avance del buque de entre 12 y 14 nudos con un deslizamiento inferior al 18%, temperatura media del agua de mar de 20 °C, temperatura media ambiente de 30 °C y temperatura media de la sala de máquinas de 37 °C.

4.3.2.3 Datos perdidos

Aunque la monitorización de los datos estaba automatizada y era continua, el almacenamiento de los datos estaba a merced de que el operario diese la orden de grabado. La orden de almacenamiento de los datos se realizó a la finalización de cada guardia, obteniendo 4 muestras diarias.

Sin embargo, a través del IAS, el propio operario realizaba la lectura y posterior transcripción de los datos, por algún motivo, entendemos que por falta de comunicación entre el personal, únicamente se obtuvo una muestra diaria de las variables (Scavenge air cooler air inlet temperatura, Turbocharger speed Rpm y SFOC), monitorizadas a través del IAS. Esta situación generó diariamente tres datos perdidos de cada una de las variables.

Siguiendo la metodología ya descrita en el apartado 4.2.2, mediante el análisis de regresión por pasos, y apoyándonos en el software SPSS, se obtuvieron diferentes modelos para cada una de las variables dependientes, en combinación con todas las variables disponibles durante la adquisición de datos. El resumen de los modelos para cada una de las variables dependientes, se muestran en las siguientes tablas a continuación:

Tabla 4.18: Modelos generados, variable dependiente: Scavenge air cooler air inlet temperatura.

Modelo	R ²	R ² Ajustado	Cambio en R ²
1	0,494	0,474	0,494
2	0,628	0,597	0,134
3	0,712	0,675	0,085
4	0,864	0,839	0,152
5	0,856	0,838	-0,008
6	0,895	0,876	0,039
7	0,881	0,866	-0,014
8	0,916	0,9	0,035
9	0,934	0,918	0,018

1.Predictores: Constante, Cyl3. Crank angle
 2.Predictores: Constante, Cyl3. Crank angle, Cyl2. Exhaust valve close delay
 3.Predictores: Constante, Cyl3. Crank angle, Cyl2. Exhaust valve close delay, P(scav)
 4.Predictores: Constante, Cyl3. Crank angle, Cyl2. Exhaust valve close delay, P(scav), Cyl6. Pcom
 5.Predictores: Constante, Cyl3. Crank angle, P(scav), Cyl6. Pcom
 6.Predictores: Constante, Cyl3. Crank angle, P(scav), Cyl6. Pcom, Cyl2. Crank angle
 7.Predictores: Constante, P(scav), Cyl6. Pcom, Cyl2. Crank angle
 8.Predictores: Constante, P(scav), Cyl6. Pcom, Cyl2. Crank angle, Cyl1. Exhaust valve close position
 9. Predictores: Constante, P(scav), Cyl6. Pcom, Cyl2. Crank angle, Cyl1. Exhaust valve close position, Cyl6. Exhaust valve close delay

Tabla 4.19: Modelos generados, variable dependiente: Exhaust gas temperature at turbine inlet.

Modelo	R ²	R ² Ajustado	Cambio en R ²
1	0,674	0,661	0,674
2	0,937	0,931	0,262
3	0,957	0,951	0,02
4	0,97	0,964	0,013
5	0,975	0,969	0,005
6	0,983	0,978	0,008
7	0,988	0,984	0,005

1.Predictores: Constante, Cyl2. Pcom
 2.Predictores: Constante, Cyl2. Pcom, P(scav)
 3.Predictores: Constante, Cyl2. Pcom, P(scav), Cyl6. Pcom
 4.Predictores: Constante, Cyl2. Pcom, P(scav), Cyl6. Pcom, Cyl3 Pcom
 5.Predictores: Constante, Cyl2. Pcom, P(scav), Cyl6. Pcom, Cyl3. Pcom, Crankshaft r.p.m.
 6.Predictores: Constante, Cyl2. Pcom, P(scav), Cyl6. Pcom, Cyl3. Pcom, Crankshaft r.p.m., Cyl6. Exhaust valve close delay
 7.Predictores: Constante, Cyl2. Pcom, P(scav), Cyl6. Pcom, Cyl3. Pcom, Crankshaft r.p.m., Cyl6. Exhaust valve close delay, Cyl1 Pcom

Tabla 4.20: Modelos generados, variable dependiente: Turbocharger Speed.

Modelo	R ²	R ² Ajustado	Cambio en R ²
1	0,733	0,722	0,733
2	0,902	0,893	0,169
3	0,928	0,918	0,026
4	0,947	0,938	0,02

1.Predictores: Constante, P(scav)
 2.Predictores: Constante, P(scav),Cyl6.Pcom
 3.Predictores: Constante, P(scav),Cyl6.Pcom,Cyl3. Exhaust valve open delay
 4.Predictores: Constante, P(scav),Cyl6.Pcom,Cyl3. Exhaust valve open delay, Cyl1 crank angle

Tabla 4.21: Modelos generados, variable dependiente: SFOC.

Modelo	R ²	R ² Ajustado	Cambio en R ²
1	0,733	0,722	0,733

1.Predictores: Constante, Cyl1 Estimate Effective Power

Para la elección del modelo, se siguió el siguiente criterio: para cada una de las variables dependientes, se observó el cambio que se iba produciendo en el coeficiente de determinación R² a medida que se iban incorporando variables al modelo. Un cambio grande en R² indico que esa variable contribuyo de forma importante a explicar lo que ocurría con la variable dependiente, una vez el cambio fue pequeño, menor a 0,1ya no se tuvieron en cuenta más variables para generar el modelo. Los modelos seleccionados para generar los datos perdidos de cada una de las variables, junto con sus coeficientes, se presentan en a continuación:

Tabla 4.22: Modelo seleccionado, variable dependiente: Scavenge air cooler air inlet Temperature.

Modelo 4	Coefficientes	R ² ajustado
Cte/. modelo	171,406	0,839
Cyl,3-Crank Angle	7,117	
Exhaust valve close delay c2	-0,803	
P (scav)	102,889	
Cyl, 6-P (com)	-1,821	

Tabla 4.23: Modelo seleccionado, variable dependiente: Exhaust gas temperature at turbine inlet.

Modelo 2	Coefficientes	R ² ajustado
Cte/. modelo	778,077	0,931
Cyl,2-p(com)	-6,64	
P(scav)	233,637	

Tabla 4.24: Modelo seleccionado, variable dependiente: Turbocharger speed.

Modelo 2	Coefficientes	R ² ajustado
Cte/. modelo	10028,76	0,893
P(scav)	4317,416	
Cyl, 6 - P (com)	-57,829	

Tabla 4.25: Modelo seleccionado, variable dependiente: SFOC.

Modelo 1	Coefficientes	R cuadrado ajustado
Cte/. modelo	31,945	0,658
Cyl,1-Estimete Effective Power	0,071	

Para todos los modelos el ajuste fue superior a 0,8, menos para la variable SFOC que fue de 0,6.

Con los modelos generados, calculamos los valores perdidos de cada una de las variables dependientes. La base de los datos adquiridos ya incluidos los valores perdidos se presentan en la tabla 4.26.

Tabla 4.26: Base de los datos adquiridos.

		Número de identificación de variable										
		1	2	3	4	5	6	7	8	9	10	11
Número de Muestra	1	58,0	66,1	2,7	144,0	402,0	10.366,0	1,6	10.601,0	110,7	138,8	157,3
	2	54,0	62,6	2,6	142,8	399,2	10.428,6	1,6	10.328,0	110,3	138,9	154,3
	3	59,0	67,6	2,7	149,9	404,9	10.592,4	1,6	10.569,0	112,2	141,1	156,6
	4	56,0	64,0	2,6	149,1	404,6	10.583,9	1,6	10.550,0	111,7	139,8	156,1
	5	57,0	65,7	2,7	142,0	402,0	10.407,0	1,6	10.460,0	110,2	139,6	157,8
	6	59,0	65,4	2,7	147,9	403,5	10.489,1	1,6	10.460,0	111,1	139,9	155,3
	7	56,0	63,9	2,6	144,5	406,5	10.480,6	1,6	10.470,0	110,3	139,5	156,0
	8	55,0	63,1	2,6	153,4	381,6	10.920,4	1,9	10.446,0	124,0	139,4	157,7
	9	58,0	66,2	2,7	157,0	378,0	11.021,0	1,9	10.440,0	124,7	139,6	157,1
	10	57,0	67,2	2,7	151,4	381,0	10.908,8	1,9	10.378,0	123,9	139,7	156,8
	11	57,0	64,9	2,6	153,6	382,3	10.926,1	1,9	10.454,0	123,6	138,4	157,9
	12	59,0	67,9	2,7	151,6	377,0	10.914,6	1,9	10.516,0	124,2	138,6	158,0
	13	56,0	64,3	2,7	157,0	380,0	11.003,6	1,9	10.619,0	124,6	138,7	158,0
	14	55,0	62,8	2,6	153,1	380,0	10.992,1	1,9	10.517,0	125,2	140,5	157,2

Tabla 4.26: Base de los datos adquiridos (continuación).

		Número de identificación de variable										
		1	2	3	4	5	6	7	8	9	10	11
Número de muestra	15	56,0	64,7	2,7	157,0	377,0	11.105,0	1,9	10.390,0	124,6	139,5	158,3
	16	55,0	63,0	2,6	151,5	375,4	10.968,9	1,9	10.724,0	124,4	139,8	162,0
	17	61,0	70,6	2,8	160,1	383,8	11.201,8	2,0	10.909,0	125,7	140,7	164,0
	18	57,0	65,8	2,6	150,5	372,6	10.785,1	1,8	10.441,0	123,9	140,4	157,9
	19	56,0	65,2	2,6	150,0	369,0	10.922,0	1,8	10.485,0	125,4	141,2	156,4
	20	56,0	64,1	2,7	151,3	366,7	10.807,8	1,8	10.560,0	124,6	140,8	158,8
	21	56,0	65,0	2,7	158,5	375,8	11.126,6	2,0	10.881,0	127,6	141,2	163,7
	22	60,0	68,7	2,7	153,4	371,4	10.894,1	1,9	10.482,0	124,4	140,3	157,7
	23	59,0	67,8	2,7	146,0	363,0	10.745,0	1,9	10.534,0	126,3	141,0	157,9
	24	57,0	64,5	2,7	147,7	367,4	10.824,8	1,9	10.522,0	125,5	140,9	158,3
	25	58,0	66,5	2,7	152,2	370,4	11.017,1	1,9	10.676,0	127,4	141,5	160,0
	26	57,0	66,0	2,7	149,4	369,8	10.959,7	1,9	10.678,0	126,8	141,7	160,1
	27	56,0	65,9	2,7	148,0	362,0	10.813,0	1,9	10.400,0	125,6	141,0	157,7
	28	56,0	63,3	2,6	147,1	366,0	10.807,4	1,9	10.483,0	126,4	141,7	158,8
	29	57,0	65,2	2,7	149,0	365,1	10.919,2	1,9	10.724,0	127,4	141,4	161,1
	30	59,0	66,0	2,7	148,7	368,1	10.899,1	1,9	10.682,0	127,6	142,1	160,1
	31	58,0	64,9	2,7	148,0	362,0	10.829,0	1,9	10.559,0	126,8	140,8	159,4
	32	57,0	64,6	2,7	149,8	362,1	10.910,7	1,9	10.646,0	127,2	141,6	159,8
	33	55,0	62,6	2,6	147,0	361,4	10.778,5	1,9	10.570,0	127,0	142,4	159,9
	34	58,0	67,6	2,7	147,8	362,1	10.870,2	1,9	10.605,0	127,4	141,7	160,2
	35	55,0	63,6	2,7	145,0	359,0	10.784,0	1,9	10.467,0	127,6	141,2	158,0
	36	56,0	65,8	2,7	151,9	362,8	10.841,7	1,9	10.467,0	126,6	141,5	156,8
	37	55,0	64,1	2,7	147,8	362,4	10.861,7	1,9	10.528,0	127,2	141,5	160,1
	38	57,0	64,9	2,7	147,6	359,1	10.856,0	1,9	10.490,0	127,4	141,4	157,3
	39	55,0	63,3	2,6	145,0	354,0	10.636,0	1,8	10.358,0	125,6	141,1	155,6
	40	59,0	66,4	2,7	146,5	360,1	10.829,7	1,9	10.699,0	127,7	142,5	161,3
	41	57,0	64,6	2,7	144,8	357,7	10.701,0	1,8	10.526,0	125,2	142,3	159,3
42	56,0	63,2	2,7	146,6	356,4	10.775,4	1,9	10.552,0	126,9	141,8	157,7	
43	55,0	62,9	2,6	142,0	354,0	10.654,0	1,8	10.330,0	126,6	141,7	158,8	
44	57,0	64,5	2,6	147,2	362,8	10.807,0	1,9	10.586,0	127,2	141,5	159,7	
45	55,0	62,8	2,6	143,8	356,4	10.695,2	1,8	10.318,0	125,4	140,7	156,6	
46	57,0	64,8	2,7	145,0	354,0	10.840,0	1,9	10.583,0	128,7	141,7	159,5	
47	58,0	66,8	2,7	149,4	361,5	10.984,7	2,0	10.722,0	129,7	141,8	161,6	

Los valores máximos, mínimos, medias y desviaciones típicas de cada una de las variables fueron calculados. Los valores, se presentan en la tabla 4.27.

Tabla 4.27: Medias y desviaciones típicas y valores máximos y mínimos de las variables del proceso de combustible. Base de datos adquiridos.

N.º	Variables	Unidades	N	Valor Min.	Valor Max.	Media(μ)	Desviación típica (σ)
1	Engine Load	%	47	54	61	56,85	1,546
2	Fuel Index	%	47	62,6	70,6	65,136	1,7644
3	Fuel plunger stroke	mA	47	2,58	2,77	2,6621	0,03581
4	Scavenge air cooler air inlet T ^a	°C	47	142	160	149,21	4,287
5	Exhaust gas T ^a at turbine inlet	°C	47	354	407	372,63	15,306
6	Turbocharger speed	r.p.m.	47	10.366	11.202	10.818,84	185,504
7	P (scav)	Bar	47	1,57	1,96	1,834	0,10934
8	Estimate Effective Power	kW	47	10.318	10.909	10.540,11	130,761
9	Pcom	Bar	47	110,18	129,72	123,8326	5,63466
10	Pmax	Bar	47	138,43	142,47	140,7391	1,08839
11	SFOC	g/kWh	47	154,28	164,01	158,5111	2,042

Con las muestras que formaban la base de datos adquiridas, se realizó un análisis univariante a través del test Grubbs, no detectando ningún dato anómalo, originado durante las mediciones.

4.3.2.4 Base de datos preliminar

El periodo de muestreo disponible a bordo fue demasiado lento, por lo que para poder aplicar el método, se hizo necesaria la utilización de una técnica de interpolación para conseguir un número mínimo de muestras que cumpliera la ecuación (4.1)

$$\text{numero de muestras} = 2p + \frac{p(p-1)}{2} \quad (4.1)$$

La técnica utilizada fue la función de interpolación cubica spline [126]. Considerando la adquisición de 2 muestras cada hora, se generaron 599 muestras, quedando así definida la base de datos preliminar. Con esto se consiguió el número de muestras necesario para validar el estudio.

Una vez definida la base de datos preliminar, se calcularon los valores máximos, mínimos, medias y desviaciones típicas de cada una de las variables, para las 599 muestras, presentadas en la tabla 4.28.

Tabla 4.28: Medias y desviaciones típicas - valores máximos y mínimos de las variables del proceso de combustible. Base de datos preliminar.

N.º	VARIABLES	UNIDADES	N	Valor Min.	Valor Max.	Media(μ)	Desviación típica (σ)
1	Engine Load	%	599	54	61	56,81	1,421
2	Fuel Index	%	599	62,4	70,7	65,095	1,6098
3	Fuel plunger stroke	mA	599	2,58	2,77	2,6613	0,0326
4	Scavenge air cooler air inlet T ^a	°C	599	142	160	149,24	4,111
5	Exhaust gas T ^a at turbine inlet	°C	599	353	408	372,43	14,926
6	Turbocharger speed	r.p.m.	599	10.366	11.206	10.821,55	174,32
7	P (scav)	Bar	599	1,55	1,96	1,8352	0,10472
8	Estimate Effective Power	kW	599	10.316	10.936	10.536,81	121,952
9	Pcom	Bar	599	109,18	129,72	123,9036	5,42306
10	Pmax	Bar	599	138,37	142,59	140,7471	1,05311
11	SFOC	g/kWh	599	154,28	164,36	158,4811	1,96348

Alternativamente, las observaciones de la base de datos adquiridos y la base de datos preliminar fueron monitorizadas, comprobando que seguían una misma tendencia, dando por valido la aplicación de la función spline en la generación de los datos.

4.3.2.5 Selección de variables

Para la selección de las variables, al igual que en el proceso de lubricación de cilindros, se realizó un análisis de regresión multivariable, entre el conjunto completo de variables que definían el proceso. Sin embargo, en este proceso los coeficientes de determinación R^2 de los modelos fueron demasiados bajos, por lo que se optó, primeramente, con las muestras generadas en la base de datos preliminar, por realizar un análisis de correlación de Pearson [104], entre las 11 variables en las que quedo definido el proceso de combustible. El análisis de correlaciones se presenta en la tabla 4.29.

Tabla 4.29: Correlación entre variables. Base de datos preliminar.

		Número de identificación de variable										
		1	2	3	4	5	6	7	8	9	10	11
Número de identificación de variable	1	1	0,88	0,75	0,245	0,114	0,171	0,096	0,415	0,012	0,047	0,277
	2	0,88	1	0,769	0,373	0,125	0,282	0,156	0,358	0,046	-0,34	0,262
	3	0,75	0,769	1	0,181	0,101	0,119	0,067	0,494	-0,03	0,142	0,384
	4	0,245	0,373	0,181	1	0,16	0,79	0,414	0,409	0,235	-0,313	0,34
	5	0,114	0,125	0,101	0,16	1	-0,39	-0,79	-0,06	-0,9	-0,693	-0,38
	6	0,171	0,282	0,119	0,79	-0,39	1	0,857	0,494	0,739	0,115	0,623
	7	0,096	0,156	0,067	0,414	-0,79	0,857	1	0,404	0,975	0,457	0,648
	8	0,415	0,358	0,494	0,409	-0,06	0,494	0,404	1	0,286	0,318	0,843
	9	0,012	0,046	-0,03	0,235	-0,9	0,739	0,975	0,286	1	0,563	0,57
	10	0,047	-0,03	0,142	-0,31	-0,69	0,115	0,457	0,318	0,563	1	0,439
	11	0,277	0,262	0,384	0,34	-0,38	0,623	0,648	0,843	0,57	0,439	1

Alternativamente se realizó un análisis de correlaciones de las 47 muestras adquiridas, mostradas en la tabla 4.30, donde, se comprobó que existía un mismo patrón con respecto a los datos adquiridos a bordo.

Tabla 4.30: Correlación entre variables. Base de datos adquiridos.

		Número de identificación de variable										
		1	2	3	4	5	6	7	8	9	10	11
Número de identificación de variable	1	1	0,886	0,756	0,256	0,141	0,16	0,08	0,433	-0,009	0,048	0,276
	2	0,886	1	0,785	0,365	0,14	0,266	0,14	0,379	0,03	-0,018	0,272
	3	0,756	0,785	1	0,191	0,111	0,129	0,07	0,493	-0,024	0,14	0,393
	4	0,256	0,365	0,191	1	0,135	0,788	0,43	0,4	0,257	-0,267	0,335
	5	0,141	0,14	0,111	0,135	1	-0,41	-0,79	-0,03	-0,897	-0,689	-0,36
	6	0,16	0,266	0,129	0,788	-0,41	1	0,86	0,471	0,751	0,167	0,61
	7	0,078	0,141	0,07	0,431	-0,79	0,864	1	0,386	0,975	0,482	0,638
	8	0,433	0,379	0,493	0,4	-0,03	0,471	0,39	1	0,265	0,312	0,833
	9	-0,01	0,03	-0,024	0,257	-0,9	0,751	0,98	0,265	1	0,581	0,885
	10	0,048	-0,018	0,14	-0,27	-0,69	0,167	0,48	0,312	0,581	1	0,445
	11	0,276	0,272	0,393	0,335	-0,36	0,61	0,64	0,833	0,559	0,445	1

Como se puede apreciar, la correlación entre variables aumenta y disminuye en la misma proporción, en la base de datos preliminar y datos adquiridos.

Una vez comprobado la existencia de un mismo patrón entre ambos análisis de correlaciones, para la elección de las variables, se utilizó la base de datos preliminar, siguiendo el siguiente criterio:

- El ajuste de los modelos de regresión tenían que tener un coeficiente de determinación R^2 ajustado mayor a 0,8.
- Las variables seleccionadas tenían que tener una correlación con al menos una de las variables no seleccionadas igual o superior a 0,49.
- Las variables no seleccionadas, variables predictivas, tenían que ser parte del modelo de regresión de al menos una de las variables seleccionadas.
- Las variables seleccionadas tenían que tener una correlación entre si menor a 0,49.

Con este criterio, se pretendió reducir las variables que definen el proceso, sin pérdida de información del mismo.

De las 11 variables en las que quedo definido el proceso de combustible, el proceso quedo monitorizado por las siguientes 3 variables (Fuel Index, Exhaust gas temperature at turbine inlet y Turbocharger speed), siguiendo el criterio mencionado anteriormente.

Finalmente, mediante el software SPSS, se comprobó el ajuste de los modelos entre las tres variables seleccionadas y sus variables predictoras, mediante un análisis de regresión, obteniendo los siguientes coeficientes de determinación R^2 ajustado, (0,8), (0,95) y (0,96) para cada modelo respectivamente. Los modelos generados, se presentan en la tabla 4.31.

Tabla 4.31: Modelos generados. Variables seleccionadas / Variables predictoras.

Modelo	Coeficientes	R ²	R ² ajust	Correlación de Pearson	Correlación Parcial
Cte/. modelo	-12,153	0,802	0,801		
Engine Load	0,787			0,88	0,718
Fuel plunger stroke	12,231			0,769	0,345

Variable dependiente: Fuel Index

Modelo	Coeficientes	R ²	R ² ajust	Correlación de Pearson	Correlación Parcial
Cte/. modelo	897,214	0,956	0,956		
P (scav)	226,836			-0,792	0,828
Pcom	-6,654			-0,9	-0,901
Pmax	-0,828			-0,693	-0,196

Variable dependiente: Exhaust gas temperature at turbine inlet

Cte/. modelo	Coeficientes	R ²	R ² ajust	Correlación de Pearson	Correlación Parcial
Constante	6579,114	0,967	0,967		
Scavenge air cooler	18,266			0,79	0,752
P (scav)	1952,267			0,857	0,534
Estimate Effective Power	-0,1			0,494	-0,0173
Pcom	-17,044			0,739	-0,3
SFOC	6,917			0,623	0,172

Variable dependiente: Turbocharger speed

4.3.2.6 Proceso de purgado de outliers

En esta fase, siguiendo el método, desarrollado en el apartado 4.2.5, se monitorizaron las 599 muestras de la base de datos preliminar de media μ y desviación típica σ , siguiendo una distribución normal, $N_p(\mu, \sigma)$, a través de la carta de control T^2 de Hotelling, siguiendo la ecuación (4.6):

$$T^2 = (X_i - \bar{X})^t \cdot S^{-1} \cdot (X_i - \bar{X}) \tag{4.6}$$

Dónde:

$X_i = (X_{i1}; X_{i2}; \dots; X_{ip})^t$, el vector de medias, $\bar{X} = (65,095 \ 372,43 \ 10.821,55)$ y la inversa de la matriz de covarianza S^{-1} es:

$$S^{-1} = \begin{pmatrix} 0,4515 & -0,0135 & -0,0016 \\ -0,0135 & 0,0057 & 0,0002 \\ -0,0016 & 0,0002 & 4,479E - 05 \end{pmatrix}$$

A continuación, se presentan en las figuras 4.12 a 4.34, los gráficos de control, hasta conseguir el HDS del funcionamiento óptimo del proceso.

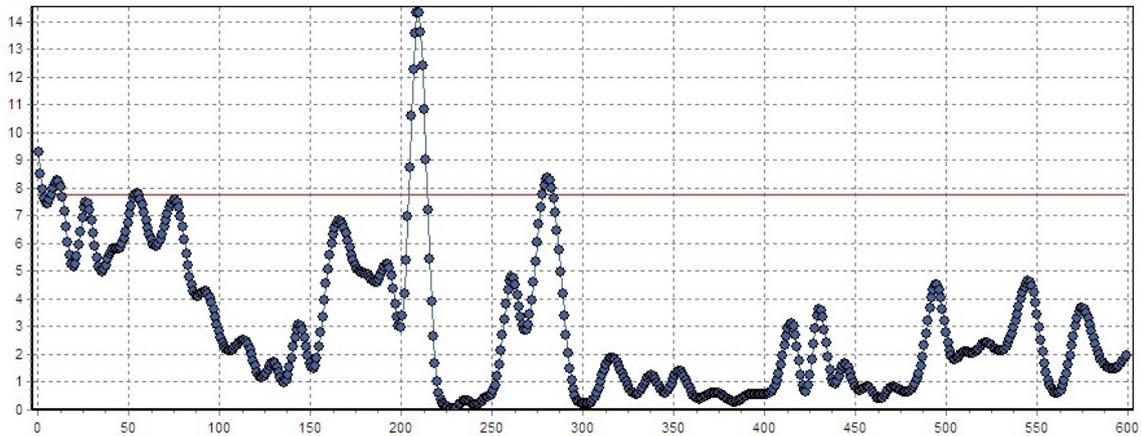


Figura 4.12: T² Control Chart – Step 1.

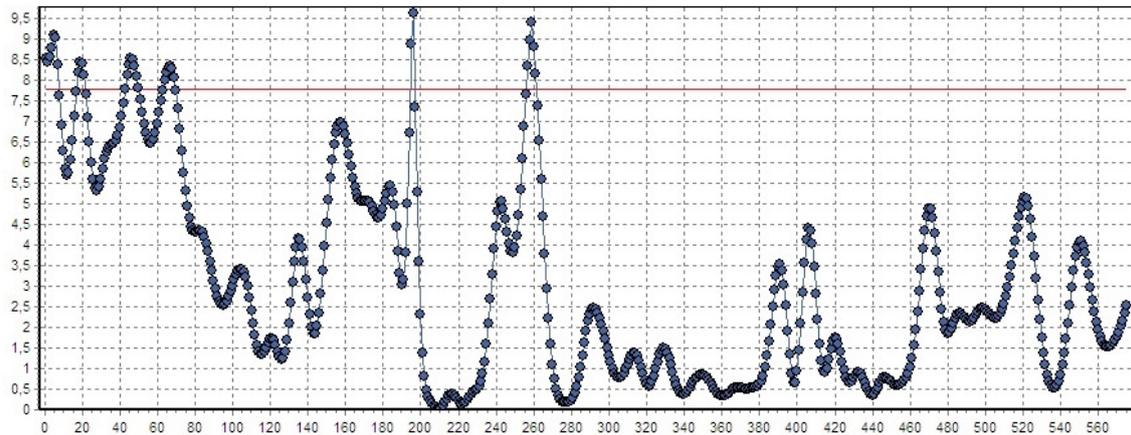


Figura 4.13: T² Control Chart – Step 2.

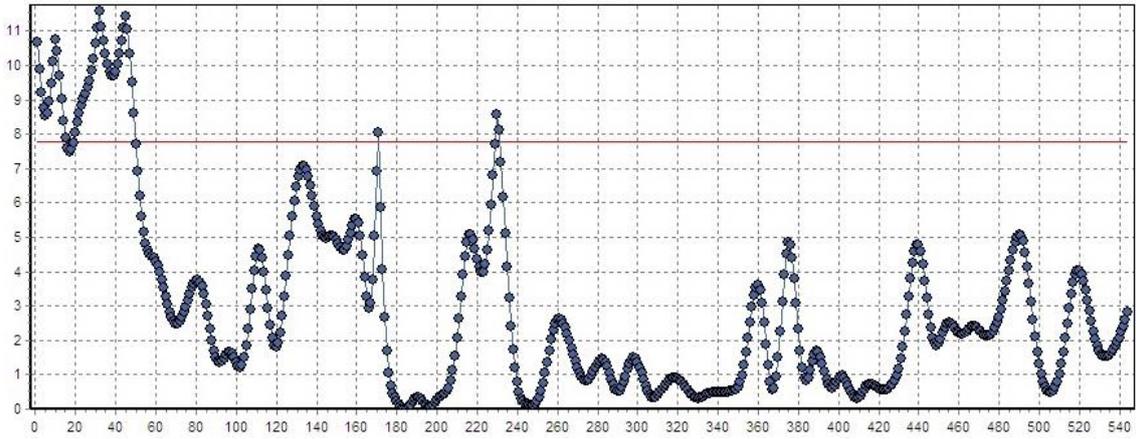


Figura 4.14: T^2 Control Chart – Step 3.

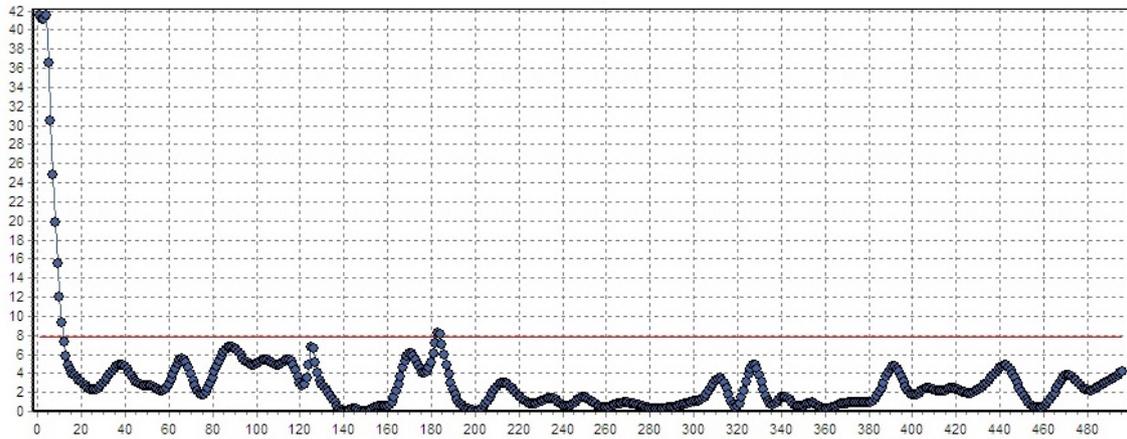


Figura 4.15: T^2 Control Chart – Step 4.

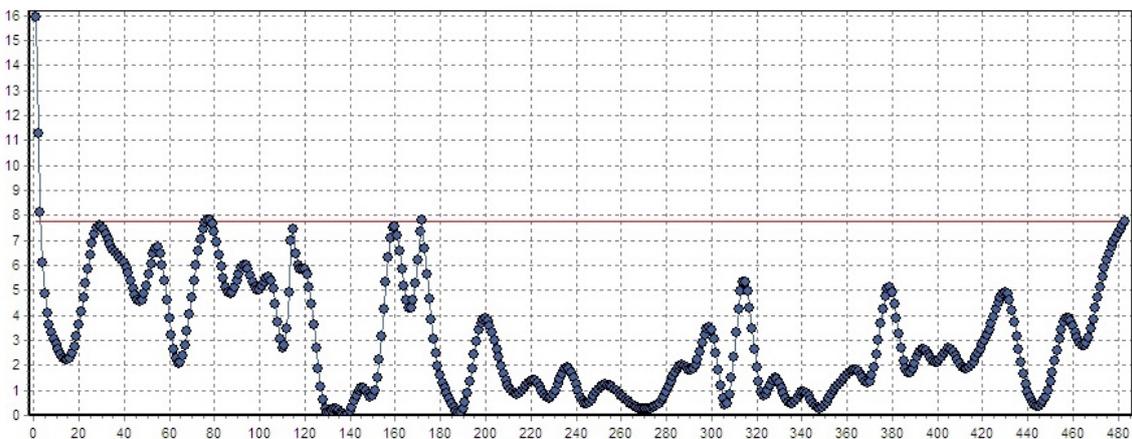


Figura 4.16: T^2 Control Chart – Step 5.

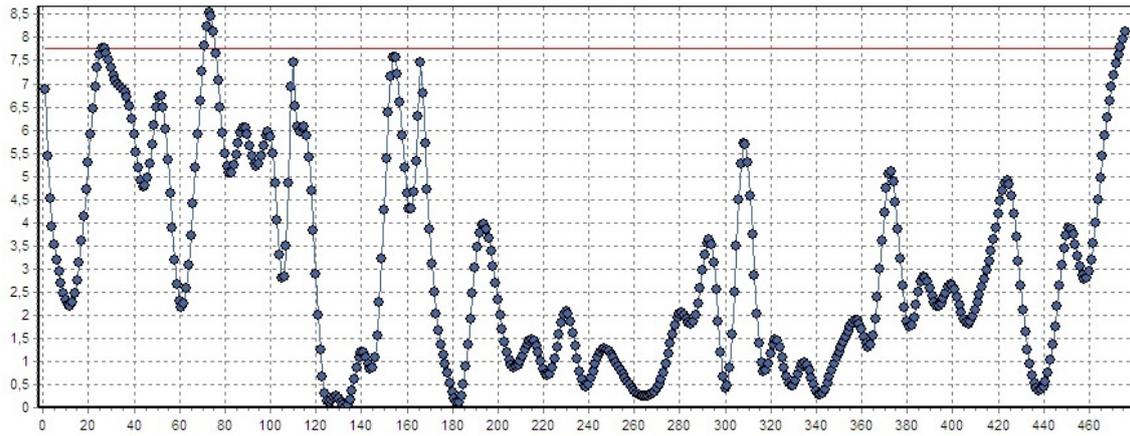


Figura 4.17: T² Control Chart – Step 6.

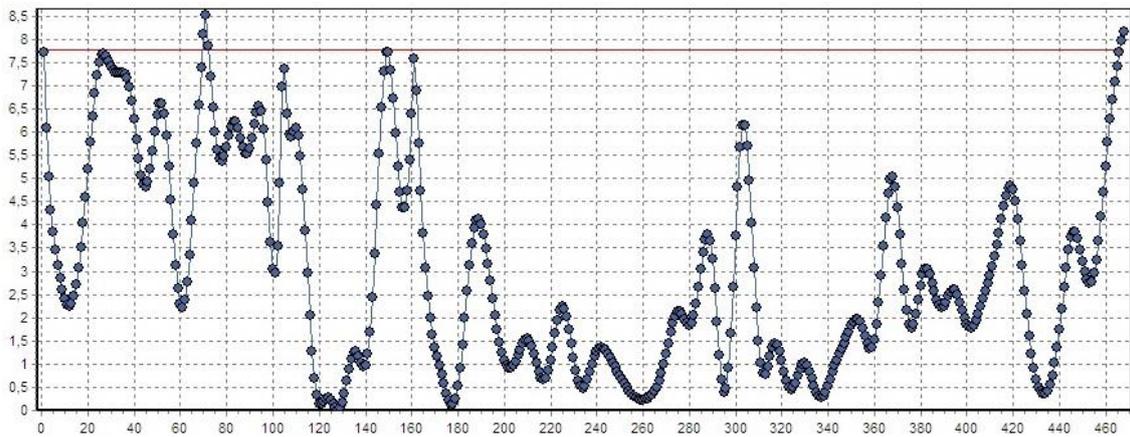


Figura 4.18: T² Control Chart – Step 7.

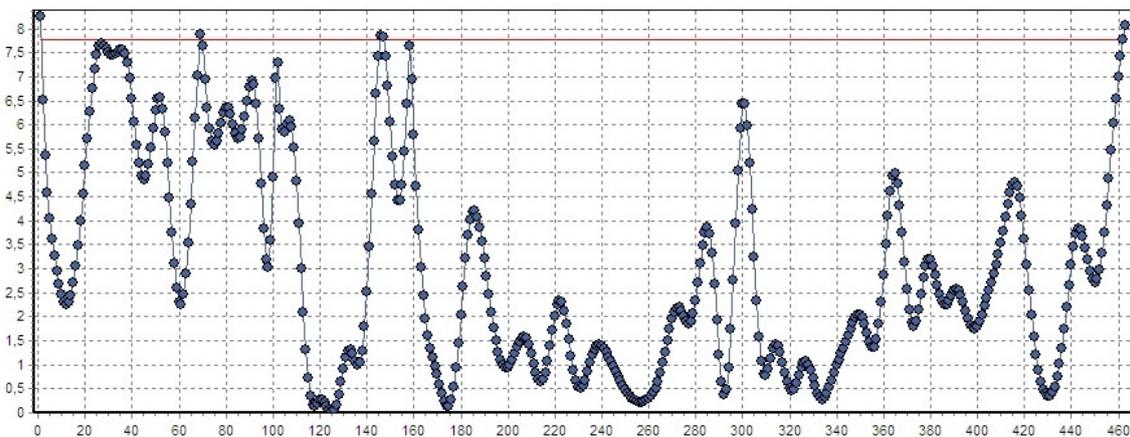


Figura 4.19: T² Control Chart – Step 8.

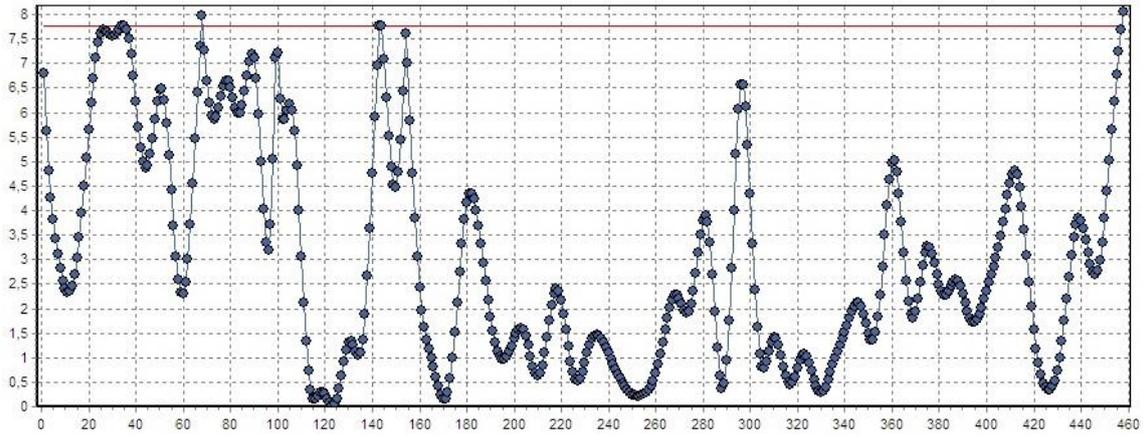


Figura 4.20: T^2 Control Chart – Step 9.

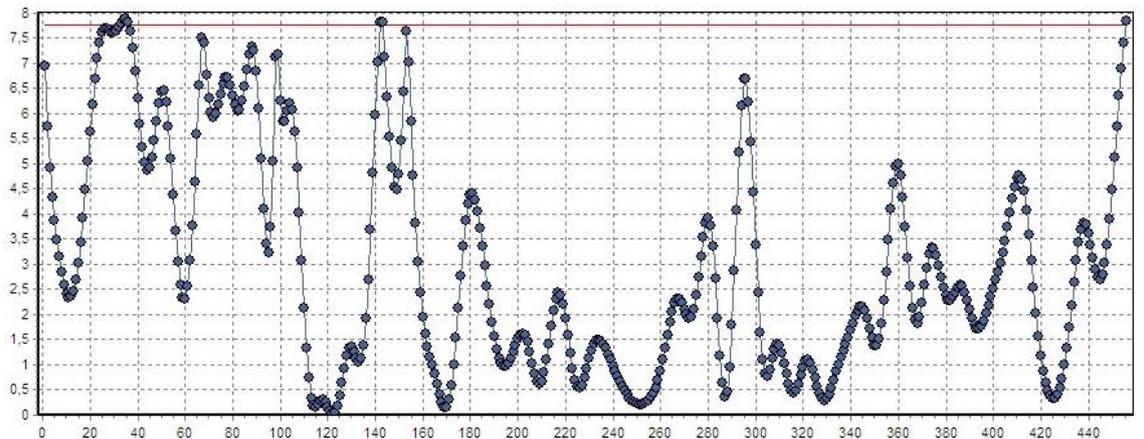


Figura 4.21: T^2 Control Chart – Step 10.

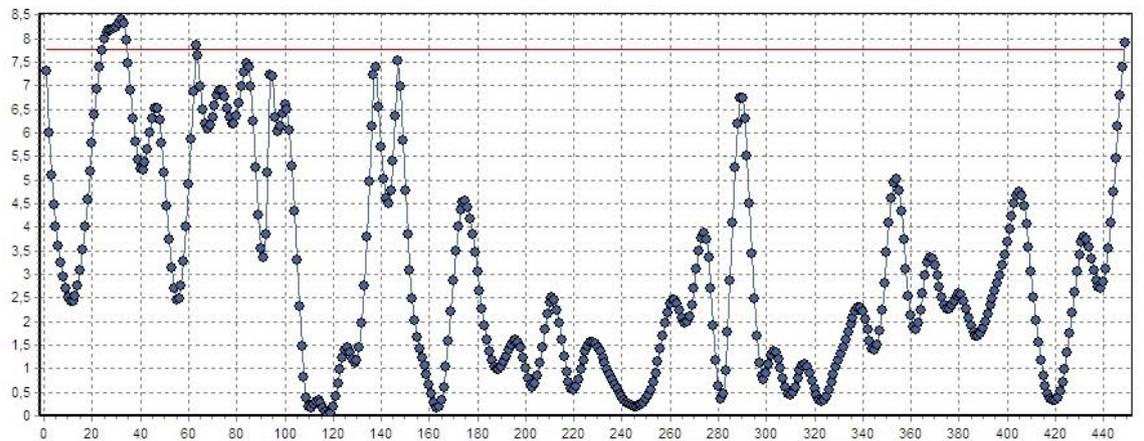


Figura 4.22: T^2 Control Chart – Step 11.

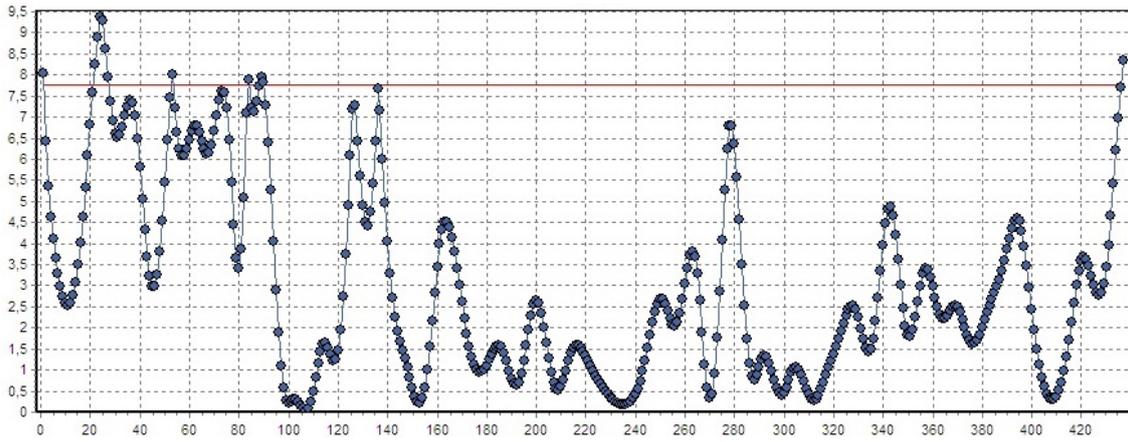


Figura 4.23: T^2 Control Chart – Step 12.

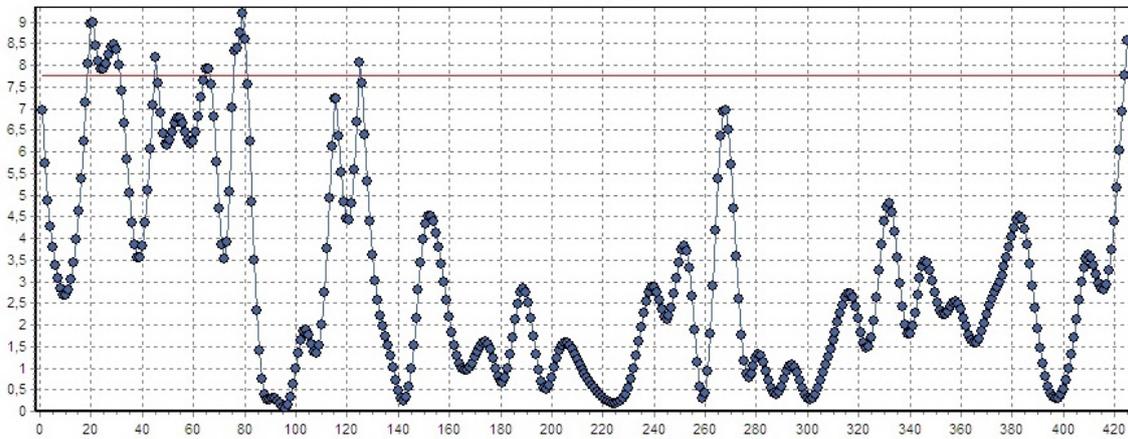


Figura 4.24: T^2 Control Chart – Step 13.

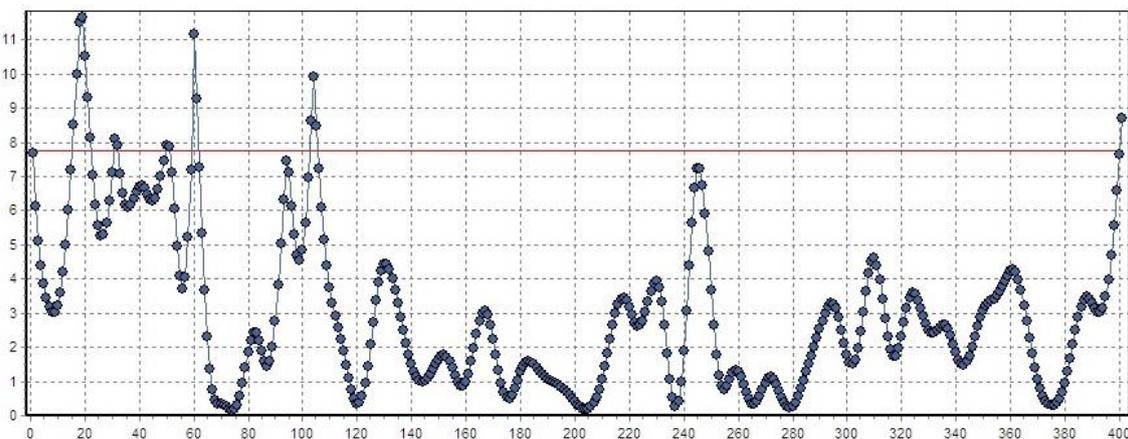


Figura 4.25: T^2 Control Chart – Step 14.

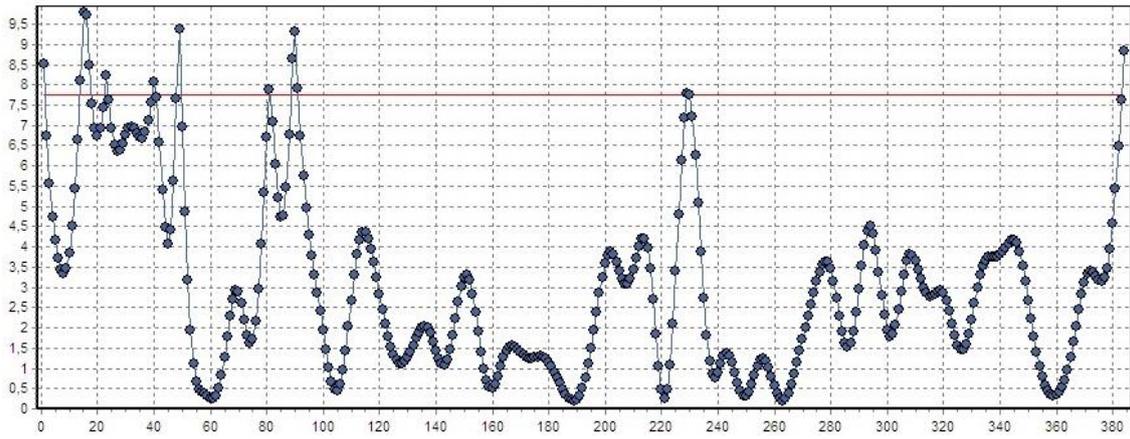


Figura 4.26: T^2 Control Chart – Step 15.

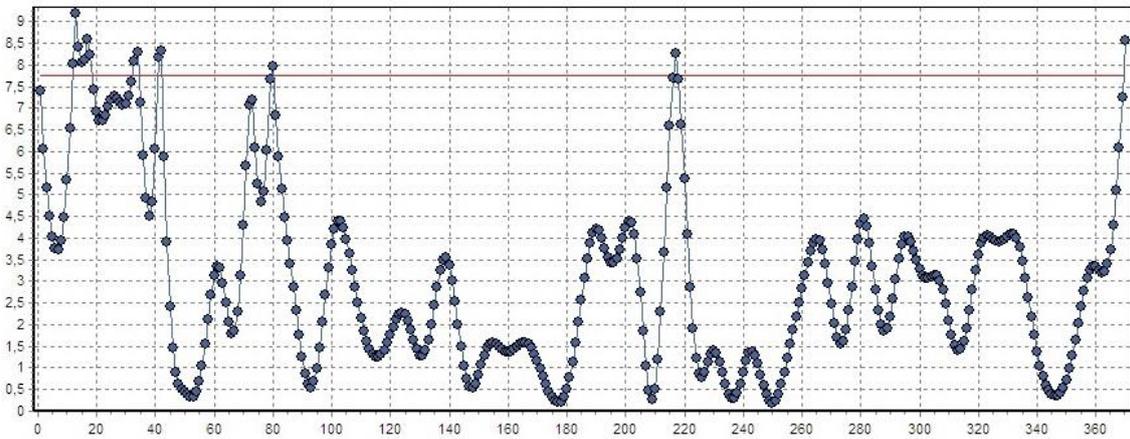


Figura 4.27: T^2 Control Chart – Step 16.

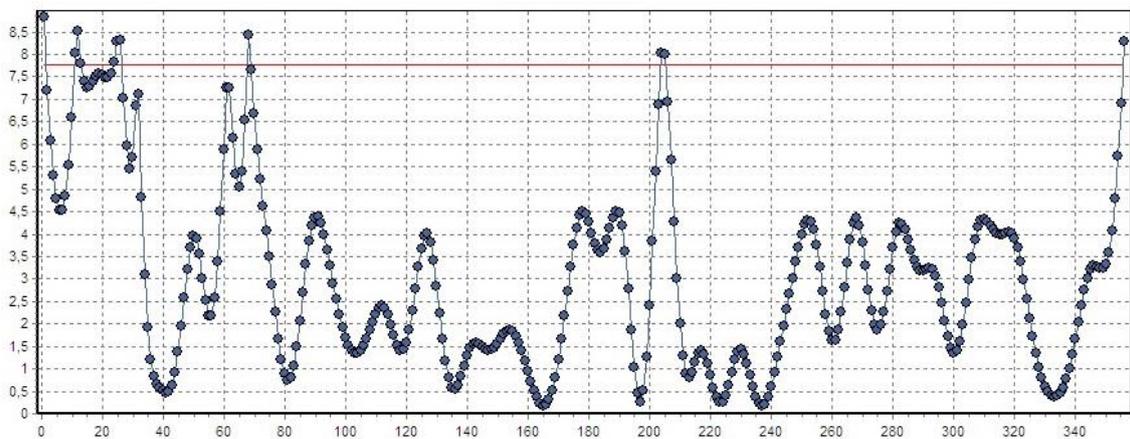


Figura 4.28: T^2 Control Chart – Step 17.

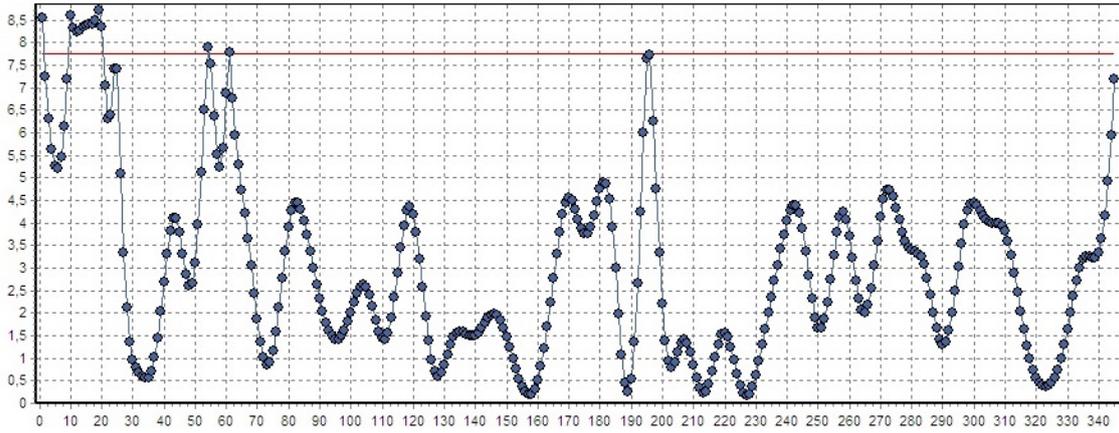


Figura 4.29: T^2 Control Chart – Step 18.

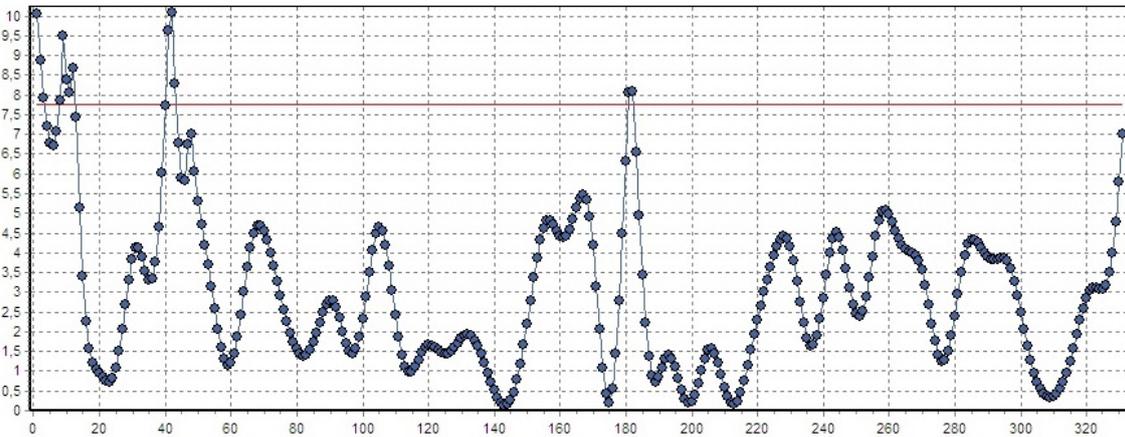


Figura 4.30: T^2 Control Chart – Step 19.

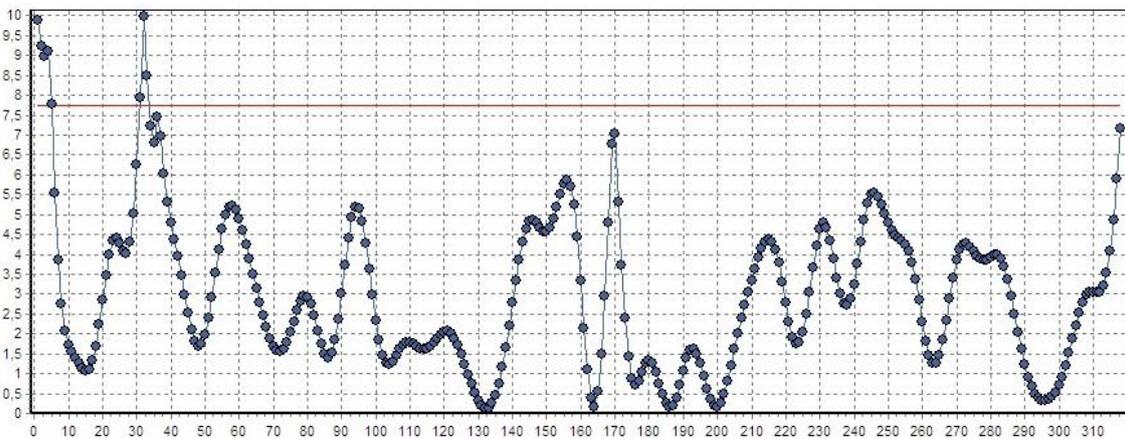


Figura 4.31: T^2 Control Chart – Step 20.

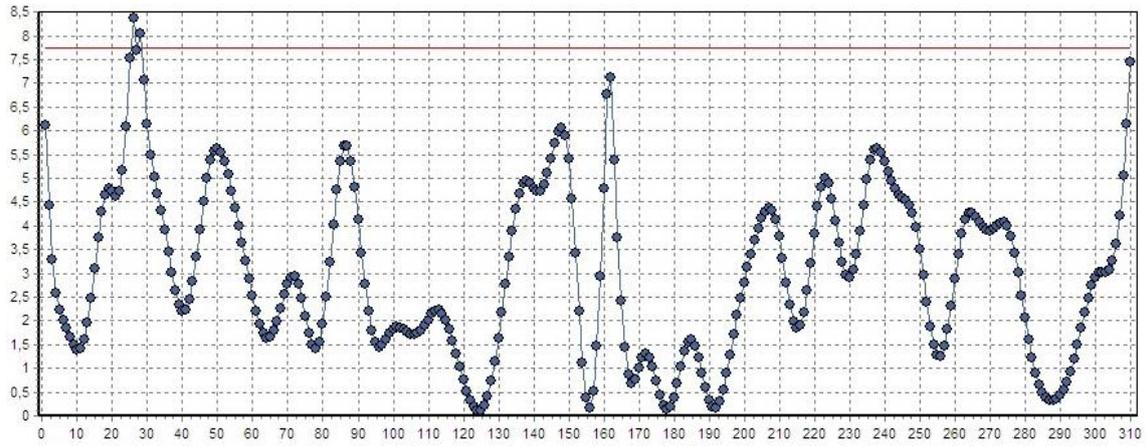


Figura 4.32: T^2 Control Chart – Step 21.

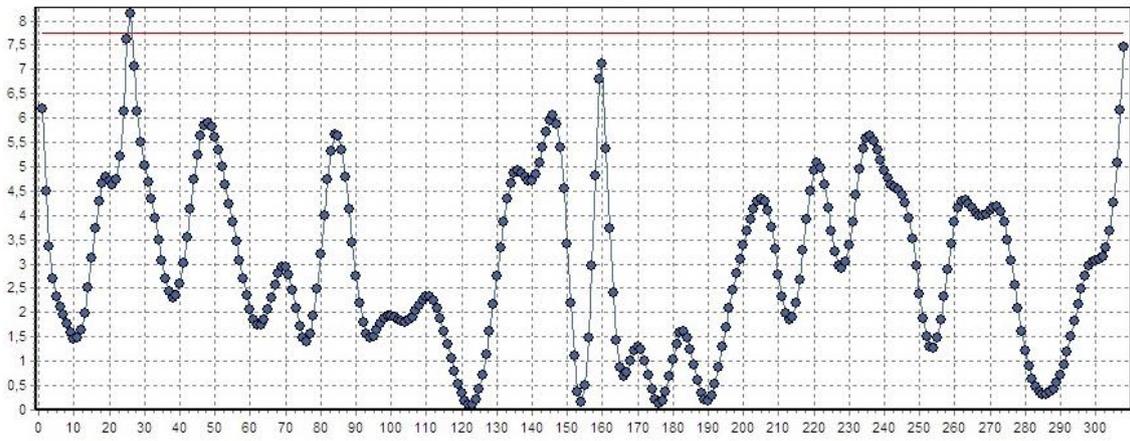


Figura 4.33: T^2 Control Chart – Step 22.

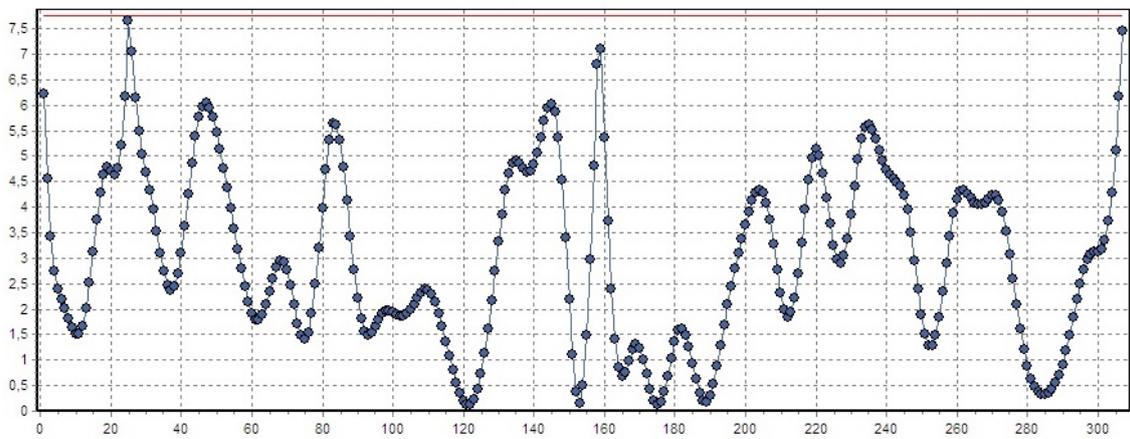


Figura 4.34: T^2 Control Chart – Step 23.

A continuación, en la tabla 4.32, se resumen los datos obtenidos en los gráficos, indicando el LCS calculado en cada paso y el número de outliers detectados hasta conseguir el HDS.

Tabla 4.32: Pasos hasta conseguir el HDS.

N.º de Observaciones	LCS	N.º outliers detectados
599	7,783	24
575	7,782	31
544	7,78	48
496	7,777	13
483	7,776	7
476	7,775	8
468	7,774	5
463	7,774	5
458	7,774	2
456	7,773	7
449	7,773	12
437	7,772	12
425	7,77	24
401	7,768	17
384	7,766	14
370	7,764	14
356	7,762	11
345	7,76	14
331	7,758	13
318	7,755	8
310	7,754	2
308	7,753	1
307	7,753	0

De la tabla se desprende un elevado número de outliers detectados, la generación de datos desde los datos adquiridos, ha podido ser la causa. Si el periodo de muestreo hubiese sido mayor, consideramos que lo ideal hubiesen sido dos muestras cada hora, hubiésemos tenido suficientes muestras disponibles para aplicar el método y el número de outliers creemos que hubiese sido menor.

4.3.2.7 Normalidad multivariante

La normalidad multivariante se comprobó mediante gráficos Q-Q, siguiendo una distribución beta.

En las figuras 4.35 y 4.36, se representa la normalidad multivariable de la base de datos preliminar, y después del proceso de purgado, HDS, respectivamente.

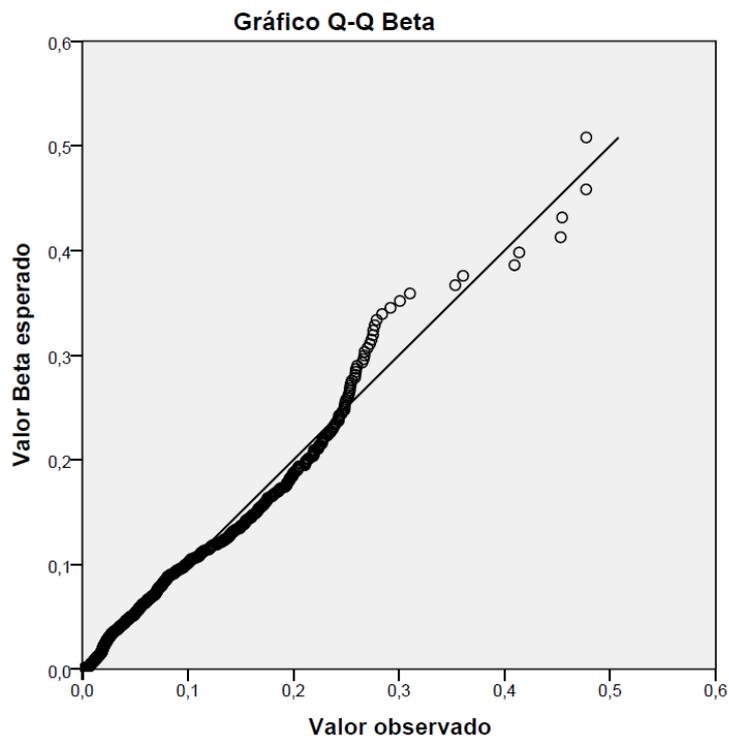


Figura 4.35: Análisis de normalidad multivariable, Base de datos preliminar.

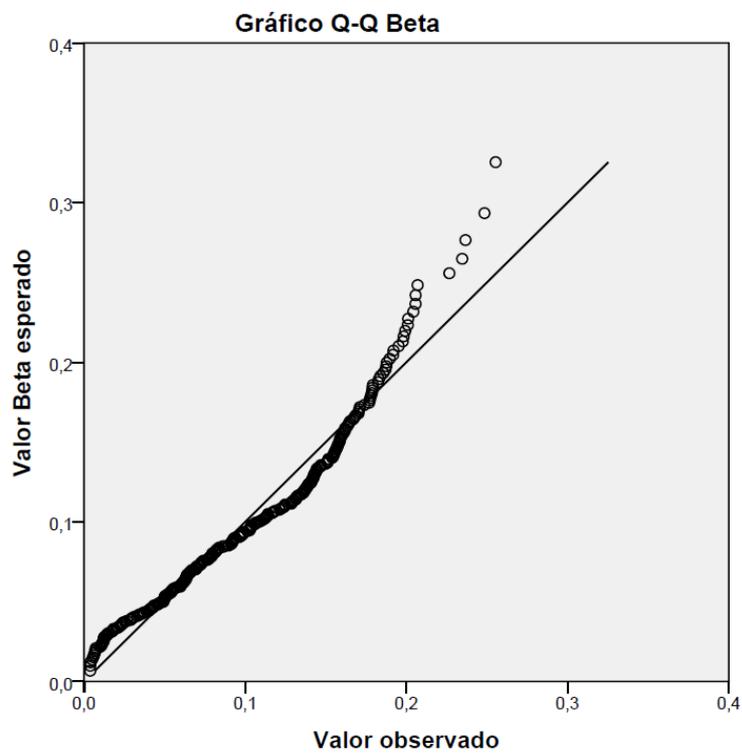


Figura 4.36: Análisis de normalidad multivariable, HDS.

4.3.2.8 Control del proceso

Una vez obtuvimos el HDS de las condiciones óptimas de funcionamiento en la Fase I, se comprobó si una nueva entrada de datos generaba alguna señal con respecto a las condiciones óptimas de funcionamiento.

La nueva entrada de datos, correspondió a muestras adquiridas durante un viaje en condición de carga, obteniendo un total de 13 muestras validas, representadas en la tabla 4.33, después de analizarlas según el criterio de las condiciones normales de operación, expuestas en el apartado 4.3.2.2.

Tabla 4.33: Nueva entrada de datos.

N.º Observación	Fuel index (%)	Exhaust gas temperature (°C)	Turbocharger speed (r.p.m.)
1	64,70	373	10.992
2	66,30	373	11.052
3	64,50	372	10.960
4	68,40	378	11.260
5	66,30	375	11.076
6	65,70	381	11.159
7	69,00	369	10.987
8	66,90	370	11.048
9	64,70	369	10.906
10	69,50	369	11.139
11	65,40	368	11.154
12	64,30	369	11.070
13	65,90	352	10.141

Los resultados obtenidos ante esta nueva entrada de datos, se muestran en el siguiente capítulo.

Capítulo 5

Resultados y Discusión

5. RESULTADOS Y DISCUSIÓN

En este apartado primero se muestran los resultados obtenidos en el proceso de lubricación de cilindros, posteriormente los obtenidos en el proceso de combustible y finalmente, se valida la metodología aplicada, comparándola con una técnica ampliamente reconocida y utilizada por numerosos autores, como es: (Hotelling + MYT).

5.1 Proceso de lubricación de cilindros

5.1.1 Aplicación estadístico T^2 de Hotelling

El proceso de lubricación de cilindros fue monitorizado por siete variables mediante la carta de control de T^2 Hotelling.

Primeramente, se calcularon los valores T^2 según la expresión (4.8), para cada una de las 23 nuevas observaciones, y se monitorizaron en una carta de control, según la figura 5.1. Con un Límite superior de control LCS = 14,59 previamente calculado, según la expresión (4.9).

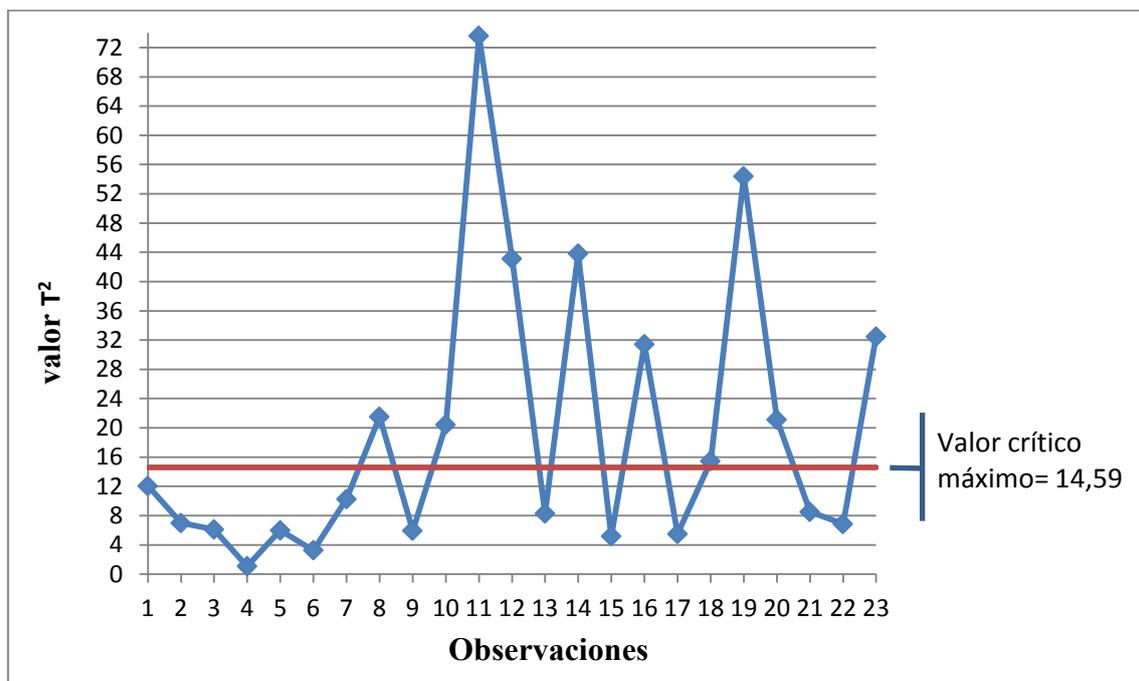


Figura 5.1: Carta de control.

De la carta de control se constata la existencia de valores de T^2 , por encima del LCS, lo cual nos indicó, que en ese intervalo de tiempo, el proceso, tuvo una desviación, con respecto a su funcionamiento de operación normal.

Esta situación, no quiere decir que el motor está fallando, si no que el proceso se ha movido desde sus condiciones de operación normal, pero si esta tendencia se repitiese en el tiempo, se deberían tomar acciones correctivas para prevenir una mal función en el proceso.

En la tabla 5.1, se muestra la cronología de las observaciones fuera de rango con respecto a sus respectivos valores T^2 .

Tabla 5.1: Cronología de observaciones.

Fecha	N.º Observaciones	Valor T^2
21/01/2016	8	21,515
23/01/2016	10	20,398
24/01/2016	11	73,573
25/01/2016	12	43,117
01/03/2016	14	43,837
03/03/2016	16	31,418
05/03/2016	18	15,437
06/03/2016	19	54,389
07/03/2016	20	21,132
13/03/2016	23	32,479

En la siguiente fase, se identificaron cuáles eran las variables que habían producido el estado fuera de rango de cada observación mediante la descomposición MYT.

5.1.2 Descomposición de los valores T^2

En esta fase, mediante la técnica de descomposición MYT, se descompuso cada valor de T^2 de cada una de las señales detectadas, valores superiores al límite superior de control, para detectar cual era la variable que más había contribuido al estado fuera de rango del proceso, los términos condicionales se calcularon siguiendo la expresión (4.14) y los términos incondicionales, se calcularon siguiendo la expresión (4.18). La

descomposición se presenta en la tabla 5.2, de la que se desprende que las variables (1, 4, 5 y 7), causaron el estado fuera de rango del proceso.

Tabla 5.2: Descomposición de los valores T².

N.º Observaciones	N.º Variable	Variables
8	7	MaximumPressure, P. max
10	4	<i>Hydraulic oil pressure After Pump 1</i>
11	4	<i>Hydraulic oil pressure After Pump 1</i>
12	4	<i>Hydraulic oil pressure After Pump 1</i>
14	5	<i>Total Oil</i>
16	5	<i>Total Oil</i>
18	4	<i>Hydraulic oil pressure After Pump 1</i>
19	7	MaximumPressure, P. max
20	5	<i>Total Oil</i>
23	1	Engine Load

En la tabla 5.2, se puede observar que las principales variables que causaron la desviación desde el modo de operación normal fueron las variables “5-Total oil” y “4-Hydraulic oil pressure After Pump 1”.

Un anómalo comportamiento en la variable 4, pudo causar fallos mecánicos, tales como desgastes o pérdida de compresión. Por otro lado, un incremento en ella pudo causar un excesivo consumo de aceite.

La desviación en la variable 5-Total Oil podría haber sido causada por un incremento en la dosificación de aceite en el cilindro, para prevenir un cambio en el azufre contenido en el último bunkering de combustible.

El método fue eficiente desde el punto de vista que detecto observaciones fuera de rango en el proceso e identificó las variables que generaron ese estado. Si el número de variables a monitorizar hubiese sido mayor, la identificación de las variables causantes de la desviación en el proceso, hubiese sido confusa e incluso imposible. Además, hubo un riesgo de no detectar cambios pequeños y progresivos en el proceso, debido a las dificultades que tienen estos gráficos de control para detectar este tipo de comportamiento.

Para resolver estas dificultades, se planteó en el capítulo anterior una metodología denominada (SSDM), la cual se ha aplicado en el proceso de combustible y cuyos resultados se muestran a continuación.

5.2 Proceso de combustible

5.2.1 Aplicación estadístico T^2 de Hotelling

El proceso de combustible del motor principal fue monitorizado por solo tres variables con baja correlación entre ellas, mediante la metodología (SSDM), combinación de técnicas multivariantes y univariantes.

Primeramente, se calcularon los valores T^2 según la expresión (4.8), para cada una de las 13 nuevas observaciones, y se monitorizaron en una carta de control, según figura 5.2. Con un Límite superior de control LCS = 7,981 previamente calculado, según la expresión (4.9).

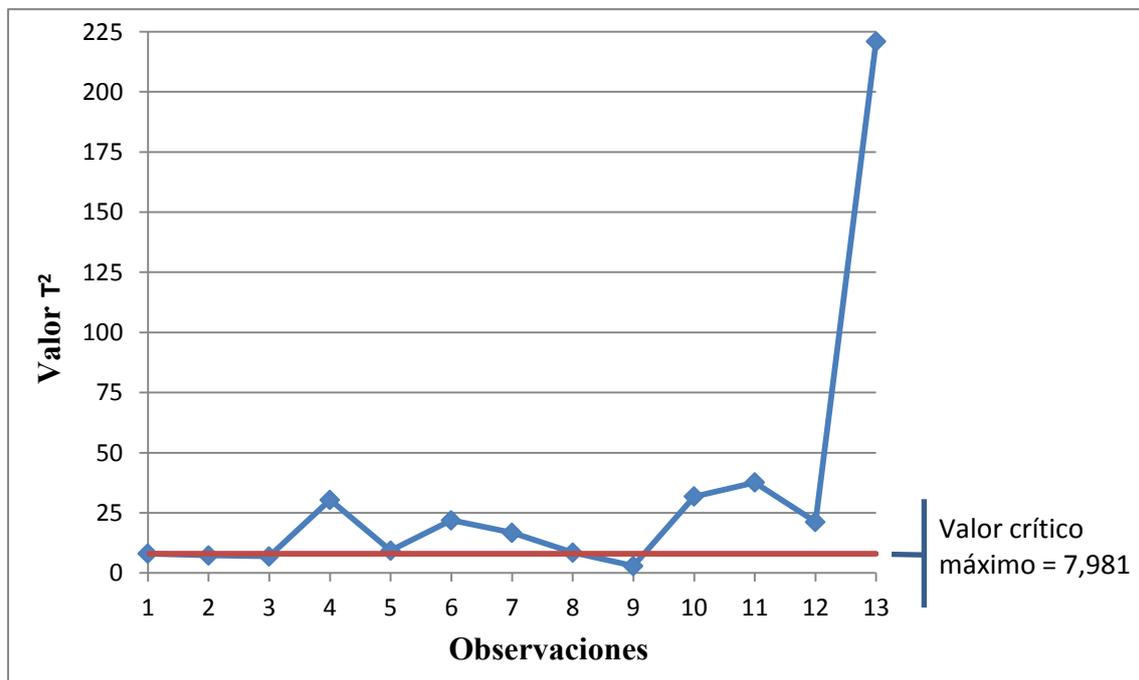


Figura 5.2: Carta de control.

De la carta de control se desprende que existen valores de T^2 , por encima del LCS, lo cual nos indicó, que en ese intervalo de tiempo, el proceso, tuvo una desviación, con respecto a su funcionamiento de operación normal.

Esta situación, no quiere decir que el motor está fallando, si no que el proceso se ha desplazado desde sus condiciones de operación normal, pero si esta tendencia se repite en el tiempo, se deberían tomar acciones correctoras para prevenir un mal funcionamiento en el proceso.

En la siguiente fase, se identificaron cuáles eran las variables que habían producido el estado fuera de rango de cada observación mediante la descomposición MYT.

5.2.2 Descomposición de los valores T^2

En esta fase, mediante la técnica de descomposición MYT, se descompuso cada valor de T^2 de cada una de las señales detectadas, valores superiores al límite superior de control, para detectar cual era la variable que más había contribuido al estado fuera de rango del proceso, los términos condicionales se calcularon siguiendo la expresión (4.14) y los términos incondicionales, se calcularon siguiendo la expresión (4.18). La descomposición se presenta en la tabla 5.3, de la que se desprende que la principal variable que causó la desviación desde su modo de operación normal fue la variable Turbocharger speed.

Tabla 5.3: Descomposición de los valores T^2 .

N.º Observaciones	Variables
1	Exhaust gas temperature
4	<i>Turbocharger speed</i>
5	<i>Turbocharger speed</i>
6	Exhaust gas temperature / <i>Turbocharger speed</i>
7	Fuel Index
8	<i>Turbocharger speed</i>
10	Fuel Index
11	<i>Turbocharger speed</i>
12	<i>Turbocharger speed</i>
13	<i>Turbocharger speed</i>

5.2.3 Suma acumulada, Cusum

En esta fase, se monitorizaron las variables predictivas de la variable Turbocharger speed, mediante la técnica de suma acumulada Cusum, para intentar detectar cambios pequeños y progresivos en el proceso, que no hubiesen sido detectados por medio de los

gráficos multivariantes. Los valores de las medias y desviaciones típicas de cada una de las variables predictivas en control se presentan en la tabla 5.4.

Tabla 5.4: Medias y desviaciones típicas de las variables predictivas.

VARIABLES	UNIDADES	Media(μ)	Desviación típica (σ)
Scavenge air cooler air inlet temperature	°C	147,35	2,24
P (scav)	Bar	1,84	0,10
Estimate Effective Power	kW	10.542,36	103,63
Pcom	Bar	124,80	5,38
SFOC	g/kWh	158,58	1,70

Se monitorizaron 61 observaciones para cada una de las variables; las 48 primeras observaciones correspondieron al ARL_0 y las 13 siguientes fueron los nuevos datos de entrada.

Se estableció un intervalo de decisión, inferior al marcado por el fabricante, un 15 % sobre el valor de la media de cada variable en condiciones óptimas de operación.

Las figuras 5.3 a 5.7, muestran los gráficos Cusum para cada una de las variables.

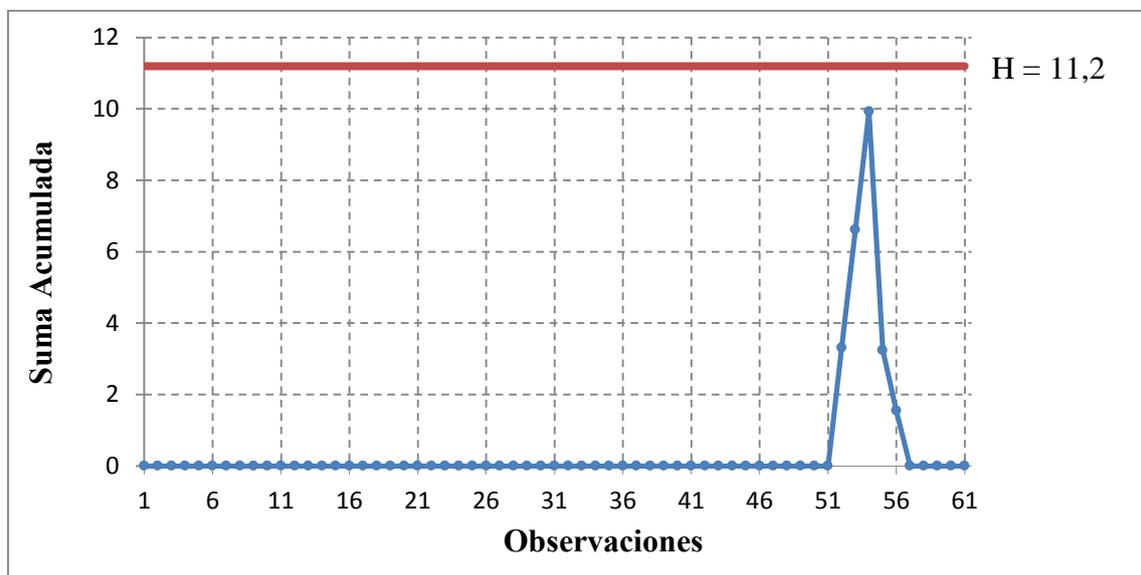


Figura 5.3: Scavenge air cooler air inlet temperature ($ARL_1=1,23$).

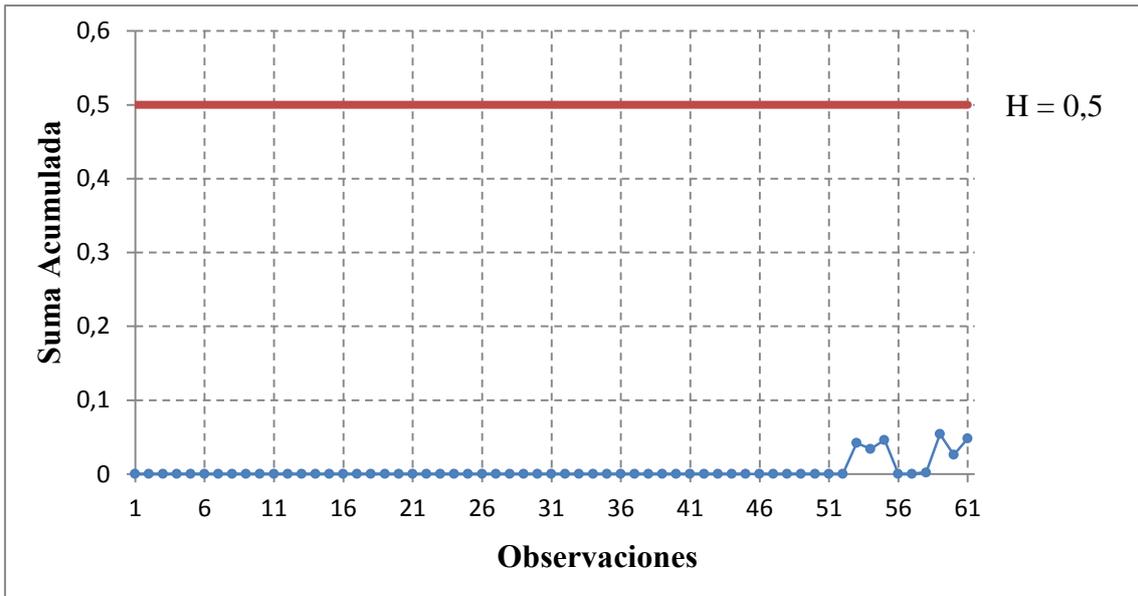


Figura 5.4: P (scav) ($ARL_1=4,29$).

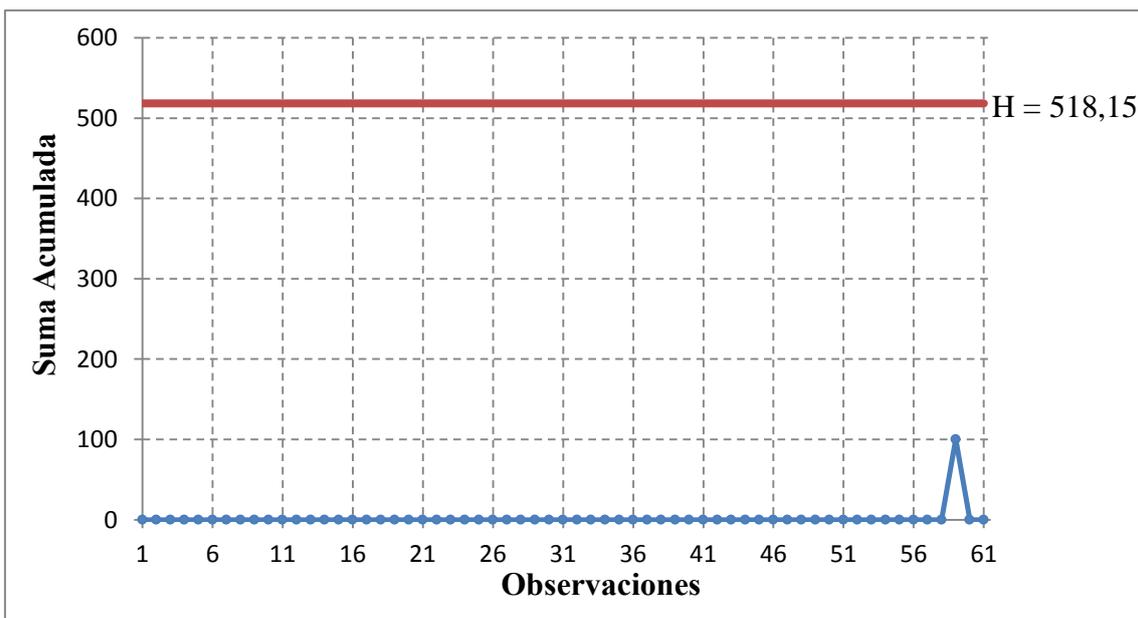


Figura 5.5: Estimate Effective Power ($ARL_1=0,79$).

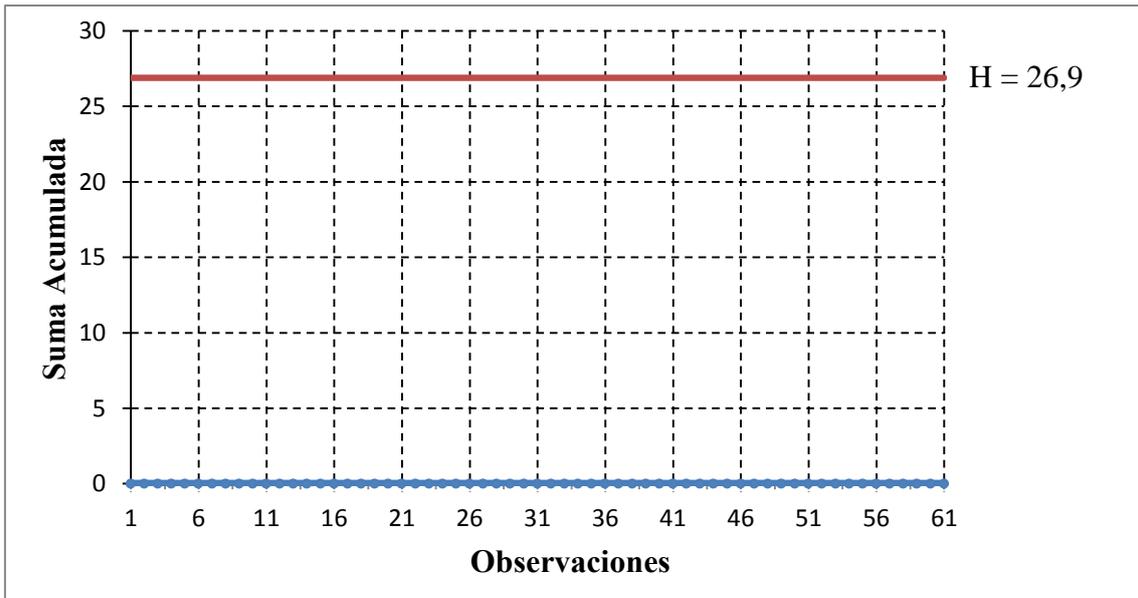


Figura 5.6: Pcom ($ARL_1=3,37$).

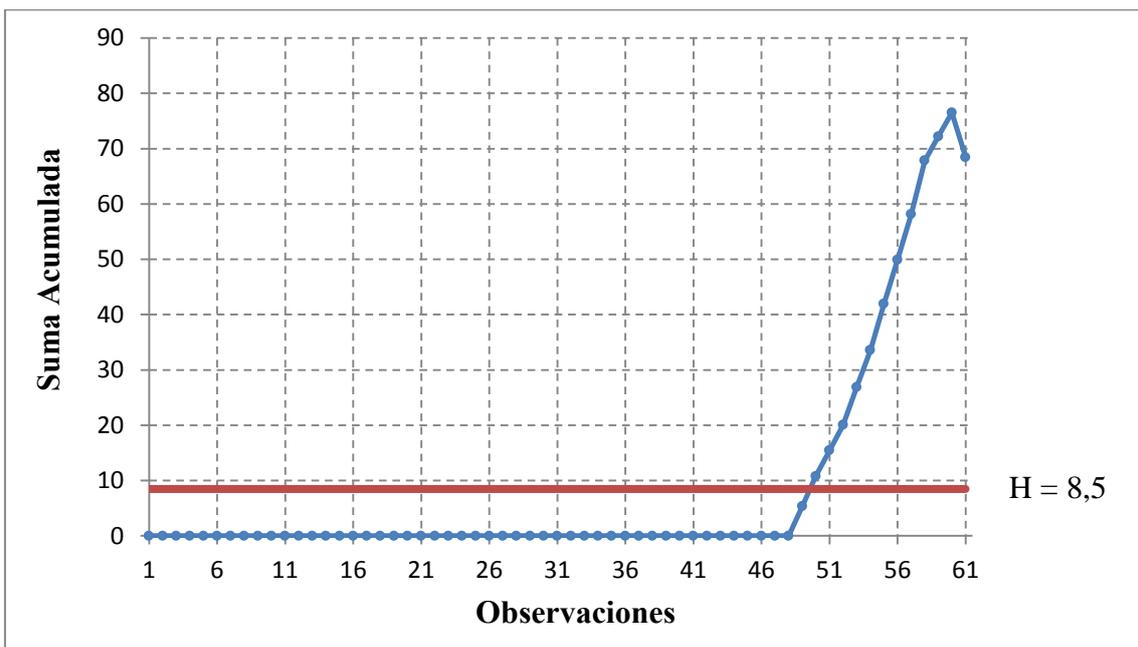


Figura 5.7: SFOC ($ARL_1=0,87$).

Se observa que la única variable que supero su intervalo de decisión fue la variable SFOC, donde la muestra 50 es $C_{50}^+ = 10,8$. Esta es la primera muestra en la que $C_i^+ > H=8,5$. Por lo que la variable estuvo fuera de rango en ese momento.

Sin embargo, el gráfico Cusum, también indica cuando comenzó a originarse el cambio. Podemos observar que la primera muestra consecutiva en la que $C_i^+ > 0$ antes de superar el valor de H, es la muestra 49, $C_{49}^+ = 5,36$, esto indicó que el desajuste en la variable podría haber comenzado en la muestra 49.

A través de un análisis de regresión multivariable, las once variables que definieron el proceso de combustible, fue reducido únicamente a tres variables, con coeficientes de determinación R^2 ajustado superiores a 0,8 entre ellas y sus variables predictivas, asegurando que la información contenida en ellas era capaz de representar al proceso completo.

Las tres variables: (Fuel Index, Exhaust gas temperature at turbine inlet y Turbucharger speed), fueron monitorizadas a través de gráficos multivariados, T^2 de Hotelling, detectando eficazmente cambios bruscos en el proceso. En la tabla 5.5, se presenta la cronología de las observaciones fuera de rango detectadas, con sus respectivos valores de T^2 .

Tabla 5.5: Cronología de observaciones fuera de rango.

Fecha	N.º Observaciones	T^2
25/08/2016	1	8.042
28/08/2016	4	30.395
29/08/2016	5	9.294
30/08/2016	6	21.884
31/08/2016	7	16.745
01/09/2016	8	8.489
03/09/2016	10	31.785
04/09/2016	11	37.641
05/09/2016	12	21.212
06/09/2016	13	221.034

Las observaciones que estaban por encima del límite establecido fueron descompuestas, identificando la variable Turbocharger speed, como la principal variable que origino el estado fuera de rango en el proceso multivariable.

Posteriormente, con el objetivo de detectar cambios pequeños y progresivos en el proceso, que no hubiesen podido ser detectados por los gráficos multivariados, cada una de las variables predictivas de la variable Turbocharger speed, fueron monitorizadas a través de gráficos univariantes Cusum.

Se estableció un intervalo de decisión menor que el que tiene establecido el fabricante del motor en su sistema de diagnóstico y alarma; el intervalo de decisión, correspondió al 15% sobre el valor medio de cada variable en óptimas condiciones de funcionamiento.

Bajo estas condiciones, la variable SFOC, excedió el límite establecido y los gráficos Cusum, detectaron cuando se produjo el comienzo de la desviación desde su operación normal antes de llegar al umbral establecido.

Por razones de operatividad del buque, la última limpieza del sistema de aire de carga en profundidad, se había realizado seis meses antes de la última toma de datos, y la limpieza del intercambiador de calor del sistema de admisión de aire, solo se pudo llevar a cabo mediante productos químicos; esta situación generó un ensuciamiento progresivo del intercambiador de calor, que redujo la eficiencia del motor.

El cambio en el valor de la presión diferencial en el intercambiador de calor, no fue suficientemente elevado, como para superar el umbral marcado por el fabricante, que provocó un incremento en el consumo de combustible de manera progresiva, y con desviaciones pequeñas con respecto a su media, para mantener la velocidad establecida del buque.

En el siguiente apartado se validarán los resultados obtenidos, comparándolos con una técnica ampliamente reconocida.

5.3 Validación de la metodología propuesta

Que mejor forma de validar los resultados obtenidos a través de la metodología desarrollada en la presente tesis doctoral, que comparándola con un método ampliamente reconocido como es: (T^2 Hotelling + MYT).

Como se ha comentado a lo largo de la presente tesis, este método (T^2 Hotelling + MYT), tiene dificultades para desarrollar toda su potencia cuando el número de variables a monitorizar es elevado.

Esta técnica la aplicaremos a las once variables que definen el proceso de combustible:

- 1- Engine Load (%).
- 2- Fuel Index (%).

- 3- Fuel Plunge Stroke (mA).
- 4- Scavenge air cooler air inlet temperatura (°C).
- 5- Exhaust gas temperature at turbine inlet (°C).
- 6- Turbocharger speed Rpm (r.p.m.).
- 7- P(scav) (Bar).
- 8- Estimate Effective Power (kW).
- 9- Pcom (Bar).
- 10- Pmax (Bar).
- 11- SFOC (g/kWh).

Si el número de variables hubiese sido mayor, los resultados obtenidos hubiesen sido tan confusos que la veracidad de los resultados hubiese sido nula y no hubiese sido posible realizar esta validación.

La primera fase, fue construir el HDS del total de las variables que definieron el proceso de combustible. Los valores máximos, mínimos, medias y desviaciones típicas de cada una de las 11 variables de la base de datos preliminar formada por 599 muestras, se presentó en la tabla 4.28.

Se utilizó la distribución beta en el proceso de purgado de outliers, producidos por errores en las medidas. Considerando un proceso de purgado donde un único vector de 599 muestras de la base de datos preliminar $X^t=(x_1, x_2, \dots, x_p)$ estaba siendo monitorizado para el control del proceso utilizando una carta de control T^2 . Nosotros asumimos que los datos seguían una distribución normal multivariante con un vector de medias desconocido μ y matriz de covarianza desconocida Σ .

Desde la base de datos preliminar, nosotros obtuvimos estimaciones de \bar{X} y S de μ y Σ , respectivamente. Nosotros comenzamos el proceso de purgado a través de la base de datos preliminar. Para un $\alpha=0,05$, todos los vectores observación cuyos valores eran menores o iguales al límite de control superior (LCS) permanecieron en la base de datos, Los valores T^2 fueron calculados mediante la ecuación (4.6):

$$T^2 = (X_i - \bar{X})^t \cdot S^{-1} \cdot (X_i - \bar{X}) \quad (4.6)$$

Donde el límite superior de control fue determinado por la ecuación (4.7):

$$LCS = \left\{ \frac{(n - 1)^2}{n} \right\} \beta_{\left\{ \alpha; \frac{p}{2}; (n-p-1)/2 \right\}} \tag{4.7}$$

Dónde $\beta_{\left\{ \alpha; p/2; (n - p - 1)/2 \right\}}$, es el ath, quantile de la distribución beta, $\beta_{\left\{ p/2; (n - p - 1)/2 \right\}}$

Si el valor de T^2 que se monitorizo para una observación, excedía del LCS, la observación era eliminada de la base de datos preliminar.

Con las restantes observaciones, calculamos un nuevo vector de medias y matriz de covarianza. Otra vez, se detectaron y eliminaron outliers, y se repitió el proceso, hasta obtener un set de observaciones homogéneo. El set final de datos fue el (HDS), del modo de operación normal del proceso, formado por 333 muestras, En la tabla 5.6, se presenta parte del HDS.

Tabla 5.6: Parte del HDS.

		Número de identificación de variables										
		1	2	3	4	5	6	7	8	9	10	11
Número de Muestra	1	55,81	64,53	2,67	145,20	401,10	10.490,52	1,59	10.395,79	110,97	139,82	154,98
	2	56,30	65,04	2,68	145,81	401,61	10.504,45	1,59	10.418,04	111,14	140,04	155,20
	3	56,81	65,55	2,69	146,44	402,14	10.518,42	1,60	10.441,52	111,31	140,25	155,43
	4	57,30	66,04	2,70	147,08	402,68	10.532,25	1,60	10.465,55	111,47	140,46	155,67
	5	57,77	66,49	2,71	147,70	403,20	10.545,74	1,61	10.489,41	111,64	140,65	155,90
	6	58,19	66,90	2,71	148,31	403,70	10.558,70	1,61	10.512,39	111,79	140,81	156,11
	7	58,54	67,23	2,72	148,88	404,16	10.570,92	1,61	10.533,79	111,93	140,95	156,29
	8	58,82	67,47	2,72	149,40	404,57	10.582,22	1,62	10.552,89	112,06	141,06	156,44
	9	59,00	67,60	2,72	149,87	404,92	10.592,39	1,62	10.569,00	112,17	141,13	156,55
	10	59,07	67,61	2,72	150,27	405,20	10.601,27	1,62	10.581,57	112,26	141,16	156,61
	11	59,04	67,52	2,71	150,59	405,40	10.608,76	1,62	10.590,72	112,32	141,15	156,63
	12	58,93	67,33	2,71	150,85	405,54	10.614,82	1,63	10.596,76	112,36	141,10	156,61
	13	58,74	67,07	2,70	151,03	405,62	10.619,40	1,63	10.599,97	112,38	141,02	156,57
	14	58,49	66,75	2,69	151,13	405,65	10.622,42	1,63	10.600,65	112,38	140,92	156,50
	15	58,20	66,38	2,68	151,16	405,63	10.623,83	1,62	10.599,08	112,36	140,80	156,42
	16	57,88	65,99	2,67	151,11	405,56	10.623,58	1,62	10.595,57	112,32	140,67	156,34
	17	57,53	65,59	2,66	150,98	405,46	10.621,61	1,62	10.590,40	112,27	140,52	156,25
	18	57,18	65,19	2,65	150,78	405,33	10.617,85	1,62	10.583,87	112,19	140,37	156,18
	19	56,84	64,82	2,64	150,49	405,17	10.612,27	1,62	10.576,28	112,10	140,22	156,12
	20	56,52	64,49	2,64	150,12	404,98	10.604,79	1,62	10.567,90	111,99	140,08	156,09
	21	56,24	64,21	2,63	149,67	404,79	10.595,35	1,61	10.559,05	111,86	139,94	156,08
	22	56,00	64,00	2,63	149,14	404,58	10.583,91	1,61	10.550,00	111,72	139,82	156,12

En la segunda fase, se comprobó si una nueva entrada de datos generaba alguna señal con respecto al HDS. Nosotros asumimos que el proceso estaba siendo monitorizado por observaciones de un simple vector de 13 nuevas muestras, $X^t=(x_1, x_2, \dots, x_p)$ en p variables para cada punto, presentadas en la tabla 5.7. Estas muestras fueron adquiridas durante un viaje en condición de carga, después de analizarlas según los criterios de las condiciones normales de operación, expuestas en el apartado 4.3.2.2

Tabla 5.7: Nueva entrada de datos.

		Número de identificación de variables										
		1	2	3	4	5	6	7	8	9	10	11
Número de Muestra	1	57	64,70	2,67	155	373	10.992	1,89	10.808	126,98	140,80	176,27
	2	57	66,30	2,73	155	373	11.052	1,89	10.782	127,17	141,80	176,35
	3	58	64,50	2,73	155	372	10.960	1,90	10.758	127,30	141,80	175,55
	4	61	68,40	2,77	160	378	11.260	2,01	11.128	130,32	142,10	175,53
	5	59	66,30	2,73	160	375	11.076	2,06	11.229	130,47	142,10	177,71
	6	60	65,70	2,75	160	381	11.159	2,01	11.144	128,45	141,40	177,63
	7	61	69,00	2,83	150	369	10.987	2,03	11.173	130,60	143,10	179,27
	8	57	66,90	2,78	155	370	11.048	1,95	10.947	128,98	142,50	178,86
	9	57	64,70	2,70	150	369	10.906	1,93	10.849	127,72	141,90	179,21
	10	62	69,50	2,75	155	369	11.139	2,02	11.267	131,63	143,20	180,57
	11	60	65,40	2,80	155	368	11.154	2,07	11.410	133,03	144,50	175,27
	12	58	64,30	2,73	155	369	11.070	1,99	11.111	130,42	142,80	175,22
	13	59	65,90	2,80	150	352	10.141	2,04	11.195	135,10	145,60	162,86

Los valores T^2 de Hotelling asociados con X , fueron calculados siguiendo la expresión (4.8):

$$T^2 = (X_i - \bar{X})^t \cdot S^{-1} \cdot (X_i - \bar{X}) \tag{4.8}$$

Donde, las estimaciones de \bar{X} y S^{-1} , son obtenidas desde el HDS. Aquí, el estadístico T^2 sigue la distribución F .

Para un $\alpha=0,05$, el LCS fue calculado siguiendo la expresión (4.9):

$$LCS = \left\{ \frac{p(n + 1)(n - 1)}{n(n - p)} \right\} F_{\{\alpha;p;(n-p)\}} \tag{4.9}$$

Dónde p , es el número de variables, n , el tamaño del HDS y $F\{\alpha; p; (n - p)\}$, es el α th, quantile de $F\{p; (n - p)\}$.

El LCS calculado fue 20,68 y los valores de T^2 para cada una de las observaciones se monitorizaron en la carta de control, según figura 5.8.

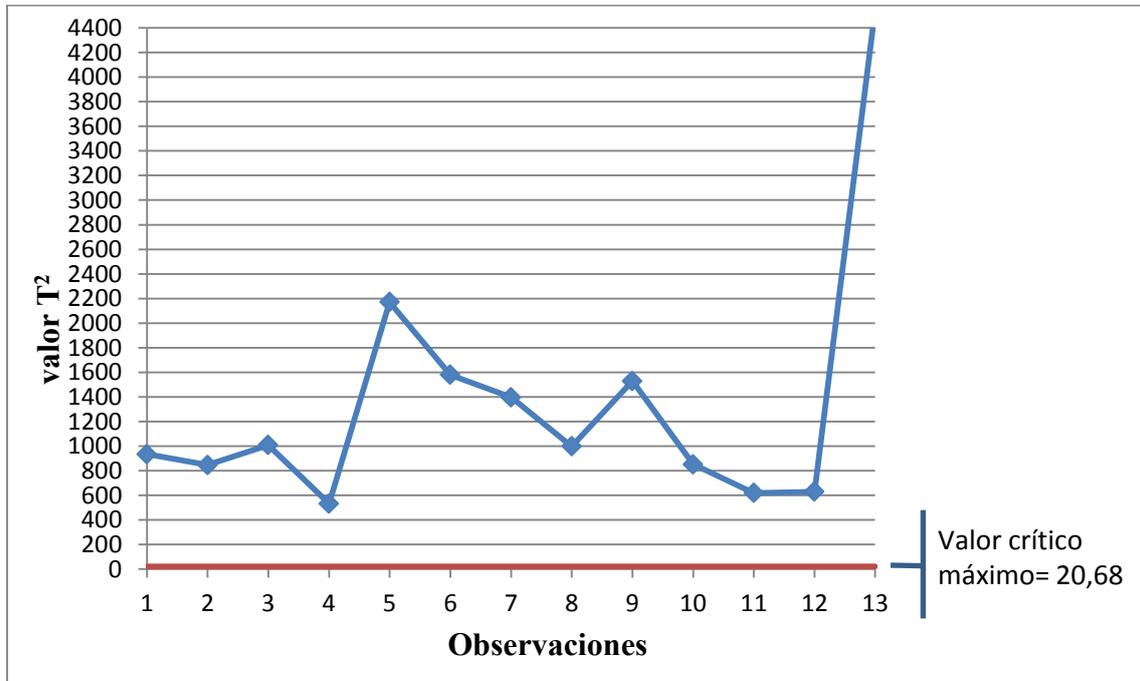


Figura 5.8: Carta de control.

La carta de control, detecto que todas las observaciones estaban fuera de rango para el LCS calculado.

Una vez detectado observaciones fuera de rango, se identificaron las variables que mayormente habían contribuido a que el valor de T^2 estuviese por encima del límite de control superior; Para ello, se descompuso cada observación mediante la técnica de descomposición MYT.

La descomposición MYT para cada una de las observaciones, se realizó a través del software Qualtstat y se presentan a continuación en las figuras 5.9 a 5.21.

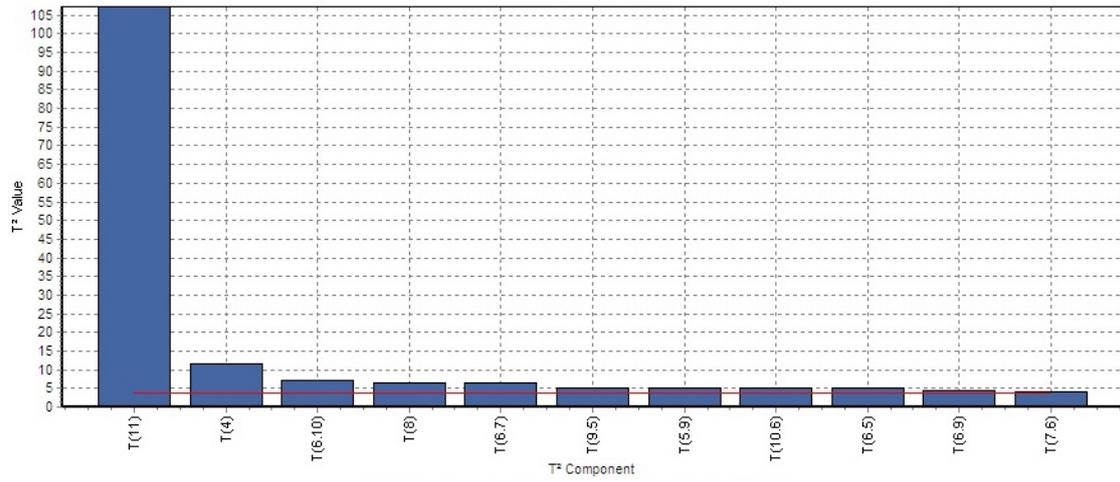


Figura 5.9: Descomposición T². Observación 1.

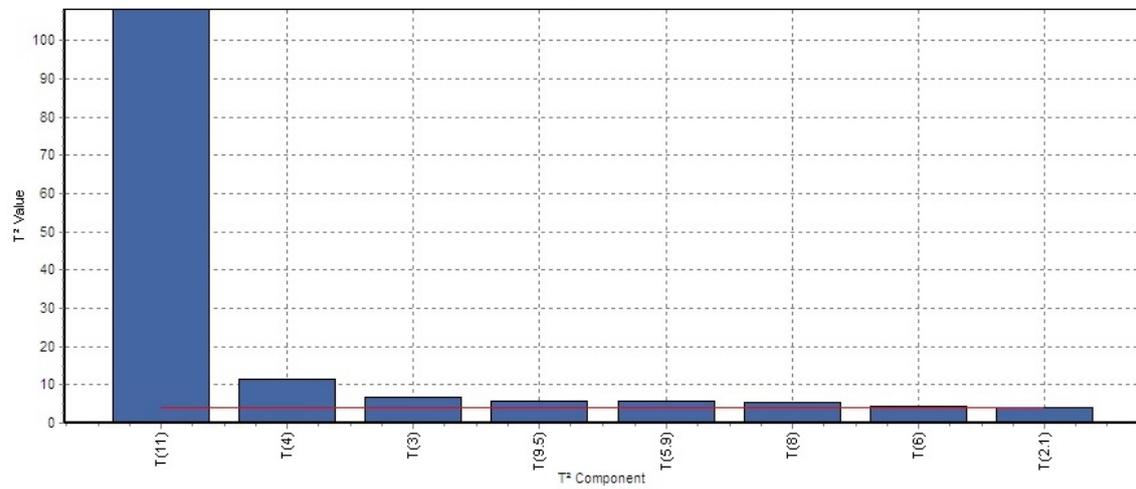


Figura 5.10: Descomposición T². Observación 2.

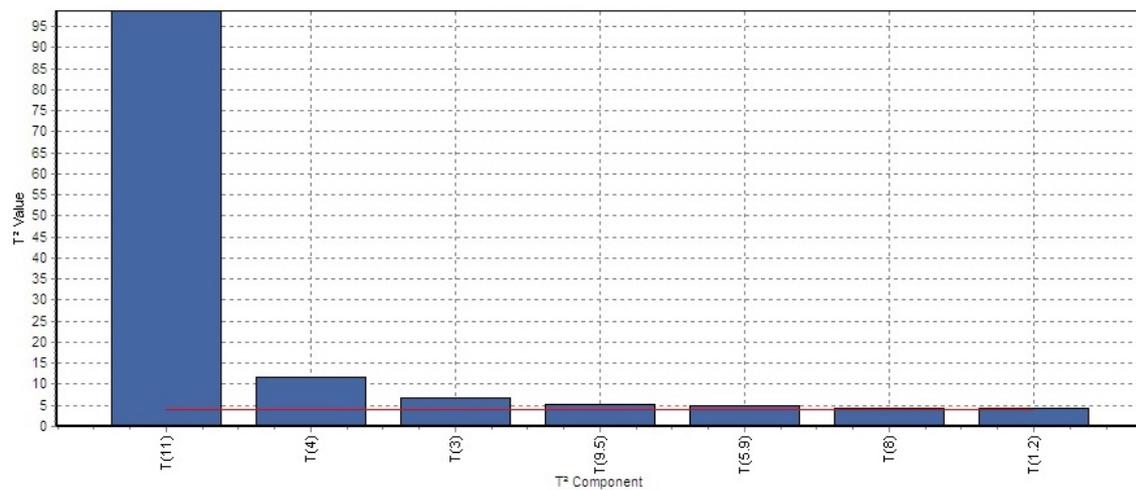


Figura 5.11: Descomposición T². Observación 3.

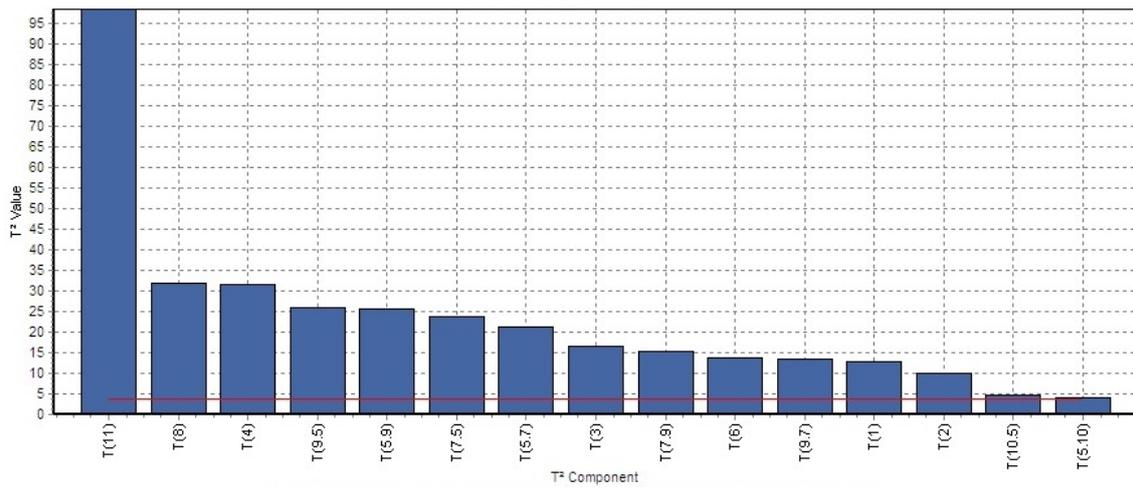


Figura 5.12: Descomposición T². Observación 4.

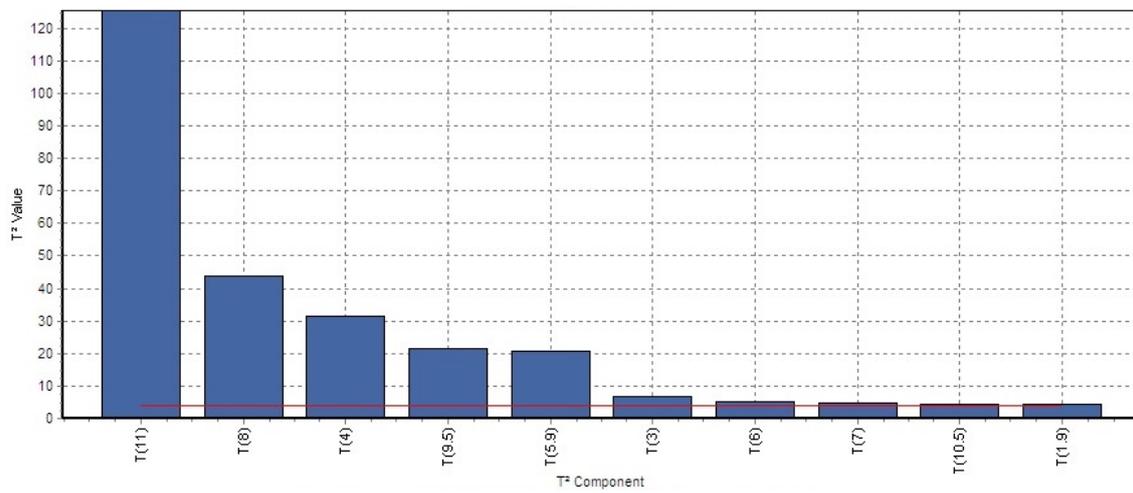


Figura 5.13: Descomposición T². Observación 5.

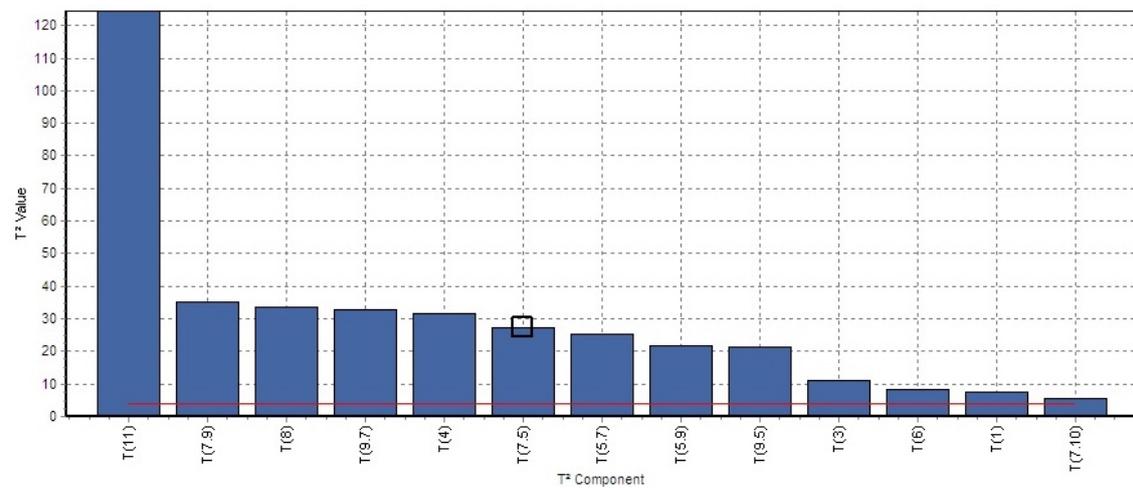


Figura 5.14: Descomposición T². Observación 6.

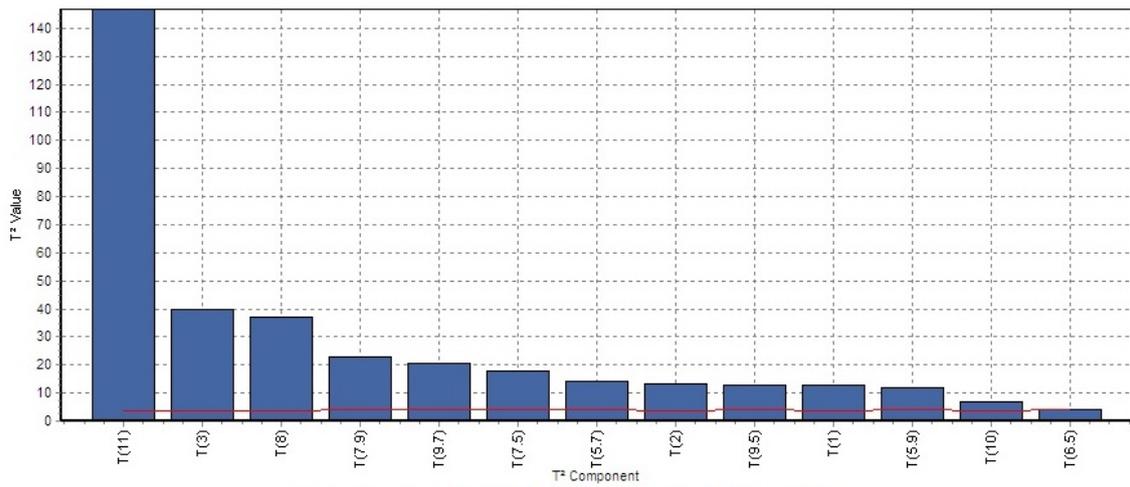


Figura 5.15: Descomposición T². Observación 7.

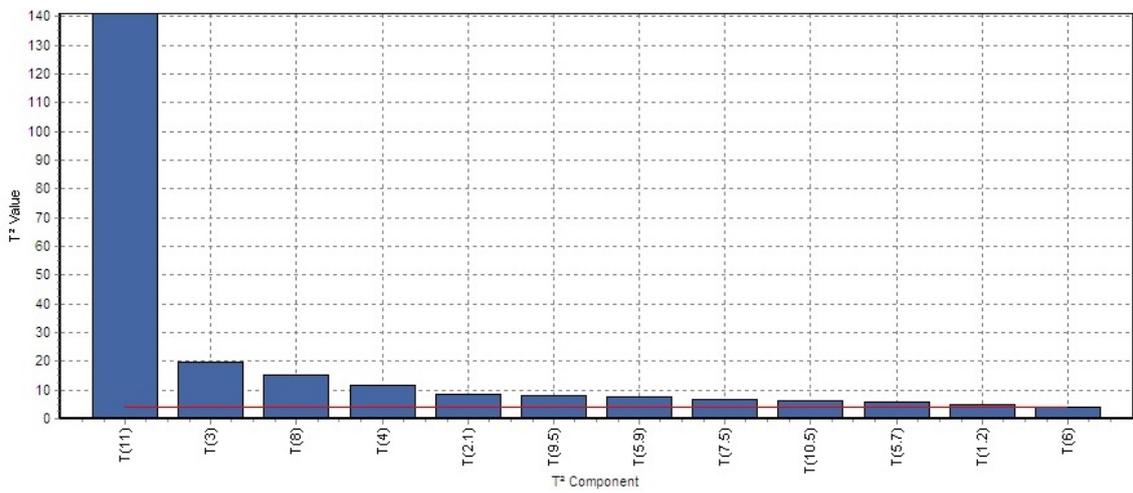


Figura 5.16: Descomposición T². Observación 8.

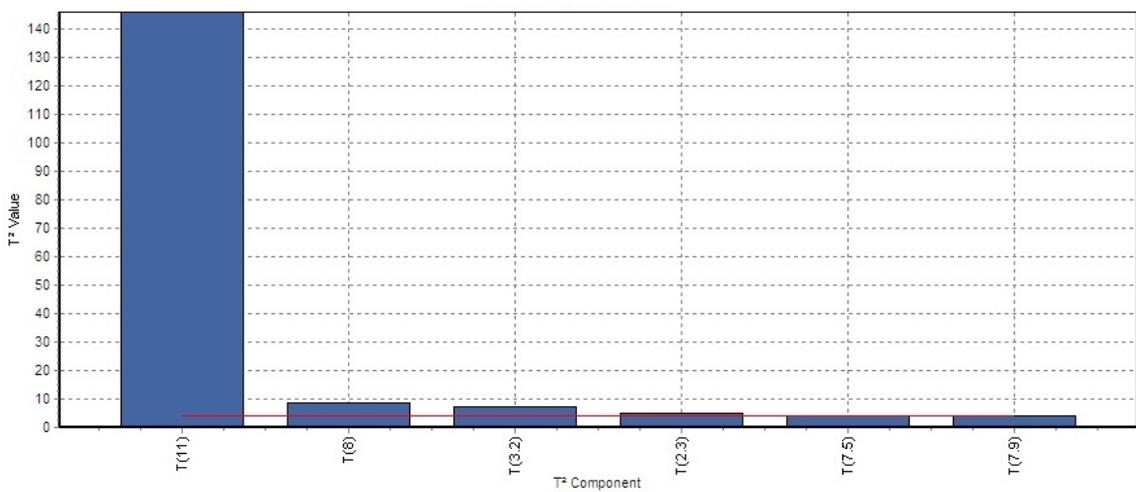


Figura 5.17: Descomposición T². Observación 9.

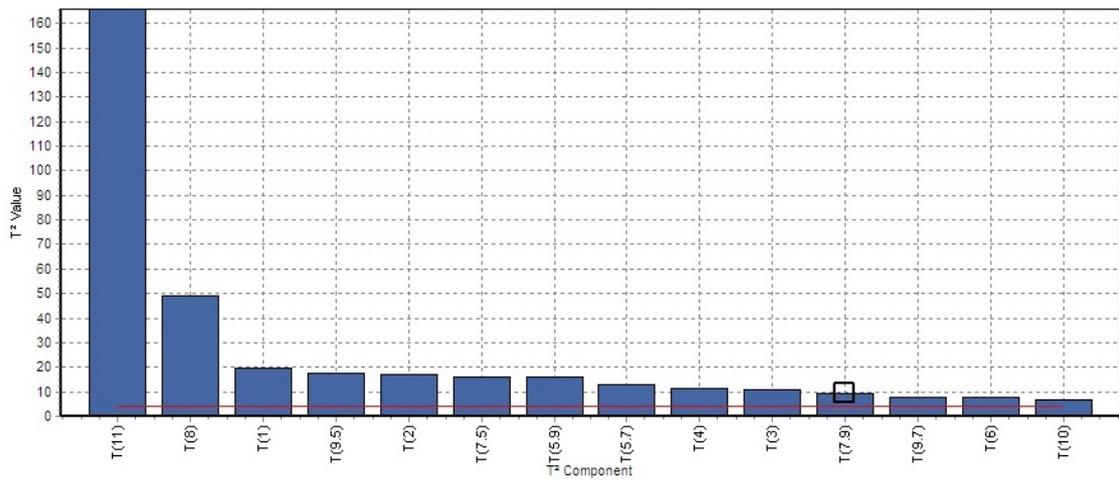


Figura 5.18: Descomposición T². Observación 10.

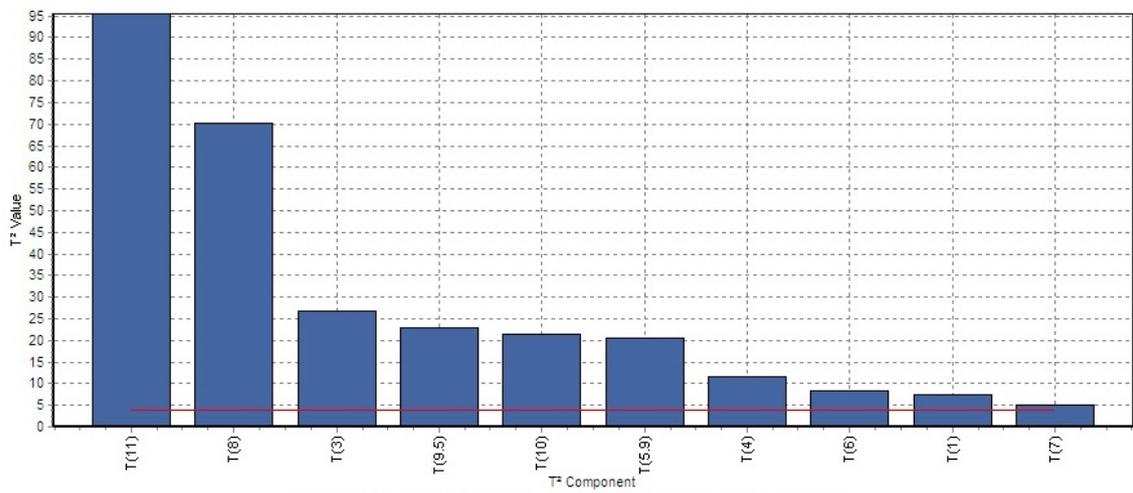


Figura 5.19: Descomposición T². Observación 11.

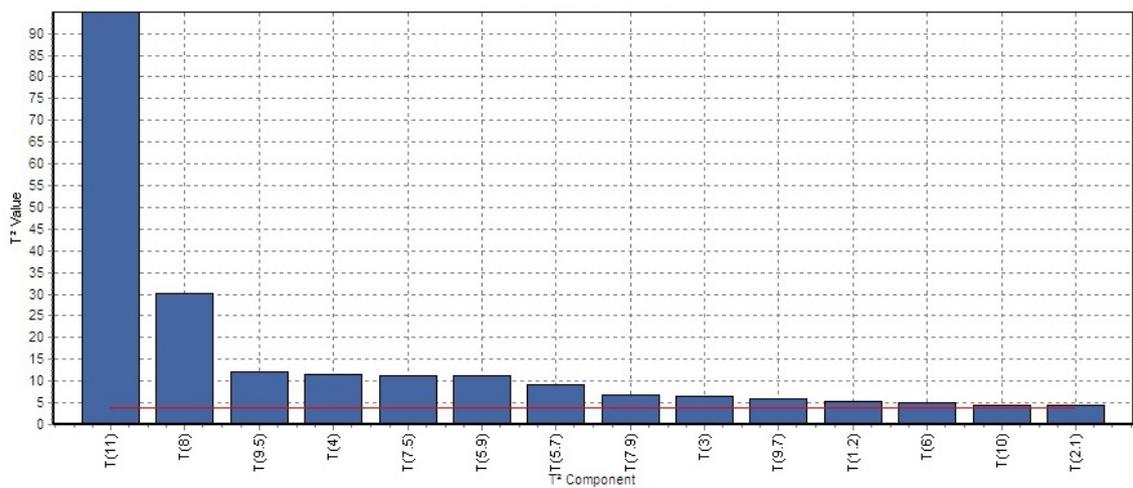


Figura 5.20: Descomposición T². Observación 12.

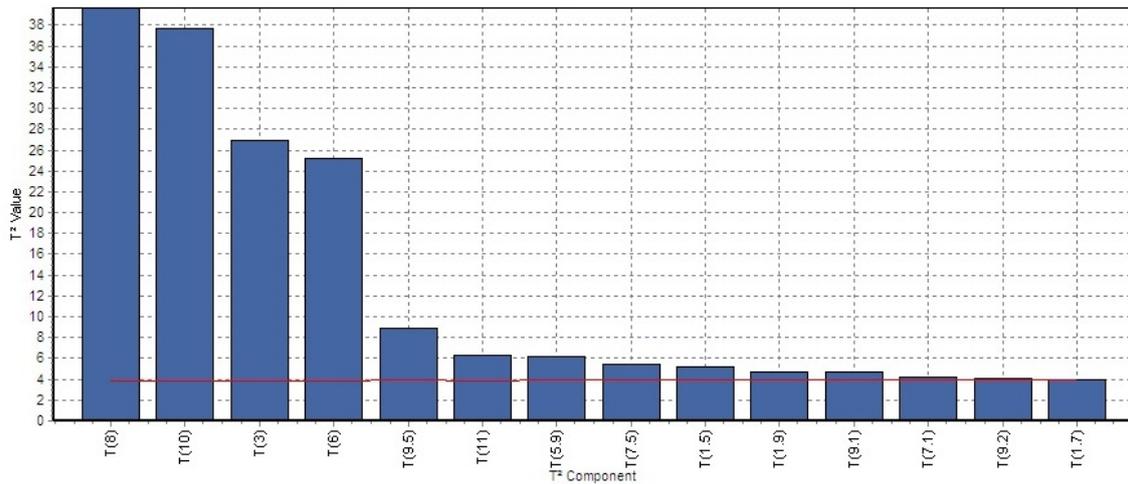


Figura 5.21: Descomposición T². Observación 13.

En síntesis, los resultados obtenidos mediante esta técnica, coinciden con los obtenidos mediante la metodología propuesta SSDM. Si bien es cierto que la técnica (T² + MYT), es más sencilla de aplicar, su efectividad depende de la cantidad de variables a monitorizar. Este hecho hace, que en este caso, donde han sido once las variables monitorizadas, en algunas gráficas, como las correspondientes a las observaciones (4, 5, 6 y 7), podemos observar que se hace confuso determinar una única variable como causante de la desviación en el proceso.

Con esta técnica (T² + MYT), conforme aumenta el número de variables se hace cada vez más inviable su uso, puesto que el número de combinaciones dentro de la descomposición crece exponencialmente y la interpretación de los resultados se hace totalmente confusa.

Sin embargo, se puede concluir que la variable 11-SFOC, al igual que con la metodología propuesta (SSDM), es la que ha contribuido con mayor peso al estado fuera de rango en el proceso.

La metodología propuesta (SSDM), es capaz de reducir el número de variables, siendo capaz, como se ha demostrado, de detectar tanto cambios elevados como pequeños en el proceso, haciendo su aplicación sencilla y eficiente, cuando es inviable la utilización de la técnica (T² + MYT).

Capítulo 6

Conclusiones generales

6. CONCLUSIONES GENERALES

6.1 Conclusiones

En el presente capítulo se recogen las principales conclusiones que se han obtenido tras el desarrollo de la presente tesis doctoral.

- La metodología desarrollada denominada “Method for detection Small and Sudden Deviations” (SSDM) por sus siglas en inglés, fue capaz de reducir el número de variables monitorizadas, sin pérdida de información del proceso, incluso cuando la correlación entre ellas era baja, mejorando la capacidad que tiene el gráfico (T^2 Hotelling + MYT), para realizar sus funciones cuando el número de variables es elevado.

Métodos convencionales de reducción de variables como PCA no son eficientes en la ejecución de esta tarea cuando las variables están pobremente correlacionadas, tal y como se ha comentado en el presente trabajo de investigación.

- Además, la nueva metodología SSDM, fue capaz de detectar eficazmente cambios de elevado y bajo valor en el proceso, dotando así al gráfico T^2 de Hotelling con capacidad de detección de desviaciones progresivas y de pequeña magnitud.

La metodología propuesta se aplicó en dos procesos del motor principal: proceso de lubricación de cilindros y proceso de combustible, cuyas conclusiones se analizan a continuación.

Proceso de lubricación de cilindros:

- El método de reducción de variables, fue capaz de reducir el número de variables monitorizadas, de diez variables en las que inicialmente estaba monitorizado a siete variables representativas del proceso.
- Las siete variables en las que quedo representado el proceso de lubricación de cilindros, se monitorizo completamente y de forma multivariable para unas condiciones determinadas de operación del buque.

- A través de la carta T^2 de Hotelling se detectaron eficazmente desviación de gran magnitud, que se habían generado, con respecto a las condiciones de operación normales del proceso. Posteriormente, se descompusieron aquellos valores T^2 que habían superado el límite de control y se identificaron las variables, “Hydraulic oil pressure after pump 1” y “Total Oil”, como las responsables de las desviaciones en el proceso.
- Un comportamiento anómalo de la variable “Hydraulic oil pressure after pump 1”, pudo causar fallos mecánicos, como desgastes o pérdida de compresión, por otro lado, la desviación en la variable “Total Oil”, podría haber sido causada por un incremento en la dosificación de aceite de cilindros, para contrarrestar el cambio en el azufre contenido en el último bunkering de combustible.
- La carta de control de Hotelling, aunque es muy eficaz en la detección de cambios elevados, no está capacitada para detectar cambios progresivos y de baja magnitud, por debajo de 1,5 desviaciones típicas, por lo que el riesgo de no detectar este tipo de cambios estuvo latente.

Proceso de combustible:

- La metodología propuesta de reducción de variables, a través del análisis de regresión múltiple y análisis de correlaciones, fue capaz de reducir las once variables, pobremente correlacionadas entre ellas, en las que estaba definido el proceso, en únicamente tres variables representativas.
Esto da una idea del grado de reducción al que se puede llegar con el método propuesto, sin embargo, si lo comparamos con la reducción a la que se llegó en el proceso de lubricación, y a la vista de los resultados, queda de manifiesto que el grado de reducción de variables está supeditado a la complejidad intrínseca del propio proceso.
- A través de la metodología desarrollada de monitorización de variables (SSDM), basada en la combinación de gráficos de control (T^2 de Hotelling y Cusum), se detectaron eficazmente tanto desviaciones de elevada magnitud como progresivas y de baja magnitud.
- El valor de la presión diferencial en el intercambiador de aire de carga, no fue lo suficientemente elevado como para superar el umbral de alarma marcado por el fabricante; la metodología SSDM, detecto una pequeña desviación en la variable

SFOC. Este incremento, pudo ser debido para contrarrestar el ensuciamiento del intercooler, y mantener la velocidad del buque.

Sin esta identificación, hubiesen tenido que esperar a que el valor de la presión diferencial estuviese por encima del valor establecido por el fabricante, resultando en un mayor ensuciamiento del intercooler y un incremento del consumo de combustible. Conocer que esta variable se estaba desviando de su valor óptimo, les ayudo a programar futuras acciones preventivas.

- La metodología propuesta (SSDM), fue validada, comparándola con una técnica ampliamente reconocida y utilizada por numerosos autores como es la técnica ($T^2 + MYT$). A la vista de los resultados se comprueba que con ambas técnicas se alcanzan los mismos resultados. Si bien es cierto, que la técnica ($T^2 + MYT$), tuvo dificultades en la identificación de la variable causante de la desviación en el proceso en alguna de las observaciones, debido a la dificultad que tienen en ejecutar esta tarea cuando el número de variables es elevado; si el número de variables a monitorizar hubiese sido mayor, la técnica ($T^2 + MYT$), hubiese sido imposible aplicarla, con resultados satisfactorios.

Muchos procesos involucrados en el funcionamiento de un motor diésel tienen decaimientos pequeños y progresivos además de bruscos y elevados, por esta razón esta metodología, tiene la ventaja que puede ser customizado para cualquier tipo de motor ya que es capaz de detectar cualquier variación en un proceso, independientemente de cual sea su magnitud.

En ocasiones, implementar una metodología cuando el buque ya está terminado genera costes añadidos de implementación, los algoritmos generados en la metodología SSDM, tienen la ventaja que se pueden implementar de forma económica y sencilla a requerimiento del armador en función de las condiciones de operación del buque.

Las repercusiones económicas procedentes de la aplicación de esta metodología, se aventuran que sean beneficiosas, aunque no han sido objetivo principal de este estudio.

Aunque existe un gran interés por parte de los armadores en implementar este método, en las oficinas centrales de la naviera, que sirviese como parte del protocolo del mantenimiento predictivo, o a bordo del buque, como complemento de los sistemas de alarmas actualmente instalados. A día de hoy, este objetivo no se ha podido llevar a cabo, por falta de financiación y disponibilidad del buque.

6.2 Líneas futuras

Las principales líneas de investigación que han quedado abiertas tras el desarrollo de la presente tesis doctoral son las siguientes:

- Implementación de la metodología a bordo del buque y estudio de los resultados obtenidos a largo plazo.
- Estudio comparativo entre la metodología SSDM y PCA, en los procesos estudiados en la presente tesis doctoral.
- Aplicación de la nueva metodología desarrollada en otros sectores industriales como el de la generación de energía eólica.
- Comprobar la efectividad del método con técnicas de detección multivariantes, como MCusum y MEWMA.

6.3 Publicaciones y Congresos

Las aportaciones derivadas de la investigación desarrollada en esta tesis, han sido las siguientes:

Publicaciones

- Boullosa, D., Larrabe, J. L., Lopez, A., & Gomez, M. A. (2017). Monitoring through T 2 Hotelling of cylinder lubrication process of marine diesel engine. *Applied Thermal Engineering*, 110, 32-38.
- Boullosa-Falces, D., Larrabe, J. L., Lopez-Arraiza, A., Menendez, J., & Solaetxe, M. A. G. (2017). Monitoring of fuel oil process of marine diesel engine. *Applied Thermal Engineering*, 127, 517-526.

Congresos

- David Boullosa-Falces, Miguel Angel Gomez Solaetxe, Juan Luis Larrabe Barrena, Alberto Lopez Arraiza. “Monitorización de las variables de salida de un aerogenerador, mediante el estadístico T2 de Hotelling”. I jornadas Doctorales de la UPV/EHU. Bilbao 2016.
- David Boullosa-Falces, Miguel Ángel Gómez Solaetxe, Zaloa Sánchez Varela. “Hotelling T2 Technique applied in a marine propulsión engine”. VII Jornadas de automática Marina (AUTOMAR’17). Castellón 2017.
- David Boullosa-Falces, Jaime Menendez, Zaloa Sanchez Varela, Juan Luis Larrabe Barrena, Alberto Lopez-Arraiza, Miguel Angel Gomez Solaetxe. “Control estadístico de los motores marinos diésel”. II Congreso internacional La industria marítima en el siglo XXI. Bilbao 2017.
- David Boullosa-Falces, Zaloa Sánchez Varela, Juan Luis Larrabe Barrena, Miguel Ángel Gómez Solaetxe. “Evaluación de incidentes durante operaciones realizadas por buques remolcadores manejadores de anclas”. II Congreso internacional La industria marítima en el siglo XXI. Bilbao 2017.
- Zaloa Sánchez Varela, David Boullosa-Falces, Juan Luis Larrabe Barrena, Miguel Ángel Gómez Solaetxe. “Formación en posicionamiento dinámico en tiempos difíciles”. II Congreso internacional La industria marítima en el siglo XXI. Bilbao 2017.

Estancia en otras universidades en labores de investigación

- Estancia de investigación pre doctoral en el departamento de Ciencias y Técnicas de la Navegación y de la construcción Naval de la Universidad de Cantabria, con una duración de un año entre (2016 y 2017).

Capítulo 7

Bibliografía

7. BIBLIOGRAFÍA

- [1] J. F. MacGregor and T. Kourti, "Statistical process control of multivariate processes," *Control Eng. Pract.*, vol. 3, (3), pp. 403-414, 1995.
- [2] M. Best and D. Neuhauser, "Walter A Shewhart, 1924, and the Hawthorne factory," *Qual. Saf. Health. Care.*, vol. 15, (2), pp. 142-143, Apr, 2006. . DOI: 15/2/142 [pii].
- [3] M. B. Hossain and M. S. Masud, "Performance of T2 Chart over \bar{X} Chart for Monitoring the Process Mean: A Simulation Study," 2016.
- [4] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, (1/2), pp. 100-115, 1954.
- [5] S. Roberts, "Control chart tests based on geometric moving averages," *Technometrics*, vol. 1, (3), pp. 239-250, 1959.
- [6] H. Hotelling, "Multivariate quality control illustrated by the air testing of sample bomb sights, techniques of statistical analysis, ch. ii," 1947.
- [7] P. C. Mahalanobis, "On the generalised distance in statistics," *Proceedings of the National Institute of Sciences of India*, 1936, pp. 49-55, 1936.
- [8] R. B. Crosier, "Multivariate generalizations of cumulative sum quality-control schemes," *Technometrics*, vol. 30, (3), pp. 291-303, 1988.
- [9] C. A. Lowry, W. H. Woodall, C. W. Champ and S. E. Rigdon, "A multivariate exponentially weighted moving average control chart," *Technometrics*, vol. 34, (1), pp. 46-53, 1992.
- [10] J. F. V. Barrio, *Control Estadístico De Los Procesos (SPC)*. FC Editorial, 2005.
- [11] D. C. Montgomery, *Introduction to Statistical Quality Control*. John Wiley & Sons (New York), 2009.
- [12] Wester Electric, *Statistical Quality Control Handbook*. Western Electric Corporation, Indianapolis, IN., 1956.

- [13] C. W. Champ and W. H. Woodall, "Exact results for Shewhart control charts with supplementary runs rules," *Technometrics*, vol. 29, (4), pp. 393-399, 1987.
- [14] L. I. Rodríguez, M. I. Rodríguez and M. A. Rodríguez, "Análisis Comparativo de los Gráficos de Control Shewhart, CUSUM y EWMA, para Procesos en Condiciones de Control y Fuera de Control," 2012.
- [15] E. Page, "Cumulative sum charts," *Technometrics*, vol. 3, (1), pp. 1-9, 1961.
- [16] W. D. Ewan, "When and how to use cu-sum charts," *Technometrics*, vol. 5, (1), pp. 1-22, 1963.
- [17] J. M. Lucas, "The design and use of V-mask control schemes," *Journal of Quality Technology*, vol. 8, (1), pp. 1-12, 1976.
- [18] D. M. Hawkins, "A CUSUM for Scale Parameter", *Journal of Quality Technology*, vol. 13, 1981.
- [19] F. Gan, "An optimal design of CUSUM quality control charts," *Journal of Quality Technology*, vol. 23, (4), pp. 279-286, 1991.
- [20] W. H. Woodall and B. M. Adams, "The statistical design of CUSUM charts," *Quality Engineering*, vol. 5, (4), pp. 559-570, 1993.
- [21] K. M. Bower, "Using CUSUM Charts To Detect Small Process Shifts," 2010.
- [22] M. Ghane, A. R. Nejad, M. Blanke, Z. Gao and T. Moan, "Statistical fault diagnosis of wind turbine drivetrain applied to a 5MW floating wind turbine," in *Journal of Physics: Conference Series*, 2016, pp. 052017.
- [23] D. Brook and D. Evans, "An approach to the probability distribution of CUSUM run length," *Biometrika*, vol. 59, (3), pp. 539-549, 1972.
- [24] D. M. HAWKINS, "Cumulative sum control charting: an underutilized SPC tool," *Quality Engineering*, vol. 5, (3), pp. 463-477, 1993.
- [25] G. A. Barnard, "Control charts and stochastic processes," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 239-271, 1959.

- [26] N. L. Johnson, "A simple theoretical approach to cumulative sum control charts," *Journal of the American Statistical Association*, vol. 56, (296), pp. 835-840, 1961.
- [27] L. C. Vance, "Average run lengths of cumulative sum control charts for controlling normal means." *J. Quality Control.*, vol. 18, (3), pp. 189-193, 1986.
- [28] D. M. HAWKINS, "A fast accurate approximation for average run lengths of CUSUM control charts," *Journal of Quality Technology*, vol. 24, (1), pp. 37-43, 1992.
- [29] D. O. Siegmund, "Sequential Analysis: Tests and Confidence Intervals Springer-Verlag," *New York*, 1985.
- [30] V. Koshti, "Cumulative sum control chart," *International Journal of Physics and Mathematical Sciences*, vol. 1, (1), pp. 28-32, 2011.
- [31] J. M. Lucas, "Combined Shewhart-CUSUM quality control schemes," *Journal of Quality Technology*, vol. 14, pp. 51-59, 1982.
- [32] S. V. Crowder, "A simple method for studying run-length distributions of exponentially weighted moving average charts," *Technometrics*, vol. 29, (4), pp. 401-407, 1987.
- [33] S. V. Crowder, "Design of exponentially weighted moving average schemes," *Journal of Quality Technology*, vol. 21, (3), pp. 155-162, 1989.
- [34] J. M. Lucas and M. S. Saccucci, "Exponentially weighted moving average control schemes: properties and enhancements," *Technometrics*, vol. 32, (1), pp. 1-12, 1990.
- [35] P. Robinson and T. Y. Ho, "Average run lengths of geometric moving average charts by numerical methods," *Technometrics*, vol. 20, (1), pp. 85-93, 1978.
- [36] J. STUART HUNTER, "A one-point plot equivalent to the Shewhart chart with Western Electric rules," *Quality Engineering*, vol. 2, (1), pp. 13-19, 1989.
- [37] J. MacGregor and T. Harris, "The exponentially weighted moving variance," *Journal of Quality Technology*, vol. 25, (2), pp. 106-118, 1993.

- [38] K. M. Bower, "Using exponentially weighted moving average (EWMA) charts," *Asia Pacific Process Engineer*, 2000.
- [39] Schilling, E.G., Nelson, P.R., "The Effect of Non-Normality on the Control Limits of X Charts". *Journal of Quality Technology*, vol. 25, 1976.
- [40] D. Hawkins and D. Olwell, "Cumulative Sum Control Charts and Charting for Quality Control," 1998.
- [41] J. Michalak, "Using Control Charts to Detect Small Process Shifts," 2005.
- [42] S. Bersimis, S. Psarakis and J. Panaretos, "Multivariate statistical process control charts: an overview," *Qual. Reliab. Eng. Int.*, vol. 23, (5), pp. 517-543, 2007.
- [43] C. Fuchs and R. S. Kenett, *Multivariate Quality Control: Theory and Applications*. CRC Press, 1998.
- [44] W. H. Woodall and M. M. Ncube, "Multivariate CUSUM quality-control procedures," *Technometrics*, vol. 27, (3), pp. 285-292, 1985.
- [45] J. J. Pignatiello and G. C. Runger, "Comparisons of multivariate CUSUM charts," *Journal of Quality Technology*, vol. 22, (3), pp. 173-186, 1990.
- [46] S. S. Prabhu and G. C. Runger, "Designing a multivariate EWMA control chart," *Journal of Quality Technology*, vol. 29, (1), pp. 8, 1997.
- [47] W. Molnau, G. Runger, D. Montgomery and K. Skinner, "A program of ARL calculation for multivariate EWMA charts," *Journal of Quality Technology*, vol. 33, (4), pp. 515, 2001.
- [48] C. Barbiero, M. Flury, A. Pagura, M. Quaglino and M. Ruggieri, "Control estadístico de procesos multivariados mediante gráficos de control multivariados T2 de Hotelling, MEWMA y MCUSUM," *Octavas Jornadas Investigaciones En La Facultad De Ciencias Económicas Y Estadística, Instituto De Investigaciones Teóricas Y Aplicada, Escuela De Estadística*, pp. 12, 2003.

- [49] C. A. Lowry and D. C. Montgomery, "A review of multivariate control charts," *IIE Transactions*, vol. 27, (6), pp. 800-810, 1995.
- [50] R. S. Sparks, "Quality control with multivariate data," *Australian & New Zealand Journal of Statistics*, vol. 34, (3), pp. 375-390, 1992.
- [51] S. J. Wierda, "Multivariate statistical process control—recent results and directions for future research," *Statistica Neerlandica*, vol. 48, (2), pp. 147-168, 1994.
- [52] T. P. Ryan, *Statistical Methods for Quality Improvement*. John Wiley & Sons, 2011.
- [53] W. H. Woodall, R. W. Hoerl, A. C. Palm and D. J. Wheeler, "Controversies and contradictions in statistical process control/Discussion/Response," *Journal of Quality Technology*, vol. 32, (4), pp. 341, 2000.
- [54] J. H. Sullivan and W. H. Woodall, "A comparison of multivariate control charts for individual observations," *Journal of Quality Technology*, vol. 28, (4), pp. 398-408, 1996.
- [55] J. A. VARGAS N, "Robust estimation in multivariate control charts for individual observations," *Journal of Quality Technology*, vol. 35, (4), pp. 367-376, 2003.
- [56] G. A. F. Seber, "Multivariate observations." *J. Wiley*, 1984.
- [57] P. Maravelakis, "An investigation of some characteristics of univariate and multivariate control charts," *Unpublished Doctoral Dissertation, Department of Statistics, Athens University of Economics and Business.*(Available from Internet: <Http://Stat-Athens.Aueb.Gr/~Jpan/Diatrives/Maravelakis/Index.Html>), 2003.
- [58] W. R. Anchiraico Agudo, *Metodos Multivariantes En Control Estadistico De La Calidad*, 2003.
- [59] F. Alt, "Multivariate Quality Control Encyclopedia of Statistical Sciences, Vol. 6, edited by NL Johnson and S. Kotz," 1985.

- [60] A. J. Hayter and T. KWOK-LEUNG, "Identification and quantification in multivariate quality control problems," *Journal of Quality Technology*, vol. 26, (3), pp. 197-208, 1994.
- [61] D. M. Hawkins, "Multivariate quality control based on regression-adjusted variables," *Technometrics*, vol. 33, (1), pp. 61-75, 1991.
- [62] M. Wade and W. Woodall, "A review and analysis of cause-selecting control charts," *Journal of Quality Technology*, vol. 25, (3), pp. 161-169, 1993.
- [63] G. C. Runger, F. B. Alt and D. C. Montgomery, "Contributors to a multivariate statistical process control chart signal," *Communications in Statistics--Theory and Methods*, vol. 25, (10), pp. 2203-2213, 1996.
- [64] B. Murphy, "Selecting Out of Control Variables With the T^2 Multivariate Quality Control Procedure," *The Statistician*, pp. 571-581, 1987.
- [65] J. E. Jackson, "Principal components and factor analysis: Part I-principal components," *Journal of Quality Technology*, vol. 12, (4), pp. 201-213, 1981.
- [66] R. Mason, N. TRACY and J. YOUNG, "A control procedure for autocorrelated multivariate processes," in *Proc. Am. Stat. Assoc*, pp. 143-145, 1997.
- [67] R. L. Mason, N. D. Tracy and J. C. Young, "Decomposition of T^2 for multivariate control chart interpretation," *Journal of Quality Technology*, vol. 27, (2), pp. 99-1108, 1995.
- [68] R. L. Mason and J. C. Young, *Multivariate Statistical Process Control with Industrial Applications*. SIAM, 2002.
- [69] J. S. Fermín, M. Valdiviezo, G. Orlandoni and S. Barreto, "Control estadístico de procesos multivariantes en la industria alimentaria: implementación a través del estadístico T^2 -Hotelling," *Agroalimentaria*, vol. 15, (28), pp. 91-105, 2009.
- [70] J. Mosquera Restrepo, J. Olaya Ochoa and R. Escobar, "Aplicación del control estadístico multivariante en un proceso de extrusión de película plástica," *Scientia Et Technica*, vol. 13, (37), 2007.

- [71] D. Boullosa, J. L. Larrabe, A. Lopez and M. A. Gomez, "Monitoring through T 2 Hotelling of cylinder lubrication process of marine diesel engine," *Appl. Therm. Eng.*, vol. 110, pp. 32-38, 2017.
- [72] M. Henneberg, B. Jørgensen and R. L. Eriksen, "Oil condition monitoring of gears onboard ships using a regression approach for multivariate T 2 control charts," *J. Process Control*, vol. 46, pp. 1-10, 2016.
- [73] T. Kourti and J. F. MacGregor, "Multivariate SPC methods for process and product monitoring," *Journal of Quality Technology*, vol. 28, (4), pp. 409-428, 1996.
- [74] A. Ferrer, "Control estadístico MegaVariante para los procesos del siglo XXI," in *Dpto De Estadística E Investigación Operativa. UPV. Congreso Nacional De Estadística E Investigación Operativa. Valencia, España, 2003*, pp. 1-15.
- [75] J. M. Fonseca, B. M. Sousa, W. E. Aguiar, A. R. Braga, A. P. Lemos, H. C. Michel and C. M. Braga, "Monitoring of a thermoelectric power plant based on multivariate statistical process control," in *Evolving and Adaptive Intelligent Systems (EAIS), 2016 IEEE Conference On*, 2016, pp. 49-56.
- [76] P. Nomikos and J. F. MacGregor, "Multivariate SPC charts for monitoring batch processes," *Technometrics*, vol. 37, (1), pp. 41-59, 1995.
- [77] J. Larrabe, M. Gómez, I. Sotés, F. Alvarez, M. Rey, V. Mielgo, I. Sotes and X. Basogain, "Vessel condition monitorization through satellite using principal component analysis," *Journal of Maritime Research*, vol. 7, (3), pp. 3-14, 2014.
- [78] Y. Wang, X. Ma and M. J. Joyce, "Reducing sensor complexity for monitoring wind turbine performance using principal component analysis," *Renewable Energy*, vol. 97, pp. 444-456, 2016.
- [79] F. Aparisi, M. A. de Luna and E. Epprecht, "Optimisation of a set of or principal components control charts using genetic algorithms," *Int J Prod Res*, vol. 48, (18), pp. 5345-5361, 2010.
- [80] I. González and I. Sánchez, "Variable selection for multivariate statistical process control," *Journal of Quality Technology*, vol. 42, (3), pp. 242, 2010.

- [81] J. C. Vega-Vilca and J. Guzmán, "Regresión PLS y PCA como solución al problema de multicolinealidad en regresión múltiple," *Revista De Matemática Teoría Y Aplicaciones*, vol. 18, (1), pp. 09-20, 2011.
- [82] S. Joe Qin, "Statistical process monitoring: basics and beyond," *J. Chemometrics*, vol. 17, (8-9), pp. 480-502, 2003.
- [83] H. Oddan, *Multivariate Statistical Condition Monitoring*. (Master's thesis, NTNU), 2017.
- [84] A. M. C. Zamarrón, E. M. Prado and F. Z. Luis, "Monitoreo y control de un proceso normal multivariado," *Conciencia Tecnológica*, (43), pp. 29-35, 2012.
- [85] D. Peña, *Análisis De Datos Multivariantes*. McGraw-Hill Madrid, 2002.
- [86] D. Boullosa-Falces, Larrabe, J. L., A. Lopez-Arraiza, J. Menendez and M. A. G. Solaetxe, "Monitoring of fuel oil process of marine diesel engine," *Appl. Therm. Eng.*, vol. 127, pp. 517-526, 2017.
- [87] R. L. Mason and J. C. Young, "Improving the sensitivity of the T2 statistic in multivariate process control," *Journal of Quality Technology*, vol. 31, (2), pp. 155, 1999.
- [88] Z. Stamboliska, E. Rusiński and P. Moczko, "Proactive condition monitoring of low-speed machines," in *Proactive Condition Monitoring of Low-Speed Machines* Anonymous Springer, 2015, pp. 53-68.
- [89] F. Aparisi and J. C. GARCIA, "Aumento de la potencia del gráfico de control multivariante T 2 de Hotelling utilizando señales adicionales de falta de control," *Estadística Española*, vol. 43, (148), pp. 171-188, 2001.
- [90] J. Botella, "Potencia de pruebas alternativas para dos muestras relacionadas con datos perdidos," *Psicothema*, vol. 14, (1), 2002.
- [91] E. Uriel and J. A. Manzano, *Análisis Multivariante Aplicado*. Paraninfo, 2002.

- [92] J. Hair, R. Anderson, R. Tathan and W. Black, "Análisis Multivariante. Quinta Edición. Prenticehall," *Iberia, Madrid, España*, 1999.
- [93] J. W. Graham, S. M. Hofer and A. M. Piccinin, "Analysis with missing data in drug prevention research," *NIDA Res. Monogr.*, vol. 142, pp. 13-13, 1994.
- [94] J. G. Orme and J. Reis, "Multiple regression with missing data," *Journal of Social Service Research*, vol. 15, (1-2), pp. 61-91, 1991.
- [95] R. G. Downey and C. V. King, "Missing data in Likert ratings: A comparison of replacement methods," *J. Gen. Psychol.*, vol. 125, (2), pp. 175-191, 1998.
- [96] J. W. Graham, "Missing data analysis: Making it work in the real world," *Annu. Rev. Psychol.*, vol. 60, pp. 549-576, 2009.
- [97] R. L. Brown, "Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 1, (4), pp. 287-316, 1994.
- [98] R. Fernández-Alonso, J. Suárez-Álvarez and J. Muñiz, "Imputación de datos perdidos en las evaluaciones diagnósticas educativas," *Psicothema*, vol. 24, (1), 2012.
- [99] F. M. Shrive, H. Stuart, H. Quan and W. A. Ghali, "Dealing with missing data in a multi-question depression scale: a comparison of imputation methods," *BMC Medical Research Methodology*, vol. 6, (1), pp. 57, 2006.
- [100] J. R. van Ginkel, K. Sijtsma, van der Ark, L. Andries and J. K. Vermunt, "Incidence of missing item scores in personality measurement, and simple item-score imputation," *Methodology*, 2010.
- [101] B. G. Tabachnick and L. S. Fidell, "LS: Using multivariate statistics", New York: Haper Collins, 1996 .
- [102] A. P. Merino and Diaz, C Miguel Angel Ruis, *Análisis De Datos Con SPSS 13 Base*. McGraw-Hill Interamericana, 2006.
- [103] B. Williams, *A Sampler on Sampling*. JSTOR, 1978.

- [104] K. Pearson, "Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical Or Physical Character*, vol. 187, pp. 253-318, 1896.
- [105] Norusis, Marija J Marija J, *SPSS: SPSS for Windows, Base System User's Guide Release 6.0*. SPSS Inc., 1993.
- [106] D. M. Hawkins, *Identification of Outliers*. Springer, 1980.
- [107] V. Barnett and T. Lewis, *Outliers in Statistical Data*. Wiley New York, 1994.
- [108] D. W. Wichern and R. A. Johnson, *Applied Multivariate Statistical Analysis*. Prentice Hall, 1992.
- [109] L. C. Hamilton, *Regression with Graphics: A Second Course in Applied Statistics*. Duxbury press, 1992.
- [110] C. Judd and G. McClelland, "Data Analysis, A Model-Comparison Approach Orlando," *Fl, Usa*, 1989.
- [111] C. Pérez López, "Técnicas de análisis multivariante de datos," *Aplicaciones Con SPSS, Madrid, Universidad Complutense De Madrid*, 2004.
- [112] F. E. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics*, vol. 11, (1), pp. 1-21, 1969.
- [113] W. Stefansky, "Rejecting outliers in factorial designs," *Technometrics*, vol. 14, (2), pp. 469-479, 1972.
- [114] R. L. Mason, Y. Chou, J. H. Sullivan, Z. G. Stoumbos and J. C. Young, "Systematic patterns in T (2) charts," *Journal of Quality Technology*, vol. 35, (1), pp. 47, 2003.
- [115] R. Gnanadesikan, *Methods for Statistical Data Analysis of Multivariate Observations*. John Wiley & Sons, 2011.

- [116] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. Prentice-Hall New Jersey, 2014.
- [117] S. C. Gutiérrez and E. D. L. Aguilera, "QQ Plot Normal. Los puntos de posición gráfica," *Iniciación a La Investigación*, (2), 2007.
- [118] R. L. Mason, N. D. Tracy and J. C. Young, "A practical approach for interpreting multivariate T2 control chart signals," *Journal of Quality Technology*, vol. 29, (4), pp. 396, 1997.
- [119] S. Agog, H. G. Dikko and O. E. Asiribo, "Decomposing Hotelling's T2 statistic using four variables," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 3, (4), pp. 11449-11454, 2014.
- [120] T. García, M. Vásquez and J. García, "Efectos sobre la calidad del papel higiénico cuando se producen outliers multivariantes no detectables por los gráficos de control univariantes," *Revista INGENIERÍA UC*, vol. 11, (2), pp. 67-77, 2004.
- [121] *MAN B&W S70ME-C8.5-TII – Project Guide, Electronically Controlled Two-Stroke Engines. MAN B&W Diesel, Copenhagen, 2016*. [Accessed 2016/02/09]. Available:
http://marine.man.eu/applications/projectguides/2stroke/content/printed/S70ME-C8_5.pdf.
- [122] *DIESEL & TURBO MAN*, 2011. [Accessed 2016/02/10]. Available:
<<http://primeserv.man.eu/docs/librariesprovider5/primeserv-documents/alpha-lubricator.pdf?sfvrsn=6>>.
- [123] *Alpha Adaptive Cylinder-oil Control. Alpha ACC*, 2002. [Accessed 2016/08/04]. Available: www.manbw.dk.
- [124] N. Ye, D. Parmar and C. M. Borrer, "A Hybrid SPC Method with the Chi-Square Distance Monitoring Procedure for Large-scale, Complex Process Data," *Qual. Reliab. Eng. Int.*, vol. 22, (4), pp. 393-402, 2006.
- [125] M. R. Spiegel and L. J. Stephens, "Estadística." 2009.

[126] S. McKinley and M. Levine, "Cubic spline interpolation," *College of the Redwoods*, vol. 45, (1), pp. 1049-1060, 1998.