

# Sistema diagnóstico para pozos de bombeo mecánico convencional usando técnicas de reconocimiento de patrones

ALEXANDER JOSÉ MOLERO MONTIEL

MÁSTER EN INGENIERÍA DE SISTEMAS Y DE CONTROL  
FACULTAD DE INGENIERÍA INFORMÁTICA  
UNIVERSIDAD COMPLUTENSE DE MADRID  
UNIVERSIDAD NACIONAL EXPERIMENTAL DE LA DISTANCIA



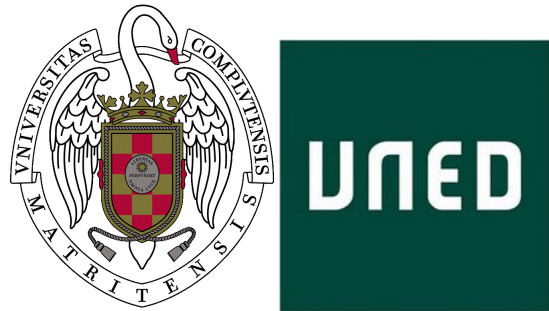
TRABAJO FIN DE MÁSTER EN INGENIERÍA DE SISTEMAS Y CONTROL

Fecha: 12 de Junio 2015

Director: Sebastián Dormido Canto

Curso académico: 2014 - 2015

Convocatoria de defensa: Junio 2015



## Autorización

Se autoriza a la Universidad Complutense de Madrid (UCM) y a la Universidad Nacional de Educación a Distancia (UNED) a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a sus autores, tanto la memoria de este Trabajo Fin de Máster, como el código, la documentación y/o el prototipo desarrollado.

Firmado: Alexander José Molero Montiel

22 de Junio de 2015

# Dedicatoria

*A Dios*

*A mi Familia (Adelso, Arelis, Adelso José y Laurimar)*

# Agradecimientos

A mi tutor Sebastián Dormido Canto, que desde el momento de inicio de la disertación siempre me asesoró. Sin su ayuda no hubiese podido llegar a buen término. Muchísimas gracias Profesor.

A la profesora Eva Besada-Portas. Sin su apoyo no hubiese podido realizar este Máster. Muchísimas gracias Profesora.

A Miguel Bravo y José Romero, que me ayudaron a recolectar y procesar los datos para iniciar este trabajo, además de muchas asesorías. Muchas gracias.

A PDVSA INTEVEP, brindándome los recursos para recolectar los datos. Muchas gracias.

# Índice general

Índice de figuras	XI
Índice de cuadros	XIV
Índice de abreviaturas	XV
Resumen	XVI
Abstract	XVII
<b>1. Introducción</b>	<b>1</b>
1.1. Introducción . . . . .	1
1.2. Motivación . . . . .	3
1.3. Objetivos . . . . .	4
1.4. Estructura de la memoria . . . . .	5
<b>2. Estado del Arte</b>	<b>6</b>
2.1. SVM . . . . .	6
2.2. Wavelet . . . . .	8
2.3. Redes Neuronales . . . . .	9
2.4. Fourier . . . . .	13
2.5. Reconocimiento de Patrones . . . . .	14
<b>3. Marco Teórico</b>	<b>21</b>
3.1. Bombeo Mecánico Convencional . . . . .	21
3.1.1. Equipos de superficie . . . . .	22
3.1.1.1. Cabezal de Pozo . . . . .	22
3.1.1.2. Unidades de bombeo . . . . .	23
3.1.1.3. Prensa-estopa . . . . .	24
3.1.1.4. Barra pulida . . . . .	24
3.1.2. Equipos de subsuelo . . . . .	24
3.1.2.1. Tubería de producción . . . . .	24
3.1.2.2. Sarta de cabillas . . . . .	24

3.1.2.3.	Bomba de subsuelo . . . . .	25
3.1.2.4.	Ancla de gas . . . . .	28
3.2.	Extractores de características . . . . .	29
3.2.1.	Descriptores de Fourier . . . . .	29
3.2.2.	<i>Wavelet</i> . . . . .	30
3.2.2.1.	Generalidades . . . . .	31
3.2.2.2.	Bases <i>Wavelet</i> . . . . .	33
3.2.2.3.	<i>Wavelet Packet</i> . . . . .	37
3.3.	Regresor: <i>Support Vector Machines</i> . . . . .	39
3.3.1.	Teoría de Aprendizaje Estadístico . . . . .	39
3.3.2.	Fundamentos de la SVM . . . . .	40
3.3.3.	Clasificación usando SVM . . . . .	41
3.3.3.1.	Caso separable . . . . .	41
3.3.3.2.	Caso No Separable . . . . .	44
3.3.3.3.	Clasificadores No Lineales . . . . .	46
3.3.3.4.	Clasificación Multi-Clase . . . . .	48
3.3.3.5.	Regresión usando SV . . . . .	48
<b>4.</b>	<b>Análisis de Resultados</b>	<b>50</b>
4.1.	Estudio de los datos . . . . .	50
4.1.1.	Datos de entrenamiento y validación . . . . .	52
4.1.2.	Datos de prueba . . . . .	52
4.2.	Extractor de Características . . . . .	53
4.2.1.	Funcionamiento del extractor donde el pozo BMC funciona (con bombeo) normal . . . . .	53
4.2.2.	Funcionamiento del extractor donde el pozo BMC funciona con crudo viscoso . . . . .	56
4.2.3.	Funcionamiento del extractor donde el pozo BMC posee filtración en la válvula viajera . . . . .	60
4.2.4.	Funcionamiento del extractor donde el pozo BMC posee filtración en la válvula fija . . . . .	64
4.2.5.	Funcionamiento del extractor donde el pozo BMC posee interferencia de gas . . . . .	68
4.2.6.	Funcionamiento del extractor donde el pozo BMC posee golpe de fluido . . . . .	72
4.2.7.	Funcionamiento del extractor donde el pozo BMC posee golpe de bomba - pistón arriba . . . . .	76
4.2.8.	Funcionamiento del extractor donde el pozo BMC posee golpe de bomba - pistón abajo . . . . .	80

4.2.9. Funcionamiento del extractor donde el pozo BMC posee falla de la caja de transmisión . . . . .	84
4.2.10. Funcionamiento del extractor donde el pozo BMC posee cabillas partidas . . . . .	88
4.2.11. Funcionamiento del extractor donde la unidad BMC funciona en un estado no determinado . . . . .	92
4.3. Desarrollo del regresor SVM . . . . .	96
<b>5. Conclusiones</b>	<b>130</b>
5.1. Aportes . . . . .	130
5.2. Limitaciones . . . . .	130
5.3. Recomendaciones y Trabajo Futuro . . . . .	131
<b>Bibliografía</b>	<b>133</b>

# Índice de figuras

3.1. Componentes principales de un sistema de levantamiento por BMC: <i>Sucker Rod Pump</i> es la bomba de subsuelo, <i>Sucker Rod String</i> es la sarta de cabillas y <i>Pumping Unit</i> es la unidad de bombeo. . . . .	22
3.2. CDS de un pozo BMC donde la caja de transmisión presenta mucho desgaste (aparición de esfuerzos vibrantes adicionales sobre la carta) . . . . .	24
3.3. CDS de un pozo BMC con cabillas partidas . . . . .	25
3.4. CDS (cartas de color azul) y CDF (cartas de color rojo) de un pozo BMC con fuga en la válvula viajera (izquierda) y fuga en la válvula fija (derecha). Los parámetros CMP y CEP son inherentes al funcionamiento del método BMC. El CMP es la Carrera Máxima del Pistón (máxima longitud de movimiento del pistón respecto al barril durante una carrera completa) y CEP es la Carrera Efectiva del Pistón (longitud de la carrera del pistón cuando la carga total de fluido esta actuando sobre la válvula fija) . . . . .	26
3.5. CDS (carta de color azul) y CDF (carta de color rojo) de un pozo BMC con golpe de pistón abajo. Golpe de pistón arriba es simétricamente opuesta. . . . .	27
3.6. CDS (carta de color azul) y CDF (carta de color rojo) de un pozo BMC con golpe de fluido. Los parámetros CMP y CEP son inherentes al funcionamiento del método BMC y están descritos en la figura 3.4. . . . .	28
3.7. CDS (carta de color azul) y CDF (carta de color rojo) de un pozo BMC con interferencia de gas. Los parámetros CMP y CEP son inherentes al funcionamiento del método BMC y están descritos en la figura 3.4. . . . .	28
3.8. Una figura y su representación como una secuencia en números complejos. Los puntos $(x_0, y_0)$ y $(x_1, y_1)$ son los dos primeros puntos (arbitrarios) de la secuencia. . . . .	29
3.9. Representaciones tiempo-frecuencia de las <i>Wavelets</i> $\psi_{u,s}$ y $\psi_{u_0,s_0}$ . Cuando $s$ disminuye, la representación temporal de la señal disminuye ( <i>support</i> ), pero la dispersión frecuencial aumenta y cubre el intervalo que esta asociado a las altas frecuencias. . . . .	32
3.10. Mosaico de las bases <i>Wavelets</i> en el plano tiempo - frecuencia . . . . .	34
3.11. <i>Wavelets</i> Daubechies $\psi$ con su respectiva función de escala $\phi$ con $p$ momentos de desvanecimiento . . . . .	36

3.12. <i>Wavelets</i> Daubechies (las dos de la izquierda) y Symmlets (las dos de la derecha) , con sus funciones de escala $\phi$ . . . . .	36
3.13. Representación en plano tiempo - frecuencia de las <i>Wavelet Packet</i> . El mosaico es obtenido por medio de la traslación en tiempo de <i>Wavelets Packets</i> cubriendo cada intervalo frecuencial. . . . .	38
3.14. Descomposición binaria de 3 niveles usando <i>Wavelet Packet</i> . En este caso las letras A de cada descomposición viene asociada a los coeficientes de aproximación (salida de un filtro pasa bajo construido con <i>Wavelets</i> ) y las letras D de cada descomposición viene asociada a los coeficientes de detalle (salida de un filtro pasa alto construido con <i>Wavelets</i> ). Los subíndices denotan el nivel de descomposición . . . . .	38
3.15. Separación de dos clases mediante un hiperplano óptimo. Las muestras sombreadas corresponden a los SV. . . . .	42
3.16. Dos posibles clasificadores lineales, mostrando distancias de separación $z_1$ y $z_2$ desde el hiperplano clasificador respectivo. Los SV son los que tienen la circunferencia externa de los clasificadores que apliquen. . . . .	43
3.17. Caso no separable, donde: los puntos simples representan que están correctamente clasificados (fuera de la banda), los puntos inscritos dentro de rombos están correctamente clasificados pero cayeron dentro de la banda y los puntos inscritos dentro de círculos están mal clasificados. . . . .	45
3.18. Arquitectura de un regresor usando SV . . . . .	49
4.1. Histograma de los grados API ( $^{\circ}$ API) de los pozos diagnosticados para realizar la extracción de características con todos los datos . . . . .	51
4.2. Espectro de la señal de carga para bombeo normal, obtenida usando la FFT (arriba), FD de la señal de carga con una cantidad de términos fijo (abajo) . . . . .	53
4.3. Señal de carga original para la clase bombeo normal (arriba) y comprimida usando Wavelet (abajo) . . . . .	54
4.4. Comparación de la señal de carga original para bombeo normal y comprimida (usando Wavelet) . . . . .	55
4.5. Centros determinados a la CDF para bombeo normal usando FCM y SOM . . . . .	56
4.6. Espectro de la señal de carga para la clase bombeo de crudo viscoso, obtenida usando la FFT (arriba), FD de la señal de carga con una cantidad de términos fijo (abajo) . . . . .	57
4.7. Señal de carga original para la clase bombeo de crudo viscoso (arriba) y comprimida usando Wavelet (abajo) . . . . .	58
4.8. Comparación de la señal de carga original para la clase bombeo de crudo viscoso y comprimida (usando Wavelet) . . . . .	59

4.9. Centros determinados a la CDF para la clase bombeo de crudo viscoso usando FCM y SOM . . . . .	60
4.10. Espectro de la señal de carga con filtración en la válvula viajera, obtenida usando la FFT (arriba), FD de la señal de carga con una cantidad de términos fijo (abajo) . . . . .	61
4.11. Señal de carga original con filtración en la válvula viajera (arriba) y comprimida usando Wavelet (abajo) . . . . .	62
4.12. Comparación de la señal de carga original con filtración en la válvula viajera y comprimida (usando Wavelet) . . . . .	63
4.13. Centros determinados a la CDF con filtración en la válvula viajera usando FCM y SOM . . . . .	64
4.14. Señal de carga original con filtración en la válvula fija (arriba) y comprimida usando Wavelet (abajo) . . . . .	66
4.15. Comparación de la señal de carga original con filtración en la válvula fija y comprimida (usando Wavelet) . . . . .	67
4.16. Centros determinados a la CDF con filtración en la válvula fija usando FCM y SOM . . . . .	68
4.17. Señal de carga original con interferencia de gas (arriba) y comprimida usando Wavelet (abajo) . . . . .	70
4.18. Comparación de la señal de carga original con interferencia de gas y comprimida (usando Wavelet) . . . . .	71
4.19. Centros determinados a la CDF con interferencia de gas usando FCM y SOM	72
4.20. Espectro de la señal de carga con golpe de fluido, obtenida usando la FFT (arriba), FD de la señal de carga con una cantidad de términos fijo (abajo)	73
4.21. Señal de carga original con golpe de fluido (arriba) y comprimida usando Wavelet (abajo) . . . . .	74
4.22. Comparación de la señal de carga original con golpe de fluido y comprimida (usando Wavelet) . . . . .	75
4.23. Centros determinados a la CDF con golpe de fluido usando FCM y SOM .	76
4.24. Señal de carga original con golpe de bomba - pistón arriba (arriba) y comprimida usando Wavelet (abajo) . . . . .	78
4.25. Comparación de la señal de carga original con golpe de bomba - pistón arriba y comprimida (usando Wavelet) . . . . .	79
4.26. Centros determinados a la CDF con golpe de bomba - pistón arriba usando FCM y SOM . . . . .	80
4.27. Señal de carga original con golpe de bomba - pistón abajo (arriba) y comprimida usando Wavelet (abajo) . . . . .	82
4.28. Comparación de la señal de carga original con golpe de bomba - pistón abajo y comprimida (usando Wavelet) . . . . .	83

4.29. Centros determinados a la CDF con golpe de bomba - pistón abajo usando FCM y SOM . . . . .	84
4.30. Señal de carga original con falla de la caja de transmisión y comprimida usando Wavelet (abajo) . . . . .	86
4.31. Comparación de la señal de carga original con falla de la caja de transmisión y comprimida (usando Wavelet) . . . . .	87
4.32. Centros determinados a la CDF con falla de la caja de transmisión usando FCM y SOM . . . . .	88
4.33. Señal de carga original con cabillas partidas y comprimida usando Wavelet (abajo) . . . . .	90
4.34. Comparación de la señal de carga original con cabillas partidas y comprimida (usando Wavelet) . . . . .	91
4.35. Centros determinados a la CDF con cabillas partidas usando FCM y SOM	92
4.36. Señal de carga original para un estado no determinado y comprimida usando Wavelet (abajo) . . . . .	94
4.37. Comparación de la señal de carga original para un estado no determinado y comprimida (usando Wavelet) . . . . .	95
4.38. Centros determinados a la CDF para un estado no determinado usando FCM y SOM . . . . .	96

# Índice de cuadros

2.1.	Aspectos resaltantes de sistemas diagnósticos usando técnicas de reconocimiento de patrones y procesamiento de señales para pozos BMC . . . . .	18
3.1.	Componentes de la bomba de subsuelo de BMC . . . . .	26
4.1.	Error cuadrático medio (MSE) y coeficiente de correlación ( $R^2$ ) obtenidos con FD como entrada al SVR por cada estado de funcionamiento. . . . .	97
4.2.	MSE y $R^2$ obtenidos con <i>Wavelet</i> como entrada al SVR por cada estado de funcionamiento . . . . .	98
4.3.	MSE y $R^2$ obtenidos con <i>Self Organizing Map</i> (SOM) como entrada al SVR por cada estado de funcionamiento . . . . .	99
4.4.	MSE y $R^2$ obtenidos con <i>Fuzzy C-Means</i> (FCM) como entrada al SVR por cada estado de funcionamiento . . . . .	100
4.5.	MSE y $R^2$ obtenidos con los grados API del crudo como entrada al SVR por cada estado de funcionamiento . . . . .	101
4.6.	MSE y $R^2$ obtenidos con FD y <i>Wavelet</i> como entrada al SVR por cada estado de funcionamiento . . . . .	102
4.7.	MSE y $R^2$ obtenidos con FD y SOM como entrada al SVR por cada estado de funcionamiento . . . . .	103
4.8.	MSE y $R^2$ obtenidos con FD y FCM como entrada al SVR por cada estado de funcionamiento . . . . .	104
4.9.	MSE y $R^2$ obtenidos con FD y API como entrada al SVR por cada estado de funcionamiento . . . . .	105
4.10.	MSE y $R^2$ obtenidos con <i>Wavelet</i> y SOM como entrada al SVR por cada estado de funcionamiento . . . . .	106
4.11.	MSE y $R^2$ obtenidos con <i>Wavelet</i> y FCM como entrada al SVR por cada estado de funcionamiento . . . . .	107
4.12.	MSE y $R^2$ obtenidos con <i>Wavelet</i> y API como entrada al SVR por cada estado de funcionamiento . . . . .	108
4.13.	MSE y $R^2$ obtenidos con SOM y FCM como entrada al SVR por cada estado de funcionamiento . . . . .	109

4.14.	MSE y $R^2$ obtenidos con SOM y API como entrada al SVR por cada estado de funcionamiento . . . . .	110
4.15.	MSE y $R^2$ obtenidos con FCM y API como entrada al SVR por cada estado de funcionamiento . . . . .	111
4.16.	MSE y $R^2$ obtenidos con FD, <i>Wavelet</i> y SOM como entrada al SVR por cada estado de funcionamiento . . . . .	112
4.17.	MSE y $R^2$ obtenidos con FD, <i>Wavelet</i> y FCM como entrada al SVR por cada estado de funcionamiento . . . . .	113
4.18.	MSE y $R^2$ obtenidos con FD, <i>Wavelet</i> y API como entrada al SVR por cada estado de funcionamiento . . . . .	114
4.19.	MSE y $R^2$ obtenidos con FD, SOM y FCM como entrada al SVR por cada estado de funcionamiento . . . . .	115
4.20.	MSE y $R^2$ obtenidos con FD, SOM y API como entrada al SVR por cada estado de funcionamiento . . . . .	116
4.21.	MSE y $R^2$ obtenidos con FD, FCM y API como entrada al SVR por cada estado de funcionamiento . . . . .	117
4.22.	MSE y $R^2$ obtenidos con <i>Wavelet</i> , SOM y FCM como entrada al SVR por cada estado de funcionamiento . . . . .	118
4.23.	MSE y $R^2$ obtenidos con <i>Wavelet</i> , SOM y API como entrada al SVR por cada estado de funcionamiento . . . . .	119
4.24.	MSE y $R^2$ obtenidos con <i>Wavelet</i> , FCM y API como entrada al SVR por cada estado de funcionamiento . . . . .	120
4.25.	MSE y $R^2$ obtenidos con SOM, FCM y API como entrada al SVR por cada estado de funcionamiento . . . . .	121
4.26.	MSE y $R^2$ obtenidos con FD, <i>Wavelet</i> , SOM y FCM como entrada al SVR por cada estado de funcionamiento . . . . .	122
4.27.	MSE y $R^2$ obtenidos con FD, <i>Wavelet</i> , SOM y API como entrada al SVR por cada estado de funcionamiento . . . . .	123
4.28.	MSE y $R^2$ obtenidos con FD, <i>Wavelet</i> , FCM y API como entrada al SVR por cada estado de funcionamiento . . . . .	124
4.29.	MSE y $R^2$ obtenidos con FD, SOM, FCM y API como entrada al SVR por cada estado de funcionamiento . . . . .	125
4.30.	MSE y $R^2$ obtenidos con <i>Wavelet</i> , SOM, FCM y API como entrada al SVR por cada estado de funcionamiento . . . . .	126
4.31.	MSE y $R^2$ obtenidos con FD, <i>Wavelet</i> , SOM, FCM y API como entrada al SVR por cada estado de funcionamiento . . . . .	127
4.32.	MSE y $R^2$ obtenidos con todas las características posibles como entrada al SVR para la fusión de los estados bomba normal y bomba funciona con crudo viscoso en bomba funciona normal (incluyendo crudo viscoso) . . . .	128

- 4.33. Cuadro resumen de los mejores regresores en función de los extractores de características y los estados de funcionamiento de BMC con los datos de E/V y P, calculando la Media Armónica (M.A) entre el  $R^2$  de V y P . . . 129

# Índice de abreviaturas

BMC: Bombeo Mecánico Convencional  
CDS: Carta Dinográfica de Superficie  
CDF: Carta Dinográfica de Fondo  
PFM: Proyecto de Fin de Máster  
CP/XP: Crudos Pesados y Extrapesados  
SVM: *Support Vector Machines*  
RBF: *Radial Basis Function*  
FD: *Fourier Descriptors*  
SOM: *Self Organizing Map*  
FCM: *Fuzzy C-Means*  
API: *American Petroleum Institute*  
BAM: *Bidirectional Associative Memories*  
ANN: *Artificial Neural Networks*  
DFT: *Discrete Fourier Transform*  
FFT: *Fast Fourier Transform*  
SV: *Support Vector*  
SVM-R: *Support Vector Machine for Regression*  
SVR: *Support Vector Regressor*  
M.A: Media Armónica

# Resumen

Se desarrolló un sistema de reconocimiento de patrones por medio de cartas dinagráficas de fondo para estimar el estado de funcionamiento de pozos de petróleo que utilicen como método de levantamiento bombeo mecánico convencional. Primero se implantó el método de las características para inferir la carta dinagráfica de fondo a partir de la carta dinagráfica de superficie, por medio de la resolución de la ecuación de onda. Luego se realizó un extractor de características usando técnicas como: descriptores de Fourier, Ondículas, Mapas Autorganizativos y C-Medias Difuso. Se identificaron los siguientes estados de funcionamiento de la bomba: bombeo normal, bombeo de fluido viscoso, filtración en la válvula viajera, filtración en la válvula fija, interferencia de gas, golpe de fluido, golpe de bomba – pistón arriba, golpe de bomba – pistón abajo, falla en caja de transmisión, cabillas partidas y otros estados no identificados en este Proyecto Final de Máster (PFM). Para realizar el entrenamiento de la máquina de aprendizaje se tienen alrededor de 467 cartas dinagráficas diagnosticadas, donde más de una carta posee múltiples estados activos (o clases). Finalmente se implementó un regresor usando Maquinas de Soporte Vectorial, en el cual se obtuvo un error cuadrático medio de  $4,35 \times 10^{-4}$  y un coeficiente de correlación de 0,9731 promedio en entrenamiento y validación y un error cuadrático medio de  $2,28 \times 10^{-2}$  y un coeficiente de correlación de 0,5672 promedio en pruebas entre los estados: interferencia de gas, golpe de fluido, bomba normal, cabillas partidas, filtración en la válvula fija y otro estado no determinado en el presente PFM .

**Palabras clave:** Sistema de Reconocimiento de Patrones, Bombeo Mecánico Convencional, Maquinas Vectores Soporte para Regresión, Mapas Autorganizativos, C-Medias Difuso, Ondícula, Descriptores de Fourier

# Abstract

A pattern recognition system was developed to estimate working status of oil wells with artificial lift method sucker rod pumping by analyzing dynamometers cards. Method of Characteristic was implemented with the purpose of solve wave equation, in order to infer downhole dynamometer card from surface dynamometer card. Feature extractor was done using techniques as Fourier Descriptor, Wavelet, Self Organizing Maps and Fuzzy C-Means. States identified were normal pump (including viscous oil), leak travelling valve, leak standing valve, plunger hitting top, plunger hitting bottom, gearbox failure, broken-dropped of sucker rod, and others states not identified in this master's degree final project. Training phase of machine learning was done using 467 dynamometer cards diagnosed, where more than one card has multiple label (or active classes). Finally, a regressor was done using support vector machines, in which average mean square error of  $4,35 \times 10^{-3}$  and correlation coefficient of 0.9731 in training and validation phase and average mean square error of  $2,28 \times 10^{-2}$  and correlation coefficient of 0.5672 were obtained among states: gas interference, fluid pound, normal pump, broken-dropped of sucker rod, leak standing valve and others states not identified in this master's degree final project.

**Keywords:** Pattern Recognition System, Sucker Rod Pumping, Support Vector Machine for Regression, Self-Organizing Map, Fuzzy C-Means, Wavelet, Fourier Descriptors

# Capítulo 1

## Introducción

### 1.1. Introducción

El petróleo y sus derivados es hoy en día la principal fuente de energía en el mundo. Esto es debido a que no se conoce otra fuente de energía con características tan atractivas: alta densidad de energía, fácil transportabilidad, relativa abundancia y riesgo considerable para su manejo. Todas las fuentes de energía tienen riesgo cuando se está manejando, pero este riesgo es mucho menor respecto a otras fuentes de energía de alta densidad, por ejemplo, la energía nuclear.

La mayoría de las cosas que compramos y consumimos vienen derivados de algún proceso de conversión del crudo. En efecto, la Administración de Información Energética de los Estados Unidos de Norte América informa que el consumo de crudo en el mundo diario es alrededor de 92.13 millones de barriles el año pasado [1], con una tendencia de ir consumiendo más energía en el transcurso del tiempo [2].

De forma general, el crudo es extraído de los yacimientos petrolíferos por la energía almacenada del petróleo comprimido. Esta energía es obtenida por la diferencia de presión entre el yacimiento y el hoyo perforado. Si esta diferencia de presión es lo suficientemente alta, entonces dicho pozo será capaz de producir crudo de forma natural. Esta diferencia de presión es conocida como *drawdown* [3].

Después de producir por un periodo de tiempo, la presión del yacimiento ira disminuyendo y como resultado, el *drawdown* será menor. Por ende, el crudo no será capaz de emerger a la superficie o en su defecto, el volumen aportado a la superficie no será sustentable económicamente. En esta etapa, se debe incorporar métodos de levantamiento artificial para resolver el impedimento descrito anteriormente. De esta forma se logra que el pozo produzca nuevamente a condiciones óptimas [3].

Uno de los objetivos principales del levantamiento artificial es maximizar la tasa de producción de los pozos. Por ende, los métodos de levantamiento artificial pueden dividirse en dos, basado en el tipo de mecanismo de levantamiento: por gas y por bombeo. Los métodos de levantamiento por gas usan gas comprimido inyectado desde la superficie a ciertos puntos en la tubería de producción (o *tubing*, como se llama en la terminología petrolero). El gas hará que la densidad de la columna de fluido sea menor, causando la reducción en la presión del hoyo y por ende incrementar producción. Los métodos de bombeo vienen dados por la inclusión de una bomba a cierta profundidad a lo largo del *tubing*. Esta bomba hará que la presión del hoyo sea menor, incrementando el *drawdown* y de esta forma incrementar producción. Los métodos por bombeo más usados son: Bombeo Electrosumergible –BES– (*Electrical Submersible Pump*), Bombeo por Cavidades Progresivas –BCP– (*Progressive Cavity Pumps*), Bombeo Mecánico Convencional –BMC– (*Sucker Rod Pump*) entre otros [3].

Para determinar el método de levantamiento artificial más apropiado de los pozos viene en función de un análisis nodal, comprendiendo algunas variables descritas a continuación: presión, profundidad, caudal, temperatura, propiedades corrosivas del crudo, presencia de gas, cantidad de sólidos en el crudo y densidad. Esta última es una de las más importantes, ya que un crudo muy viscoso hace difícil su bombeo desde el yacimiento hasta la superficie [3].

Los métodos de levantamiento artificial más apropiados para CP/XP por excelencia en función del caudal de aporte del yacimiento son: BCP y BMC [3].

Venezuela es el país con mayor cantidad de reservas probadas de crudo en el mundo [4, 5], con 297.735 millones de barriles ( $4,72 \times 10^{10} m^3$ ) desglosado en: condensados, livianos y medianos en 22.794 millones de barriles y pesados y extrapesados en 274.941 millones de barriles [5].

BMC fue el primer método de levantamiento artificial de crudo usado en la industria petrolera, ocasionando de esta manera que ha sido el más estudiado y por sus características sea el método con mayor adaptabilidad a las condiciones de los pozos. Se puede colocar casi en cualquier tipo de pozo productor de petróleo, aunque esto no implica que sea la opción más rentable y beneficiosa económicamente, teniendo en cuenta que otros métodos de levantamiento artificial puede aportar mayor caudal de producción [6].

En los pozos BMC puede instalarse instrumentos en la unidad de superficie para verificar el correcto funcionamiento de este método de levantamiento. En este caso a través de un sensor de posición (inclinómetro) y esfuerzo (carga) se realiza una gráfica de

carga vs. posición llamada Carta Dinagráfica de Superficie (CDS) o *Surface Dynamometer Card* [7].

La CDS representa los esfuerzos de la barra pulida vs. la posición en la unidad de superficie en una gráfica  $X - Y$ . Sin embargo, se desea medir los esfuerzos que está sometida la bomba, ya que es el equipo que tiene contacto directo con el crudo. Si no se tiene instalada una instrumentación de posición y esfuerzo en la bomba no se podría estimar la carta dinagráfica en la bomba para hacer diagnóstico del funcionamiento de dicha bomba. Esta carta dinagráfica es comúnmente llamada Carta Dinagráfica de Fondo (CDF) o *Downhole Dynamometer Card* o *Downhole Dynagraph* [8, 9].

Se prefiere estimar la CDF a través de la CDS, dado que es más fácil instrumentar la unidad de superficie que en fondo de pozo. Por ende se han desarrollado diversos métodos descritos en [10, 11, 12] para determinar la CDF a través de la CDS. La dificultad de determinar la CDF a través de la CDS viene dado por la resolución de la ecuación de onda a través de diferentes métodos: Fourier [10], Características [11] y Diferencias Finitas [12].

Una vez obtenida la CDF se desarrollara un sistema de reconocimiento de patrones para realizar diagnóstico sobre pozos BMC a partir de la CDF, obtenida usando la técnica descrita en [11], que será el objeto del presente Proyecto de Fin de Máster (PFM).

## 1.2. Motivación

PDVSA, como empresa operadora de petróleo de mayor nivel accionario en Venezuela requiere optimizar los procesos relacionados a la producción de crudo, así como incrementar la confiabilidad de las operaciones. INTEVEP (Instituto de Tecnología Venezolano del Petróleo), filial de PDVSA, tiene como misión generar soluciones tecnológicas en la cadena de valor del negocio petrolero.

INTEVEP debe desarrollar sistema que diagnostiquen las condiciones de funcionamiento de los pozos, incluyendo las fallas para todos los métodos de levantamiento de artificial en tiempo real, de manera de tomar acciones preventivas y correctivas con la finalidad de apoyar los procesos de optimización de producción de crudo.

De igual manera, existe una tecnología desarrollada en INTEVEP que consiste en un sistema operativo empotrado (Emdebian [13]) y tiempo real (Xenomai [14]) basado en Software Libre (Linux [15]) con la finalidad de tener un Controlador Automata Programable enfocado en los procesos de producción y refinación del petróleo.

En el presente caso, el desarrollo de un sistema de reconocimiento de patrones que diagnostique si un pozo está funcionando bien o presenta fallas en tiempo real para pozos con método de levantamiento artificial tipo BMC a través de la CDF, sería muy oportuno para la producción de CP/XP, teniendo en cuenta que esos tipos de crudo representan más del 92 % de las reservas probadas en el país, además de apalancar tecnologías como la descrita anteriormente, ya que el desarrollo de este sistema se implementaría en ese tipo de tecnología.

Finalmente la realización de este PFM apalanca el desarrollo del país, ya que no se contrataría para desarrollar este tipo de sistemas o compraría soluciones comerciales, sino que se contaría con uno propio para apoyar de esta manera diagnosticar de una manera más rápida las fallas que poseen dichos pozos, disminuyendo la producción diferida y por consiguiente, aumentando la producción de crudo.

### 1.3. Objetivos

Realizar un sistema diagnóstico por medio de regresión de los estados de funcionamiento para pozos con método de producción BMC. Se consideran características que puedan describir variaciones en el tiempo y la frecuencia, a saber: descriptores de Fourier y Wavelet [16], usando como algoritmo de aprendizaje Máquinas de Vectores Soporte (SVM o *Support Vector Machines*) [17]. El análisis y desarrollo de esta propuesta será por medio del entorno de desarrollo integrado MATLAB usando la librería LIBSVM [18] para el entrenamiento, validación y prueba de la SVM, con la finalidad de implementarlo sobre sistemas de adquisición datos tiempo real.

Con la finalidad de lograr el objetivo planteado se debe realizar lo siguiente:

- Implementar el Método de las Características según [11] para determinar CDF en función de la CDS.
- Procesar y clasificar de las cartas dinagráficas adquiridas de forma manual, provenientes de diferentes campos petroleros de Venezuela.
- Realizar el estado de arte de los diferentes métodos de extracción de características utilizadas para diagnóstico sobre BMC usando cartas dinagráficas.
- Diseñar e implementar como método de extracción características: Wavelet y Descriptor de Fourier a las CDF procesadas.
- Entrenar, probar y validar el regresor SVM variando el método de extracción de las características.

## 1.4. Estructura de la memoria

La estructura de la memoria viene dada por capítulos, donde el capítulo 1 es la introducción.

El capítulo 2 consiste en el estado del arte de los sistemas diagnóstico para pozos BMC usando técnicas de reconocimiento de patrones y procesamiento de señales.

El capítulo 3 estudia los fundamentos teóricos de los métodos de extracción de características y de la técnica de regresión usada en el presente PFM: SVM.

El capítulo 4 muestra los resultados de las diferentes configuraciones de características usadas para determinar el que obtuviese mejores métricas de regresión.

El capítulo 5 muestran las conclusiones y los trabajos futuros para mejorar el desempeño del regresor desarrollado en el presente PFM.

# Capítulo 2

## Estado del Arte

El presente capítulo muestra los avances referentes a sistemas diagnóstico para pozos BMC usando cartas dinagráficas. En efecto, se muestran las últimas y mas significativas publicaciones que se han realizado en función de diversas técnicas de reconocimiento de patrones y procesamiento de señales.

### 2.1. SVM

El trabajo descrito en [19] desarrolla un sistema de diagnóstico de fallas para BMC usando SVM. Usan alrededor de 50 a 100 cartas dinagráficas por estado. Solo trabajan con tres estados de funcionamiento: bombeo normal, suministro insuficiente de fluido y otros tipos de fallas. La cantidad de puntos en la carta dinagráfica varía de 200 a 300. La carta dinagráfica se normaliza entre  $[0,1]$  tanto en posición como en carga. La cantidad de características calculadas son 13 y obtenidas de la siguiente manera: 5 puntos de desplazamiento y 5 pendientes obtenidos de los 5 puntos de mayor curvatura, 2 valores de cargas promedio correspondientes a la carga ascendente y descendente y el área del pentágono formado por los 5 puntos de curvaturas, formando la sucesión de 4 triángulos. Obtenidas las características se normaliza entre  $[-1,1]$  debido a que en  $[0,1]$  no se obtenía buenos resultados al realizar el modelo de clasificación. Se hace una búsqueda tipo malla o rejilla de los parámetros que rigen el SVM. En este caso costo y gamma. En la data de entrenamiento obtienen más del 87 % de clases correctamente clasificadas. En la data de validación obtienen más del 91 % de clases correctamente clasificadas.

En [20] se divide la carta dinagráfica en 4 partes para luego aplicar momentos invariantes. Una vez extraídas las características por momentos invariantes se usa un SVM para realizar la clasificación. Los parámetros inherentes a la SVM como la penalidad del error (tradicionalmente denominado el factor  $C$ ) y el parámetro  $\gamma$  inherente al kernel (tipo *Radial Basis Function* -RBF- o gaussiano) son ajustados por medio de la optimización por enjambre de partículas (*Particle Swarm Optimization* - PSO). La aplicación de los

momentos invariantes viene dado por el rango del momento geométrico y el momento geométrico central. Se divide en 4 partes la carta dinagráfica y se determina hasta el tercer rango el momento geométrico y el momento geométrico central, generando 7 características por cada parte, haciendo un total de 28 características. El conjunto de entrenamiento viene dado de la siguiente manera:

- Falla tipo 1: 14 cartas dinagráficas de bomba normal
- Falla tipo 2: 20 cartas dinagráficas de filtración en la válvula viajera
- Falla tipo 3: 24 cartas dinagráficas de baja entrada de fluido
- Falla tipo 4: 10 cartas dinagráficas de cabillas partidas
- Falla tipo 5: 10 cartas dinagráficas de crudo viscoso
- Falla tipo 6: 10 cartas dinagráficas de fuga en la válvula viajera
- Falla tipo 7: 10 cartas dinagráficas de golpes de la bomba en posición de punto muerto de fondo
- Falla tipo 8: 10 cartas dinagráficas de golpes de la bomba en posición de punto muerto de tope
- Falla tipo 9: 10 cartas dinagráficas de fuga en la válvula fija
- Falla tipo 10: 10 cartas dinagráficas de producción de arena

Se utilizan 88 cartas para el entrenamiento y 40 para validación. Por medio de la aplicación de PSO se ajustan los parámetros  $C$  y  $\gamma$  de manera de obtener hasta un 98 % de acierto en la etapa de validación.

El trabajo mostrado en [21] versa sobre un sistema de reconocimiento de patrones donde se diagnostica fallas en pozos de BMC basado en 4 condiciones (4 clases: bomba normal, presencia de gas, cabillas partidas y deficiencia de liquido –bajo nivel de fluido–) de tipo uno contra todos (one vs. all). El modelo de clasificación usado fue SVM y como método de extracción de características fue de Métodos de Momentos Invariantes. Con la finalidad de que cada valor de las características tuviese la misma importancia se normalizo entre cero y uno. El tamaño del conjunto de entrenamiento fue de 400 cartas dinagráficas de superficie (100 cartas por cada clase). No muestran figuras de mérito sobre el desempeño del clasificador.

El trabajo descrito en [22] tiene como objetivo ejecutar *clustering* de forma automática, a través de la técnica de clustering espectral (*Spectral Clustering* – SC). SC usa dentro

de su algoritmo K-Medias para determinar los centros. Como bien se sabe una de las dificultades de aplicar esta técnica es la alta sensibilidad a la ubicación de los centros iniciales. Por ende, se utiliza el SC para clustering y para la ubicación de los centros se considera una mejora de la técnica heurística de optimización *Black Hole*, llamada *Fast Black Hole* (FBH). De esta manera la ubicación de los centros se trata como un problema de optimización a ser resuelta usando FBH. Luego se hace clustering de forma automática y conociendo el numero de clases a priori se puede realizar un proceso de diagnóstico de fallas. Se tiene 228 cartas dinagráficas con 11 estados: 24 de funcionamiento normal, 20 de filtración en la válvula viajera, 24 de baja entrada de fluido, 20 de cabillas partidas, 20 de crudo viscoso, 20 de fuga en la válvula viajera, 20 de golpeo arriba de bomba (*upstroke*), 20 de golpeo arriba de bomba (*downstroke*), 20 de fuga en la válvula fija, 20 de producción de arena y 20 de pistón va fuera del cilindro. Se utiliza un extractor de características como “*Freeman chain code-based method*”, donde de obtiene 10 autovectores por el cambio de curvatura y 2 autovectores por máxima y mínima carga, obteniendo un total de 12 autovectores. Se obtienen buenos resultados en el proceso de hacer clustering de forma automática.

Se muestra en [23] que a través de la regularización de los pesos en una tarea se puede llevar acabo aprendizaje multi tarea. De forma específica, con esta técnica se puede optimizar una SVM con diferentes tipos de kernel de manera de mejorar las capacidades de regularización. Usando esta técnica se obtienen mejores resultados respecto a aprendizajes multitareas simples y aprendizajes mono tarea simples. Las métricas usadas en este apartado aparte de los errores son *Precision* y *Recall*, donde esta técnica ofrece mejores prestaciones respecto a las anteriores citadas. Se muestran 4 estados de clasificación: cabillas partidas, fuga, bomba normal y baja entrada de fluido. Toma como variables de entrada diferentes características de la carga: área de la señal de carga, pico de carga, mínimo de carga y ejecución diaria; por un periodo de tiempo. Sin embargo, al comparar estas técnicas no se realizó una optimización de la SVM en determinar sus parámetros, sino que lo hacen con los parámetros fijos de la SVM. Es una técnica interesante de implementar para mejorar el desempeño de la SVM.

## 2.2. Wavelet

El trabajo que se muestra en [24] justifica el uso de análisis Wavelet en detrimento de Descriptores de Fourier (*Fourier Descriptors* - FD) ya que este último funciona para señales estacionarias. También de existir otro tipo de fenómenos (fugas, fallas, picos de carga) estos se distribuirán a lo largo del espectro de frecuencia, haciendo más difícil la detección de fallas usando FD. En una primera instancia se utiliza la transformada Wavelet para reducir el ruido que presenta las señales de esfuerzo (carga) y desplazamiento

(posición). Es decir, se usa la transformada Wavelet como filtro. Luego se realiza una descomposición usando Wavelet a 20 términos (normalizados entre  $[0,1]$ ), sirviendo de entrada a una red neuronal (*Artificial Neural Network* - ANN) con 11 salidas (diferentes tipos de falla). Para validar la propuesta toman diferentes tipos de data para clasificar y así validar la propuesta. Se nota que para una cantidad pequeña de estados (a clasificar) la estimación no es tan buena respecto a otros métodos.

En [25] se presenta un sistema de reconocimiento de patrones donde se diagnostica fallas en pozos de BMC basado en 4 condiciones (4 clases: bomba normal, presencia de gas, presencia de arenas y deficiencia de líquido –bajo nivel de fluido–). Se entrenó una red neuronal tipo RBF con cuatro salidas, donde la salida más cercana a uno se corresponde con el estado presente en la bomba. El método de extracción de características usado fue Wavelet Packet a través de una descomposición en tres niveles, generando 8 autovectores de energía que serán la entrada a la red RBF. La regla de aprendizaje de la red RBF fue la del método de cálculo directo de los centros. Se normalizo la carta dinagráfica entre 0 y 1 para obtener las características, de igual manera que los autovectores. El tiempo de muestreo fue de alrededor de 100 muestras por ciclo de movimiento (stroke). La máxima carga fue de alrededor de 77.18 kN y la mínima de 39.64 kN. No indica el tamaño del conjunto de datos (entrenamiento, validación y pruebas). No muestran figuras de mérito sobre el desempeño del clasificador.

## 2.3. Redes Neuronales

Los trabajos desarrollados en [26, 27, 28] consiste en el entrenamiento de una red neuronal artificial para reconocimiento de patrones de cartas dinagráficas de fondo en sistemas BMC. Se identifican 13 estados: operación normal, golpe de fluido, filtración en la válvula viajera, pistón atascado, fuga en válvula fija, fuga en la válvula viajera, golpe arriba (del plunger), golpe abajo (del plunger), tubería desanclada, cabillas partidas, bomba desgastada, barril atascado y barril desgastado. Se obtuvieron 6101 cartas de 100 puntos cada una: 1843 cartas de operación normal, 4123 de golpe de fluido, 15 de filtración en la válvula viajera, 78 fuga de la válvula fija y 42 de golpeo arriba del pistón. La extracción de las características se hace a través de promediar en 16 puntos la curva ascendente y 16 puntos la curva descendente. La red fue construida con 32 entradas, 16 neuronas en la capa oculta y 8 neuronas en la capa de salida (operación normal, golpeo de fluido, filtración en la válvula viajera, pistón atascado, fuga en la válvula fija, fuga en la válvula viajera, golpeo del plunger arriba y golpeo del plunger abajo) con una función de transferencia sigmoideal en las ultimas capas. En la etapa de validación se obtuvieron el 100% de resultados correctos. En la etapa de prueba de las 6101 se obtuvieron los siguientes resultados: 11 cartas no fueron clasificadas desde que las neuronas en la capa

de salida tuvieron valores por debajo de los valores de normalización (0,1) y 69 cartas fueron incorrectamente clasificadas. En efecto, de estas 69, 30 fueron correspondientes a golpeo arriba del plunger, clasificándola como normal. Se obtuvo un error del 1.30 % de las 6101 (80 cartas).

En [29] se comparan tres métodos de extracción de características: FD, Vector Geométrico de Momento (*Geometric Moment Vector* – GMV, también llamado Método de los Momentos Invariantes) y Matriz de Escala de Grises Estadístico (*Gray Level Matrix Statistics* – GLMX). Experimentos numéricos muestran que FD requiere menos tiempo de ejecución y memoria aunada a una posible pérdida de información debido al número no óptimo de descriptores. GMV requiere mayor tiempo de ejecución y memoria. GLMX ofrece una dimensión baja del vector de características con mayor tiempo de ejecución y memoria. En efecto, se trabajaron con los siguientes estados de operación: golpeo arriba de la bomba, golpeo abajo de la bomba, cabillas partidas, insuficiencia de liquido, fuga en la válvula fija, fuga en la válvula viajera, fuga en ambas válvulas (fija y viajera), pistón atascado, gas lock, golpe de fluido, arenamiento y plunger fuera del barril. FD uso 10 términos, GMV uso 7 términos y GLMX uso 6 términos. El tiempo de ejecución (de menor a mayor) de estos métodos fue: FD, GLMX y GMV. El uso en memoria (de menor a mayor) de estos métodos fue: GLMX, GMV, FD. La invariancia rotacional (característica inherente) a los métodos como FD y GMV puede generar características no adecuadas para un sistema de reconocimiento de patrones. A pesar de lo interesante del artículo no muestra un proceso de clasificación tomando como entradas estas características.

Se construye un algoritmo de reconocimiento de patrones en el trabajo realizado en [30] para pozos BMC basados en redes neuronales con retropropagación (*Back Propagation - BP*), aplicada al monitoreo económico automático y sistemas de control en un campo petrolero. Se aplicó FD a las muestras de entrada con la finalidad de reducir la dimensión de la data, incrementar la velocidad de entrenamiento, mejorar la tasa de reconocimiento y el reconocimiento en tiempo real. Se toman en consideración los siguientes estados de funcionamiento: 1. Crudo parafínico en la sarta de cabillas y en la tubería de producción 2. Parte superior de los polos se rompe 3. Presencia de gas 4. Fuga en la succión de la bomba 5. Succión y escupiendo al mismo tiempo 6. Parte inferior de los polos se rompe 7. Arenamiento 8. Baja entrada de fluido 9. Bomba funcionando normal 10. Pistón atascado 11. Pistón se encuentra excesivamente arriba 12. Fuga de válvulas por presencia de crudo parafínico 13. Trabajo normal 14. Fuga en la bomba. Se hace uso de las transformadas seno y coseno (de Fourier) truncadas a 10 términos normalizados por FD para extraer las características de funcionamiento del estado actual de pozos BMC. Se trunca a los 10 primeros términos por que generalmente toda la información referente a las cartas dinagráficas se encuentran de forma significativa en los primeros términos. La red neuronal

tiene 10 neuronas de entrada, 18 neuronas en la capa oculta y 4 neuronas en la capa de salida, teniendo  $2^4=16$  salidas discretas posibles, la cual se descarta las salidas  $[0\ 0\ 0\ 0]$  y  $[1\ 1\ 1\ 1]$ , conformando los 14 estados descritos. El algoritmo de entrenamiento usado fue descenso de gradiente con término de momento. El tiempo de muestreo fue de alrededor de 360 muestras por ciclo de movimiento (stroke). No indica el tamaño del conjunto de datos (entrenamiento, validación y pruebas). No muestran figuras de mérito sobre el desempeño del sistema de reconocimiento de patrones.

En [31] se construye una red neuronal artificial usando mapas autorganizativos (*Self Organizing Map* – SOM) para hacer clustering sobre las cartas dinagráficas, de manera de obtener buenas capacidades de clasificación y generalización, y por ende, realizar un diagnóstico de fallas automático sobre el sistema de bombeo mecánico convencional. Se obtuvieron 7887 cartas dinagráficas, de las cuales fueron 1550 seleccionadas para entrenamiento y 6337 para validación y prueba del modelo. Se dividieron de la siguiente forma: 3051 cartas con operación normal, 3563 cartas con golpe de fluido, 393 cartas con presencia de gas, 564 cartas con depósito de parafinas y 316 con golpe de fondo (golpe abajo). Se tiene 216 muestras en un *stroke*. Se tiene desde 54, 72, 108 y 216 neuronas y desde épocas de entrenamiento 1, 10, 50, 100, 200, 300, 400 y 500, donde se eligieron 216 neuronas de entrada para mayor exactitud en la estimación y 100 épocas ya que al pasar de esta cantidad el modelo no aprendía significativamente. Se usaron 5 neuronas en la capa oculta (competitiva). Con fines comparativos también se entrenó una red neuronal tipo retropropagación (*Back Propagation* - BP) con 20 neuronas en la capa oculta, 5 neuronas en la capa de salida, 216 neuronas en la capa de entrada, 400 épocas. Se compara los resultados obtenidos entre BP y SOM, donde SOM tiene mayor exactitud, mejor estabilidad y mayor capacidad de generalización de los datos.

Los trabajos [32, 33] presentan descriptores gráficos (tanto en el contorno / borde como en la región interna de la imagen) y algunas herramientas para el cálculo de distancias y estadística (correlación de Pearson). Los descriptores de borde utilizados son: centroide, curvatura, K-Curvatura y FD. Se utilizaron 1500 cartas dinagráficas para identificar los cuatro estados de operación más común: bomba normal, *gas lock*, golpe de fluido y fuga de válvulas (fija y viajera). El descriptor de borde es generado para cada carta dinagráfica utilizada como patrón. Luego de esto, se repite lo anterior para cada carta dinagráfica adquirida de campo. Luego de esto, se calcula la distancia Euclidea entre cada descriptor de borde de las cartas patrones. El resultado del cálculo de la distancia es comparada con una función minimizadora. El valor más bajo se enlaza con el patrón más cercano de la carta de campo. Dado los resultados obtenidos por FD y K-Curvatura tienen el mejor desempeño a excepción de los casos: golpe de fluido y *gas lock*, usando tanto distancia Euclidea como correlación de Pearson. Dado los descriptores de borde usados

tanto con Distancia Euclidea como con correlación de Pearson el mejor es FD. Centroide también es bueno si se usa con correlación de Pearson. Finalmente hacen una prueba de la invariancia de las características usadas. En efecto, la traslación y el escalamiento no tienen impedimento alguno en ejecutarse. Pero el punto de partida y la rotación presentan problemas de invariancia. En efecto, una carta dinagráfica en un estado determinado si se ejecuta una rotación puede quedar en otro estado (no conserva el mismo). De igual forma el punto de partida al rotarse puede quedar en otro punto de partida, no en el mismo al realizarse la rotación. Para solventar estos problemas se proponen modificaciones al descriptor. Se usa una función de tercer orden para reconocer *Gas Lock* de golpe de fluido, obteniéndose buenos resultados. Se usa el valor absoluto para subsanar el problema del punto de partida. Se obtuvieron los siguientes resultados (aciertos): 86.60 % usando FD con distancia Euclidea, 84.55 % con centroide y correlación de Pearson, 83.12 % con descriptor de Fourier y correlación de Pearson.

Se utiliza FD y una red neuronal en [34] para clasificar cartas dinagráficas de balancines. Se normaliza los datos referentes a la carta dinagráfica. Se analizan diferentes técnicas para la extracción de características como: algoritmos para el ajuste de curvas, técnicas de descomposición y descriptores analíticos. En este último apartado se usa la densidad espectral de potencia como característica, además de propiedades geométricas intrínsecas: área, perímetro, relación de delgadez y relación área – perímetro. El tiempo de adquisición fue ajustado de manera tal que la cantidad de puntos de la carta dinagráfica fuese  $2^8=256$  puntos. Lo expuesto en este artículo fue la propuesta, no se muestran resultados concluyentes.

Propone [35] hacer un análisis segmentado en cuatro porciones de la carta dinagráfica: mientras sube, se ejecuta en la subida, mientras baja y se ejecuta en la bajada. No describe el tipo de clasificador ni características geométricas determinadas, solo resalta los siguientes aspectos a considerar para extraer las características: las cartas dinagráficas no deben ser sujetas a rotación, debe tener un contorno cerrado y periódico, el proceso de normalización elimina los efectos de traslación y escala, no debe existir otra señal que se superponga de manera de crear interferencia, los patrones de las cartas dinagráficas representan una condición de la bomba (no tiene invariancia en la forma y presenta características específicas en diferentes formas de la carta) y cada carta dinagráfica puede representar más de una condición de bombeo. El base de datos de cartas dinagráficas están compuesta por: 1863 bombeo normal, 4120 de golpe de fluido, 17 de filtración en la válvula viajera, 46 fuga de la válvula fija, 84 fuga de la válvula viajera y 48 golpeo arriba del pistón.

En [36] se describe un sistema diagnóstico usando como características propiedades

geométricas de la carta dinográfica: normalización del perímetro, espesor, relación área / perímetro, carga media y distancia entre cada punto del down (o up) stroke y el centroide. Se usa como clasificador una red neuronal retropropagada con 15 neuronas en la capa oculta. El entrenamiento arroja 76.7 porcentaje de acierto (no indica con cuantas cartas se utilizó para el entrenamiento) e incluyendo 200 muestras el porcentaje de acierto esta alrededor del 90 %. Se diagnostican 10 estados: bomba normal, llenado incompleto, influencia de gas, golpeo del plunger a la válvula fija, influencia de crudo pesado, manga confusa, fuga de la válvula fija, fuga de la válvula viajera, fuga en el plunger desde el barril y cabillas partidas.

## 2.4. Fourier

Ademas de los artículos de FD analizados con anterioridad en [37] se muestra el uso diferencias finitas para resolver la ecuación de propagación de onda. Las características extraídas son de índole geométrico, a saber: área de la curva, perímetro de la curva, abscisa del centro, ordenada del centro, área de la esquina superior izquierda, área de la esquina superior derecha, área de la esquina inferior izquierda y área de la esquina inferior derecha. Se determinan 18 tipos de fallas como: baja entrada de fluido, fluido viscoso, producción de arena, presencia de gas, bloqueo por gas, fuga en la válvula fija, fuga en la válvula viajera, fuga en ambas válvulas, plunger trabajando fuera del barril, etc. Se muestra una red neuronal utilizada como método de clasificación, ajustando el número de neuronas según el teorema de Kolmogorov, usando 15 neuronas en la capa oculta. No muestran las figuras de mérito básicas para determinar si se está realizando una buena clasificación: error de entrenamiento, error de validación, falsos positivos, etc.

En [38, 39] se utiliza FD como método extractor de características. Se utilizaron 102 cartas dinográficas para identificar los cuatro estado de operación más común: bomba normal, golpe de fluido, fuga de válvulas (fija y viajera) y golpe en la bomba. El descriptor de borde es generado para cada carta dinográfica utilizada como patrón. Luego de esto, se repite lo anterior para cada carta dinográfica adquirida de campo. Luego de esto, se calcula la correlación de Pearson entre cada descriptor de borde de las cartas patrones. El resultado del cálculo de la correlación es comparada con una función minimizadora. El valor más bajo se enlaza con el patrón más cercano de la carta de campo. Dado los resultados obtenidos con FD y correlación de Pearson obtuvo el 89 % de clasificación correcta.

Los trabajos descritos en [40, 41] utilizan FD y Análisis de Componentes Principales (*Principal Component Analysis* – PCA) para identificar los modos de operación de los sistemas BMC de acuerdo a la carta dinográfica, a saber: operación normal, golpe de

fluido, fuga de válvulas (fija y viajera), columna móvil (falla en el ancla de la tubería), golpeo de la bomba e filtración en la válvula viajera. Lo interesante de estos trabajos es que la energía se concentra muchas veces en las primeras componentes y que representando las dos primeras componentes se puede crear un plano para realizar posteriormente clustering y/o clasificación.

## 2.5. Reconocimiento de Patrones

Se desarrolla en [42] un sistema basado en conocimiento para diagnosticar problemas en fondo de pozo y sugerir acciones correctivas. Se aplicó aproximación estructural de reconocimiento de patrones para detectar los siguientes estados: pump-off, tubería desanclada, plunger y válvula viajera desgastada, fuga válvula fija, golpeo arriba, golpeo abajo y fricción; sistema experto basado en reglas para detectar los siguientes estados: bomba dañada, gas lock o cabillas partidas. Se usa arboles cuaternarios (*quadtree*) para comparar cartas dinagráficas tomando como referencia el API *Catalog of Analog Computer Dynamometer Cards*. Hasta el momento de la publicación de este artículo solo se tenía el desarrollo de un prototipo, pero no se había hecho ni pruebas de campo ni desarrollo de la versión final. Posee 95 % de exactitud pero no suministra información sobre el tamaño de los conjuntos de entrenamiento, validación y pruebas.

Se desarrolla en [43] un sistema de reconocimiento de patrones a través de la medición de la desviación de la carta adquirida respecto a un conjunto de cartas almacenadas, tomando como referencia las expuestas por [8]. Se genera una librería de cartas dinagráficas tanto en la carrera superior como en la carrera inferior, llamándose *upstroke library* y *downstroke library*. Esas librerías son clasificadas de acuerdo a ciertos subgrupos exhibiendo cierto grado de similitud. Esta clasificación luego es usada para diagnosticar diversos tipos de problemas en BMC. Se extraen características de índole geométrico como: centroide, área, perímetro, índice de figura (shape index) y momentos de segundo orden. Una carta desconocida se coteja con una clase de librería de cartas dinagráficas que se encuentre dentro de un umbral específico de cada librería. Luego, si la carta es mapeada bajo ciertas condiciones específicas a un subgrupo de una clase seleccionada, este subgrupo se guarda como posible diagnóstico. Esto se aplica tanto a las cartas en carrera ascendente, descendente y total. Los diagnósticos: global, ascendente y descendente son analizados para emitir un diagnóstico final. Se realiza un procesamiento para identificar cuatro regiones de la carta para determinar: anclamiento, fuga de válvulas y buen funcionamiento de la bomba. Por ende de forma secuencial o jerárquica primero se clasifica a partir de la extracción de las características de índole geométrico de las cartas dinagráficas de forma global, ascendente y descendente por medio de comparación respecto a un umbral definido. Luego se ejecuta el procesamiento en la carrera ascendente y descendente para emitir un

diagnóstico final.

En [44] se muestra un algoritmo que automáticamente calcula la CDF a partir de la CDS y clasifica de la CDF tomando en consideración un conjunto de patrones de referencia relacionadas a condiciones anormales de operación de la bomba. El sistema propuesto está actualmente funcionando en diferentes campos petroleros de Brasil. Se realiza la inferencia de la CDF a partir de un análisis de flujo laminar y Newtoniano en el anular (entre la tubería de producción y la sarta de cabillas). Primero se ejecuta una compresión de la CDF con la finalidad de reducir la complejidad computacional. La técnica computacional utilizada es interpolación neurodifusa. Luego, la extracción de características se ejecuta a través de programación lineal difusa para calcular el grado de satisfacción de las posiciones de contorno de la CDF comprimida o CDFC  $-B_i-$ . Esta caracterización se lleva a cabo a través de la codificación del comportamiento de la CDFC. La codificación viene dada por las siguientes áreas: soporte (que debe contener puntos  $-S_j-$ ) y restrictiva (que no debe contener puntos  $-R_j-$ ) para pertenecer a un patrón específico. Dada una función de membresía  $u(B_i, S_j)$  y  $u(B_i, R_j)$  se dice que pertenece al patrón  $J$  si  $u(B_i, S_j)$  tiende a 1 y  $u(B_i, R_j)$  a 0. Diferentes regiones permitidas  $S_j$  y prohibidas  $R_k$  pueden tener relevancias diferentes  $W_j, W_k$  en la caracterización de un patrón dado. Por ende el grado de satisfacción puede ser calculada como una diferencia pesada entre  $W_j$  con  $u(B_i, S_j)$  y  $W_k$  con  $u(B_i, R_k)$ . Dada la extracción de las características de la CDFC se utiliza una red neuronal para llevar a cabo la clasificación de las CDF. El artículo muestra los resultados obtenidos respecto a un solo pozo y se enfoca más en la acción de control relacionado al sistema SICAD que a la efectividad del sistema de clasificación.

Presenta [45] un desarrollo para reconocer fallas de BMC a partir de los diferentes modelos (aproximaciones) en el reconocimiento de patrones: determinístico, estadístico, secuencial, una extensión de la secuencial y sintáctica o estructural. No se incluye la aproximación por redes neuronales. Se prueba el método tipo rejilla (modelo determinístico), con la finalidad de eliminar la información que sea irrelevante en el proceso de reconocimiento de patrones. En este caso es conocer la forma de la carta dinográfica a través de una función de similitud – distancia mejor descrita en este artículo. También se prueba el método de posición basado en los FD, limitando la cantidad de puntos de la carta a una cantidad de coeficiente finita, menor que la cantidad de puntos de la carta original a través de transformadas integrales propias de la serie de Fourier. De igual forma se prueba el método de curvatura basada en FD, procediendo de una manera similar a los FD, pero probando con una función diferente en la expansión de la serie. El último método probado es el *Attributed-String-Matching* que es una aproximación estructural. Provee de operaciones como: cambiar, insertar y eliminar símbolos (primitivas) previamente definidos. Luego se calcula la distancia del patrón de referencia respecto al patrón definido por el método

*Attributed-String-Matching*. La distancia trata de ser mínima planteando una función de optimización. Este método tiene la desventaja de no ser inmune al ruido y la solución sería combinar más símbolos. Se utilizaron 28 cartas dinagráficas para clasificar 7 estados de operación: bomba poco usada, bomba completa, filtración en la válvula viajera, golpe severo de fluido, tubería desanclada, golpe moderado de fluido y otra forma de bomba completa. Los métodos que obtuvieron mejor ajuste y tiempo de computo fueron el método de la rejilla y el método de posición basado en FD, luego el método de curvatura basada en los FD y finalmente el que ofreció peor ajuste fue *Attributed-String-Matching*.

El sistema de reconocimiento de patrones propuesto en [46] tuvo una exactitud de alrededor de 82-86 % de éxito, con un 11-15 % de falsas alarmas. Se obtuvo data de 100 pozos. Mencionan que a pesar de tener mucha data de diferentes variables (cartas dinagráficas, frecuencia, run time, spm, llenado de la bomba, torque, potencia, eficiencia de la bomba, etc.) esto no garantiza que pueda diseñar e implementar con éxito un sistema de reconocimiento de patrones. En este trabajo consideran que con la carta dinagráfica basta para la detección de fallas en los pozos. Otro apartado a considerar es determinar cuál es el conjunto de características correctas a elegir. Se menciona el área, tiempo en máxima carga durante la carrera ascendente (upstroke), etc. El tiempo en máxima carga durante upstroke es relevante porque este es mayor cuando esta bomba en funcionamiento normal respecto a la bomba cuando tiene presencia de arenas. Para determinar la máxima carga se construye una función de máxima verosimilitud basada en los datos de carga. Una de las principales ventajas de esta tecnología es su buen desempeño, incluso hasta en presencia de ruido. Ya realizado el procesamiento de las características se emplea de un clasificador estadístico, sin darle mayor importancia a este apartado.

En [47] se propone un sistema de inteligencia artificial para diagnóstico BMC compuesto por dos subsistemas basados en redes neuronales: uno de reconocimiento de patrones, que suministra información relevante para activar las redes neuronales de alto nivel, llamado Redes de Razonamiento Experto, que tienen el objetivo de simular la toma de decisión de los expertos. Una vez determinado el diagnóstico el sistema de Redes de Razonamiento Experto es capaz de regular la tasa de bombeo por medio de control difuso. Todo el análisis se hace con la carta dinagráfica de fondo. El sistema de reconocimiento de patrones realiza una disminución de la cantidad de puntos para solo analizar los más significativos. Dicha reducción se lleva a cabo por medio de un análisis geométrico específico y un análisis probabilístico de concordancia entre los patrones de referencia y los patrones adquiridos. Los patrones de referencia pueden ser reales o alterados mientras tenga correspondencia física. El sistema de Redes de Razonamiento Experto hace uso aparte de las cartas dinagráficas de fondo de la Carta de Toque bajo una arquitectura distribuida para toma de decisiones usando lógica difusa. No existen métricas de validación y pruebas del sistema

de diagnóstico, aunque muestran que han implementado varios módulos con éxito.

Finalmente en [48] se realiza una propuesta de usar como extractor de característica la extracción de los bordes, usando como método de clasificación redes neuronales –no supervisadas– tipo BAM (*Bidirectional Associative Memories*) con aprendizaje tipo Hebb. No muestra resultados del entrenamiento ni de la validación.

En las páginas siguientes se muestran el cuadro 2.1 donde se resume los aspectos resaltantes de sistemas diagnóstico usando técnicas de reconocimiento de patrones y procesamiento de señales para pozos BMC.

CUADRO 2.1: ASPECTOS RESALTANTES DE SISTEMAS DIAGNÓSTICOS USANDO TÉCNICAS DE RECONOCIMIENTO DE PATRONES Y PROCESAMIENTO DE SEÑALES PARA POZOS BMC

Fuente	Método de extracción y cantidad características	Clasificador	Cantidad de estados detectados	Exactitud	Tamaño del conjunto de Entrenamiento (E) y Validación (V)
[19]	Geométrico: 13 (5 de desplazamiento, 5 pendientes (de los puntos de mayor curvatura), 2 valores de carga específicos y un área específica	SVM	3	87% E, 91%V	50 a 100 cartas por estado
[20]	Momentos invariantes: 28. La CDF se divide en 4 secciones y se aplica dicho método para generar 7 momentos por cada sección	SVM	10	98% V	88 para E y 44 V
[21]	Momentos invariantes: 7	SVM	4	-	400 (100 por cada clase)
[23]	Tendencias cortas y largas de variables como carga, ciclos, ejecución diaria, etc. (no se especifica mucho las características)	WTRML, RTML y SVM	4	WRTL mayor que el resto con $P^1$ y $R^2$	300 pozos por 100 instantes cada uno
[22]	<i>Freeman chain code-based method</i> : 12 autovectores determinados de la siguiente forma: 10 autovectores por el cambio de curvatura y 2 autovectores por máxima y mínima carga	SC - FBH Aprendizaje No Supervisado	11	91.67%	228
[24]	Wavelet: 20 términos	ANN	11	-	-
[25]	Wavelet Packet: 8 términos	RBF	4	-	-

CONTINUACIÓN DEL CUADRO 2.1: ASPECTOS RESALTANTES DE SISTEMAS DIAGNÓSTICOS USANDO TÉCNICAS DE RECONOCIMIENTO DE PATRONES Y PROCESAMIENTO DE SEÑALES PARA POZOS BMC

Fuente	Método de extracción y cantidad características	Clasificador	Cantidad de estados detectados	Exactitud	Tamaño del conjunto de Entrenamiento (E) y Validación (V)
[26] [27] [28]	Promediación 32 puntos: 16 en la carrera ascendente y 16 en la carrera descendente	ANN	8	100 % E, 98.7 %V	6181
[29]	FD, GMV Y GLMX	-	12	-	-
[30]	FD: 10 términos	ANN	4	-	-
[31]	SOM	ANN	5	SOM 99.7% vs. BP 93.39% en promedio	7887 (1550 E y 6337 V)
[32] [33]	Centroide, curvatura, k-curvatura y FD	Distancia Euclidea y Correlación de Pearson	-	88.60 %	1500
[34]	FD, algoritmos de ajuste de curvas técnicas de descomposición y descriptores analíticos: densidad espectral de potencia, área, perímetro, relación de delgadez, relación área / perímetro	ANN	-	-	-
[35]	No se indican las características usadas: análisis segmentado de la carta (subida / bajada), no debe ser sujeta a rotación, debe tener un contorno cerrado y periódico	-	-	-	1883 bomba normal, 4129 golpe de fluido 17 filtración en la válvula viajera, 46/84 fuga de válvula fija/viajera y 48 golpeo arriba del pistón
[36]	Propiedades geométricas: perímetro (normalizado), espesor, relación área / perímetro, carga media y distancia entre cada punto del down (o up) stroke y el centroide	ANN	10	76.7 % E, 90 %V	-

FINALIZACIÓN DEL CUADRO 2.1: ASPECTOS RESALTANTES DE SISTEMAS DIAGNÓSTICOS USANDO TÉCNICAS DE RECONOCIMIENTO DE PATRONES Y PROCESAMIENTO DE SEÑALES PARA POZOS BMC

Fuente	Método de extracción y cantidad características	Clasificador	Cantidad de estados detectados	Exactitud	Tamaño del conjunto de Entrenamiento (E) y Validación (V)
[37]	Propiedades geométricas: área, perímetro, abscisa del centro, ordenada del centro, área de la esquina superior/inferior derecha/izquierda (4 regiones)	ANN	18	-	-
[38]	FD	Correlación de Pearson	-	-	-
[40, 41]	FD y PCA	Clustering	-	-	-
[42]	Aproximación estructural	Quadtree	10	95 %	-
[43]	Propiedades geométricas: centroide área, perímetro, índice de figura ( <i>shape index</i> ) y momentos de segundo orden	Desviación estándar, grado de similaridad	4	-	-
[44]	Compresión de la carta por interpolación neurodifusa y comparación de posiciones	ANN	-	-	-
[45]	Rejilla, FD de posición, FD de curvatura y <i>Attributed String Matching</i>	-	7	-	-
[46]	Tiempo en máxima carga durante la carrera ascendente ( <i>upstroke</i> )	Estadístico (no definido)	2	82 % al 86 %	-
[47]	Reducción de puntos de la CDF por análisis geométrico	ANN	-	-	-
[48]	Extracción de bordes	ANN tipo BAM (Hebb)	-	-	-

# Capítulo 3

## Marco Teórico

Este capítulo describe los fundamentos teóricos del método de levantamiento artificial, los métodos de extracción de características y el regresor a usarse en la presente disertación.

### 3.1. Bombeo Mecánico Convencional

El Bombeo Mecánico Convencional (BMC, cuyo nombre en inglés es *Sucker Rod Pumping*) es un método de levantamiento artificial cuya principal característica es la de utilizar una unidad de bombeo para transmitir movimiento a la bomba de subsuelo a través de una serie (sarta) de cabillas mediante la energía suministrada por un motor (que puede ser de combustión o eléctrico) [3, 6, 7, 8, 49, 50].

El principio de funcionamiento del BMC se fundamenta en la aplicación de una fuerza torsional que convierte el movimiento rotacional del motor –caja de engranaje, a través del sistema biela - manivela de la unidad de superficie–, en un movimiento recíprocante, aplicándose a una sarta de cabillas con el propósito de accionar la bomba de subsuelo [50].

En la figura siguiente se muestran los componentes principales de este método de levantamiento.

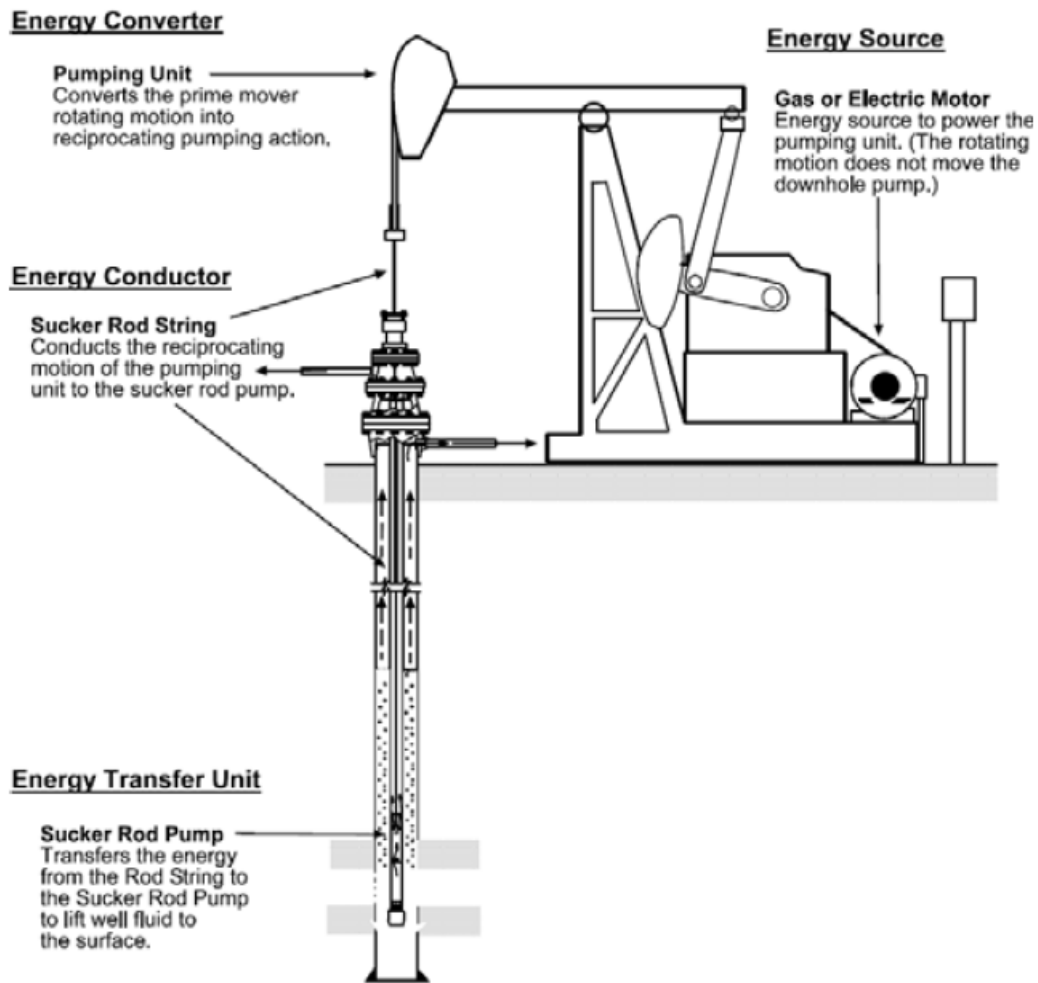


Figura 3.1: Componentes principales de un sistema de levantamiento por BMC: *Sucker Rod Pump* es la bomba de subsuelo, *Sucker Rod String* es la sarta de cabillas y *Pumping Unit* es la unidad de bombeo.

Fuente: [51]

Los equipos que constituyen los métodos de levantamiento artificial están divididos en: subsuelo y superficie. En el caso de BMC estos equipos se especifican según [52]. A continuación se describen como están constituidos.

### 3.1.1. Equipos de superficie

#### 3.1.1.1. Cabezal de Pozo

Esta formado por un conjunto de piezas, que tienen como finalidad mantener el pozo bajo control en la superficie durante la producción y permiten la salida de crudo hasta el centro de recolección de crudo (esta es llamada en la terminología petrolera estación de flujo) [49, 50].

### 3.1.1.2. Unidades de bombeo

La unidad de bombeo (balancín) es una maquina integrada y articulada, situada entre el motor y cabezal del pozo, que proporciona el movimiento apropiado para accionar la sarta de cabillas y estas a la bomba de subsuelo (equipos de subsuelo a describirse en la próxima subsección). La función principal de la unidad de bombeo es proporcionar el movimiento reciprocante apropiado, con el propósito de accionar la sarta de cabillas y estas a la bomba de subsuelo. Mediante la acción de correas y engranajes se logra reducir las velocidades de rotación. El movimiento rotatorio resultante se transforma en uno reciprocante, a través de la manivela, biela y la propia unidad de bombeo [49, 50].

Las unidades de bombeo mecánico se clasifican en:

- Unidades API
  - Convencionales
  - Unitorque
  - Balanceada por aire
- Unidades no API
  - Hidráulica
  - Tipo *winch*

Las unidades convencionales, unitorque y balanceada por aire cumplen las especificaciones del Instituto Americano del Petróleo (*American Petroleum Institute - API*). Estas unidades tienen el mismo principio de operación, su clasificación se hace con base a la geometría y la forma de contrabalanceo de las mismas [49, 50].

La unidad de bombeo posee un sistema motriz, en los cuales se convierte la energía de rotación transmitida por el motor, en movimiento reciprocante. Se hace uso de una caja de transmisión (engranajes) para transformar la velocidad del motor principal (alta revolución) a la velocidad adecuada para la unidad de bombeo [49, 50].

Si la caja de transmisión presenta mucho desgaste, este estado de funcionamiento (no deseado) muestra una CDS característica, observándose en la figura 3.2.

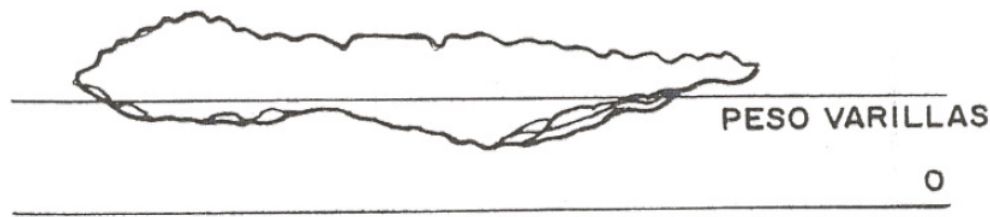


Figura 3.2: CDS de un pozo BMC donde la caja de transmisión presenta mucho desgaste (aparición de esfuerzos vibrantes adicionales sobre la carta)

Fuente: [53]

### 3.1.1.3. Prensa-estopa

Es el componente mas importante en el sistema de cabezal. Es una cámara cilíndrica que contiene los elementos de empaque que se ajustan alrededor de la barra pulida para efectuar el sello. En la parte superior se encuentra una cámara llena de aceite que sirve como medio de lubricación de la barra pulida [49, 50].

### 3.1.1.4. Barra pulida

Esta pieza soporta el peso de la sarta de cabillas, de la bomba y del fluido dentro de la tubería de producción (estos tres elementos se describirán en la subsección de equipos de subsuelo) en la carrera ascendente del balancín. Por esta razón, son fabricadas de materiales muy resistentes y, como su nombre lo indica, su superficie es completamente pulida a fin de permitir el libre paso, a través de las empacaduras que tiene el prensa-estopa sin dañarlas [49, 50].

El extremo inferior de la barra pulida se acopla a la primera sección de las cabillas y en la parte superior va conectada a las guayas del balancín por medio de un porta-vástago o un elevador [49, 50].

## 3.1.2. Equipos de subsuelo

### 3.1.2.1. Tubería de producción

Es una serie de tubos que se usa para transportar el fluido (crudo) y, al mismo tiempo, sirve de guía a la sarta de cabillas que acciona la bomba [49, 50].

### 3.1.2.2. Sarta de cabillas

Son unas piezas metálicas cilíndricas, de diámetros variables y longitudes de 25 y 30 pies, usadas para transmitir el movimiento de la unidad de bombeo hasta la bomba de

subsuelo [49, 50].

Las principales funciones de la sarta de cabillas son: transferir energía, soportar cargas y accionar la bomba. Todo esto es posible, ya que las cabillas sirven de conexión entre la bomba de subsuelo y la unidad de bombeo instalada en superficie. La barra pulida es la primera cabilla del sistema [49, 50].

Si existen cabillas partidas, este estado de funcionamiento (no deseado) muestra una CDS característica, observándose en la figura 3.3.

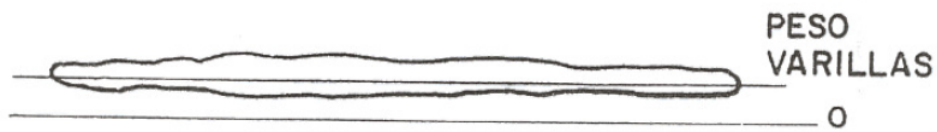


Figura 3.3: CDS de un pozo BMC con cabillas partidas  
Fuente: [53]

### 3.1.2.3. Bomba de subsuelo

Es un equipo de desplazamiento positivo desde el fondo hasta la superficie, que funciona mediante presiones diferenciales del sistema sobre sus componentes, razón por la cual es conocida también como bomba de fuerza [49, 50].

El siguiente cuadro muestra los componentes que conforman una bomba de subsuelo de BMC [49, 50].

CUADRO 3.1: COMPONENTES DE LA BOMBA DE SUBSUELO DE BMC

Componente	Descripción
Barril o cilindro de la bomba	Es una pieza cilíndrica pulida donde se almacena el fluido, dentro del cual se mueve el pistón. El diámetro del barril representa el diámetro nominal de la bomba
Pistón (o émbolo)	Es el elemento movable dentro de la bomba. Su diámetro determina la capacidad de desplazamiento y su resistencia es menor que la del cilindro. El diámetro del pistón, representa el diámetro de desplazamiento de la bomba
Válvula viajera	Esta ubicada en el pistón, de tipo bola y asiento, permite el desplazamiento del fluido desde el barril hacia la tubería de producción a través del pistón
Válvula fija	Es de tipo bola y asiento, facilita la entrada del fluido desde el pozo hasta el barril de la bomba
Zapata (o anclaje)	Es la combinación de partes reunidas inferiormente para obtener el anclaje de la bomba y efectuar un sello hermético

Si la válvula fija o viajera presentan fuga exhiben CDS/CDF características, mostrándose en la figura 3.4.

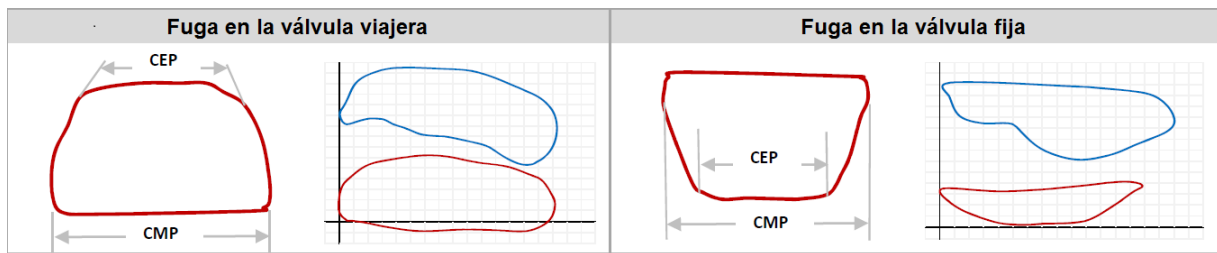


Figura 3.4: CDS (cartas de color azul) y CDF (cartas de color rojo) de un pozo BMC con fuga en la válvula viajera (izquierda) y fuga en la válvula fija (derecha). Los parámetros CMP y CEP son inherentes al funcionamiento del método BMC. El CMP es la Carrera Máxima del Pistón (máxima longitud de movimiento del pistón respecto al barril durante una carrera completa) y CEP es la Carrera Efectiva del Pistón (longitud de la carrera del pistón cuando la carga total de fluido esta actuando sobre la válvula fija)

Fuente: [54]

De la misma forma si existe golpeo arriba o abajo se debe a mal espaciamento (volumen mínimo entre la válvula fija y el pistón) de la bomba o porque la bomba tiene una relación de compresión<sup>1</sup> inadecuada o espacio inutilizado para el asentamiento de la bomba. El golpeo arriba o abajo del pistón también tienen CDS/CDF características que representa un estado no deseado de funcionamiento en pozos BMC, identificándose porque existe un cambio repentino en la carga, ya sea en la carrera ascendente o descendente respectivamente, mostrándose en la figura 3.5 solo el golpeo abajo del pistón.

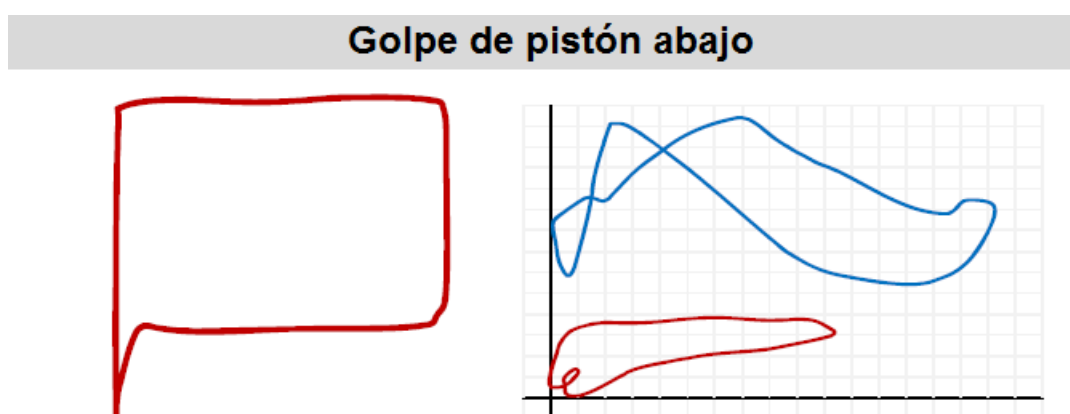


Figura 3.5: CDS (carta de color azul) y CDF (carta de color rojo) de un pozo BMC con golpe de pistón abajo. Golpe de pistón arriba es simétricamente opuesta.

Fuente: [54]

El golpe de fluido ocurre cuando la bomba no está completamente llena durante la carrera ascendente y al comienzo de la carrera descendente el pistón se desliza por un medio gaseoso hasta tocar el fluido en la cámara de la bomba, lo cual abre repentinamente transfiriendo la carga (fluido y peso de la sarta de cabilla) a la tubería, disminuyendo violentamente la carga, lo cual genera una onda de choque que se transfiere al sistema. Este estado de funcionamiento no deseado tiene CDS/CDF características, mostrándose en la figura 3.6.

<sup>1</sup>La relación de compresión se define como el cociente entre el volumen de barrido (volumen de líquido desplazado en consideraciones donde el fluido es totalmente líquido) más el volumen del espaciamento más el volumen del espacio muerto (volumen entre el punto muerto de descenso y el final de la completación) entre el volumen de espaciamento más el volumen de espacio muerto.

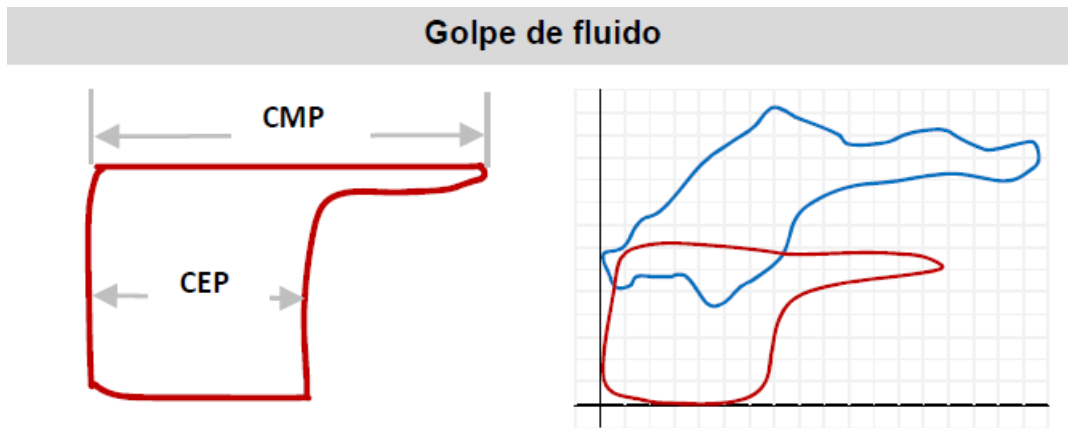


Figura 3.6: CDS (carta de color azul) y CDF (carta de color rojo) de un pozo BMC con golpe de fluido. Los parámetros CMP y CEP son inherentes al funcionamiento del método BMC y están descritos en la figura 3.4.

Fuente: [54]

#### 3.1.2.4. Ancla de gas

Consiste en un tubo ranurado o perforado, colocado bajo la zapata (de anclaje). Su objetivo es la separación gas / líquido antes de la entrada del fluido a la bomba, obteniéndose así mayor eficiencia volumétrica. Existen diferentes tipos de anclas, con la finalidad de eliminar la interferencia de gas en el funcionamiento de la bomba. Este estado no es deseado en el funcionamiento de la bomba, mostrándose en las CDS/CDF de la figura 3.7 [49, 50]. .

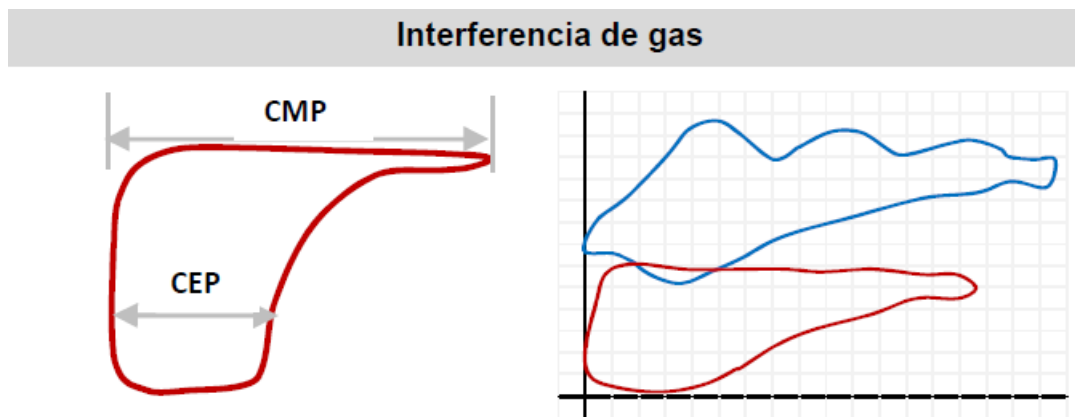


Figura 3.7: CDS (carta de color azul) y CDF (carta de color rojo) de un pozo BMC con interferencia de gas. Los parámetros CMP y CEP son inherentes al funcionamiento del método BMC y están descritos en la figura 3.4.

Fuente: [54]

## 3.2. Extractores de características

### 3.2.1. Descriptores de Fourier

Los descriptores de Fourier (FD) están descritos en un índole teórico en [55, 56] y algunas de las primeras implementaciones prácticas para reconocimiento de patrones en [57, 58].

Sea una figura de  $K$  puntos en el plano  $xy$  como se muestra en la figura 3.8.

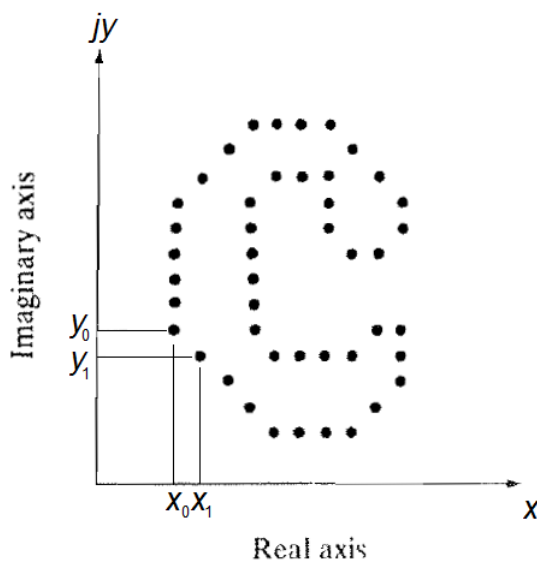


Figura 3.8: Una figura y su representación como una secuencia en números complejos. Los puntos  $(x_0, y_0)$  y  $(x_1, y_1)$  son los dos primeros puntos (arbitrarios) de la secuencia.

Fuente: [61]

Empezando en un punto arbitrario  $(x_0, y_0)$  y los pares coordenados  $(x_0, y_0), (x_1, y_1), \dots, (x_{K-1}, y_{K-1})$  se encuentran atravesando la frontera de la figura en dirección contraria de las agujas del reloj. Estas coordenadas pueden ser expresadas en la forma  $x(k) = x_k$  y  $y(k) = y_k$ . Con esta notación, la frontera misma puede ser representada como una secuencia de coordenadas  $s(k) = [x(k), y(k)]$ , para  $k = 0, 1, 2, \dots, K - 1$ . Además, cada par coordenada puede ser tratado como un número complejo de la siguiente forma

$$s(k) = x(k) + jy(k)$$

con  $k = 0, 1, 2, \dots, K - 1$ . Esto es, el eje  $x$  es tratado como el eje real y el eje  $y$  es tratado como el eje imaginario de una secuencia de números complejos. Este tratamiento reduce la representación de una señal 2-D a una tipo 1-D.

La Transformada Discreta de Fourier (*Discrete Fourier Transform* - DFT) [55, 56, 59]

de la señal  $s(k)$  es

$$a(u) = \sum_{k=0}^{K-1} s(k) e^{-j2\pi uk/K}$$

con  $u = 0, 1, 2, \dots, K - 1$ . Los coeficientes complejos  $a(u)$  son llamados descriptores de Fourier de la frontera. La transformada inversa de Fourier de esos coeficientes recuperan la señal  $s(k)$ , como se muestra a continuación

$$s(k) = \frac{1}{K} \sum_{u=0}^{K-1} a(u) e^{-j2\pi uk/K}$$

con  $k = 0, 1, 2, \dots, K - 1$ . Supóngase, sin embargo, que solo se desea escoger los primeros  $P$  coeficientes de Fourier. Esto es equivalente a realizar  $a(u) = 0$  para  $u > P - 1$  en la ecuación anterior, arrojando la siguiente aproximación de  $s(k)$

$$\hat{s}(k) = \frac{1}{P} \sum_{u=0}^{P-1} a(u) e^{-j2\pi uk/P}$$

Esta reducción de términos obedece a que los términos superiores de la DFT corresponden a los de alta frecuencia y dan detalles de la figura, mientras que los términos inferiores de la DFT corresponden a los de baja frecuencia y suministran los rasgos generales de la figura. Por ende, a medida de que  $P$  es mas pequeño, se pierde mayor detalle en la frontera [60, 61, 62].

De igual forma también se puede usar las transformadas seno y coseno discretas para generar FD [55, 56, 59]. Para un computo mas eficiente se recurre al uso de la Transformada Rápida de Fourier (*Fast Fourier Transform* - FFT) [62].

Una de las ventajas de la FD es que es muy fácil de tratar la señal por efectos de: rotación, traslación, escalamiento y variación del punto de inicio [55, 56, 59]. La desventaja de usar FD es que parte del hecho que la señal a analizar sea estacionaria [55, 56, 59], cosa que no es común en la señal de posición y carga (sobretudo en esta última), recurriéndose a usar técnicas como la que se muestran a continuación.

### 3.2.2. *Wavelet*

Las *Wavelets* son formas de ondas a diferentes escalas que mide la variación de la señal a estudiar [16]. Primero se realiza un tratamiento en tiempo continuo para luego extenderlo al discreto. La estructura de la transformada *Wavelet* se estructura por las siguientes subsecciones.

### 3.2.2.1. Generalidades

Las principales aplicaciones de las *Wavelets* son para compresión de datos (reducción de la dimensionalidad), disminución de ruido y análisis de señales.

Sea la *Wavelet* denotada por la función continua  $\psi(t)$ . Tiene las siguientes propiedades básicas:

- Tiene media cero

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (3.1)$$

Comúnmente la función  $\psi(t)$  es llamada *Wavelet* madre.

- Puede ser escalada por el parámetro  $s$  y trasladada por el parámetro  $u$

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) \quad (3.2)$$

Comúnmente la función  $\psi_{u,s}(t)$  es llamada *Wavelet* hijos.

- Se aplica a una función  $f(t)$  a la escala  $s$  y en la posición  $u$  por medio de la siguiente correlación

$$Wf(u, s) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \psi^*\left(\frac{t-u}{s}\right) dt \quad (3.3)$$

De esta propiedad se deriva que la transformada *Wavelet* puede detectar y caracterizar fenómenos transitorios a través de procedimientos de *zooming* a través de escalamiento (esto es llamado análisis multiresolución), como se muestra en la figura 3.16. Dado que la señal  $\psi(t)$  tiene media cero y suponiendo que sea una señal real los coeficientes  $Wf(u, s)$  miden la variación de  $f$  en la vecindad de  $u$  cuyo tamaño es proporcional a  $s$ . Transiciones nítidas de la señal crean grandes amplitudes en los coeficientes *Wavelets*.

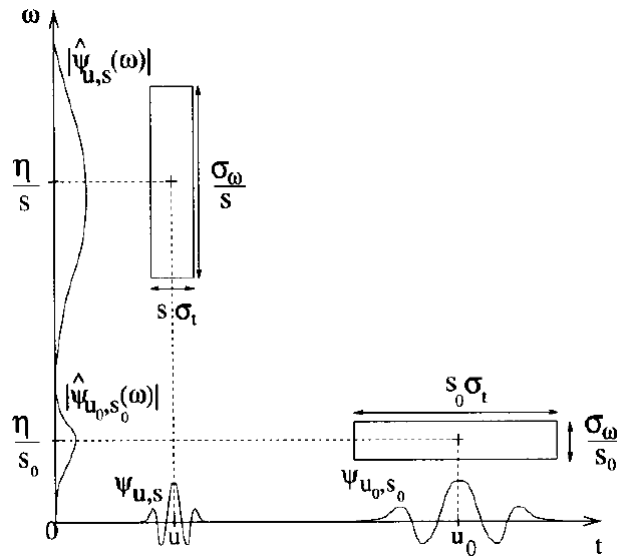


Figura 3.9: Representaciones tiempo-frecuencia de las *Wavelets*  $\psi_{u,s}$  y  $\psi_{u_0,s_0}$ . Cuando  $s$  disminuye, la representación temporal de la señal disminuye (*support*), pero la dispersión frecuencial aumenta y cubre el intervalo que esta asociado a las altas frecuencias.

Fuente: [16]

- Si  $Wf(u, s)$  es conocida solo para  $s < s_0$  y se requiere recuperar  $f$  entonces se necesita información correspondiente a  $Wf(u, s)$  para  $s > s_0$ . Esto se puede hacer por medio de la introducción de la función de escala  $\phi$ , que se correspondería a la agregación de *Wavelets* a escalas mayores que uno. Similarmente a la ecuación 3.3 se tiene

$$Lf(u, s) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \phi\left(\frac{t-u}{s}\right) dt \quad (3.4)$$

Similar a  $Wf(u, s)$ ,  $Lf(u, s)$  puede ser escalada por el parámetro  $s$  y trasladada por el parámetro  $u$

$$\phi_{u,s}(t) = \frac{1}{\sqrt{s}} \phi\left(\frac{t-u}{s}\right) \quad (3.5)$$

Comúnmente la función  $\phi_{u,s}(t)$  es llamada *Wavelet* padre.

Analizando la transformada *Wavelet* esta arroja una representación de dos dimensiones a partir de una señal de una dimensión  $f(t)$  (como se quiere analizar en el presente PFM). De esto se infiere que existe cierta redundancia que puede ser reducida e incluso removida por completo por medio de: la construcción de funciones bases eficientes (i.e. ortogonales: *Wavelet Packet* y Cosenos Locales) o del submuestreo de los parámetros de la transformada, dando origen a las tramas *Wavelet* o *Frames Wavelet*.

### 3.2.2.2. Bases *Wavelet*

Tomando en cuenta el escalamiento y traslado se puede formar la base ortonormal en  $\mathbf{L}^2(\mathbb{R})^2$ :

$$\{\psi_{j,n}(t) = a^{-j/2}\psi(a^{-j}t - nb)\}_{(j,n) \in \mathbb{Z}^2} \quad (3.6)$$

Con  $a > 1$ ,  $b > 0$  ( $a, b \in \mathbb{R}$ ). Cualquier señal de energía finita  $f$  puede ser descompuesta sobre una base ortogonal (usando una base tipo *Wavelet* en este caso)  $\{\psi_{j,n}\}_{(j,n) \in \mathbb{Z}^2}$

$$f = \sum_{j=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \psi_{j,n} \quad (3.7)$$

Como  $\psi(t)$  tiene media cero, cada suma parcial

$$d_j(t) = \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \psi_{j,n}(t) \quad (3.8)$$

Puede ser interpretada como variaciones detalladas a escala  $a^j$ . Estas capas de detalle son agregadas a todas los escalamientos de forma progresiva para mejorar la aproximación de  $f$ , hasta recuperar o reconstruir en su totalidad  $f$ . Es relativamente común encontrar  $a = 2$  para hacer descomposiciones binarias o en alta y baja frecuencia, o aproximaciones de detalle y gruesa, respectivamente.

Si  $f$  tiene variaciones suaves, se puede obtener una aproximación precisa cuando se eliminan los detalles de escala finos. Esto se hace truncando la suma de la ecuación 3.7:

$$f_J = \sum_{j=J}^{+\infty} d_j(t) \quad (3.9)$$

En la figura siguiente se muestra una representación tiempo - frecuencia de un mosaico (de bases) *Wavelet*, cubriendo el plano tiempo - frecuencia.

---

<sup>2</sup>Espacio que denota las funciones de energía finita, esto es  $\int |f(t)|^2 dt < +\infty$

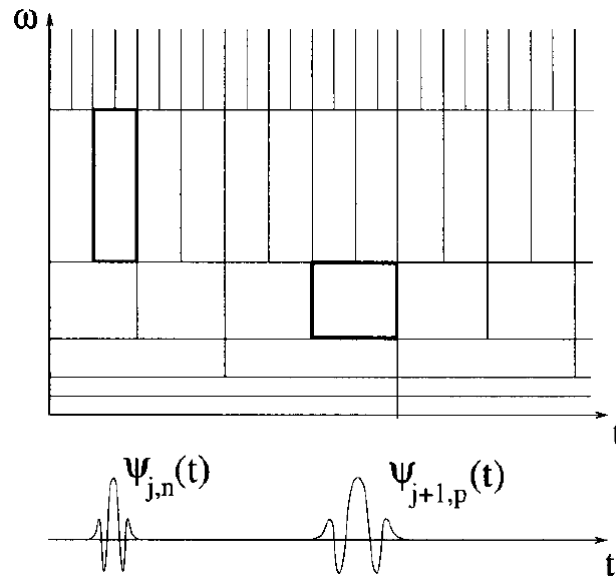


Figura 3.10: Mosaico de las bases *Wavelets* en el plano tiempo - frecuencia  
Fuente: [16]

El diseño óptimo de *Wavelets* tiene como objetivo producir el máximo número de coeficientes *Wavelets*  $\langle f, \psi_{j,n} \rangle^3$  que sean cercanos a cero. Una función  $f$  tiene una cantidad pequeña de coeficientes *Wavelets* no despreciables si los coeficientes de alta escala (alta resolución) son pequeños. Esto depende mayormente de la regularidad de  $f$ , el número de momentos de desvanecimiento de  $\psi$ , etc. Este último término se define como si la *Wavelet*  $\psi$  tiene  $p$  términos de desvanecimiento si

$$\int_{-\infty}^{\infty} t^k \psi(t) dt = 0 \quad \text{para } 0 \leq k < p \quad (3.10)$$

La definición mostrada en la ecuación 3.10 muestra que la *Wavelet*  $\psi$  es ortogonal a cualquier polinomio de grado  $p - 1$ .

### Bases Continuas

A continuación se muestran algunas bases continuas usadas para *Wavelets*:

- Haar

---

<sup>3</sup>El producto interno  $\langle f, \psi_{j,n} \rangle$  se define como  $\int_{-\infty}^{\infty} f(t) a^{-j/2} \psi(a^{-j}t - nb) dt$

Se define la *Wavelet* tipo Haar de la siguiente manera

$$\psi(t) = \begin{cases} 1 & \text{si } 0 \leq t < 1/2 \\ -1 & \text{si } 1/2 \leq t < 1 \\ 0 & \text{el resto} \end{cases} \quad (3.11)$$

- Sombrero Mexicano (*Mexican Hat*)

La *Wavelet* tipo sombrero mexicano (normalizado) viene dada por la siguiente ecuación

$$\psi(t) = \frac{2}{\pi^{1/4} \sqrt{3\sigma}} \left( \frac{t^2}{\sigma^2} - 1 \right) \exp \left( -\frac{t^2}{\sigma^2} \right) \quad (3.12)$$

- Shannon

La *Wavelet* tipo Shannon viene dada por la siguiente ecuación

$$\psi(t) = \frac{\sin(2\pi(t - 1/2))}{2\pi(t - 1/2)} - \frac{\sin(\pi(t - 1/2))}{\pi(t - 1/2)} \quad (3.13)$$

- Daubechies

Daubechies dedujo [16] que una *Wavelet* con  $p$  momentos de desvanecimiento que genera una base ortonormal  $\mathbf{L}^2(\mathbb{R})$ , tiene un tamaño (de soporte) mínimo igual a  $[-p + 1, p]$ . El soporte correspondiente a la función de escala  $\phi$  es  $[0, 2p - 1]$ . Bajo ciertas consideraciones descritas en profundidad en [16] se puede construir filtros espejos conjugados que soportan este tipo de *Wavelets*. En la figura 3.11 se muestra la *Wavelet* Daubechies con su función de escala  $\phi$  con  $p$  momentos de desvanecimiento.

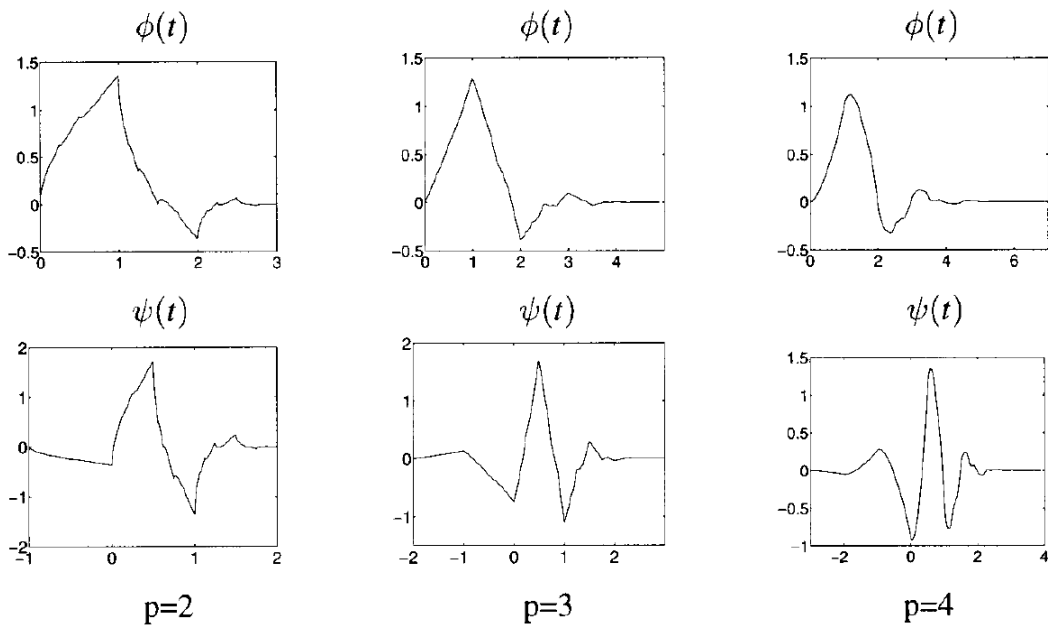


Figura 3.11: *Wavelets* Daubechies  $\psi$  con su respectiva función de escala  $\phi$  con  $p$  momentos de desvanecimiento

Fuente: [16]

- Symmlets

La *Wavelet* Symmlets se origina de forma simétrica modificando las condiciones de fase de los filtros que conforman la base *Wavelet* de Daubechies. En la figura siguiente se muestra las *Wavelets* Daubechies y Symmlets con su función de escala  $\phi$  con 8 momentos de desvanecimiento.

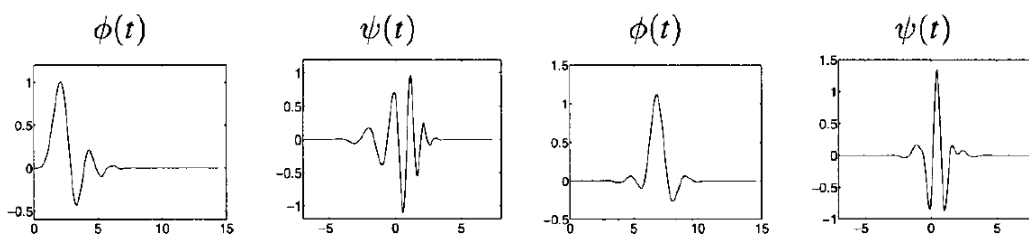


Figura 3.12: *Wavelets* Daubechies (las dos de la izquierda) y Symmlets (las dos de la derecha) , con sus funciones de escala  $\phi$ .

Fuente: [9]

- Coeiflets

La *Wavelet* tipo Coeiflets se origina añadiendo las siguientes condiciones a la función de escala

$$\int_{-\infty}^{\infty} \phi(t) dt = 1 \text{ y } \int_{-\infty}^{\infty} t^k \phi(t) dt = 0 \text{ para } 1 \leq k < p \quad (3.14)$$

La condición anterior requiere incrementar el soporte de  $\psi$ , resultando que tenga tamaño  $3p - 1$  en vez de  $2p - 1$  como la de Daubechies.

### Bases Discretas

Sea una función  $f(t)$  una señal en tiempo continuo uniformemente muestreada a intervalos  $N^{-1}$  sobre el intervalo  $[0, 1]$ . La transformada *Wavelet* solo puede ser calculada a escalas  $N^{-1} < s < 1$ . Si se desea realizar mediante cálculo discreto, es fácil de normalizar el tiempo de muestro a 1 y considerar la señal dilatada  $f(t) = f(N^{-1}t)$ . Un cambio de variable en la transformada *Wavelet* integral (ecuación 3.3) muestra que

$$Wf(u, s) = N^{-1/2}Wf(Nu, Ns) \quad (3.15)$$

Para simplificar notación, se concentran en  $f$  y denota  $f[n] = f(n)$  la señal discreta de tamaño  $N$ . Esta transformada discreta *Wavelet* se calcula en escalas  $s = a^j$ , con  $a = 2^{1/\nu}$ , donde  $\nu$  suministra una escala intermedia en la octava  $[2^j, 2^{j+1})$ .

Sea  $\psi(t)$  una *Wavelet* cuyo soporte esta incluido en  $[-K/2, K/2]$ . Para  $2 \leq a^j \leq NK^{-1}$ , entonces se define una *Wavelet* discreta escala por  $a^j$  por la siguiente ecuación

$$\psi_j[n] = \frac{1}{\sqrt{a^j}}\psi\left(\frac{n}{a^j}\right) \quad (3.16)$$

La transformada discreta *Wavelet* tiene  $Ka^j$  valores diferentes a cero en  $[-N/2, N/2]$ .

Si se desea tratar desde un punto de vista enteramente discreto considérese la señal  $f[n]$  y la *Wavelet*  $\psi_j[n]$  como señales discretas periódicas de periodo  $N$ . La transformada discreta *Wavelet* puede ser escrita como una convolución circular  $\bar{\psi}_j[n] = \psi_j^*[-n]$ :

$$Wf(n, a^j) = \sum_{m=0}^{N-1} f[m]\psi_j^*[m-n] = f[n] \otimes \bar{\psi}_j[n] \quad (3.17)$$

#### 3.2.2.3. *Wavelet Packet*

La finalidad de *Wavelet Packet* es diseñar un árbol de banco de filtros que incluya las *Wavelets* y filtros espejos conjugados. La división del eje de la frecuencia de las *Wavelets Packet* es implementada por una secuencia apropiada de convoluciones iteradas con filtros espejos conjugados. La rápida descomposición por *Wavelets Packet* es implementada con

bancos de filtros discretos.

En la figura 3.13 se muestra la división del eje frecuencial en intervalos separados de varios tamaños.

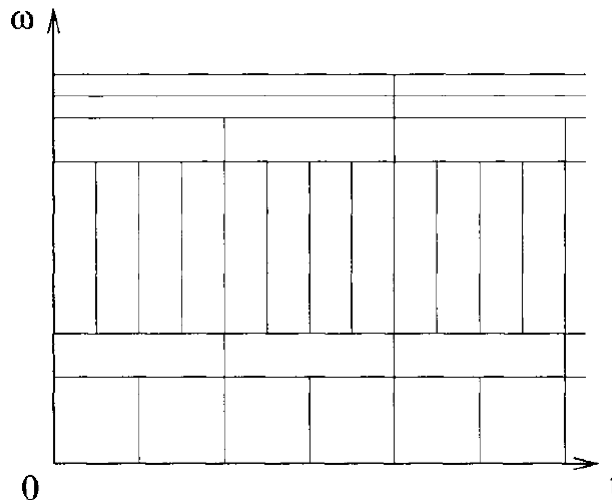


Figura 3.13: Representación en plano tiempo - frecuencia de las *Wavelet Packet*. El mosaico es obtenido por medio de la traslación en tiempo de *Wavelets Packets* cubriendo cada intervalo frecuencial.

Fuente: [9]

En la figura siguiente se muestra la descomposición obtenida por *Wavelets Packets*, de ocho niveles de aproximación con descomposición binaria.

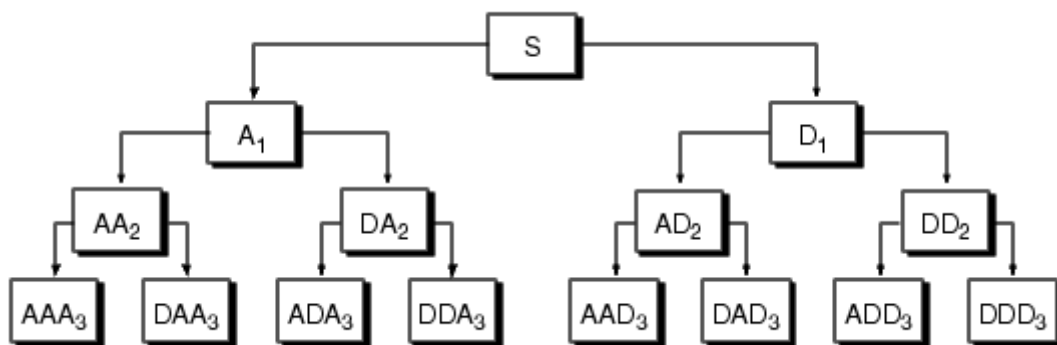


Figura 3.14: Descomposición binaria de 3 niveles usando *Wavelet Packet*. En este caso las letras A de cada descomposición viene asociada a los coeficientes de aproximación (salida de un filtro pasa bajo construido con *Wavelets*) y las letras D de cada descomposición viene asociada a los coeficientes de detalle (salida de un filtro pasa alto construido con *Wavelets*). Los subíndices denotan el nivel de descomposición

Fuente: [67]

### 3.3. Regresor: *Support Vector Machines*

Para el desarrollo del presente apartado se analizará diferentes tópicos para explicar la motivación de las Máquinas Vectores Soporte (*Support Vector Machines* - SVM).

#### 3.3.1. Teoría de Aprendizaje Estadístico

La teoría de aprendizaje estadístico (*Statistical Learning Theory* - SLT), también conocida como la teoría Vapnik - Chervonenkis (VC), es la teoría actual para realizar estimaciones estadísticas con un número finito de muestras. El enfoque conceptual de la SLT es diferente de la estadística clásica, donde el objetivo de la teoría VC es imitar el sistema, más que identificarlo. Por ende, el marco teórico de la VC es apropiado para muchas aplicaciones donde el objetivo práctico es alcanzar una buena generalización (del sistema desconocido) más que identificación del sistema exacta. Note que el objetivo posterior (de la identificación de sistemas) puede ser ni realista, para muchos problemas, debido a la maldición de la dimensionalidad [60].

La teoría VC clásica comprende [60]:

1. Condiciones de consistencia y convergencia del principio inductivo de la Minimización del Riesgo Empírico (*Empirical Risk Minimization* - ERM): de este apartado se desprende que el riesgo funcional debe tender al riesgo empírico [60].
2. Límites en la habilidad de generalización de las máquinas de aprendizaje en esas condiciones: se introduce el concepto de dimensión VC<sup>4</sup> y su correcto cálculo (o estimación), ya que si la dimensión VC es finita entonces es una condición necesaria y suficiente para rápida convergencia y consistencia del ERM. En otras palabras, las máquinas de aprendizaje bajo esta filosofía minimizan un límite superior al riesgo esperado, a diferencia del ERM que minimiza el error sobre los datos de entrenamiento.
3. Principios de inferencia inductiva basados en las limitaciones del principio ERM: ya que los límites de generalización pueden hallarse basados en la entropía VC [60], pero esta depende de las propiedades de distribuciones de probabilidad (desconocidas la mayoría de las veces). Una de las formas de solucionar la limitación anterior sería haciendo uso de la dimensión VC y construcción de las funciones de crecimiento<sup>5</sup>, justificando un nuevo principio inductivo llamado Minimización del Riesgo Estructural (*Structural Risk Minimization* - SRM). Dicho de otra forma, ERM funciona bien

---

<sup>4</sup>Es el número máximo de muestras donde todas las posibles categorizaciones binarias pueden ser inducidas (sin error) por un conjunto de funciones, en el caso específico de clasificación

<sup>5</sup>Máximo número de dicotomías que puede ser inducido sobre una muestra de tamaño  $n$  usando la función que minimice el riesgo funcional para un conjunto dado

cuando la relación del número de muestras y dimensión VC es grande y por ende el riesgo empírico tiende al riesgo verdadero. Pero si la relación anterior es pequeña (ej, menor a 20) se usa SRM ya que muestra un mecanismo formal de elegir un modelo óptimo complejo para un conjunto de muestras finitas.

4. Métodos constructivos para implementar máquinas de aprendizaje usando SRM: en este caso, las SVM definen una estructura especial donde cada elemento es indexado por el tamaño de un margen (para problemas de clasificación) y de forma más general, por un parámetro de una función de pérdida adaptativa basada en un margen.

La teoría VC clásica expone la base para construir máquinas de aprendizaje usando SRM, siendo el presente caso el desarrollo de las SVM.

### 3.3.2. Fundamentos de la SVM

Las SVM fusionan cuatro conceptos diferentes [60]:

1. Implementación del principio SRM: define una estructura especial en un conjunto de clases de equivalencia, descrito en el punto 4 de la teoría clásica de VC.
2. Mapeo de las entradas en un espacio de mayor dimensión usando un conjunto de funciones bases definidas a priori. Es común en las aplicaciones de reconocimiento de patrones para mapear los vectores de entrada a un conjunto nuevo de variables (características), que son seleccionadas de acuerdo a suposiciones sobre el problema de aprendizaje. Estas características, más allá de entradas originales, son usadas por el algoritmo de aprendizaje. Este tipo de selección de características a menudo tiene el objetivo de controlar la complejidad de los esquemas de aproximación, donde la complejidad es dependiente de la dimensionalidad de la entrada. La selección de características capitaliza la redundancia de la data con la finalidad de reducir la complejidad del problema. Esto es en contraste de la aproximación SVM que no tiene restricción en el número de funciones bases (características) usado con la finalidad de construir un mapeo de alta dimensión de las variables de entrada.
3. Funciones lineales con restricciones de la complejidad son usadas para aproximar o discriminar las muestras de entrada en un espacio de entrada alta dimensionalidad. Las SVM usan estimadores lineales para realizar la aproximación. Otras aproximaciones de aprendizaje como ANN dependen de aproximaciones no lineales directamente en el espacio de entrada. Estimadores no lineales pueden potencialmente suministrar una representación más compacta de la función de aproximación, sin embargo tienen dos desventajas: pérdida de la medición de la complejidad y de las aproximaciones de optimización, que suministra una solución óptima global.

Estimaciones exactas del modelo de complejidad pueden ser obtenidas con estimadores lineales. Las aproximaciones de optimización existen tal que pueden suministrar el mínimo global riesgo empírico de las funciones lineales. Por esa razón las SVM usan estimadores **lineales** en espacio de características de alta dimensionalidad.

4. Teoría de la dualidad de la optimización: es usada para la estimación de los parámetros del modelo (en un espacio de características de alta dimensionalidad) computacionalmente tratable. En teoría de optimización, el problema de optimización tiene una forma dual si las funciones de costo y las restricciones son estrictamente convexas. Resolviendo el problema dual es equivalente a resolver el problema original (o primal). Para la SVM, un problema de optimización cuadrática debe ser resuelto para determinar los parámetros de las funciones lineales bases de la expansión. Para espacios de características de alta dimensionalidad, un gran número de parámetros puede hacer el problema sea intratable. Sin embargo, en su forma dual el problema es práctico para resolver, así se escale en tamaño con el número de muestras de entrenamiento. La función de aproximación lineal corresponde a la solución dual y es dada por la representación de un Kernel más que la representación de la función base típica. La solución en la representación del Kernel es escrita como una suma pesada o ponderada de unos vectores soporte (*Support Vector* - SV). Los SV son un subconjunto de la data de entrenamiento en la solución al problema de aprendizaje.

Dada la fusión de las ideas descritas anteriormente, la aproximación SVM combina las siguientes ideas principales: margen, representación por Kernel y dualidad.

### 3.3.3. Clasificación usando SVM

Cuando se expone el estudio de clasificadores, es normal partir de los casos más simples. En la presente disertación se comenzará desde el caso separable y lineal, aumentando la complejidad progresivamente.

#### 3.3.3.1. Caso separable

Teniendo una data con solo dos clases (-1 y 1), la finalidad de la SVM es encontrar el hiperplano óptimo de separación entre las clases. Para el caso bidimensional el hiperplano puede interpretarse como el lugar geométrico (línea recta) que equidista de las rectas definidas por las muestras de entrenamiento (sombreadas), tal como se muestra en la figura 3.15. [61, 62]

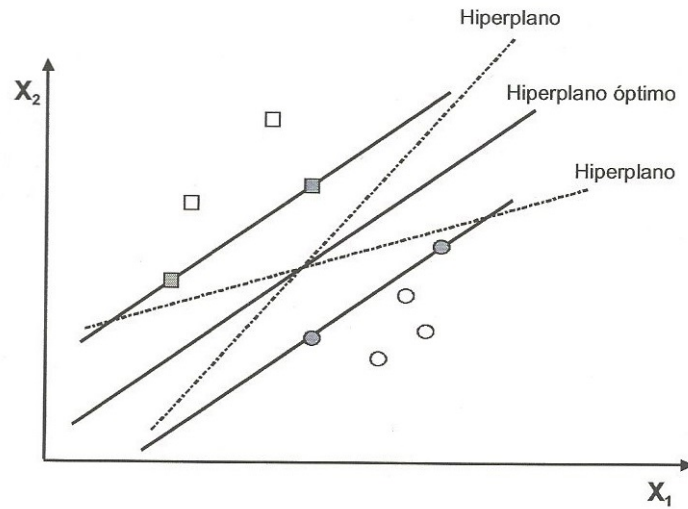


Figura 3.15: Separación de dos clases mediante un hiperplano óptimo. Las muestras sombreadas corresponden a los SV.

Fuente: [62]

Expresado de forma matemática, el problema consiste en diseñar un hiperplano de separación tal que

$$g(x) = wx + b = 0 \quad (3.18)$$

De manera que el clasificador haga lo propio con todos los vectores de entrenamiento. Dada la propia naturaleza de la data (separable) el kernel sugerido es lineal, como se puede notar en la ecuación 3.18. [64, 65]

Para calcular  $g(x)$  se debe calcular  $(w, b)$ . Se desea que el margen de separación entre clases sea el máximo posible, ya que de esta manera se garantiza que el error de clasificación sea mínimo entre clase. Ya que la distancia de un punto a un hiperplano es  $\frac{|g(x)|}{\|w\|}$  y sabiendo que si un punto pertenece a  $\omega_1$  (la zona con  $g(x) \geq 1$ ) y sería igual a 1, caso contrario si el punto pertenece a  $\omega_2$  (la zona con  $g(x) \leq -1$ ) y sería igual a -1 entonces se puede establecer el margen de la siguiente forma [61, 62]

$$\frac{|-1|}{\|w\|} + \frac{|1|}{\|w\|} = \frac{2}{\|w\|}$$

La figura 3.16 muestra el caso de dos clasificadores lineales  $g_1(x)$  y  $g_2(x)$ , donde tienen distancia de separación  $2z_1$  y  $2z_2$  respectivamente.

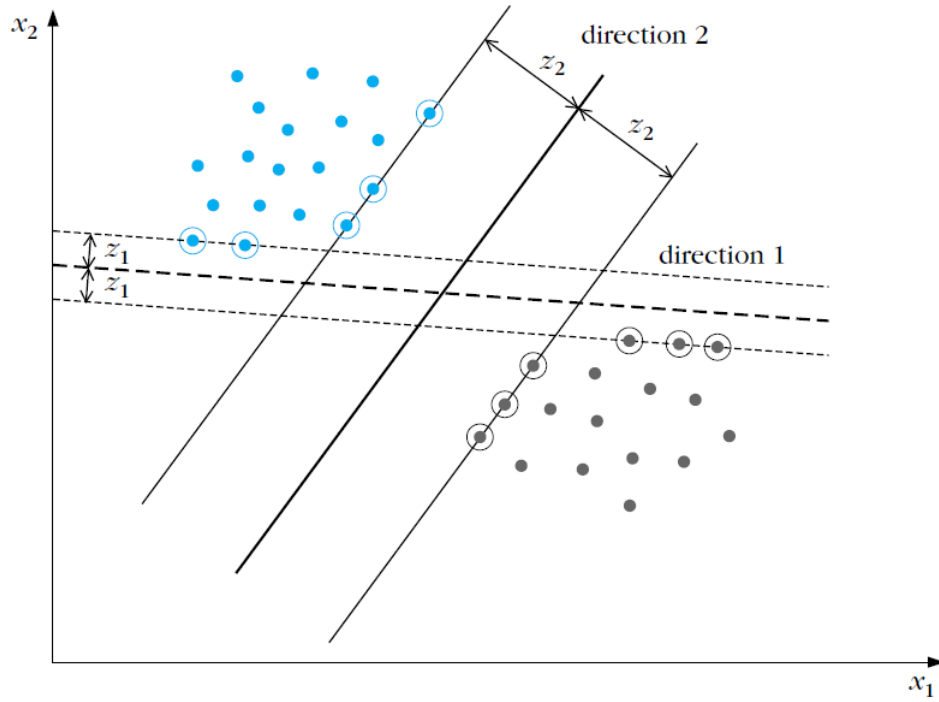


Figura 3.16: Dos posibles clasificadores lineales, mostrando distancias de separación  $z_1$  y  $z_2$  desde el hiperplano clasificador respectivo. Los SV son los que tienen la circunferencia externa de los clasificadores que apliquen.

Fuente: [61]

Maximizar el margen  $\frac{2}{\|w\|}$  es lo mismo que minimizar  $\|w\|$  o que también cualquiera de las potencias de  $\|w\|$ . Sin embargo, ya que existen funcionales cuadráticos y valiéndose de la condiciones de Karush-Kuhn-Tucker (KKT) se puede plantear el problema de encontrar  $(w, b)$  de la siguiente forma [61, 62]

$$\text{minimizar } J(w, b) = \frac{1}{2} \|w\|^2 \quad (3.19)$$

$$\text{sujeto a } y_i (wx_i + b) \geq 1, \quad i = 1, 2, \dots, N \quad (3.20)$$

Las ecuaciones 3.19 y 3.20 representan la formulación primal y es un problema de optimización no lineal (cuadrático) sujeto a restricciones lineales, cumpliendo las condiciones KKT. Aplicando las funciones de Lagrange se obtiene

$$w = \sum_{i=1}^N \lambda_i y_i x_i \quad (3.21)$$

con

$$\sum_{i=1}^N \lambda_i y_i = 0, \quad \lambda_i \geq 0, \quad i = 1, 2, \dots, N \quad (3.22)$$

La ecuación 3.22 permite obtener  $w$ . Para calcular  $w_0$

$$b = y_{SV} - \sum_{SV} \lambda_i y_i (x_i \cdot x_{SV}) \quad (3.23)$$

Una vez calculado  $b$  se tiene determinado  $g(x)$ . Si se desea analizar una muestra, se toma el signo de  $g(x)$ : si es de signo positivo entonces la muestra pertenece a  $w_1$ , caso contrario, si el signo es negativo entonces la muestra pertenece a  $w_2$ . De esta manera se tiene caracterizado el clasificador.

Prestando atención en la función de costo expresado en la ecuación 3.19, esta función es estrictamente convexa, ya que la correspondiente matriz Hessiana es definida positiva [61]. Las restricciones son funciones de desigualdades lineales. Estas dos condiciones garantizan que cualquier mínimo local es también global y único. Es decir, que el clasificador (hiperplano) óptimo basado en SVM para el caso lineal separable es único.

### 3.3.3.2. Caso No Separable

Cuando las clases no son separables, parte de lo expuesto anteriormente no es válido. Sin embargo la idea del margen sigue vigente y en este caso se crea una banda de clasificación. Ahora los vectores de entrenamiento poseen uno de los siguientes comportamientos:

- Vectores caen fuera de la banda y son correctamente clasificados. Estos vectores cumplen con la restricción de la ecuación 3.20. Dichos vectores son representados por puntos simples en la figura 3.17.
- Vectores que caen dentro de la banda y son correctamente clasificados. Estos vectores satisfacen la restricción  $0 \leq y_i (wx + b) < 1$ . Dichos vectores son representados por puntos inscritos dentro de rombos en la figura 3.17.
- Vectores que están mal clasificados. Cumplen con la restricción  $y_i (wx_i + b) < 0$  y están representados por puntos inscritos dentro de circunferencias en la figura 3.17.

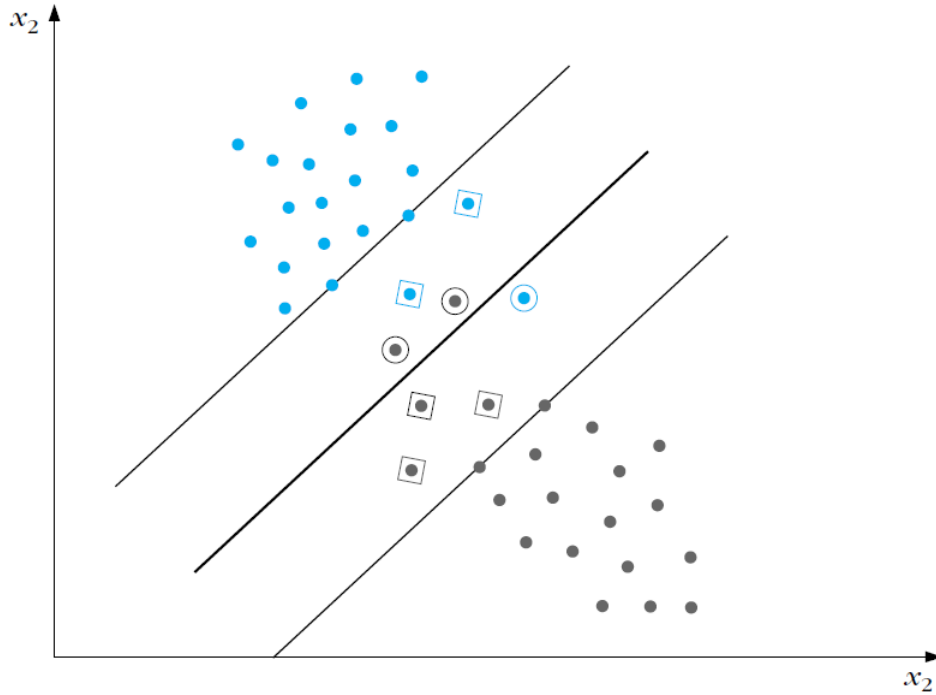


Figura 3.17: Caso no separable, donde: los puntos simples representan que están correctamente clasificados (fuera de la banda), los puntos inscritos dentro de rombos están correctamente clasificados pero cayeron dentro de la banda y los puntos inscritos dentro de círculos están mal clasificados.

Fuente: [61]

Los tres comportamientos citados pueden ser tratados bajo un simple tipo de restricciones por la introducción de un nuevo conjunto de variables  $\xi_i$ , expresado en la siguiente ecuación:

$$y_i (wx + b) \geq 1 - \xi_i \quad (3.24)$$

El primer comportamiento corresponde a  $\xi_i = 0$ , el segundo a  $0 < \xi_i \leq 1$  y el tercero  $\xi_i > 1$ . Las variables  $\xi_i$  son llamadas variables estacionarias. El objetivo ahora es hacer que el margen sea el máximo posible pero al mismo tiempo mantener los puntos con  $\xi_i > 0$  tan pequeño como sea posible. Esto es matemáticamente

$$J(w, b, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N I(\xi_i) \quad (3.25)$$

Donde  $\xi$  es el vector de parámetros  $\xi_i$ . De la ecuación 3.25 se especifica el termino  $I(\xi_i)$

$$I(\xi_i) = \begin{cases} 1 & \xi_i > 0 \\ 0 & \xi_i = 0 \end{cases} \quad (3.26)$$

El parámetro  $C$  es una constante positiva que controla la influencia relativa de  $\xi_i$  (constante de Regularización lo suficientemente grande [62]). Como la función  $I()$  es discontinua, se replantea la función de optimización de la siguiente manera

$$\text{minimizar } J(w, b, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (3.27)$$

$$\text{sujeto a } y_i (wx_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, N \quad (3.28)$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, N \quad (3.29)$$

El problema expresado en las ecuaciones 3.27, 3.28 y 3.29 es un problema de programación convexa, derivándose su solución en [61, 62].

El termino cuadrático del lado derecho de la ecuación 3.27 trata de maximizar el margen de separación mientras que el segundo termino trata de introducir el error de entrenamiento para que si la data no es linealmente separable [66, 67], se clasifique lo mejor posible, aunque se trata que este error sea mínimo [66]. Esto sirve de motivación para la próxima subsubsección.

### 3.3.3.3. Clasificadores No Lineales

Dependiendo de la complejidad de la data es probable que no pueda separarse a través de un hiperplano. Por ende, se hace uso de un Kernel para transformar el espacio de entrada a otro espacio de alta dimensionalidad, en el que los datos son separables linealmente [62]. Este ultimo espacio es llamado espacio de características [63, 65].

En vez de usar una función lineal  $g(x) = wx + b$  se usa una función Kernel  $k(x, x')$  de la siguiente forma

$$k(x, x') = \sum_{j=1}^m g_j(x) \cdot g_j(x') = \langle g(x), g(x') \rangle \quad (3.30)$$

con  $m$  finito. Como se puede notar la ecuación 3.30 se tiene un producto interno de la función  $g(x)$ , llamada función base. Para que un Kernel sea admisible como tal debe cumplirse las condiciones para estos, descritos en profundidad en [60, 65]. Una de las condiciones mas importantes a cumplirse es la de Mercer [60, 65], descrita de la siguiente forma:

$$\int \int k(x, x') \varphi(x) \varphi(x') dx dx' > 0, \quad \text{para todo } \varphi \neq 0, \quad \int \varphi^2(x) dx < \infty \quad (3.31)$$

Algunos ejemplos de las funciones bases y Kernels mas utilizados son las siguientes

[60, 62]:

- Kernels polinomiales de grado  $q$

$$k(x, x') = [\langle x \cdot x' \rangle + 1]^q \quad (3.32)$$

- Funciones base radiales de la forma

$$g(x) = \text{sign} \left( \sum_{i=1}^n \alpha_i \exp \left\{ \frac{|x - x'|^2}{\sigma^2} \right\} \right) \quad (3.33)$$

con Kernel de Base Radial

$$k(x, x') = \exp \left\{ -\frac{|x - x'|^2}{\sigma^2} \right\} \quad (3.34)$$

Este Kernel también es llamado Kernel Gaussiano.

- Funciones base de Redes Neuronales de la forma

$$g(x) = \text{sign} \left( \sum_{i=1}^n \alpha_i \tanh \{v \langle x \cdot x' \rangle + a\} + b \right) \quad (3.35)$$

y el Kernel con Producto Interno

$$k(x, x') = \tanh \{v \langle x \cdot x' \rangle + a\} \quad (3.36)$$

- Expansión de Fourier con funciones base

$$g(x) = v_0 + \sum_{j=1}^q (v_j \cos(jx) + w_j \sin(jx)) \quad (3.37)$$

y con Kernel

$$k(x, x') = \frac{\sin \left( q + \frac{1}{2} \right) (x - x')}{\sin \left( \frac{x - x'}{2} \right)} \quad (3.38)$$

Una vez resuelto el Kernel mas apropiado, la función de clasificación (o discriminante) se plantea como [60]:

$$D(x) = \sum_{i=1}^n \alpha_i^* y_i k(x, x') + b \quad (3.39)$$

Los parámetros  $\alpha_i^*$ ,  $i = 1, 2, \dots, N$  son la solución del siguiente problema de optimización cuadrático:

Minimizar

$$\frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j k(x_i, x_j) - \sum_{i=1}^N \alpha_i \quad (3.40)$$

Sujeto a las restricciones

$$\begin{aligned} \sum_{i=1}^N y_i \alpha_i &= 0 \\ 0 &\leq \alpha_i \leq C/n, \quad i = 1, 2, \dots, N \end{aligned} \quad (3.41)$$

En la literatura de optimización las ecuaciones 3.40, 3.41 representan el Problema Dual y por razones de practicidad la solución de la formulación dual es mas fácil de resolver que la primal [64, 66].

El parámetro  $b$  viene dado de la siguiente forma

$$b^* = y_{SV} - \sum_{i=1}^n \alpha_i^* y_i k(x, x_{SV}) \quad (3.42)$$

### 3.3.3.4. Clasificación Multi-Clase

La teoría desarrollada anteriormente es solo para clasificación binaria. Para problemas de clasificación con  $M$  clases, existen diferentes configuraciones [62]:

- Uno contra todos (*one-against-all*): descompone el problema de clasificación en construir una función discriminante de una clase respecto a las  $M-1$  clases restantes.
- Una frente a una (*one-versus-one*): descompone el problema de  $M$  clases en  $\frac{1}{2}M(M-1)$  problemas binarios. Cuando se debe clasificar una nueva muestra se pasa por todos los clasificadores y añade un voto a la clase ganadora para cada caso, resultando la clase ganadora la que mas votos suma.
- Aproximación directa: se propone una modificación de la función de optimización binaria para generalizar la clasificación para el numero deseado de clases  $M$

$$\text{minimizar } J(w, b, \xi) = \frac{1}{2} \sum_{m=1}^M \|w_m\|^2 + C \sum_{m=1}^M \sum_{m \neq y_i} \xi_i^m \quad (3.43)$$

$$\text{sujeto a } w_{y_i} x_i + b_{y_i} \geq w_m x_i + b_m + 2 - \xi_i^m \quad (3.44)$$

$$\xi_i^m \geq 0, \quad \{(x_i, y_i) : i \in I\}, \text{ con } I = \{1, 2, \dots, N_c\} \quad (3.45)$$

### 3.3.3.5. Regresión usando SV

Si se desea usar SVM para regresión el tratamiento es muy similar Comúnmente hacer regresión usando SV es denominado SVM-R o SVR. El patrón de entrada (datos a la cual

sirven para obtener la predicción) es mapeado a un espacio de características  $\Phi$ . Por ende, los productos internos son realizados con las imágenes de los patrones de entrenamiento bajo el mapeo  $\Phi$ . Esto se corresponde con evaluar funciones kernel  $k$  en los puntos  $k(x_i, x)$ . Finalmente los puntos productos internos son sumados usando los pesos  $v_i = \alpha_i - \alpha_i^*$ . Esto, además del término constante  $b$  conlleva la predicción de la salida. Este proceso se muestra en la figura siguiente.

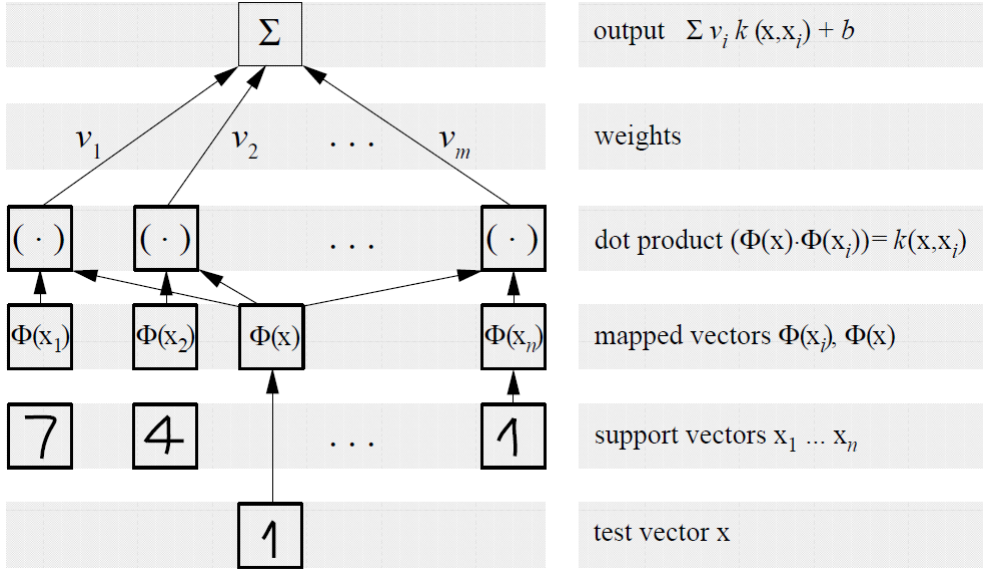


Figura 3.18: Arquitectura de un regresor usando SV  
Fuente: [65]

El proceso descrito anteriormente es muy similar a la implementación de una red neuronal con tres capas ocultas, con la diferencia que en el caso de SV los pesos de la capa de entrada son predeterminados por los patrones de entrenamiento [65].

Para labores de clasificación o regresión se varían los parámetros de la SVM:  $C$ ,  $\xi$  y los inherentes al kernel utilizado, con la finalidad de que esta máquina de aprendizaje se ajuste mejor a los datos de entrenamiento.

# Capítulo 4

## Análisis de Resultados

El presente capítulo muestra los resultados obtenidos por los diferentes tipos de extractores de características aplicados a CDF y el regresor SVM utilizado para determinar en que clase (estado) se encuentra operando el pozo BMC.

Al inicio del presente PFM el objetivo era realizar un clasificador, pero como se muestra en la próxima sección, muchas cartas dinagráficas no pertenecían en la misma proporción a una clase sino un porcentaje específico a determinada clase. Por ende, muchas cartas dinagráficas no pertenecían totalmente a una clase, limitando así la aplicación del clasificador multi-clase. De esta manera se procedió a determinar un regresor SVM con coeficiente de correlación ( $R^2$ ) lo mas cercano posible a uno y error cuadrático medio (*Mean Square Error* - MSE) lo mas pequeño posible variando las características del regresor por cada clase (estado).

### 4.1. Estudio de los datos

Dada la recolección de las CDF de referencia, existen 467 diagnosticadas, distribuyéndose de la siguiente forma:

- 188 cartas donde la bomba funciona normal
- 389 cartas donde la bomba funciona con CP/XP
- 34 cartas con filtración en la válvula viajera
- 22 cartas con filtración en la válvula fija
- 103 cartas con interferencia de gas
- 68 cartas con golpe de fluido
- 6 cartas con golpe de bomba – pistón arriba

- 132 cartas con golpe de bomba – pistón abajo
- 204 cartas con falla en caja de transmisión
- 5 cartas con cabillas partidas
- 84 cartas con otro estado no determinado en el presente PFM.

Los pozos donde se obtuvieron las CDF no representan todos los tipos de crudos (en referencia al API), como se puede ver en el siguiente histograma.

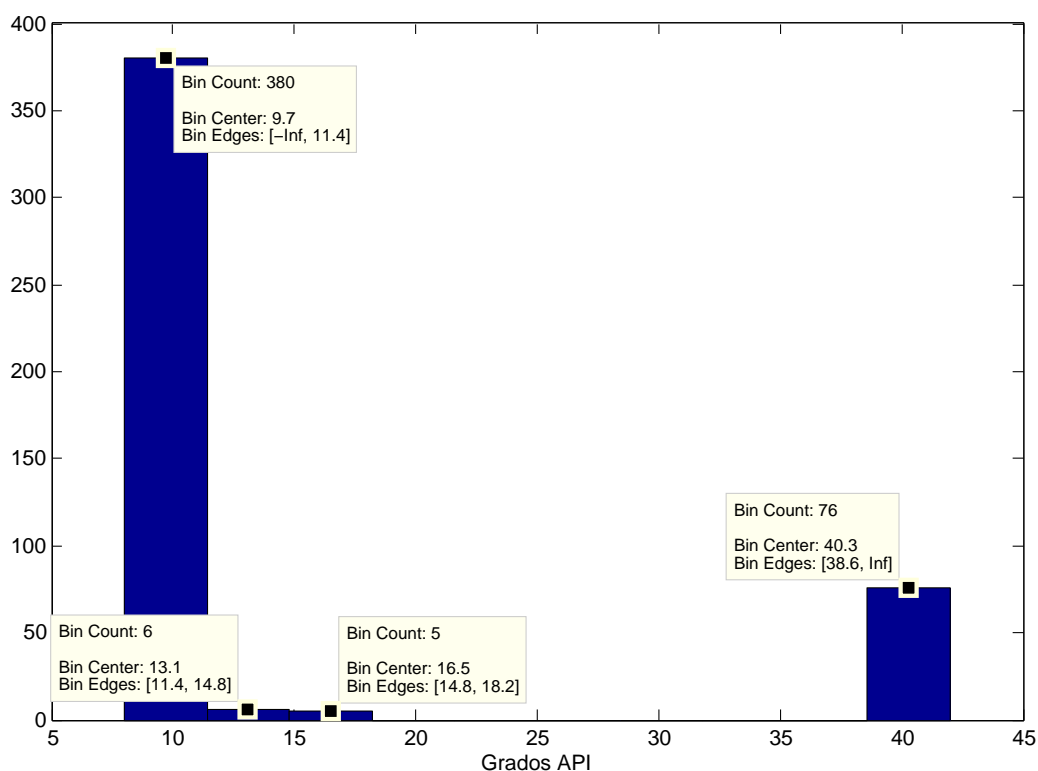


Figura 4.1: Histograma de los grados API (° API) de los pozos diagnosticados para realizar la extracción de características con todos los datos

De la figura 4.1 se puede determinar que existen 380 CDF entre 8 y 11.4 °API<sup>1</sup> (crudos pesados y extrapesados), 6 CDF entre 11.4 y 14.8 ° API (crudo pesado), 5 CDF entre 14.8 y 18.2 °API (crudo pesado) y 76 CDF entre 38.6 y 42 °API (crudo liviano), demostrando que existe un vacío entre 18.2 y 38.6 °API (desde una parte de crudo pesado, pasando por todo el rango de crudo mediano hasta parte de crudo liviano).

A continuación se presentará como fue la partición de los conjuntos de datos de entrenamiento, validación y pruebas

<sup>1</sup>Los grados API (° API) es una medida de densidad, definida como  $(141,5/GE) - 131,5$ , donde  $GE$  es la gravedad específica del crudo a 60 °F. Según el API los crudos se clasifican de la siguiente manera: crudos livianos de 30 a más de 40 °API, crudos medianos de 29.9 a 22 °API, crudos pesados (CP) de 21.9 a 10° API y crudos extrapesados (XP) menores a 10° API.

#### 4.1.1. Datos de entrenamiento y validación

Se tienen 453 CDF para realizar el entrenamiento y validación del SVR. Estas están distribuidas de la siguiente manera:

- 182 cartas donde la bomba funciona normal
- 380 cartas donde la bomba funciona con CP/XP
- 32 cartas con filtración en la válvula viajera
- 20 cartas con filtración en la válvula fija
- 101 cartas con interferencia de gas
- 66 cartas con golpe de fluido
- 4 cartas con golpe de bomba – pistón arriba
- 130 cartas con golpe de bomba – pistón abajo
- 201 cartas con falla en caja de transmisión
- 3 cartas con cabillas partidas
- 80 cartas con otro estado no determinado en el presente PFM.

#### 4.1.2. Datos de prueba

Se tienen 14 CDF para realizar la prueba del SVR. Estas están distribuidas de la siguiente manera:

- 6 cartas donde la bomba funciona normal
- 9 cartas donde la bomba funciona con CP/XP
- 2 cartas con filtración en la válvula viajera
- 2 cartas con filtración en la válvula fija
- 2 cartas con interferencia de gas
- 2 cartas con golpe de fluido
- 2 cartas con golpe de bomba – pistón arriba
- 2 cartas con golpe de bomba – pistón abajo
- 3 cartas con falla en caja de transmisión
- 2 cartas con cabillas partidas
- 4 cartas con otro estado no determinado en el presente PFM.

## 4.2. Extractor de Características

Los extractores de características utilizados en la presente disertación son FD y *Wavelet*. Sin embargo, como aporte adicional se usará dos técnicas de clustering: *Fuzzy C-Means* (FCM) [61] y *Self Organizing Maps* (SOM) [61]. A continuación se muestran casos ejemplos por estados usando cada tipo de extractor en función del estado a determinar usando SVM-R. Cada una de estas técnicas utilizaron 8 términos.

### 4.2.1. Funcionamiento del extractor donde el pozo BMC funciona (con bombeo) normal

Esta carta se corresponde con un pozo cuyo crudo es de 40 ° API. A continuación se muestra los FD obtenidos para un caso ejemplo donde la bomba esta funcionando normal.

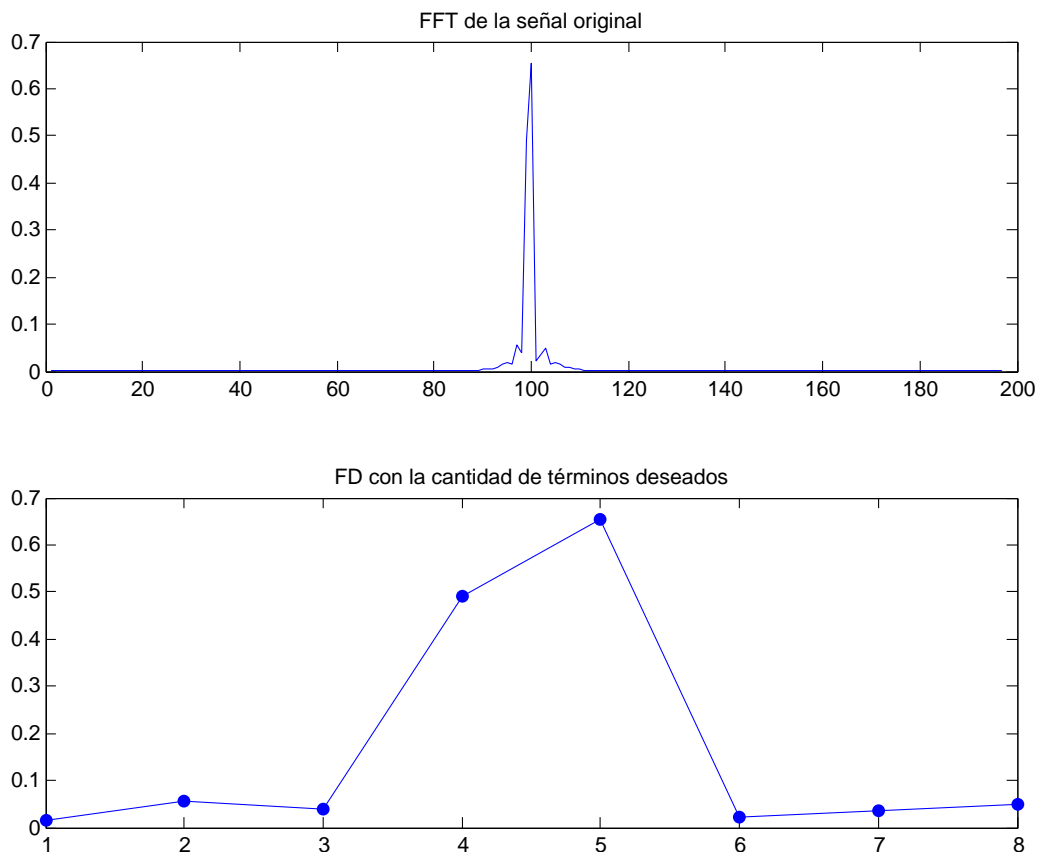


Figura 4.2: Espectro de la señal de carga para bombeo normal, obtenida usando la FFT (arriba), FD de la señal de carga con una cantidad de términos fijo (abajo)

De la misma forma se muestra la señal de carga comprimida usando *Wavelet*.

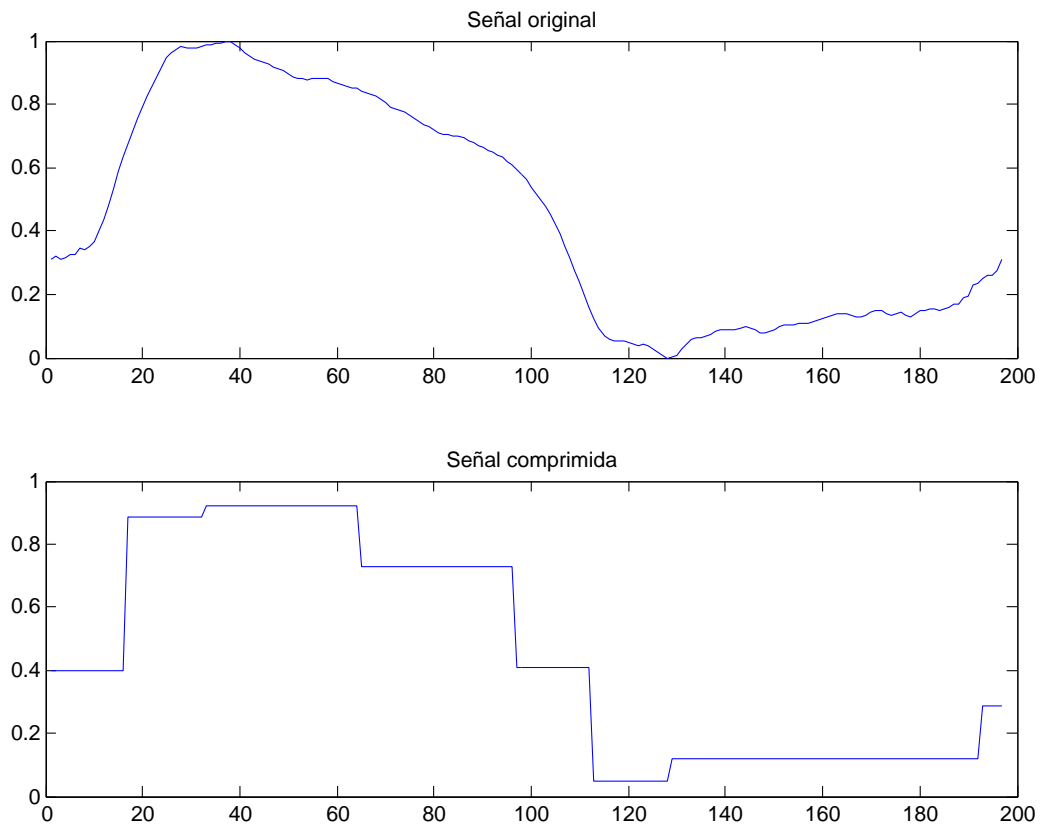


Figura 4.3: Señal de carga original para la clase bombeo normal (arriba) y comprimida usando Wavelet (abajo)

Usando la compresión por *Wavelet* se muestra como quedaría la CDF usando dicha compresión.

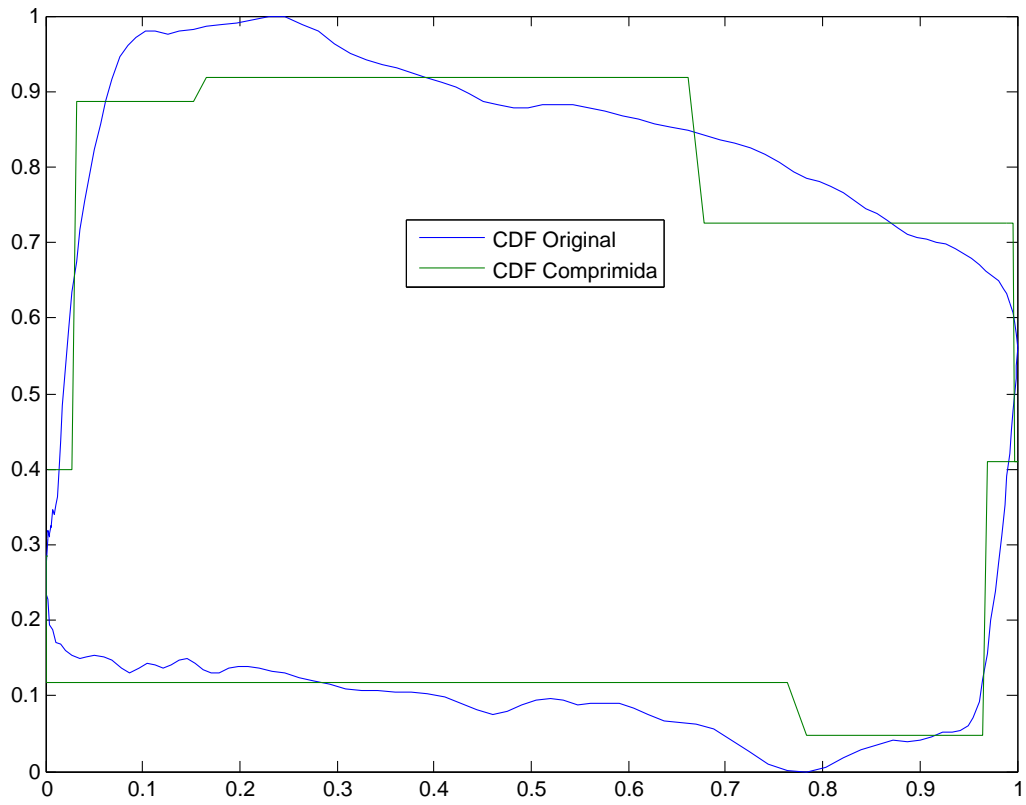


Figura 4.4: Comparación de la señal de carga original para bombeo normal y comprimida (usando Wavelet)

También se muestran los centros calculados por FCM y SOM.

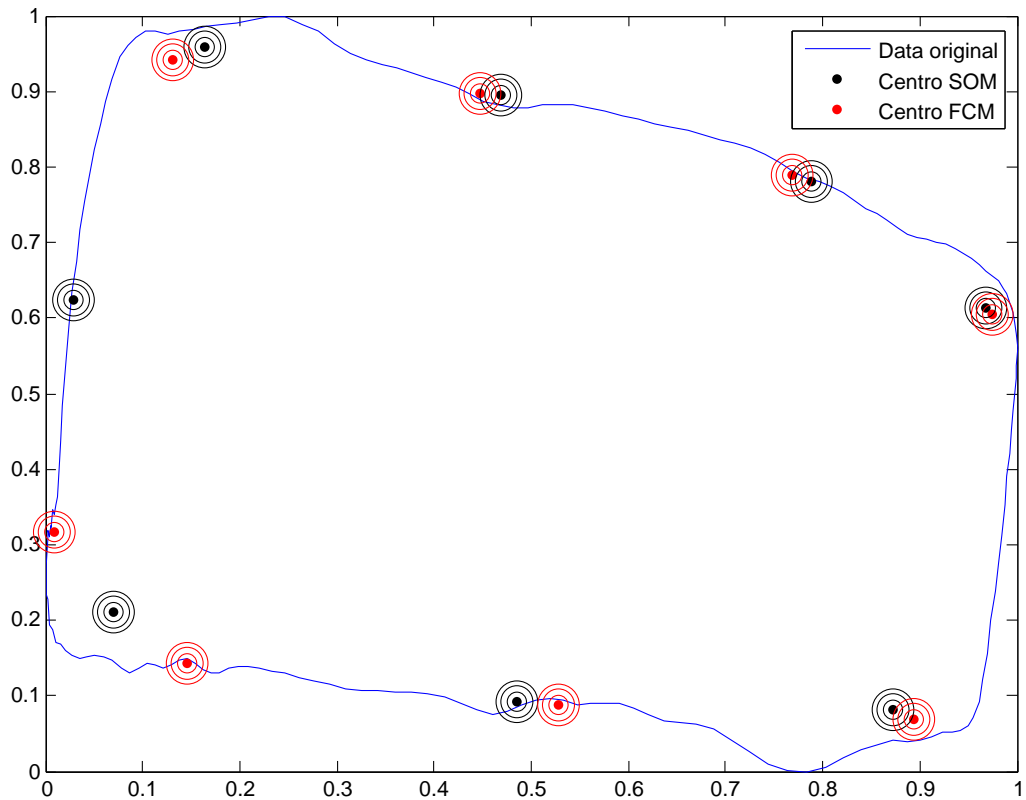


Figura 4.5: Centros determinados a la CDF para bombeo normal usando FCM y SOM

#### 4.2.2. Funcionamiento del extractor donde el pozo BMC funciona con crudo viscoso

La próxima carta se corresponde con un pozo cuyo crudo es de 11 ° API. Este caso es una condición especial de bombeo normal, ya que el pozo produce crudo viscoso. A continuación se muestra los FD obtenidos para un caso ejemplo donde la bomba esta funcionando en un estado normal.

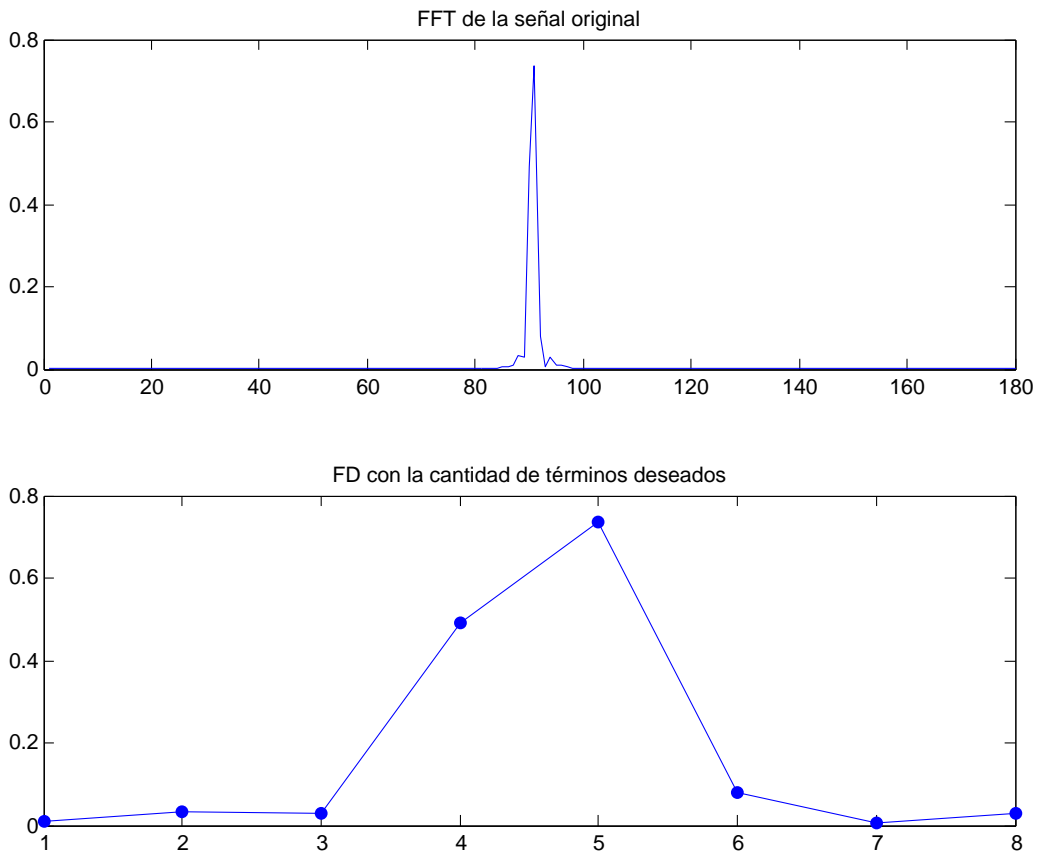


Figura 4.6: Espectro de la señal de carga para la clase bombeo de crudo viscoso, obtenida usando la FFT (arriba), FD de la señal de carga con una cantidad de términos fijo (abajo)

De la misma forma se muestra la señal de carga comprimida usando *Wavelet*.

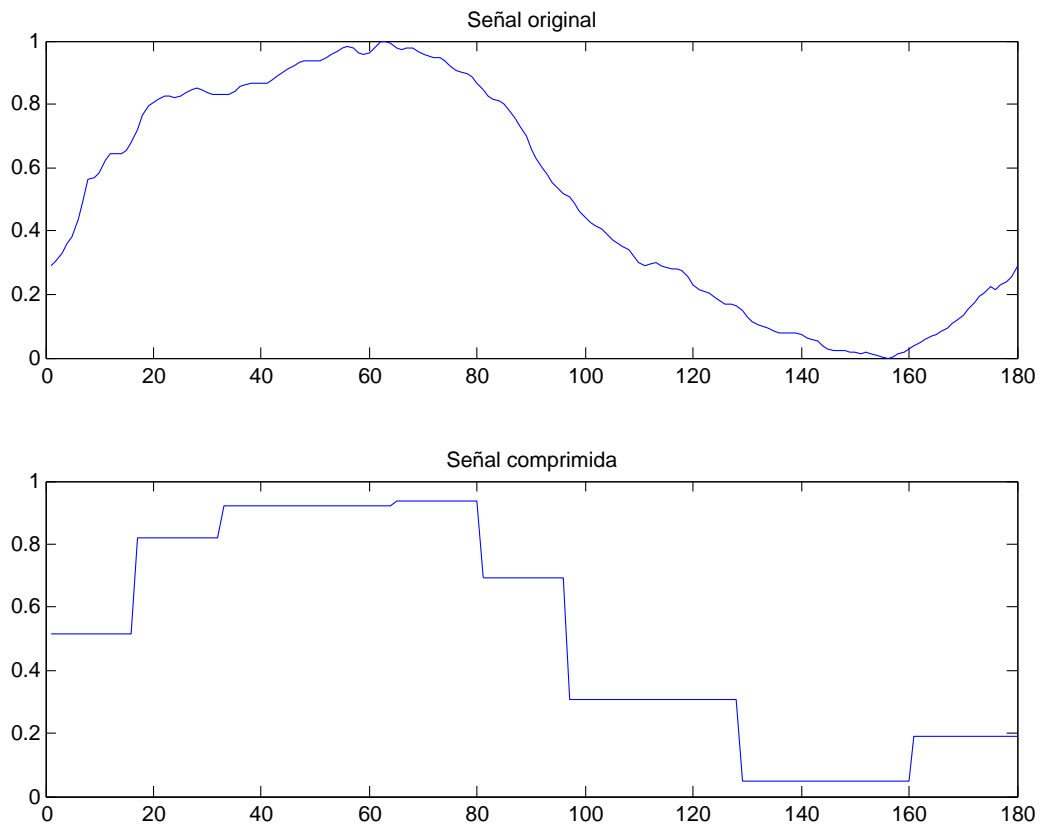


Figura 4.7: Señal de carga original para la clase bombeo de crudo viscoso (arriba) y comprimida usando Wavelet (abajo)

Usando la compresión por *Wavelet* se muestra como quedaría la CDF usando dicha compresión.

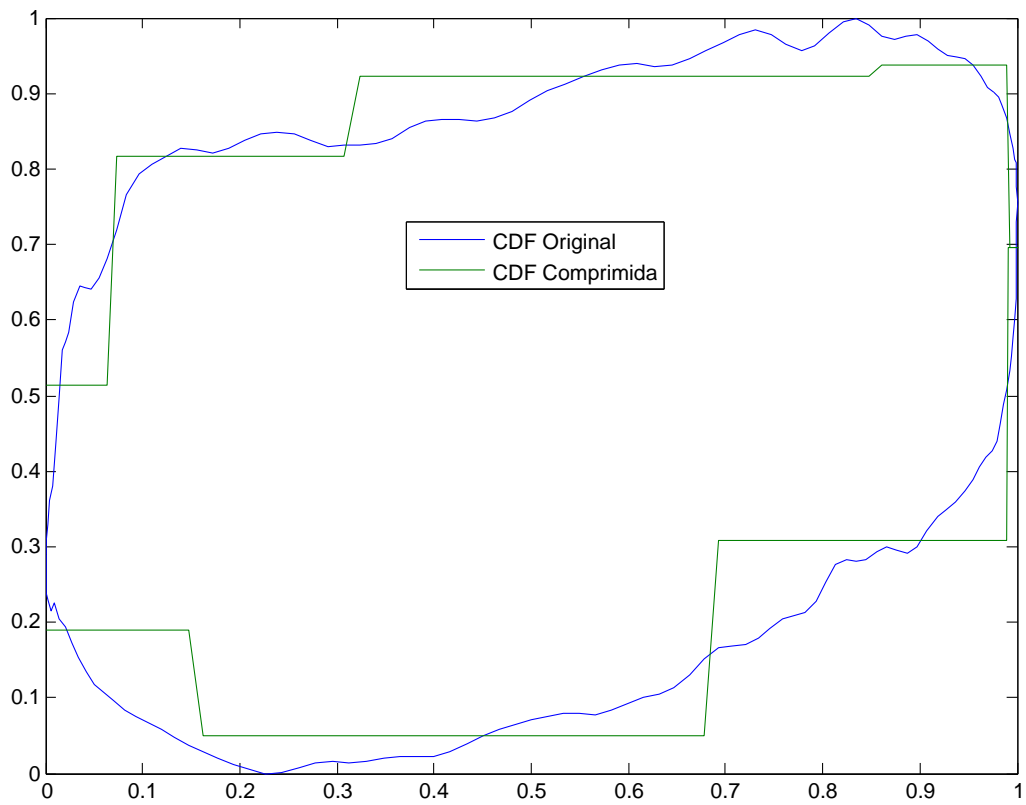


Figura 4.8: Comparación de la señal de carga original para la clase bombeo de crudo viscoso y comprimida (usando Wavelet)

También se muestran los centros calculados por FCM y SOM.

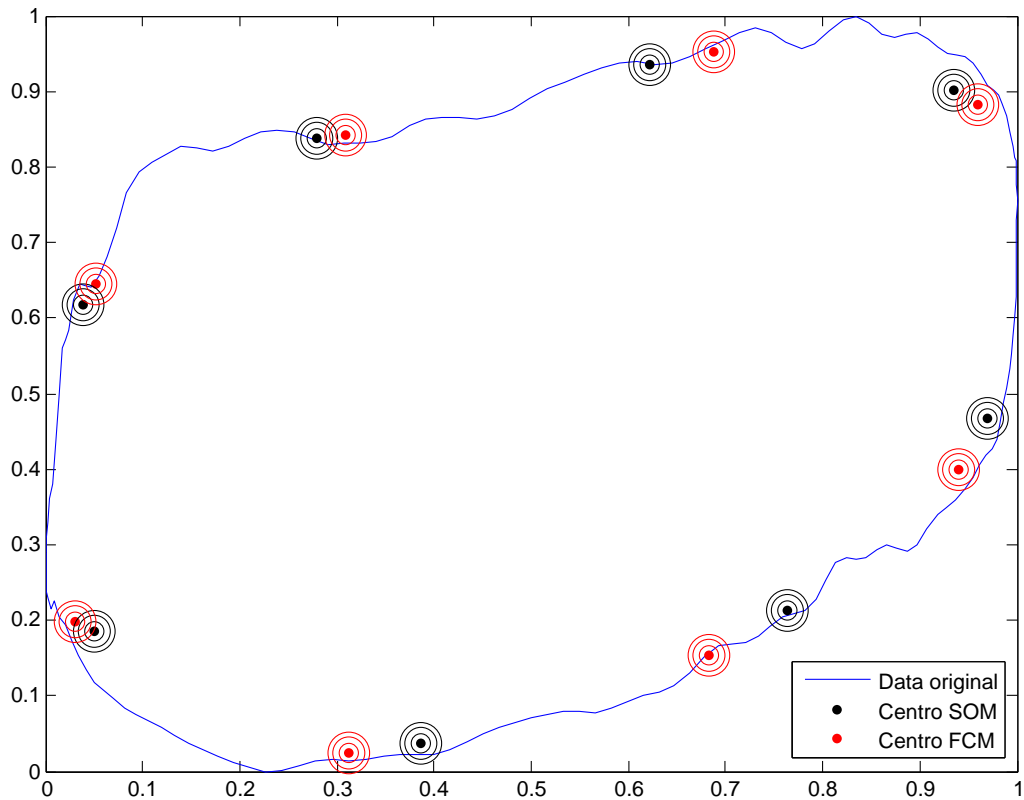


Figura 4.9: Centros determinados a la CDF para la clase bombeo de crudo viscoso usando FCM y SOM

### 4.2.3. Funcionamiento del extractor donde el pozo BMC posee filtración en la válvula viajera

La próxima carta se corresponde con un pozo cuyo crudo es de 40 ° API. A continuación se muestra los FD obtenidos para un caso ejemplo donde la bomba esta funcionando con filtración en la válvula viajera.

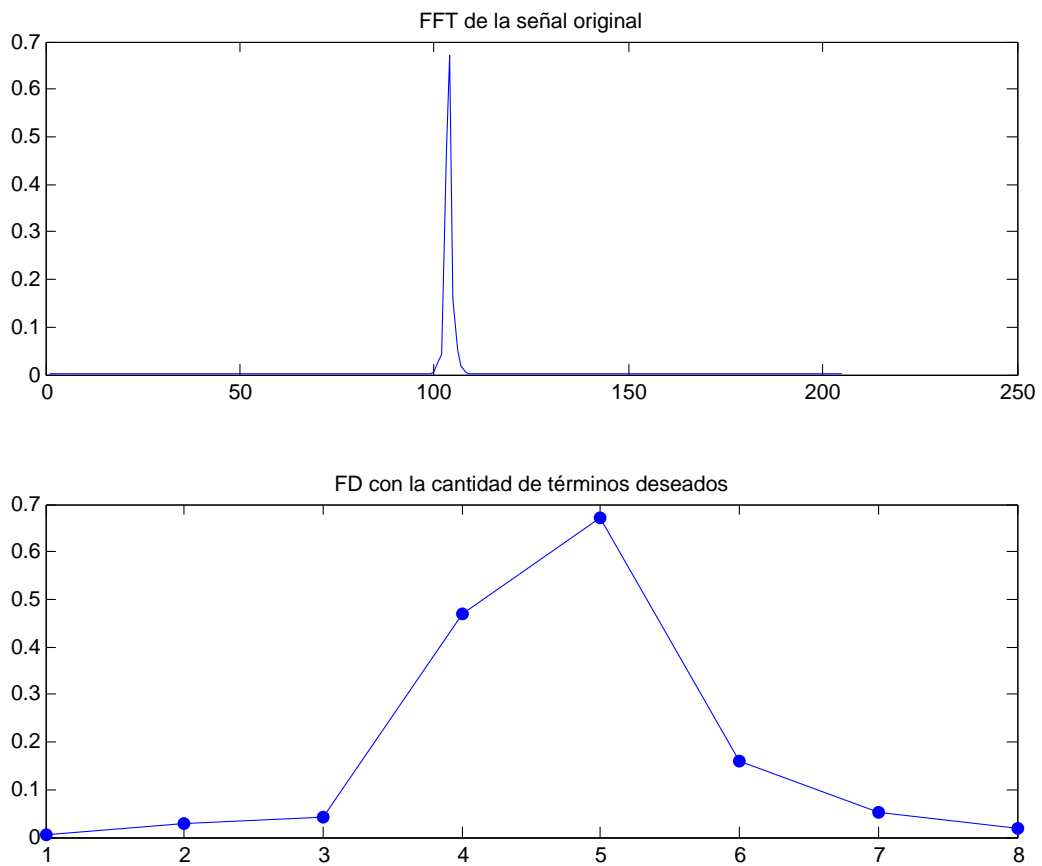


Figura 4.10: Espectro de la señal de carga con filtración en la válvula viajera, obtenida usando la FFT (arriba), FD de la señal de carga con una cantidad de términos fijo (abajo)

De la misma forma se muestra la señal de carga comprimida usando *Wavelet*.

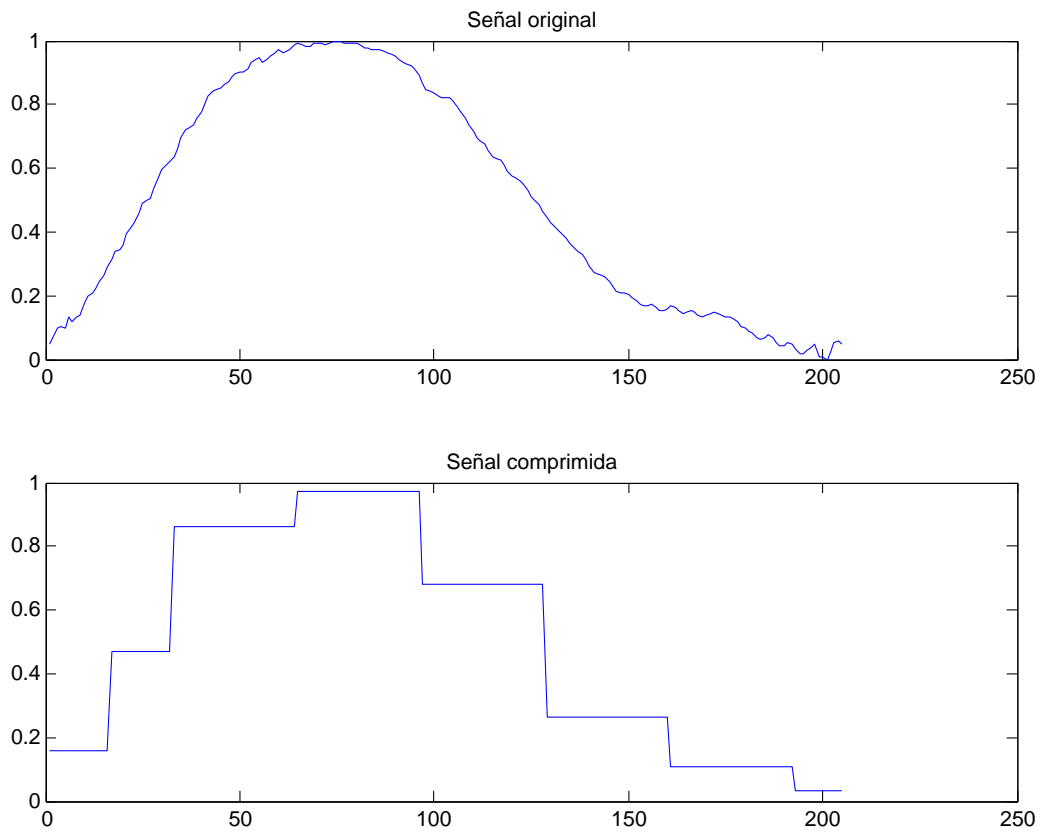


Figura 4.11: Señal de carga original con filtración en la válvula viajera (arriba) y comprimida usando Wavelet (abajo)

Usando la compresión por *Wavelet* se muestra como quedaría la CDF usando dicha compresión.

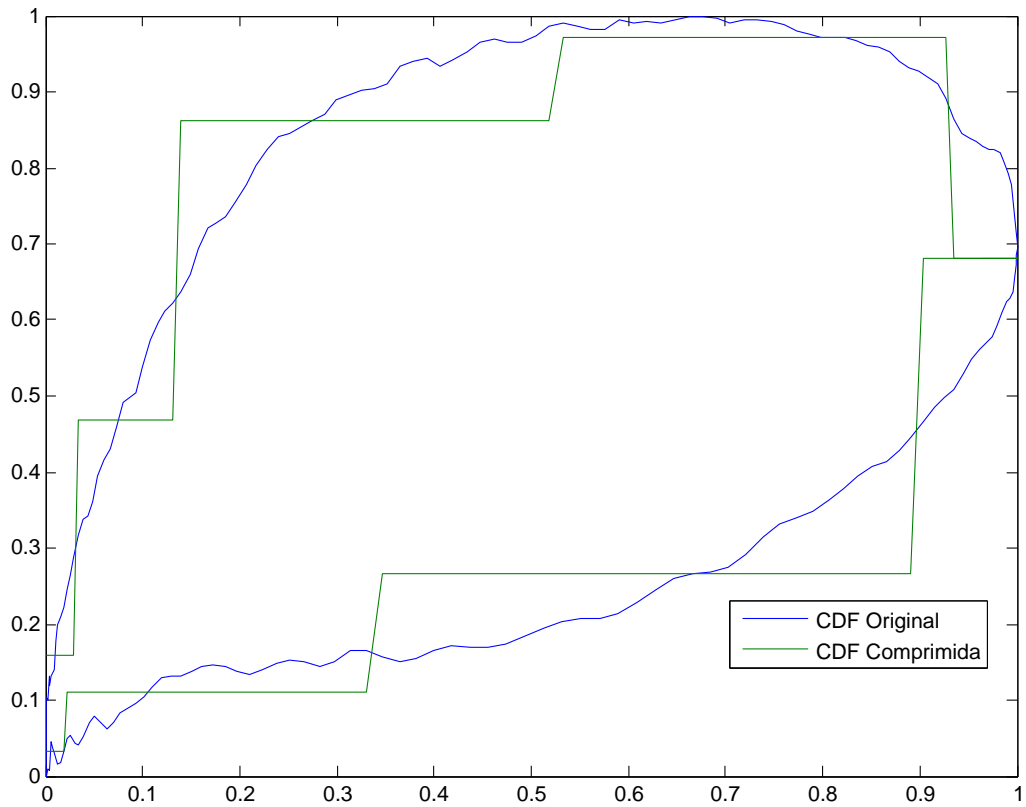


Figura 4.12: Comparación de la señal de carga original con filtración en la válvula viajera y comprimida (usando Wavelet)

También se muestran los centros calculados por FCM y SOM.

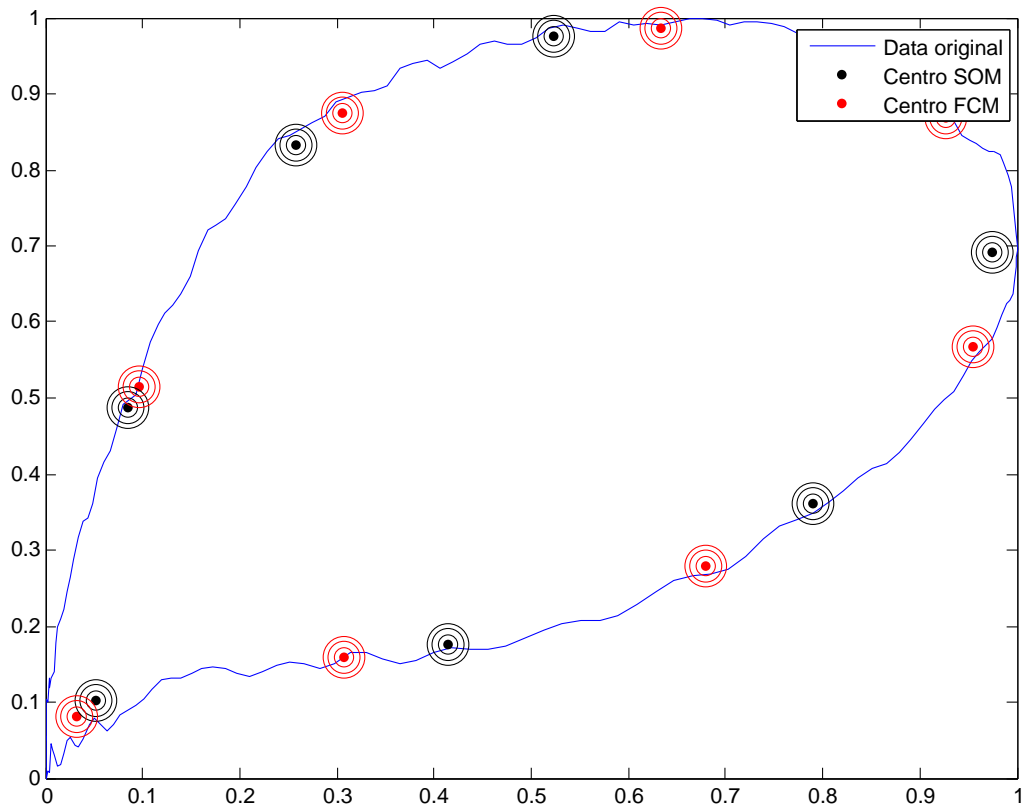
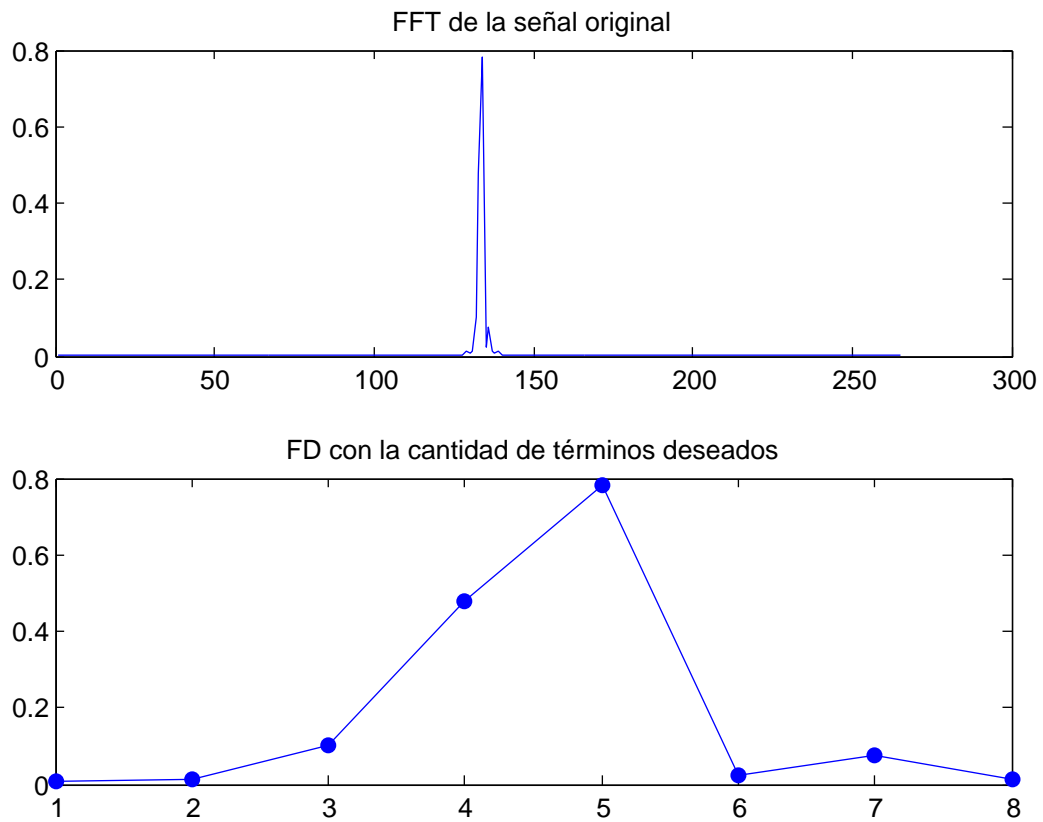


Figura 4.13: Centros determinados a la CDF con filtración en la válvula viajera usando FCM y SOM

#### 4.2.4. Funcionamiento del extractor donde el pozo BMC posee filtración en la válvula fija

La próxima carta se corresponde con un pozo cuyo crudo es de 10 ° API, presenta 80 % de filtración en la válvula fija y 20 % bombeo normal. A continuación se muestra los FD obtenidos.



Espectro de la señal de carga con filtración en la válvula fija, obtenida usando la FFT (arriba), FD de la señal de carga con una cantidad de términos fijo (abajo)

De la misma forma se muestra la señal de carga comprimida usando *Wavelet*.

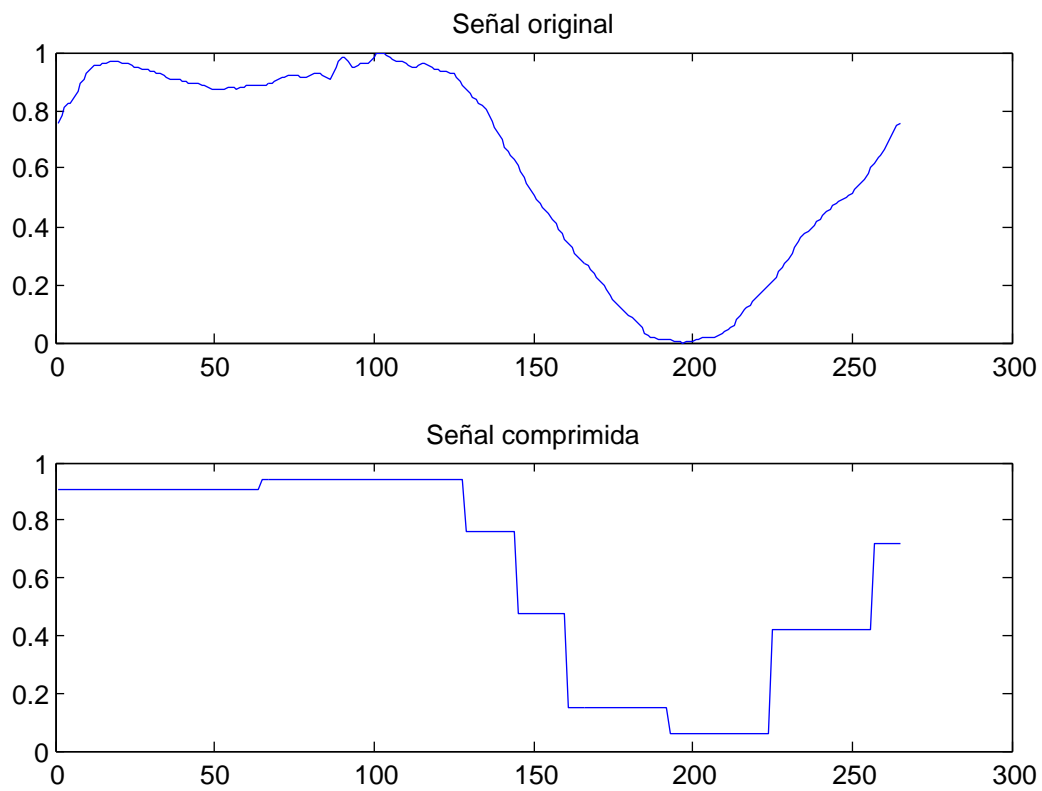


Figura 4.14: Señal de carga original con filtración en la válvula fija (arriba) y comprimida usando Wavelet (abajo)

Usando la compresión por *Wavelet* se muestra como quedaría la CDF usando dicha compresión.

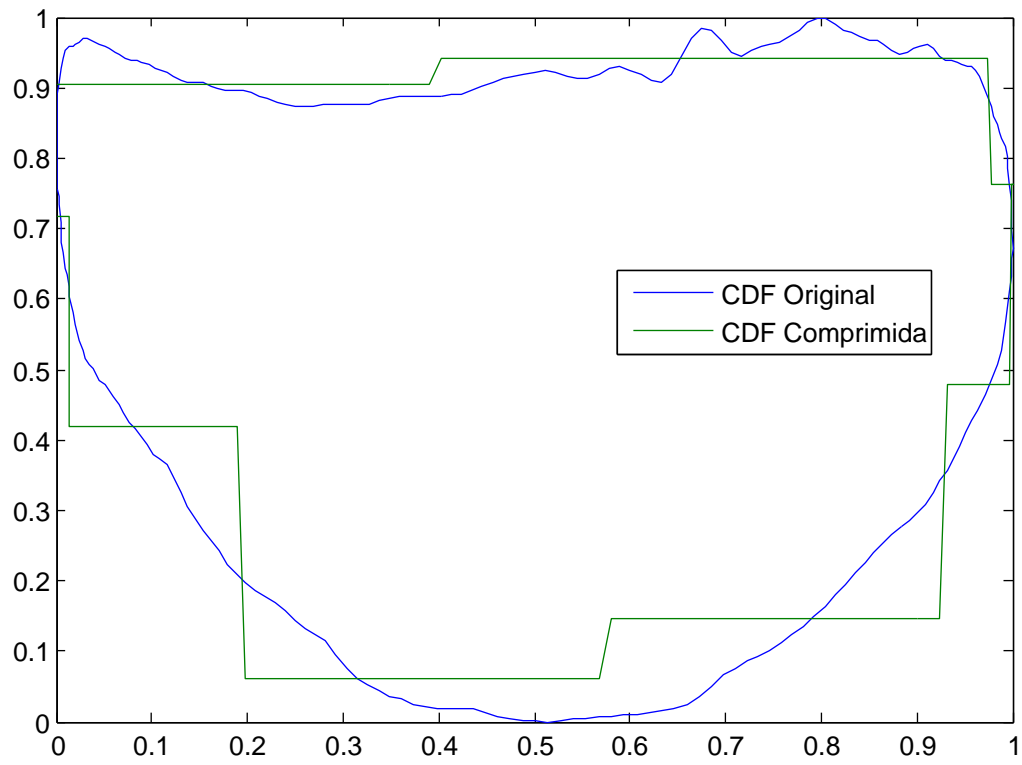


Figura 4.15: Comparación de la señal de carga original con filtración en la válvula fija y comprimida (usando Wavelet)

También se muestran los centros calculados por FCM y SOM.

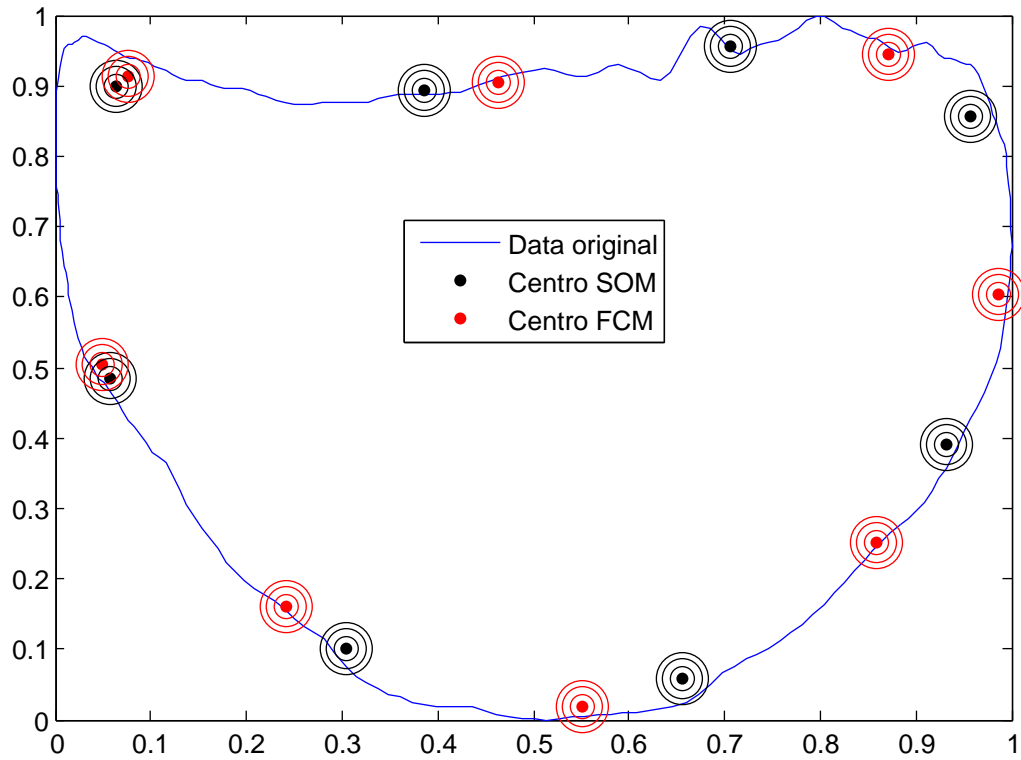
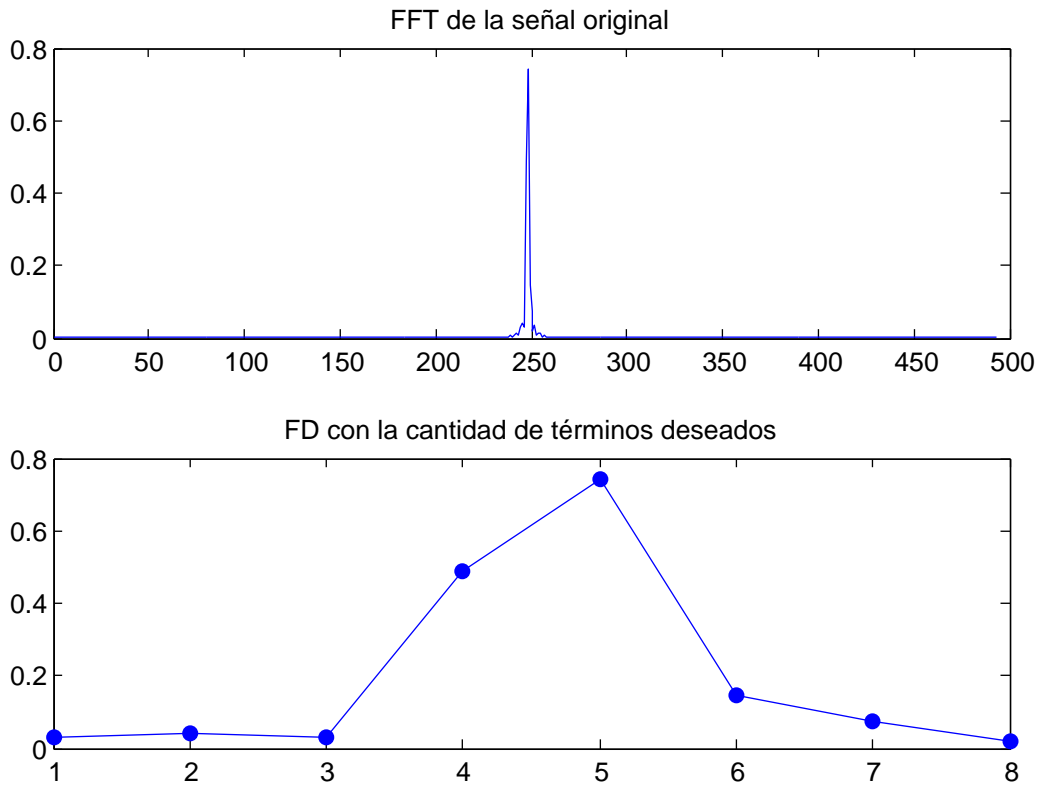


Figura 4.16: Centros determinados a la CDF con filtración en la válvula fija usando FCM y SOM

#### 4.2.5. Funcionamiento del extractor donde el pozo BMC posee interferencia de gas

La próxima carta se corresponde con un pozo cuyo crudo es de 40 ° API. A continuación se muestra los FD obtenidos.



Espectro de la señal de carga con interferencia de gas, obtenida usando la FFT (arriba),  
 FD de la señal de carga con una cantidad de términos fijo (abajo)

De la misma forma se muestra la señal de carga comprimida usando *Wavelet*.

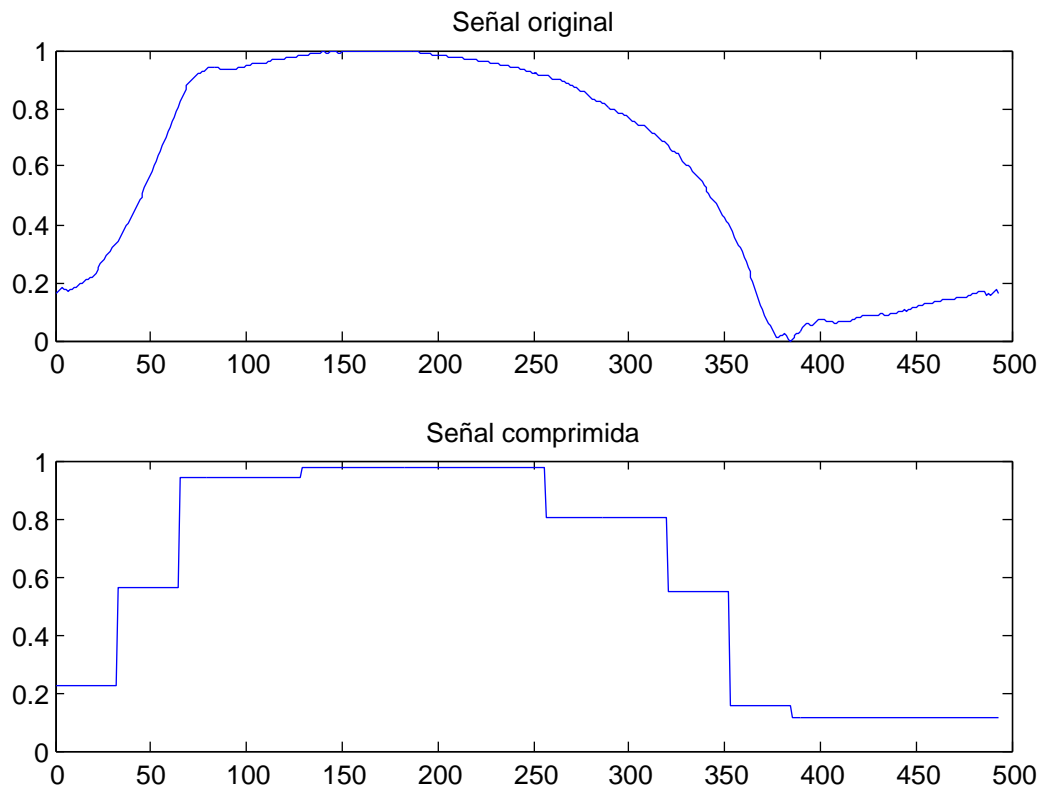


Figura 4.17: Señal de carga original con interferencia de gas (arriba) y comprimida usando Wavelet (abajo)

Usando la compresión por *Wavelet* se muestra como quedaría la CDF usando dicha compresión.

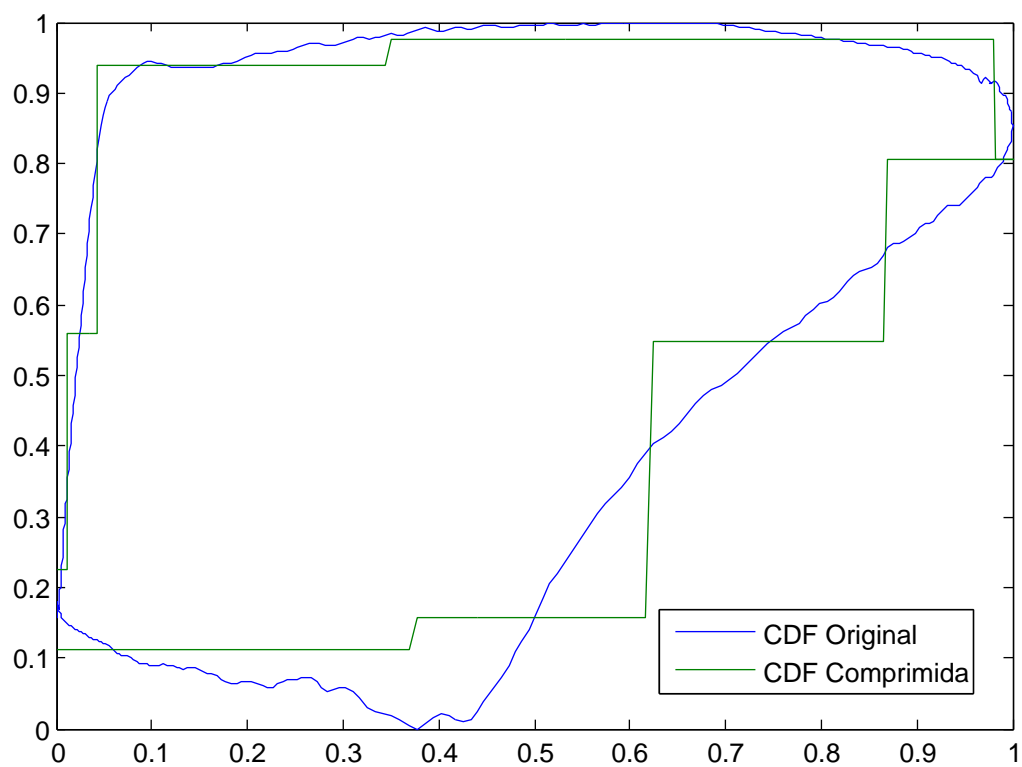


Figura 4.18: Comparación de la señal de carga original con interferencia de gas y comprimida (usando Wavelet)

También se muestran los centros calculados por FCM y SOM.

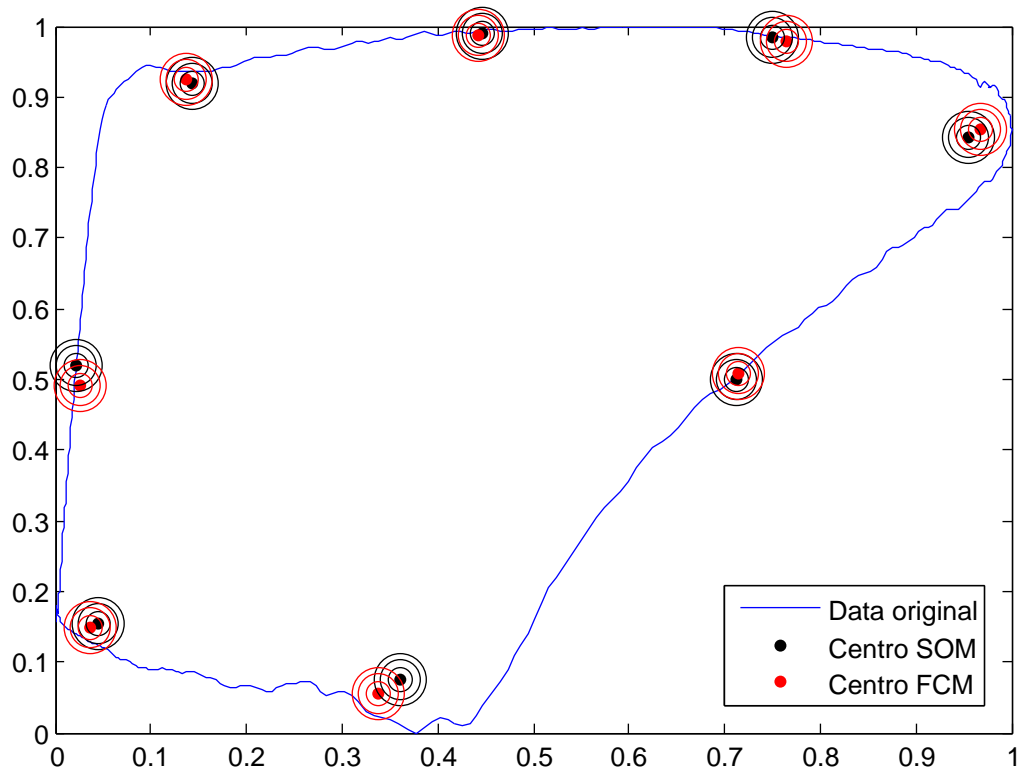


Figura 4.19: Centros determinados a la CDF con interferencia de gas usando FCM y SOM

#### 4.2.6. Funcionamiento del extractor donde el pozo BMC posee golpe de fluido

Esta carta se corresponde con un pozo cuyo crudo es de 9 ° API, presenta 90% de golpe de fluido y 10% de falla en la caja de transmisión. A continuación se muestra los FD obtenidos para este caso.

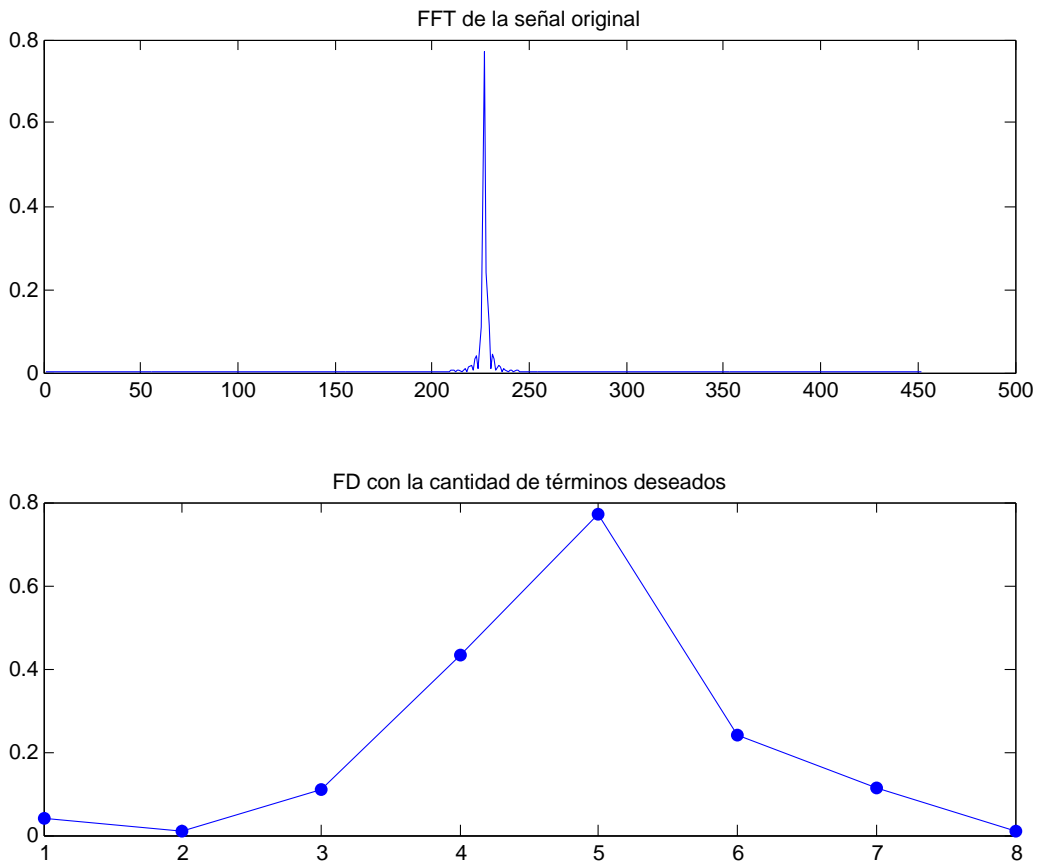


Figura 4.20: Espectro de la señal de carga con golpe de fluido, obtenida usando la FFT (arriba), FD de la señal de carga con una cantidad de términos fijo (abajo)

De la misma forma se muestra la señal de carga comprimida usando *Wavelet*.

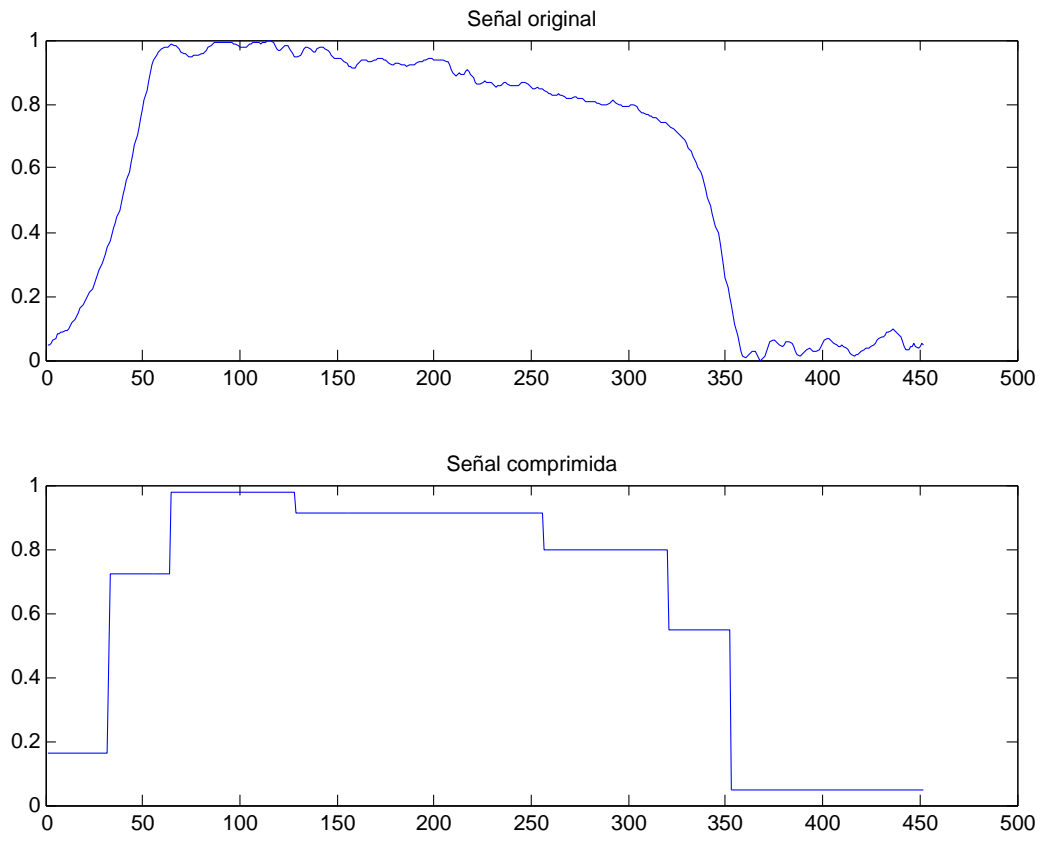


Figura 4.21: Señal de carga original con golpe de fluido (arriba) y comprimida usando Wavelet (abajo)

Usando la compresión por *Wavelet* se muestra como quedaría la CDF usando dicha compresión.

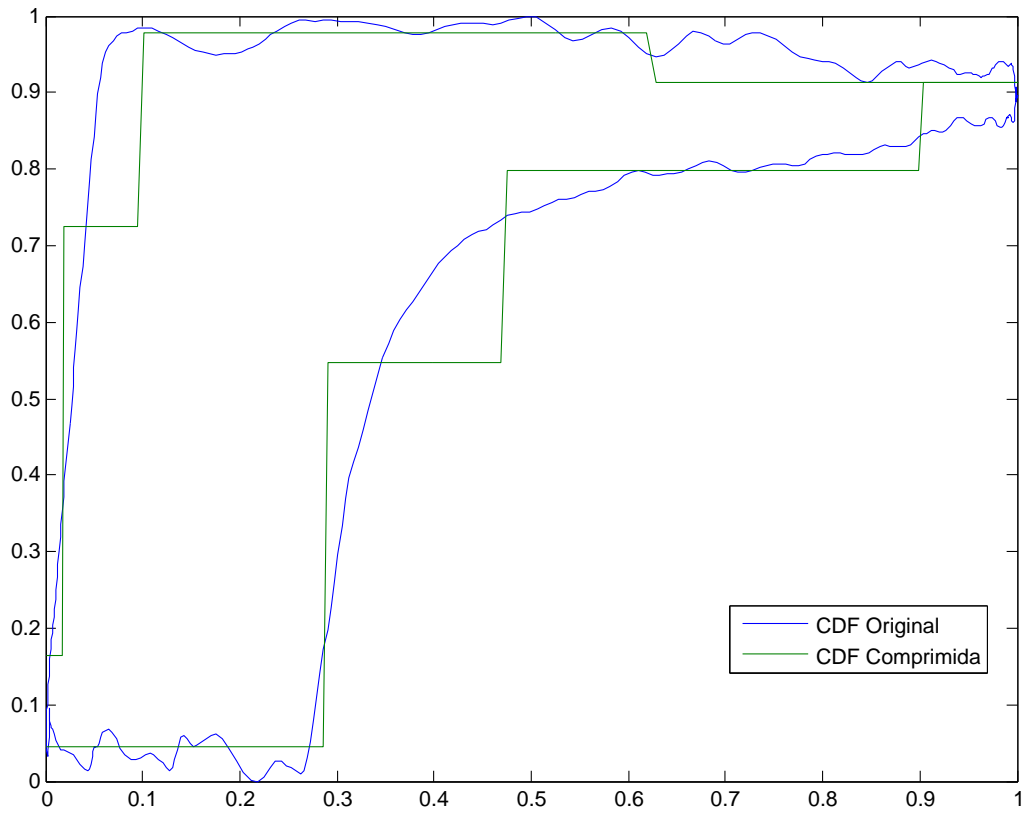


Figura 4.22: Comparación de la señal de carga original con golpe de fluido y comprimida (usando Wavelet)

También se muestran los centros calculados por FCM y SOM.

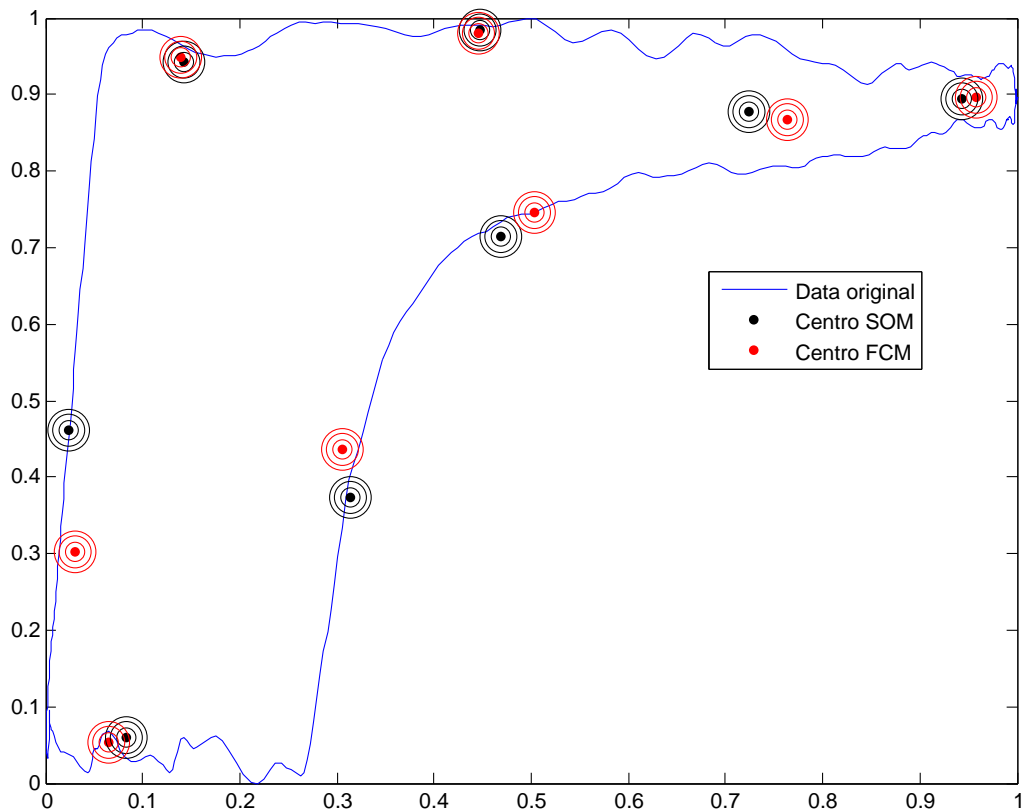
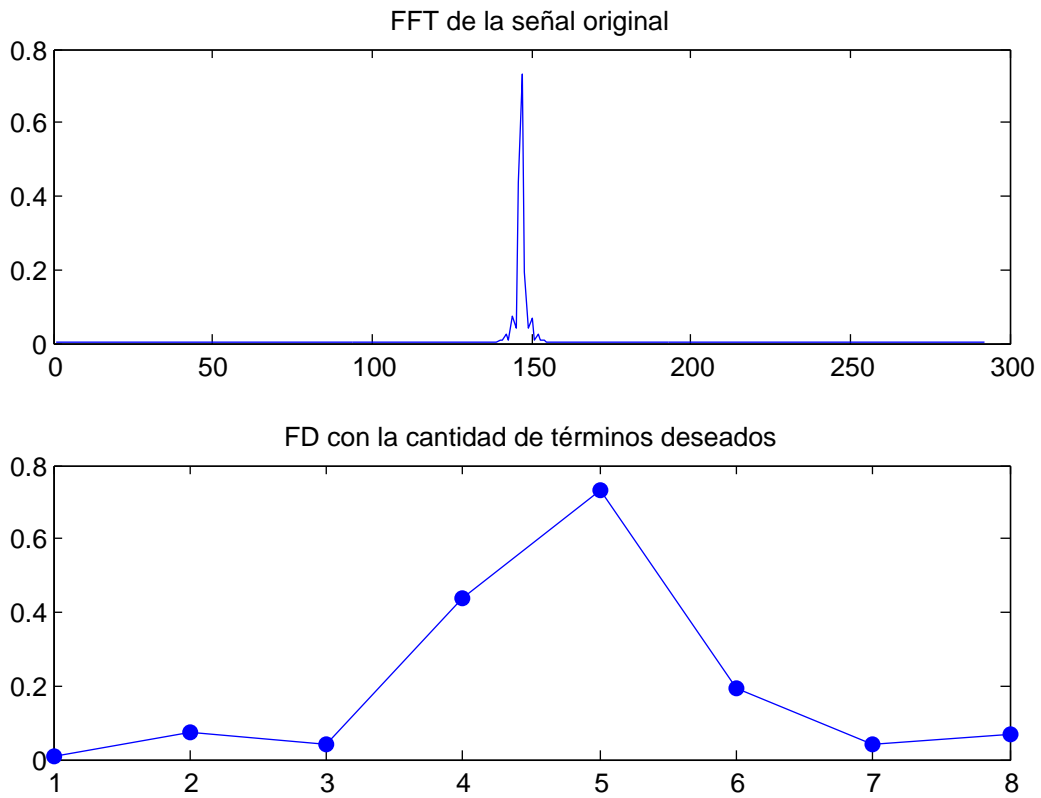


Figura 4.23: Centros determinados a la CDF con golpe de fluido usando FCM y SOM

#### 4.2.7. Funcionamiento del extractor donde el pozo BMC posee golpe de bomba - pistón arriba

La próxima carta se corresponde con un pozo cuyo crudo es de 9 ° API, presenta 30 % de golpe de bomba - pistón arriba, 40 % de bomba normal y 30 % de golpe de bomba - pistón abajo. A continuación se muestra los FD obtenidos.



Espectro de la señal de carga con golpe de bomba - pistón arriba, obtenida usando la FFT (arriba), FD de la señal de carga con una cantidad de términos fijo (abajo)

De la misma forma se muestra la señal de carga comprimida usando *Wavelet*.

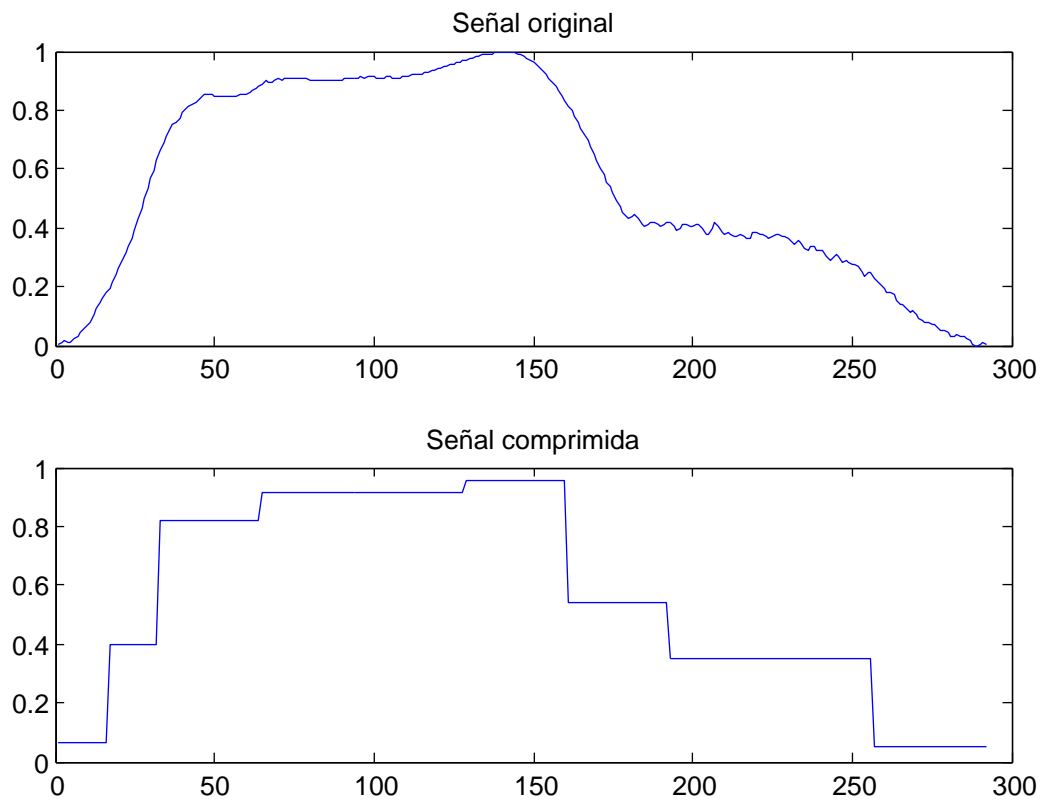


Figura 4.24: Señal de carga original con golpe de bomba - pistón arriba (arriba) y comprimida usando Wavelet (abajo)

Usando la compresión por *Wavelet* se muestra como quedaría la CDF usando dicha compresión.

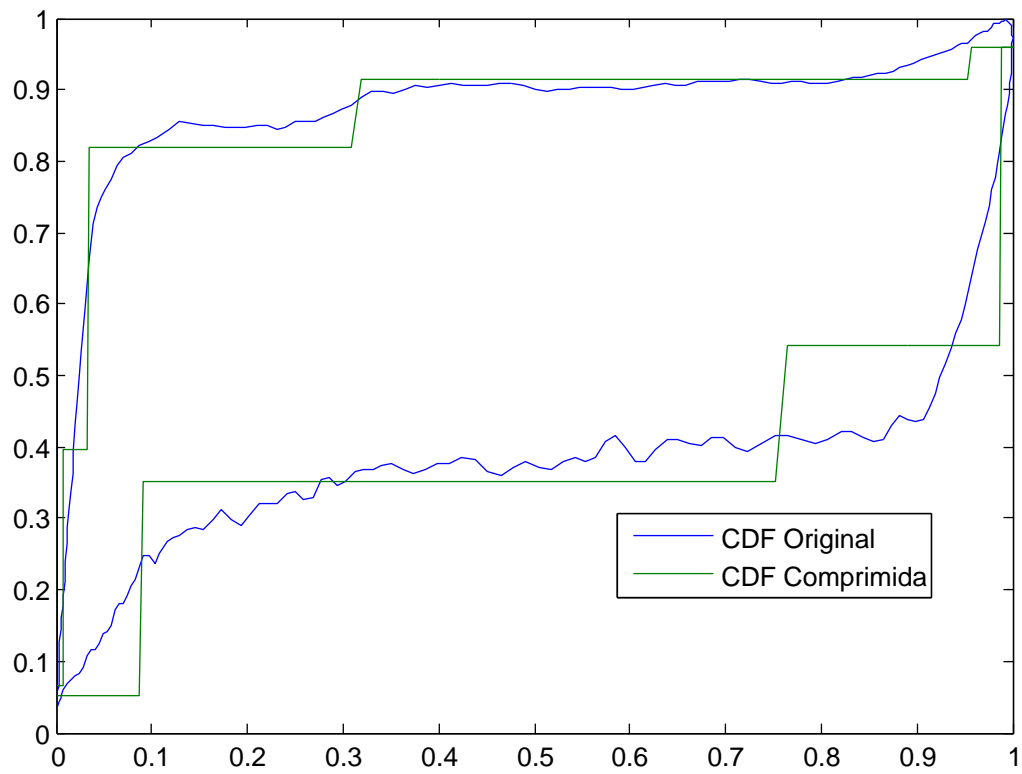


Figura 4.25: Comparación de la señal de carga original con golpe de bomba - pistón arriba y comprimida (usando Wavelet)

También se muestran los centros calculados por FCM y SOM.

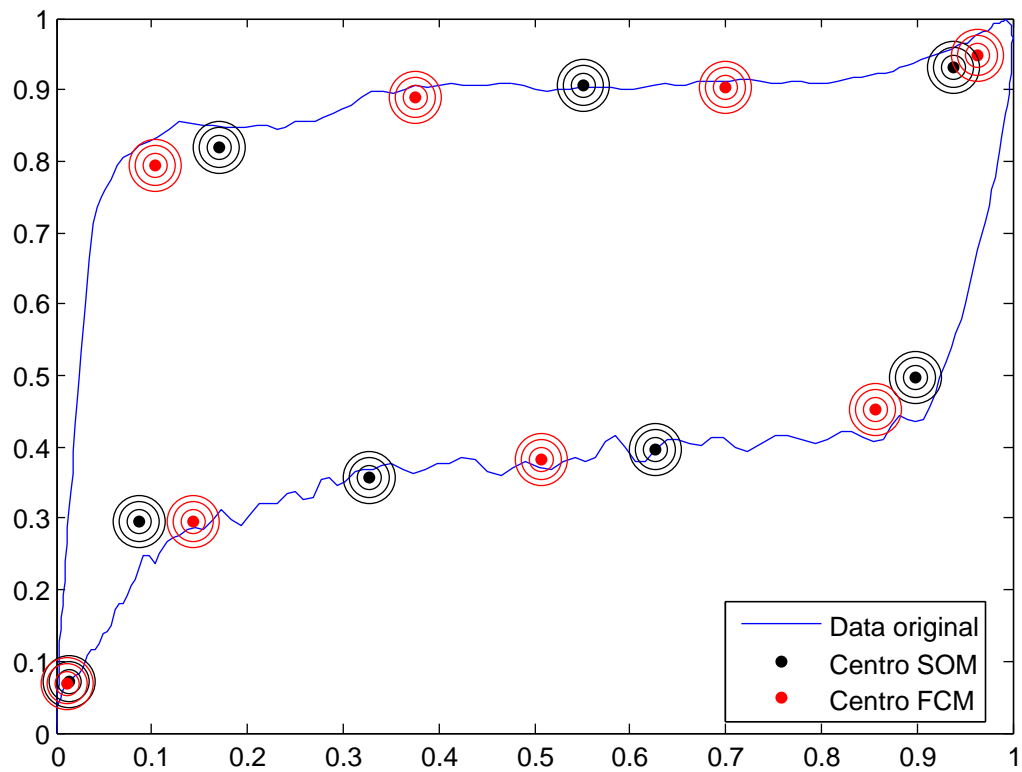
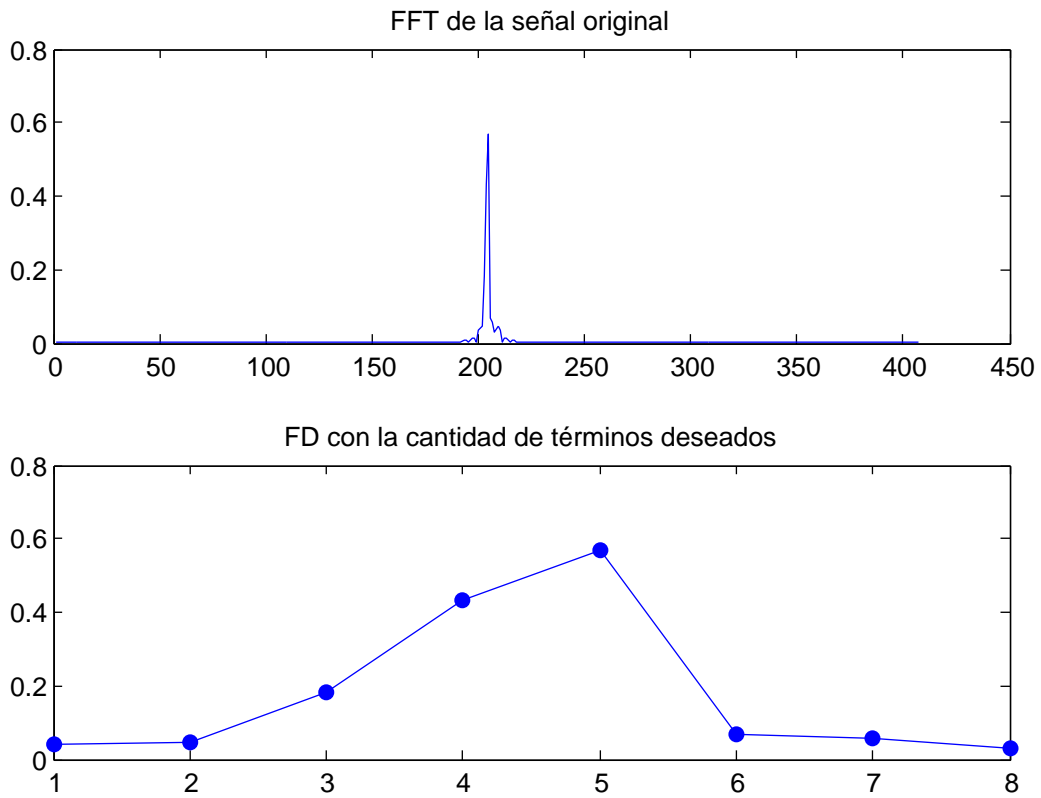


Figura 4.26: Centros determinados a la CDF con golpe de bomba - pistón arriba usando FCM y SOM

#### 4.2.8. Funcionamiento del extractor donde el pozo BMC posee golpe de bomba - pistón abajo

La próxima carta se corresponde con un pozo cuyo crudo es de 10 ° API, presenta 45 % de golpe de bomba - pistón abajo y 55 % de bomba normal. A continuación se muestra los FD obtenidos.



Espectro de la señal de carga con golpe de bomba - pistón abajo, obtenida usando la FFT (arriba), FD de la señal de carga con una cantidad de términos fijo (abajo)

De la misma forma se muestra la señal de carga comprimida usando *Wavelet*.

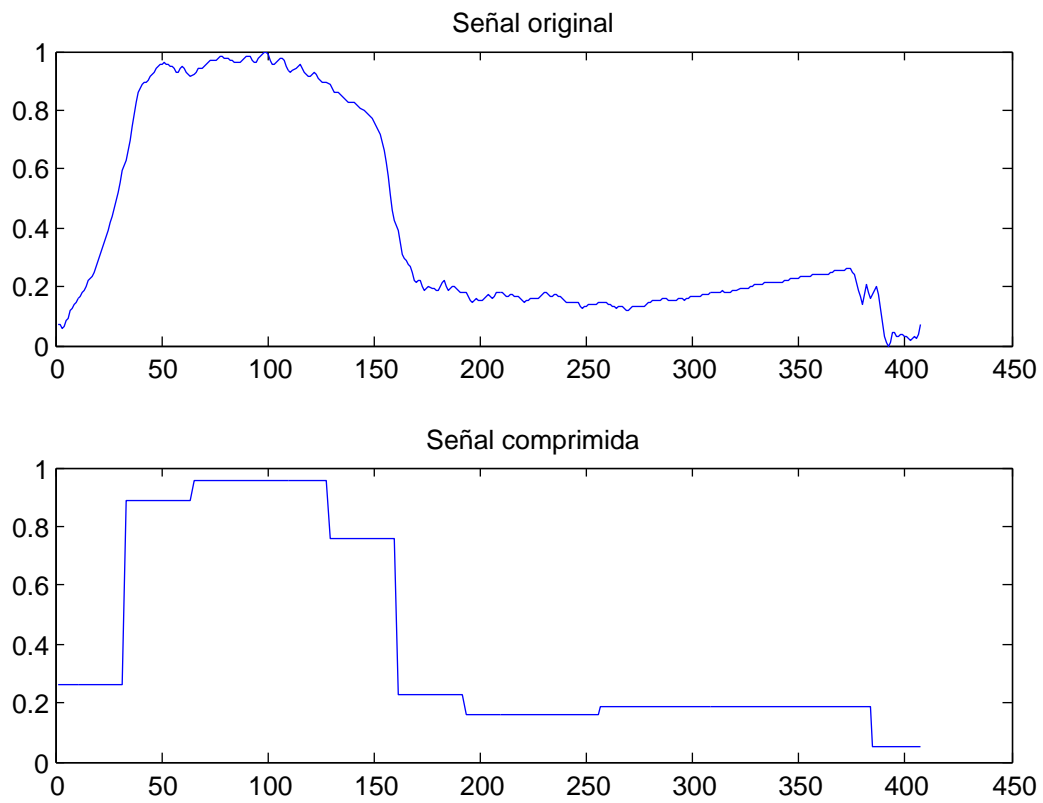


Figura 4.27: Señal de carga original con golpe de bomba - pistón abajo (arriba) y comprimida usando Wavelet (abajo)

Usando la compresión por *Wavelet* se muestra como quedaría la CDF usando dicha compresión.

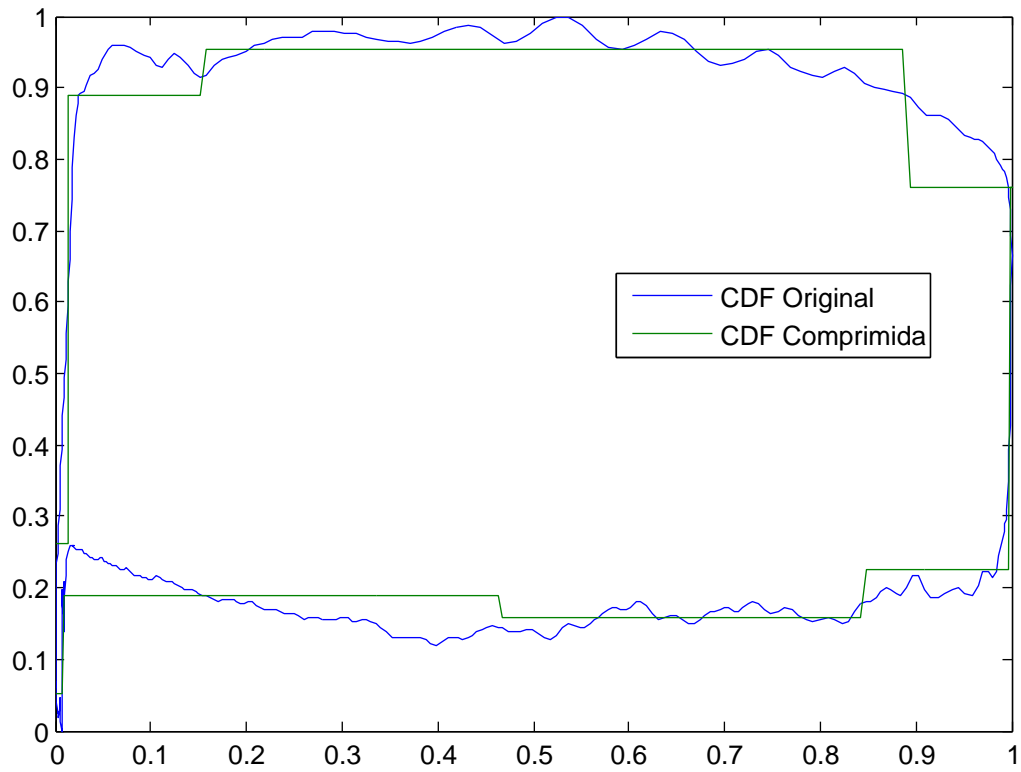


Figura 4.28: Comparación de la señal de carga original con golpe de bomba - pistón abajo y comprimida (usando Wavelet)

También se muestran los centros calculados por FCM y SOM.

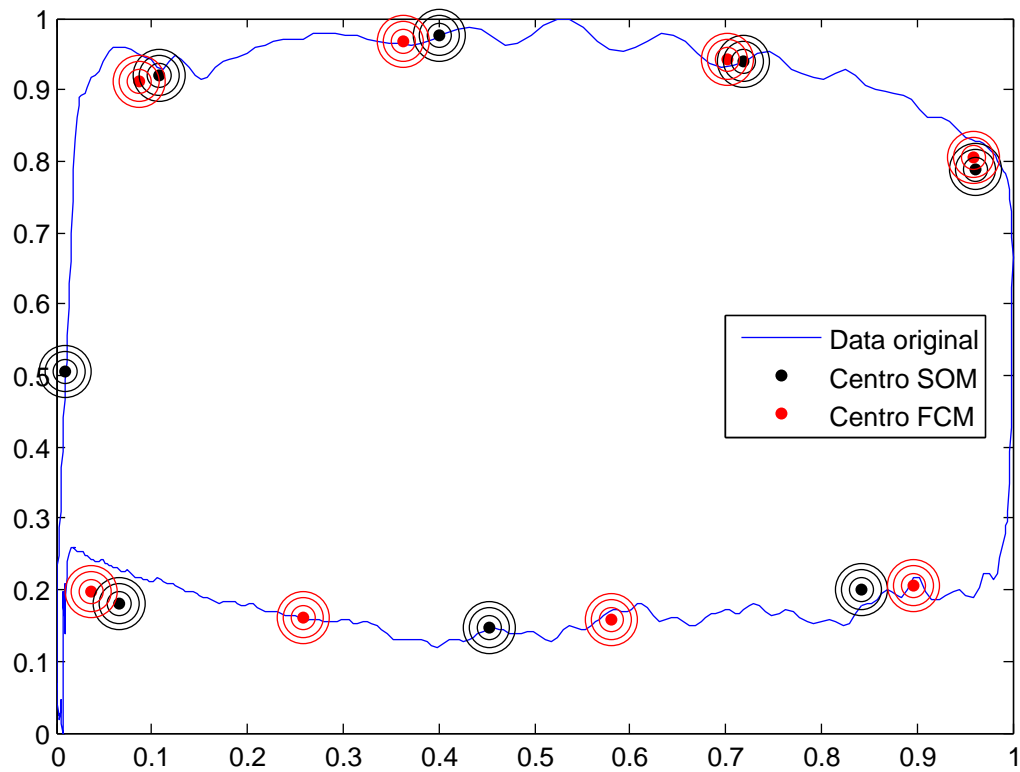
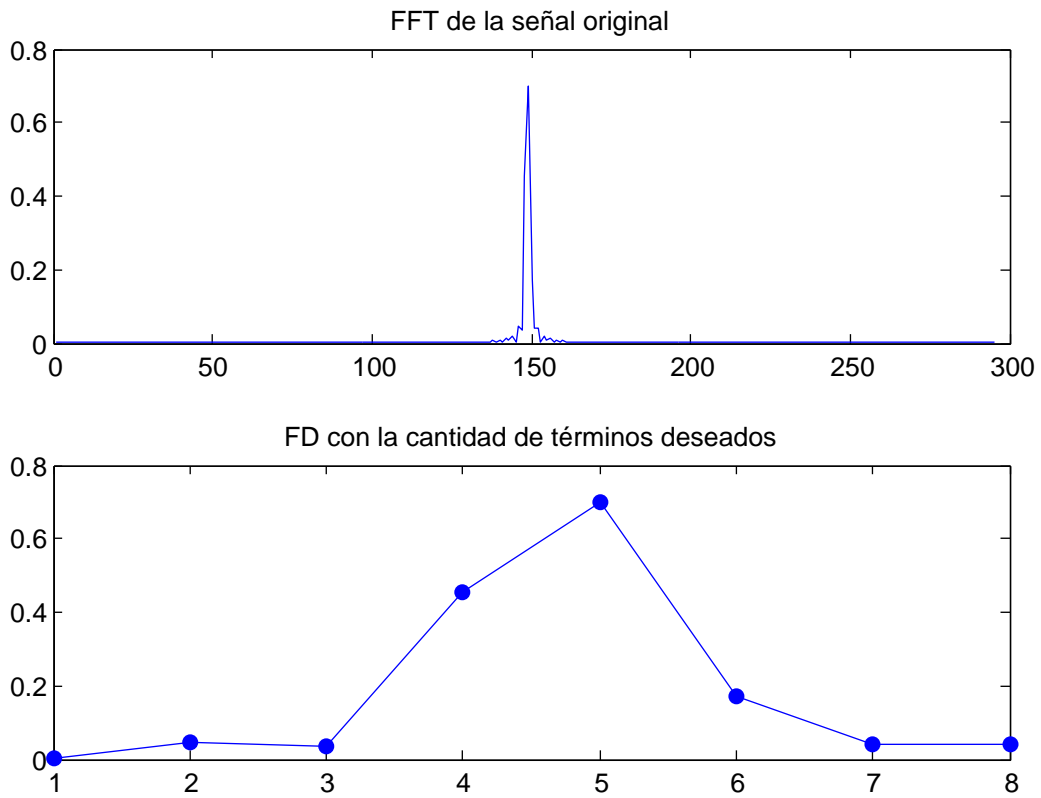


Figura 4.29: Centros determinados a la CDF con golpe de bomba - pistón abajo usando FCM y SOM

#### 4.2.9. Funcionamiento del extractor donde el pozo BMC posee falla de la caja de transmisión

La próxima carta se corresponde con un pozo cuyo crudo es de 11 ° API, presenta 70 % de falla de la caja de transmisión y 30 % de bomba normal. A continuación se muestra los FD obtenidos.



Espectro de la señal de carga con falla de la caja de transmisión, obtenida usando la FFT (arriba), FD de la señal de carga con una cantidad de términos fijo (abajo)

De la misma forma se muestra la señal de carga comprimida usando *Wavelet*.

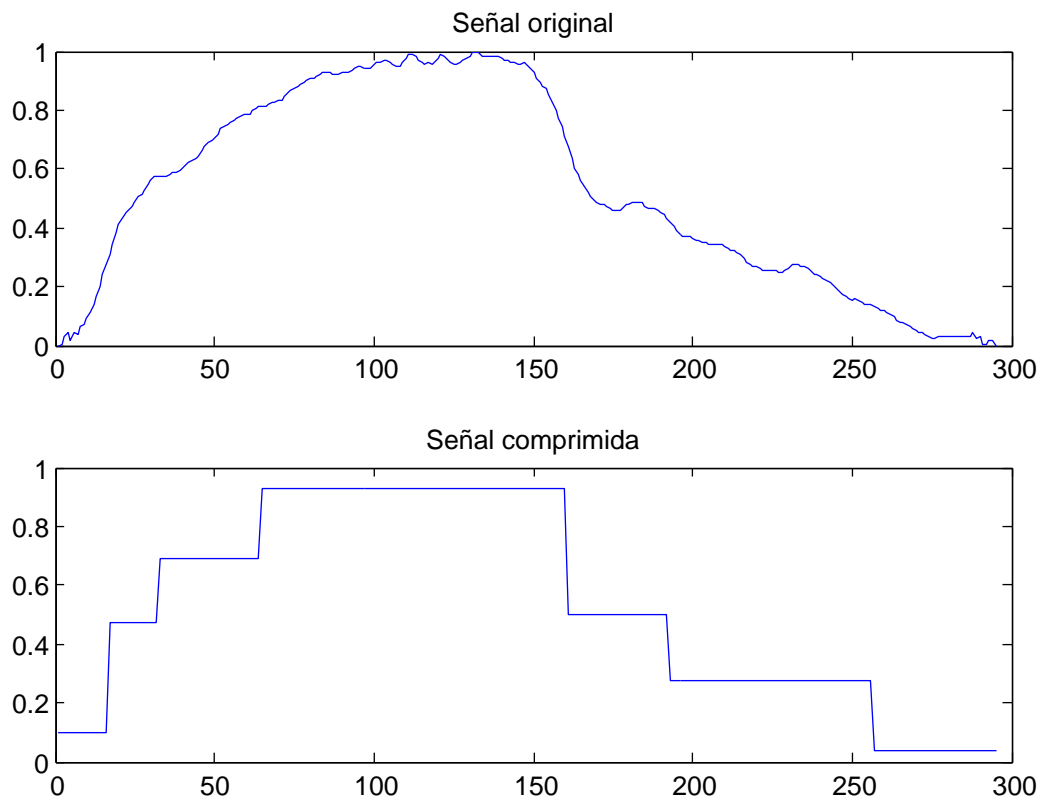


Figura 4.30: Señal de carga original con falla de la caja de transmisión y comprimida usando Wavelet (abajo)

Usando la compresión por *Wavelet* se muestra como quedaría la CDF usando dicha compresión.

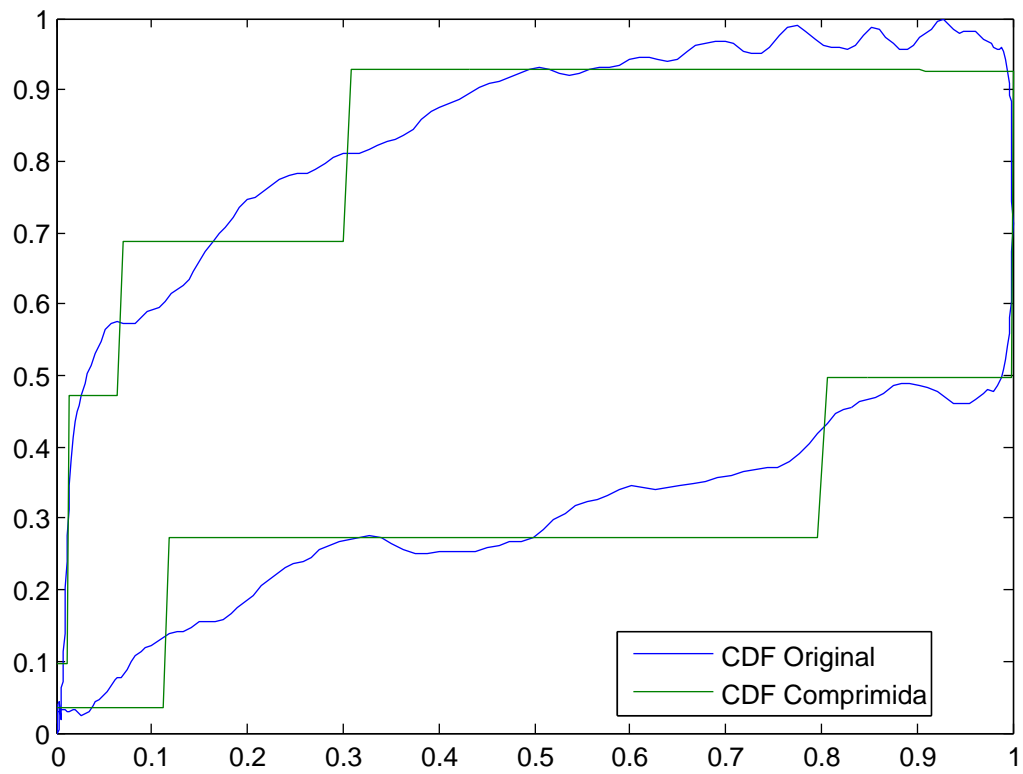


Figura 4.31: Comparación de la señal de carga original con falla de la caja de transmisión y comprimida (usando Wavelet)

También se muestran los centros calculados por FCM y SOM.

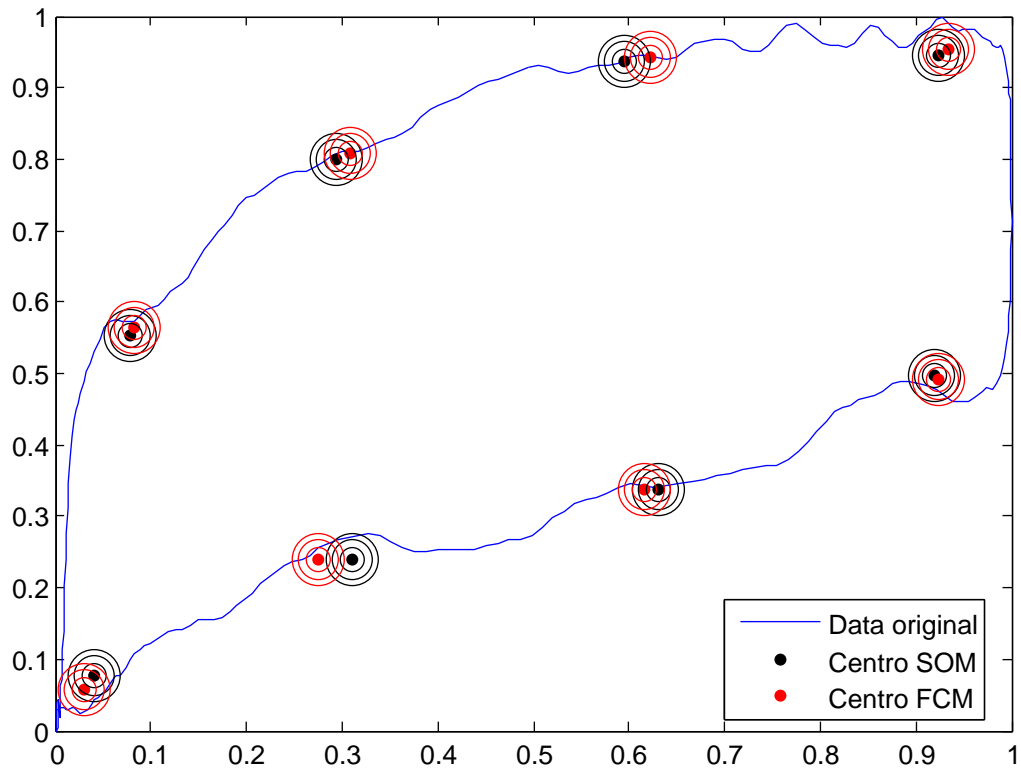
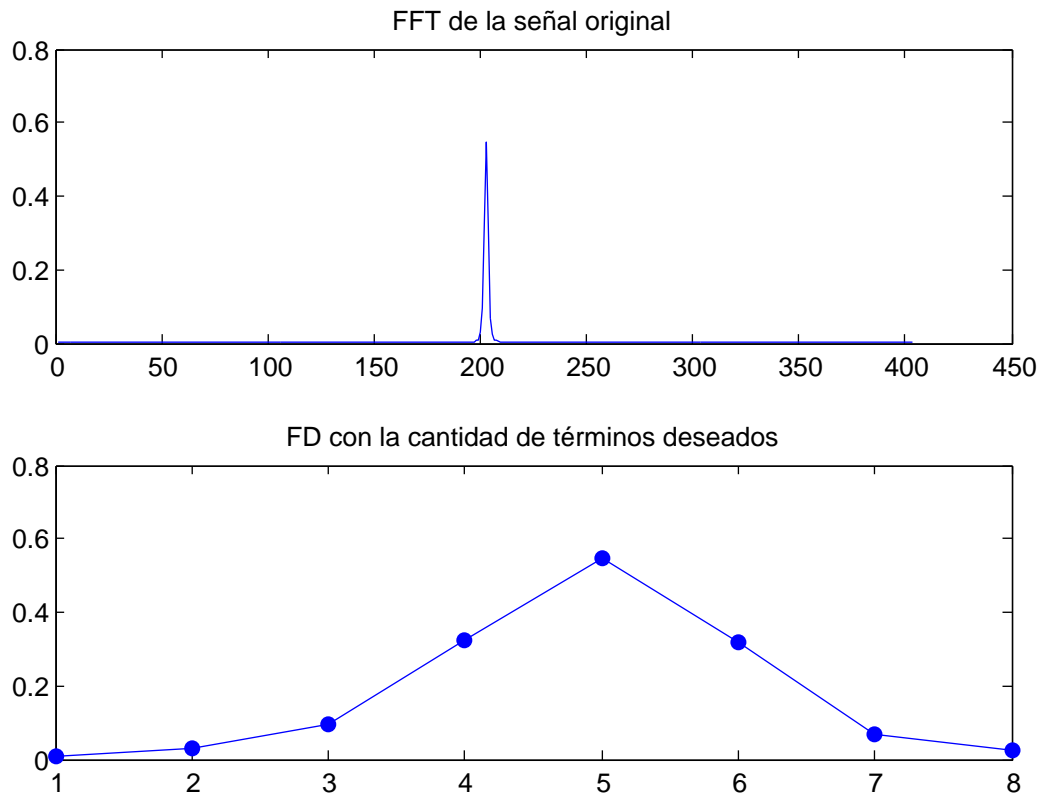


Figura 4.32: Centros determinados a la CDF con falla de la caja de transmisión usando FCM y SOM

#### 4.2.10. Funcionamiento del extractor donde el pozo BMC posee cabillas partidas

La próxima carta se corresponde con un pozo cuyo crudo es de 10 ° API, presenta 60 % de cabillas partidas y 40 % en otro estado no determinado en el presente PFM. A continuación se muestra los FD obtenidos.



Espectro de la señal de carga con cabillas partidas, obtenida usando la FFT (arriba), FD de la señal de carga con una cantidad de términos fijo (abajo)

De la misma forma se muestra la señal de carga comprimida usando *Wavelet*.

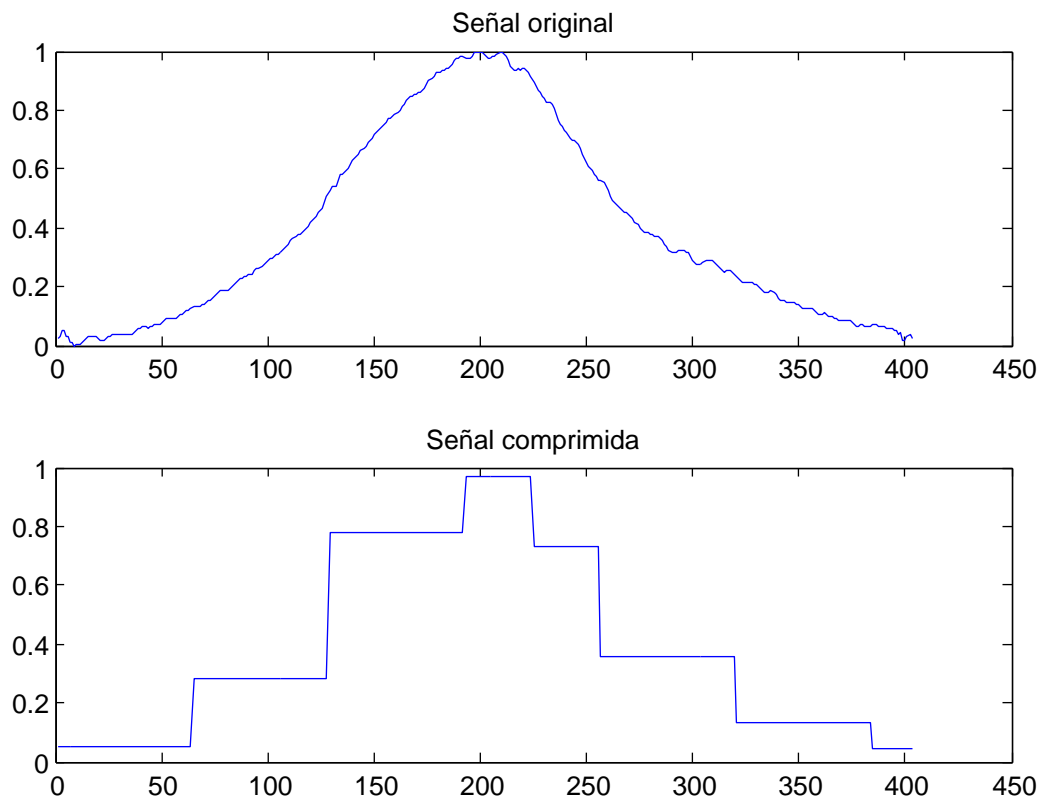


Figura 4.33: Señal de carga original con cabillas partidas y comprimida usando Wavelet (abajo)

Usando la compresión por *Wavelet* se muestra como quedaría la CDF usando dicha compresión.

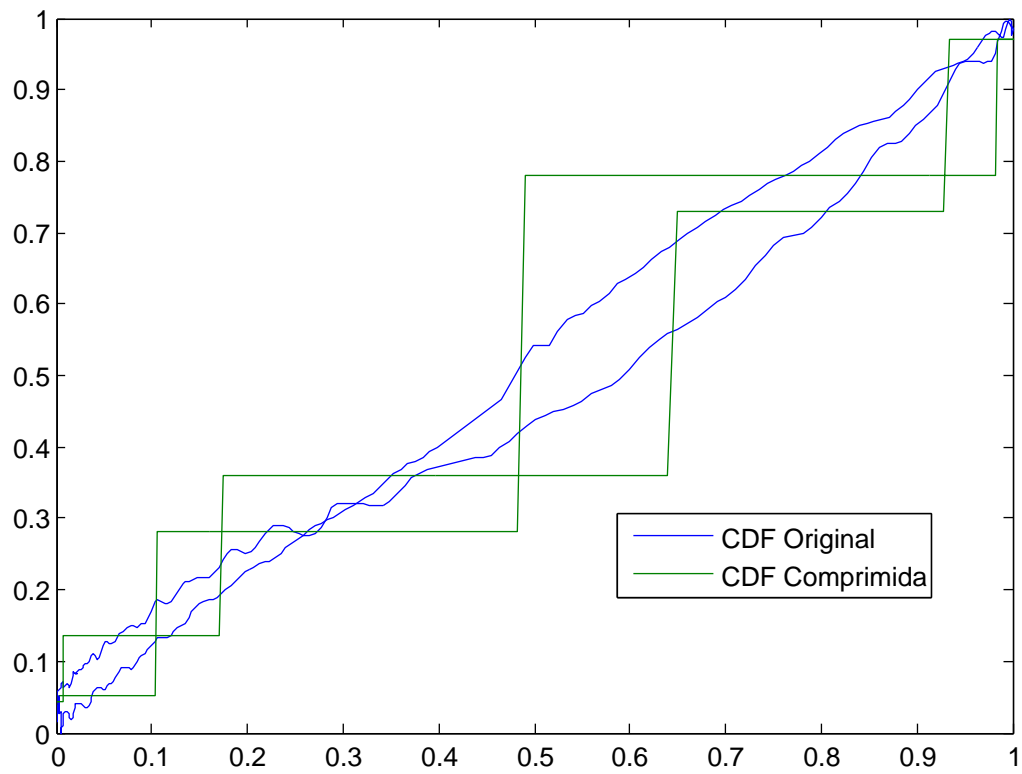


Figura 4.34: Comparación de la señal de carga original con cabillas partidas y comprimida (usando Wavelet)

También se muestran los centros calculados por FCM y SOM.

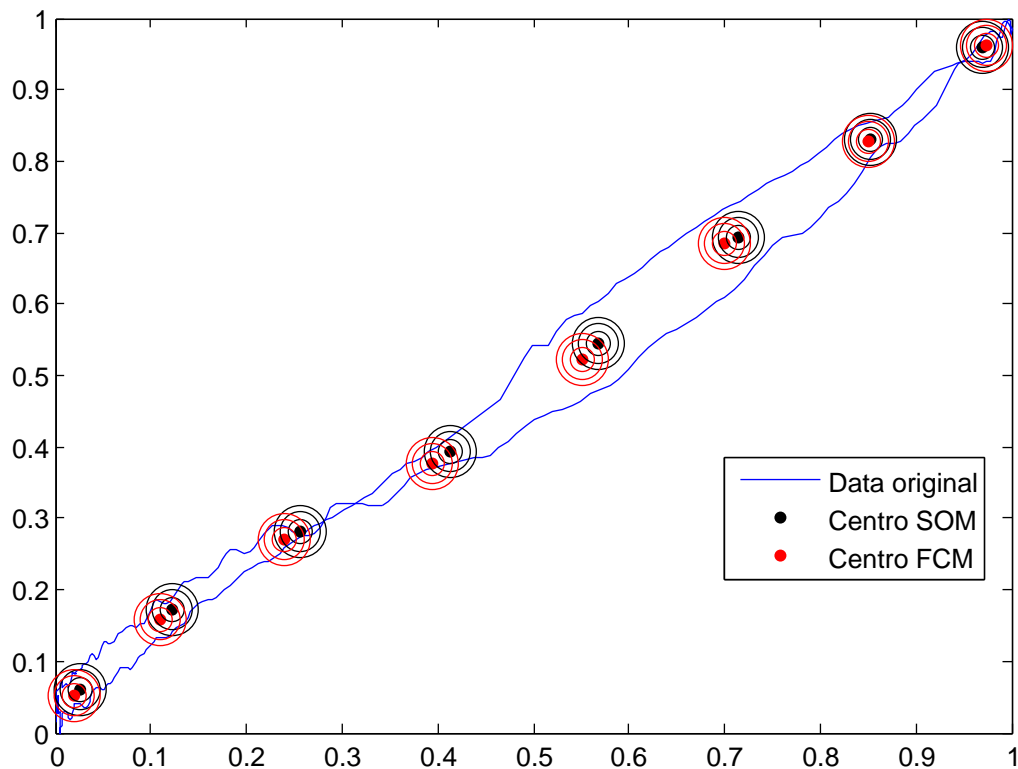
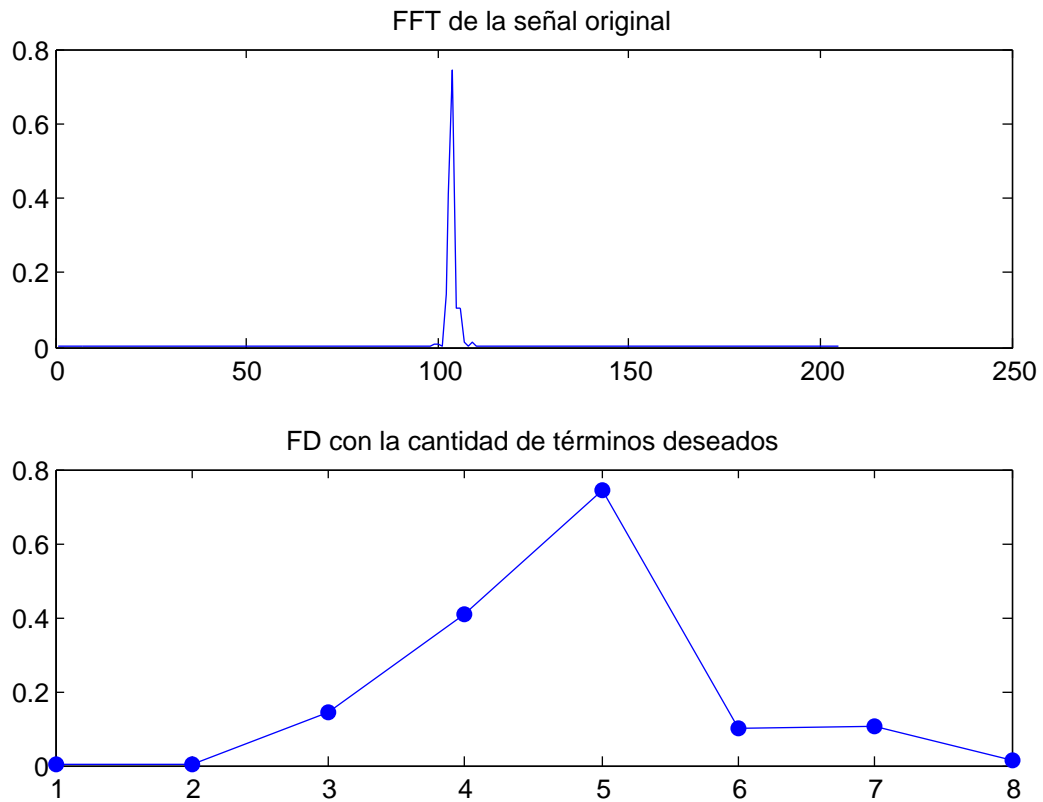


Figura 4.35: Centros determinados a la CDF con cabillas partidas usando FCM y SOM

#### 4.2.11. Funcionamiento del extractor donde la unidad BMC funciona en un estado no determinado

La próxima carta se corresponde con un pozo cuyo crudo es de 40 ° API, presenta 100 % de un estado determinado en el presente PFM. A continuación se muestra los FD obtenidos.



Espectro de la señal de carga para un estado no determinado, obtenida usando la FFT (arriba), FD de la señal de carga con una cantidad de términos fijo (abajo)

De la misma forma se muestra la señal de carga comprimida usando *Wavelet*.

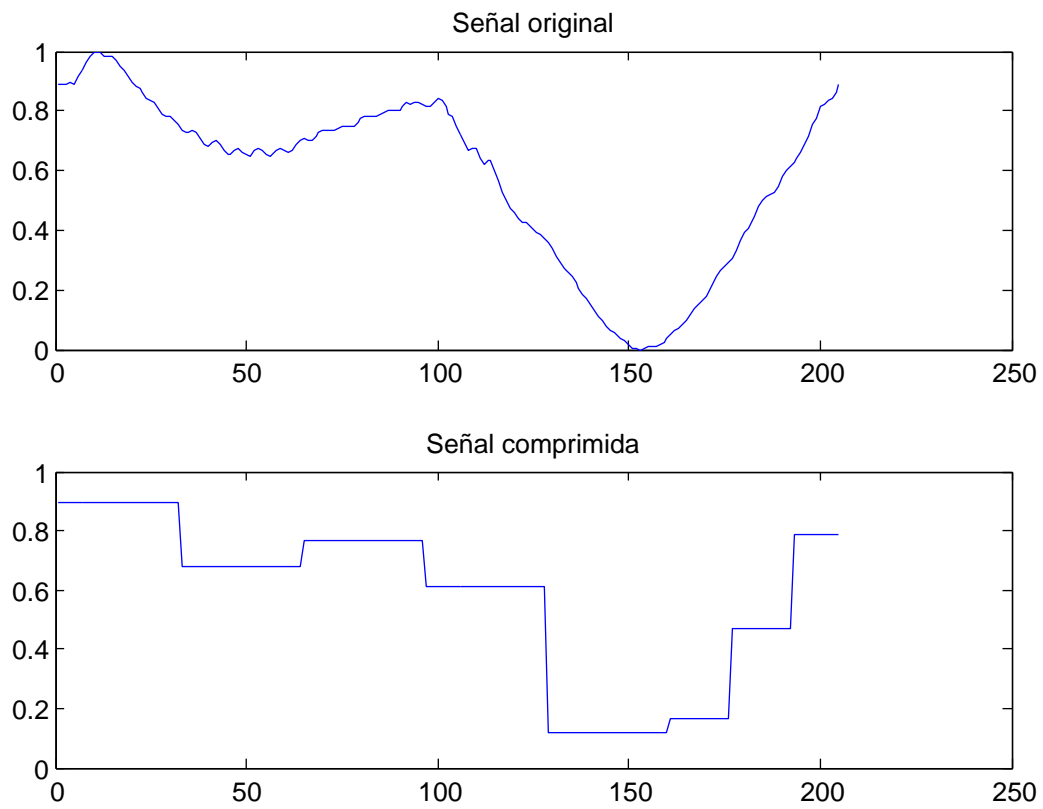


Figura 4.36: Señal de carga original para un estado no determinado y comprimida usando Wavelet (abajo)

Usando la compresión por *Wavelet* se muestra como quedaría la CDF usando dicha compresión.

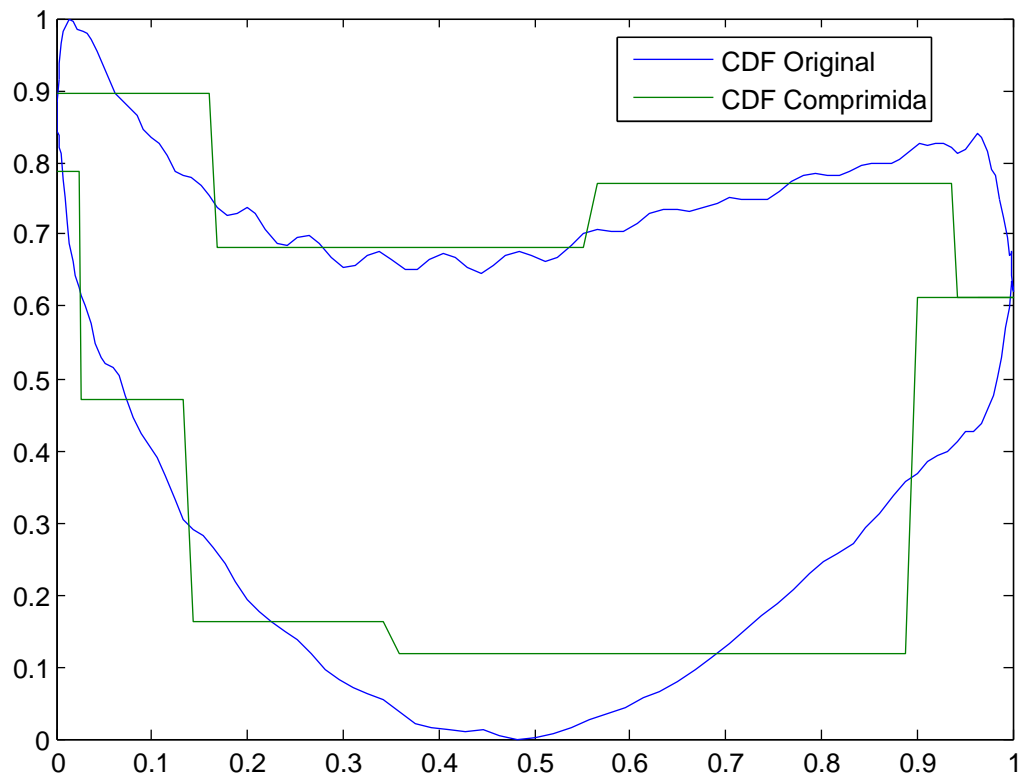


Figura 4.37: Comparación de la señal de carga original para un estado no determinado y comprimida (usando Wavelet)

También se muestran los centros calculados por FCM y SOM.

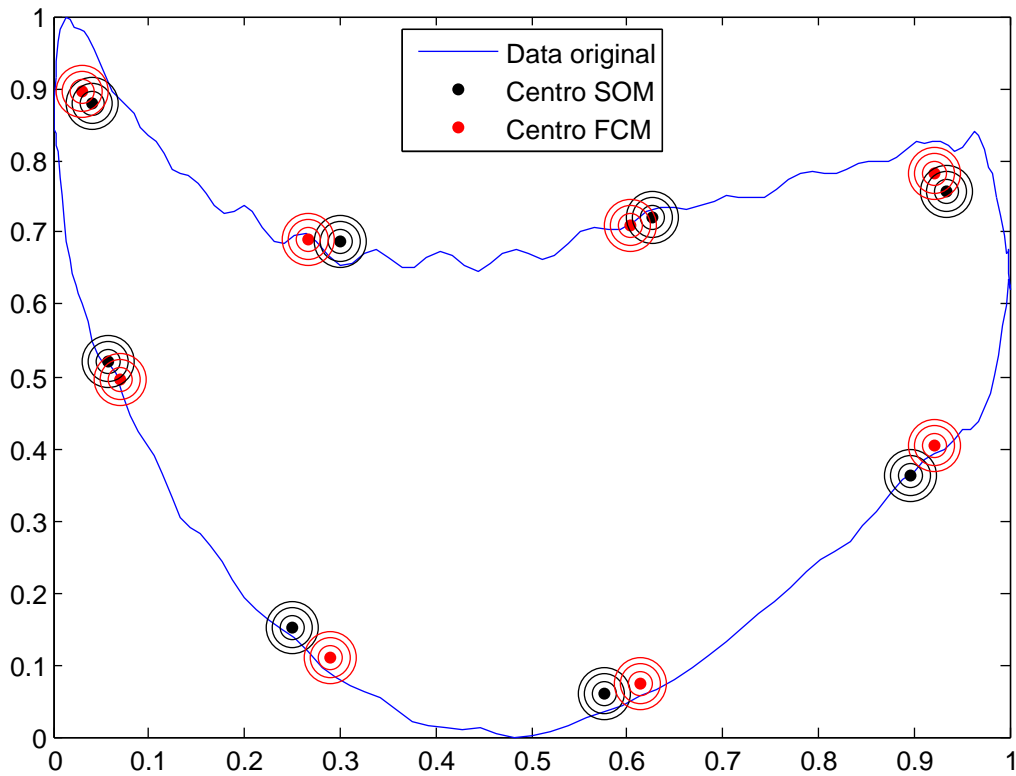


Figura 4.38: Centros determinados a la CDF para un estado no determinado usando FCM y SOM

### 4.3. Desarrollo del regresor SVM

Para el desarrollo del regresor se varió la entrada (extractor de características) y de salida se tiene un valor representativo entre 0 y 1, donde 0 es que no existe pertenencia de ese estado en las 453 CDF utilizadas en la generación de cada estado y 1 que existe una pertenencia total de ese estado en las 453 CDF utilizadas en el entrenamiento y validación de cada estado. De la misma forma se utilizó la modalidad de  $\epsilon$ -SVR, descrito en [18]. En cada estado se realizó lo siguiente:

- Se calculó  $C$  en función de la ecuación 9.61 de [60].
- Se varió  $\xi$  (ecuaciones 3.27, 3.28 y 3.29) en el rango  $[0, 20]$ , con paso aritmético de 1.
- Se utilizó un Kernel tipo RBF (ecuación 3.34), según las recomendaciones sugeridas por [64, 66].
- Se varió  $\sigma$  (ecuación 3.34) en el rango  $[2^{-20}, 2^{20}]$ , con paso geométrico de 2.

Se realiza una validación cruzada utilizando 80% de los datos para entrenamiento (363 CDF) y 20% para validación (90 CDF). Por cada estado se almacenó el MSE y  $R^2$ ,

variando los parámetros  $\xi$  y  $\sigma$  por entrada y el que obtuvo menor MSE de validación es el modelo SVR elegido para posteriormente realizar pruebas. Este procedimiento se realizó por cada estado y entrada.

Desde el cuadro 4.1 hasta el cuadro 4.31 se muestran el MSE y  $R^2$  de entrenamiento/validación (E/V) y prueba en función del estado por cada tipo de entrada (características extraídas).

CUADRO 4.1: ERROR CUADRÁTICO MEDIO (MSE) Y COEFICIENTE DE CORRELACIÓN ( $R^2$ ) OBTENIDOS CON FD COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO.

Variable de entrada y cantidad de términos a la entrada del SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
FD con 8 términos (8 términos en total)	Bomba normal	0.0437549	0.378296	0.022331	0.841041
	Bomba funciona con crudo viscoso	0.0666932	0.286443	0.114221	0.00153029
	Filtración en la válvula viajera	0.000178662	0.99076	0.0382361	0.140617
	Filtración en la válvula fija	0.00176446	0.797136	0.0352551	0.0142077
	Interferencia de gas	0.000825769	0.985479	0.0116899	0.428518
	Golpe de fluido	0.0179732	0.653839	0.022738	0.513059
	Golpe de bomba pistón arriba	0.00153079	0.0587179	0.00913279	0.507146
	Golpe de bomba pistón abajo	4.60551E-8	0.999994	0.00818497	0.0281862
	Falla en caja de transmisión	0.0125166	0.336919	0.0287548	0.0124357
	Cabillas partidas	5.33919E-8	0.999954	0.022327	0.134686
	Otro estado no determinado en el presente PFM	0.0189302	0.51351	0.0361045	0.00527153

CUADRO 4.2: MSE Y  $R^2$  OBTENIDOS CON *Wavelet* COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

Variable de entrada y cantidad de términos a la entrada del SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
<i>Wavelet</i> con 8 términos (8 términos en total)	Bomba normal	0.000353659	0.994575	0.0578793	-1 (IND)
	Bomba funciona con crudo viscoso	0.0651654	0.298382	0.111031	0.144737
	Filtración en la válvula viajera	0.0151111	0.297387	0.0535092	0.0615485
	Filtración en la válvula fija	0.00642511	0.236501	0.0353225	0.000114597
	Interferencia de gas	0.00232748	0.959059	0.0236888	0.260661
	Golpe de fluido	0.0193066	0.643913	0.0325558	0.147278
	Golpe de bomba pistón arriba	0.00156613	0.131148	0.00929217	0.00310369
	Golpe de bomba pistón abajo	0.0043103	0.506981	0.00602233	0.255911
	Falla en caja de transmisión	0.00941512	0.492499	0.0275793	0.0082139
	Cabillas partidas	0.00116922	0.0116512	0.0228759	0.0767958
	Otro estado no determinado en el presente PFM	0.0139694	0.637623	0.0139541	0.446577

CUADRO 4.3: MSE Y  $R^2$  OBTENIDOS CON *Self Organizing Map* (SOM) COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

Variable de entrada y cantidad de términos a la entrada del SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
SOM con 8 términos (8 términos en total)	Bomba normal	0.00011077	0.998313	0.0562767	0.200752
	Bomba funciona con crudo viscoso	0.0824707	0.124369	0.0661113	0.0564744
	Filtración en la válvula viajera	1.52491E-07	0.999992	0.046757	0
	Filtración en la válvula fija	3.18919E-8	0.999996	0.0335601	0.0401255
	Interferencia de gas	0.0105392	0.824897	0.0291062	0.0134758
	Golpe de fluido	0.0107299	0.790106	0.0366238	0.0822489
	Golpe de bomba pistón arriba	0.000971901	0.847036	0.00945055	0.0166999
	Golpe de bomba pistón abajo	4.57152E-8	0.999994	0.00819684	0.00242884
	Falla en caja de transmisión	4.77593E-8	0.999997	0.0266707	0.00303665
	Cabillas partidas	0.00116975	1.98826E-5	0.0228866	0.450504
	Otro estado no determinado en el presente PFM	0.0228281	0.471078	0.0238934	0.0973204

CUADRO 4.4: MSE Y  $R^2$  OBTENIDOS CON *Fuzzy C-Means* (FCM) COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

Variable de entrada y cantidad de términos a la entrada del SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
FCM con 8 términos (8 términos en total)	Bomba normal	5.24952E-8	0.999999	0.0585282	0.140625
	Bomba funciona con crudo viscoso	7.96245E-8	0.999999	0.0889669	0.211237
	Filtración en la válvula viajera	5.87118E-08	0.999997	0.0472666	0.0278346
	Filtración en la válvula fija	4.74375E-08	0.999994	0.0335673	0.025352
	Interferencia de gas	3.34548E-8	0.999999	0.0222751	0.0470062
	Golpe de fluido	4.68147E-8	0.999999	0.0390316	0.318715
	Golpe de bomba pistón arriba	9.59185E-08	0.999994	0.00908569	0.032611
	Golpe de bomba pistón abajo	5.78762E-08	0.999993	0.00813209	0.080242
	Falla en caja de transmisión	2.8448E-8	0.999998	0.0262465	0.0144592
	Cabillas partidas	0.00116883	0.0105173	0.0228792	0.152277
	Otro estado no determinado en el presente PFM	4.88295E-8	0.999999	0.0188123	0.285233

CUADRO 4.5: MSE Y  $R^2$  OBTENIDOS CON LOS GRADOS API DEL CRUDO COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

Variable de entrada y cantidad de términos a la entrada del SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
Grados API del crudo (1 término en total)	Bomba normal	0.0704303	0.144269	0.0557071	0.411637
	Bomba funciona con crudo viscoso	0.0791429	0.164726	0.055504	0.242278
	Filtración en la válvula viajera	0.0203973	0.0573739	0.0514286	0.0870265
	Filtración en la válvula fija	0.00766554	0.047146	0.0357143	0.0698149
	Interferencia de gas	0.0672517	0.00273831	0.0207143	0.0281533
	Golpe de fluido	0.0554084	5.26887E-6	0.0435714	0.516825
	Golpe de bomba pistón arriba	0.00161148	0.00116662	0.00928571	0.03939
	Golpe de bomba pistón abajo	0.0102312	0.0138607	0.00930604	0.178731
	Falla en caja de transmisión	0.0226015	0.00201637	0.0271243	0.0232273
	Cabillas partidas	0.00116998	0.00104477	0.0228571	0.0494845
	Otro estado no determinado en el presente PFM	0.040915	0.00161954	0.0299989	0.00295398

CUADRO 4.6: MSE Y  $R^2$  OBTENIDOS CON FD Y *Wavelet* COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
FD y <i>Wavelet</i> (8 términos cada variable; 16 términos en total)	Bomba normal	0.0470695	0.357579	0.0420119	0.461663
	Bomba funciona con crudo viscoso	0.0639058	0.309836	0.116527	0.0155276
	Filtración en la válvula viajera	0.000158706	0.991766	0.04674	-6.85323E-17
	Filtración en la válvula fija	0.00180017	0.791187	0.0329628	0.0732373
	Interferencia de gas	0.00654639	0.881933	0.00617619	0.718049
	Golpe de fluido	0.00901099	0.833508	0.0146738	0.619385
	Golpe de bomba pistón arriba	0.00152988	0.0869631	0.00914192	0.434741
	Golpe de bomba pistón abajo	0.00395993	0.557241	0.00527764	0.343654
	Falla en caja de transmisión	0.0092994	0.489461	0.0319089	0.000876513
	Cabillas partidas	1.33277E-7	0.999936	0.020136	0.463398
	Otro estado no determinado en el presente PFM	0.0147581	0.618216	0.0293458	0.0545482

CUADRO 4.7: MSE Y  $R^2$  OBTENIDOS CON FD Y SOM COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

<b>Variables de entrada y cantidad de términos a la entrada del SVR</b>	<b>Estado</b>	<b>MSE de E/V</b>	<b><math>R^2</math> de E/V</b>	<b>MSE de P</b>	<b><math>R^2</math> de P</b>
FD y SOM (8 términos cada variable; 16 términos en total)	Bomba normal	0.0490047	0.303825	0.0414399	0.648319
	Bomba funciona con crudo viscoso	0.0645096	0.303323	0.105623	0.000108274
	Filtración en la válvula viajera	9.74316E-08	0.999995	0.0467436	0.00865642
	Filtración en la válvula fija	5.52191E-08	0.999993	0.0325519	0.368425
	Interferencia de gas	0.00102569	0.982088	0.017659	0.0656279
	Golpe de fluido	0.00196943	0.965096	0.0252396	0.333371
	Golpe de bomba pistón arriba	0.00146561	0.161848	0.00889929	0.15429
	Golpe de bomba pistón abajo	0.00498072	0.451197	0.00668635	0.18107
	Falla en caja de transmisión	2.85194E-8	0.999998	0.0266689	0.000106977
	Cabillas partidas	0.000407032	0.967866	0.0226536	0.00569379
	Otro estado no determinado en el presente PFM	0.00943964	0.752137	0.0475094	0.0158507

CUADRO 4.8: MSE Y  $R^2$  OBTENIDOS CON FD Y FCM COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

<b>VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR</b>	<b>Estado</b>	<b>MSE de E/V</b>	<b><math>R^2</math> de E/V</b>	<b>MSE de P</b>	<b><math>R^2</math> de P</b>
FD y FCM (8 términos cada variable; 16 términos en total)	Bomba normal	0.0457015	0.383277	0.0367206	0.559694
	Bomba funciona con crudo viscoso	0.0650959	0.296842	0.111338	0.0312852
	Filtración en la válvula viajera	6.14831E-8	0.999997	0.0478852	0.168868
	Filtración en la válvula fija	0.00504547	0.674175	0.0372733	0.0515997
	Interferencia de gas	0.00835647	0.878223	0.0234189	0.0513477
	Golpe de fluido	0.0102915	0.847601	0.0198348	0.586478
	Golpe de bomba pistón arriba	0.0012504	0.341326	0.00882465	0.265465
	Golpe de bomba pistón abajo	2.58542E-08	0.999997	0.0081813	0.00510875
	Falla en caja de transmisión	2.81649E-08	0.999998	0.0262869	0.0189439
	Cabillas partidas	0.00113667	0.850979	0.0228217	0.0334636
	Otro estado no determinado en el presente PFM	0.0148506	0.676314	0.0262132	0.0612749

CUADRO 4.9: MSE Y  $R^2$  OBTENIDOS CON FD Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
FD y API (8 términos la primera variable, 1 término la segunda; 9 términos en total)	Bomba normal	0.0317855	0.537021	0.0180965	0.736121
	Bomba funciona con crudo viscoso	0.0471067	0.49447	0.083312	0.124036
	Filtración en la válvula viajera	0.00045211	0.976778	0.0386086	0.145328
	Filtración en la válvula fija	0.00105837	0.866458	0.0269964	0.465877
	Interferencia de gas	0.0236798	0.580423	0.0132624	0.359688
	Golpe de fluido	0.0152962	0.713273	0.0236961	0.437709
	Golpe de bomba pistón arriba	0.00147161	0.0914372	0.00913838	0.597603
	Golpe de bomba pistón abajo	0.00376636	0.579779	0.00768593	0.158494
	Falla en caja de transmisión	0.0138259	0.253247	0.0301512	0.0014753
	Cabillas partidas	5.62867E-8	0.999952	0.0223598	0.0955485
	Otro estado no determinado en el presente PFM	0.0170369	0.571193	0.0386618	0.00224559

CUADRO 4.10: MSE Y  $R^2$  OBTENIDOS CON *Wavelet* Y SOM COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
<i>Wavelet</i> y SOM (8 términos cada variable; 16 términos en total)	Bomba normal	5.42463E-8	0.999999	0.0586224	0.00404791
	Bomba funciona con crudo viscoso	0.0746325	0.201549	0.0915592	0.00184995
	Filtración en la válvula viajera	1.31089E-7	0.999993	0.0467545	0.0127712
	Filtración en la válvula fija	5.02884E-8	0.999993	0.0329018	0.148365
	Interferencia de gas	0.000365755	0.993536	0.0127337	0.681893
	Golpe de fluido	0.0153472	0.712499	0.025368	0.376479
	Golpe de bomba pistón arriba	0.000323437	0.959121	0.00879156	0.0091283
	Golpe de bomba pistón abajo	0.00535055	0.411612	0.0063896	0.22483
	Falla en caja de transmisión	0.000133963	0.992129	0.0261943	0.0209663
	Cabillas partidas	0.00115715	0.122507	0.0228001	0.00539911
	Otro estado no determinado en el presente PFM	5.47487E-5	0.998547	0.0345729	0.0128989

CUADRO 4.11: MSE Y  $R^2$  OBTENIDOS CON *Wavelet* Y FCM COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
<i>Wavelet</i> y FCM (8 términos cada variable; 16 términos en total)	Bomba normal	4.69727E-8	0.999999	0.0578085	4.41343E-13
	Bomba funciona con crudo viscoso	0.0768768	0.191394	0.106646	0.226158
	Filtración en la válvula viajera	6.37621E-8	0.999997	0.04722956	0.0305811
	Filtración en la válvula fija	3.7414E-5	0.995566	0.037586	0.0276809
	Interferencia de gas	3.41974E-5	0.999384	0.0256456	0.0229826
	Golpe de fluido	0.0138171	0.796446	0.0444657	0.0405089
	Golpe de bomba pistón arriba	0.00122923	0.745249	0.00909492	0.0111932
	Golpe de bomba pistón abajo	6.91892E-8	0.999992	0.00818965	-7.07148E-17
	Falla en caja de transmisión	1.95602E-5	0.998875	0.028653	0.000183822
	Cabillas partidas	0.00116897	0.0017746	0.0228347	0.00585213
	Otro estado no determinado en el presente PFM	0.0058477	0.87322	0.0198161	0.182109

CUADRO 4.12: MSE Y  $R^2$  OBTENIDOS CON *Wavelet* Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
<i>Wavelet</i> y API (8 términos la primera variable, 1 término la segunda; 9 términos en total)	Bomba normal	0.000423043	0.993512	0.0502539	0.183188
	Bomba funciona con crudo viscoso	0.0607445	0.351817	0.0745745	0.0590068
	Filtración en la válvula viajera	0.000158564	0.991773	0.0467293	0.0148084
	Filtración en la válvula fija	0.00497287	0.346155	0.0334364	0.139227
	Interferencia de gas	0.000862567	0.984886	0.0150139	0.285263
	Golpe de fluido	0.0010441	0.980212	0.0302996	0.268054
	Golpe de bomba pistón arriba	0.00155071	0.116056	0.00929721	0.00473567
	Golpe de bomba pistón abajo	0.00369111	0.576955	0.00883367	0.0540989
	Falla en caja de transmisión	0.000910951	0.946333	0.0274415	0.0148209
	Cabillas partidas	0.00205907	0.019605	0.00103742	0.03011
	Otro estado no determinado en el presente PFM	0.0320687	0.198156	0.0279236	0.0330903

CUADRO 4.13: MSE Y  $R^2$  OBTENIDOS CON SOM Y FCM COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	ESTADO	MSE DE E/V	$R^2$ DE E/V	MSE DE P	$R^2$ DE P
SOM y FCM (8 términos cada variable; 16 términos en total)	Bomba normal	4.67234E-8	0.999999	0.0578085	-4.4553E-12
	Bomba funciona con crudo viscoso	1.3296E-7	0.999999	0.0785948	-4.996E-15
	Filtración en la válvula viajera	1.52464E-7	0.999992	0.0467563	0
	Filtración en la válvula fija	1.08986E-7	0.999985	0.0336044	0.000450276
	Interferencia de gas	1.20423E-5	0.999785	0.024392	0.0236721
	Golpe de fluido	1.27333E-6	0.999974	0.0379949	0.00443274
	Golpe de bomba pistón arriba	1.76003E-7	0.999952	0.00915727	0.0514054
	Golpe de bomba pistón abajo	3.16226E-8	0.999996	0.00819103	0.00150975
	Falla en caja de transmisión	4.68541E-8	0.999997	0.0267448	0.00396286
	Cabillas partidas	0.00116972	0.000167426	0.022866	0.16884
	Otro estado no determinado en el presente PFM	1.347953E-5	0.999635	0.0270697	0.00149442

CUADRO 4.14: MSE Y  $R^2$  OBTENIDOS CON SOM Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
SOM y API (8 términos la primera variable, 1 término la segunda; 9 términos en total)	Bomba normal	0.000350419	0.994638	0.0370733	0.59371
	Bomba funciona con crudo viscoso	0.067818	0.280644	0.0586988	0.217417
	Filtración en la válvula viajera	5.10938E-8	0.999997	0.0462903	0.0654075
	Filtración en la válvula fija	5.446e-8	0.999993	0.0333268	0.0764156
	Interferencia de gas	0.00055476	0.990152	0.0233532	0.237225
	Golpe de fluido	0.00752493	0.852569	0.0360726	0.107518
	Golpe de bomba pistón arriba	0.000905806	0.880699	0.009528	0.0436904
	Golpe de bomba pistón abajo	7.44502E-8	0.999991	0.0081979	0.00785909
	Falla en caja de transmisión	3.78034E-8	0.999998	0.0266615	0.0456911
	Cabillas partidas	0.00117044	1.98182E-5	0.0228771	0.0295538
	Otro estado no determinado en el presente PFM	8.40144E-8	0.999998	0.0224938	0.586241

CUADRO 4.15: MSE Y  $R^2$  OBTENIDOS CON FCM Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
FCM y API (8 términos la primera variable, 1 término la segunda; 9 términos en total)	Bomba normal	0.0537269	0.313391	0.0479087	0.322383
	Bomba funciona con crudo viscoso	0.0700765	0.270921	0.0882825	0.0228188
	Filtración en la válvula viajera	6.52474E-5	0.996757	0.0500577	0.00913015
	Filtración en la válvula fija	3.60541E-8	0.999995	0.033544	0.0306271
	Interferencia de gas	5.60651E-8	0.999999	0.022876	0.128071
	Golpe de fluido	5.77584E-8	0.999999	0.0376986	0.393173
	Golpe de bomba pistón arriba	1.03985E-7	0.999936	0.00908654	0.0119309
	Golpe de bomba pistón abajo	4.47943E-8	0.999994	0.00814981	0.0659313
	Falla en caja de transmisión	3.01975E-8	0.999998	0.025991	0.0492615
	Cabillas partidas	0.00116983	6.19411E-6	0.0228473	0.0243394
	Otro estado no determinado en el presente PFM	4.43371E-8	0.999999	0.0215233	0.0657374

CUADRO 4.16: MSE Y  $R^2$  OBTENIDOS CON FD, *Wavelet* Y SOM COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
FD, <i>Wavelet</i> y SOM (8 términos cada una de las variables; 24 términos en total)	Bomba normal	0.0444587	0.377732	0.0426526	0.540217
	Bomba funciona con crudo viscoso	0.0629951	0.319792	0.11232	0.0234638
	Filtración en la válvula viajera	1.06944E-7	0.999994	0.0467469	0.00982961
	Filtración en la válvula fija	0.0040289	0.648856	0.03457	0.179693
	Interferencia de gas	0.0231993	0.657382	0.0109415	0.445999
	Golpe de fluido	0.00354337	0.935838	0.0273973	0.403817
	Golpe de bomba pistón arriba	0.001212	0.491934	0.00891373	0.0688807
	Golpe de bomba pistón abajo	0.00306094	0.675059	0.00605232	0.250771
	Falla en caja de transmisión	0.000692506	0.96124	0.025931	0.0333033
	Cabillas partidas	0.000169186	0.986802	0.0224395	0.00141861
	Otro estado no determinado en el presente PFM	0.0114083	0.723818	0.0313949	0.066247

CUADRO 4.17: MSE Y  $R^2$  OBTENIDOS CON FD, *Wavelet* Y FCM COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
FD, <i>Wavelet</i> y FCM (8 términos cada una de las variables; 24 términos en total)	Bomba normal	0.0476107	0.366626	0.038401	0.519421
	Bomba funciona con crudo viscoso	0.0640509	0.308305	0.110872	0.0280705
	Filtración en la válvula viajera	0.0105607	0.683186	0.0543802	0.0214731
	Filtración en la válvula fija	0.0019672	0.797001	0.0452599	0.0575787
	Interferencia de gas	0.0149004	0.799843	0.0173098	0.138069
	Golpe de fluido	0.0149876	0.756796	0.0219065	0.470341
	Golpe de bomba pistón arriba	0.000913639	0.572759	0.00863609	0.178333
	Golpe de bomba pistón abajo	0.00420494	0.581396	0.00716927	0.12606
	Falla en caja de transmisión	0.00916205	0.520258	0.0294387	0.00278172
	Cabillas partidas	0.00116881	0.0337513	0.0228468	0.249491
	Otro estado no determinado en el presente PFM	0.0160931	0.617009	0.0319317	0.0782946

CUADRO 4.18: MSE Y  $R^2$  OBTENIDOS CON FD, *Wavelet* Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

<b>VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR</b>	<b>Estado</b>	<b>MSE de E/V</b>	<b><math>R^2</math> de E/V</b>	<b>MSE de P</b>	<b><math>R^2</math> de P</b>
FD, <i>Wavelet</i> y API (8 términos las dos primeras variables, 1 término la última; 17 términos en total)	Bomba normal	0.00630787	0.90479	0.036985	0.410672
	Bomba funciona con crudo viscoso	0.0426255	0.544208	0.0919487	0.082952
	Filtración en la válvula viajera	0.00015861	0.991771	0.046744	0.00373491
	Filtración en la válvula fija	0.00117521	0.855594	0.0277399	0.326609
	Interferencia de gas	0.0129514	0.775771	0.00762656	0.633495
	Golpe de fluido	0.0071011	0.871236	0.0178967	0.554829
	Golpe de bomba pistón arriba	0.00152316	0.0870712	0.00912345	0.324881
	Golpe de bomba pistón abajo	0.00350392	0.611289	0.00576493	0.283915
	Falla en caja de transmisión	0.00886354	0.524378	0.0338537	0.0125167
	Cabillas partidas	5.38596E-8	0.999954	0.0200751	0.439397
	Otro estado no determinado en el presente PFM	0.0124145	0.682663	0.0326549	0.0457039

CUADRO 4.19: MSE Y  $R^2$  OBTENIDOS CON FD, SOM Y FCM COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
FD, SOM y FCM (8 términos cada una de las variables; 24 términos en total)	Bomba normal	0.0460107	0.387756	0.0406749	0.58424
	Bomba funciona con crudo viscoso	0.0608063	0.342467	0.11392	0.0159959
	Filtración en la válvula viajera	1.52476E-7	0.999992	0.0467568	-5.13992E-17
	Filtración en la válvula fija	0.00544385	0.718044	0.0365533	0.0496574
	Interferencia de gas	0.000422017	0.993105	0.0167259	0.219384
	Golpe de fluido	0.000833848	0.983965	0.0174462	0.6116
	Golpe de bomba pistón arriba	0.00133109	0.388093	0.0087404	0.321895
	Golpe de bomba pistón abajo	2.87847E-8	0.999997	0.00819489	0.00470034
	Falla en caja de transmisión	1.48578E-7	0.999991	0.0266538	0.000322522
	Cabillas partidas	0.00108897	0.882402	0.0228957	0.0184047
	Otro estado no determinado en el presente PFM	0.0153432	0.738187	0.0230626	0.101353

CUADRO 4.20: MSE Y  $R^2$  OBTENIDOS CON FD, SOM Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

<b>VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR</b>	<b>Estado</b>	<b>MSE de E/V</b>	<b><math>R^2</math> de E/V</b>	<b>MSE de P</b>	<b><math>R^2</math> de P</b>
FD, SOM y API (8 términos las dos primeras variables, 1 término la última; 17 términos en total)	Bomba normal	0.0353845	0.4858	0.0198646	0.706142
	Bomba funciona con crudo viscoso	0.0466701	0.500228	0.080679	0.101971
	Filtración en la válvula viajera	5.20994E-8	0.999997	0.0466992	0.0242548
	Filtración en la válvula fija	0.0054144	0.598633	0.0363962	0.0631703
	Interferencia de gas	0.0195008	0.710946	0.0126657	0.342143
	Golpe de fluido	0.0207067	0.618427	0.0250831	0.502002
	Golpe de bomba pistón arriba	0.00145728	0.172883	0.00887246	0.165305
	Golpe de bomba pistón abajo	0.00461521	0.505311	0.00776598	0.0859115
	Falla en caja de transmisión	5.28646E-8	0.999997	0.0265865	5.41823E-5
	Cabillas partidas	0.000328282	0.97308	0.022492	0.00206586
	Otro estado no determinado en el presente PFM	0.00776181	0.801921	0.0500128	0.0171742

CUADRO 4.21: MSE Y  $R^2$  OBTENIDOS CON FD, FCM Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

<b>VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR</b>	<b>Estado</b>	<b>MSE de E/V</b>	<b><math>R^2</math> de E/V</b>	<b>MSE de P</b>	<b><math>R^2</math> de P</b>
FD, FCM y API (8 términos las dos primeras variables, 1 término la última; 17 términos en total)	Bomba normal	0.0330217	0.536666	0.0173366	0.732223
	Bomba funciona con crudo viscoso	0.0492854	0.475845	0.08377575	0.0591837
	Filtración en la válvula viajera	0.0111202	0.618825	0.051927	0.00115165
	Filtración en la válvula fija	0.00384772	0.713362	0.0371485	0.011734
	Interferencia de gas	0.00648827	0.901291	0.024911	0.0545143
	Golpe de fluido	0.00859216	0.857229	0.022541	0.444419
	Golpe de bomba pistón arriba	0.00121161	0.366713	0.00871509	0.326739
	Golpe de bomba pistón abajo	5.03953E-8	0.999994	0.00828196	0.00267491
	Falla en caja de transmisión	5.4402E-8	0.999997	0.0263847	0.0751798
	Cabillas partidas	0.0011323	0.853681	0.0228196	0.0244757
	Otro estado no determinado en el presente PFM	0.0128899	0.726508	0.0247391	0.0669244

CUADRO 4.22: MSE Y  $R^2$  OBTENIDOS CON *Wavelet*, SOM Y FCM COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
<i>Wavelet</i> , SOM y FCM (8 términos cada una de las variables; 24 términos en total)	Bomba normal	4.69882E-8	0.999999	0.0578085	0
	Bomba funciona con crudo viscoso	0.070444	0.259483	0.0952043	0.0615651
	Filtración en la válvula viajera	1.52492E-7	0.999992	0.0466757	-5.13992E-17
	Filtración en la válvula fija	5.22056E-8	0.999993	0.0340015	3.1319E-5
	Interferencia de gas	5.49803E-8	0.999999	0.0202137	0.156864
	Golpe de fluido	0.0121782	0.828996	0.036659	0.0756851
	Golpe de bomba pistón arriba	6.92463E-7	0.999919	0.00925093	0.106832
	Golpe de bomba pistón abajo	1.39272E-5	0.998392	0.00661532	0.295907
	Falla en caja de transmisión	5.11069E-8	0.999997	0.0276877	0.00689677
	Cabillas partidas	5.88774E-8	0.999995	0.0225855	7.71455E-9
	Otro estado no determinado en el presente PFM	3.38575E-5	0.999184	0.0234025	0.0887096

CUADRO 4.23: MSE Y  $R^2$  OBTENIDOS CON *Wavelet*, SOM Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
<i>Wavelet</i> , SOM y API (8 términos las dos primeras variables, 1 término la última; 17 términos en total)	Bomba normal	0.0360646	0.522919	0.0586426	0.120662
	Bomba funciona con crudo viscoso	0.0545208	0.415044	0.0624948	0.212866
	Filtración en la válvula viajera	9.73377E-8	0.999995	0.046742	0.0165939
	Filtración en la válvula fija	0.00477033	0.532197	0.0358074	0.0275188
	Interferencia de gas	0.0268809	0.633007	0.0186188	0.0809613
	Golpe de fluido	0.0126796	0.768494	0.0246638	0.363246
	Golpe de bomba pistón arriba	0.000971198	0.785788	0.00902843	0.00300814
	Golpe de bomba pistón abajo	0.00504301	0.445351	0.0062503	0.23591
	Falla en caja de transmisión	0.000131456	0.992253	0.0257828	0.0353116
	Cabillas partidas	0.00117025	0.000625736	0.0228562	0.122248
	Otro estado no determinado en el presente PFM	3.7147E-5	0.999008	0.0286663	0.00118749

CUADRO 4.24: MSE Y  $R^2$  OBTENIDOS CON *Wavelet*, FCM Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
<i>Wavelet</i> , FCM y API (8 términos las dos primeras variables, 1 término la última; 17 términos en total)	Bomba normal	0.0409703	0.454069	0.0372174	0.422903
	Bomba funciona con crudo viscoso	0.0569363	0.393839	0.0839191	0.0133552
	Filtración en la válvula viajera	4.17661E-8	0.999998	0.0490662	0.00400324
	Filtración en la válvula fija	0.00307089	0.679698	0.0354026	0.00106404
	Interferencia de gas	2.62588E-5	0.999527	0.0261781	0.0236795
	Golpe de fluido	0.0116834	0.806973	0.0492118	0.0200829
	Golpe de bomba pistón arriba	0.000440132	0.88565	0.00935032	0.016046
	Golpe de bomba pistón abajo	0.00561509	0.414322	0.00815248	0.0178124
	Falla en caja de transmisión	3.52976E-8	0.999804	0.0257431	0.0317696
	Cabillas partidas	0.00116961	0.00124332	0.0228612	0.0944711
	Otro estado no determinado en el presente PFM	0.000164569	0.995621	0.0177039	0.241344

CUADRO 4.25: MSE Y  $R^2$  OBTENIDOS CON SOM, FCM Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

<b>VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR</b>	<b>Estado</b>	<b>MSE de E/V</b>	<b><math>R^2</math> de E/V</b>	<b>MSE de P</b>	<b><math>R^2</math> de P</b>
SOM, FCM y API (8 términos las dos primeras variables, 1 término la última; 17 términos en total)	Bomba normal	0.000125582	0.998086	0.0465136	0.396518
	Bomba funciona con crudo viscoso	0.0600852	0.365805	0.0757254	0.0766014
	Filtración en la válvula viajera	1.52476E-7	0.999992	0.0467568	4.11194E-17
	Filtración en la válvula fija	4.19254E-8	0.999994	0.0333308	0.051246
	Interferencia de gas	8.1066E-6	0.999855	0.0258387	0.024388
	Golpe de fluido	7.60006E-7	0.999985	0.0364622	0.0512172
	Golpe de bomba pistón arriba	5.20541E-5	0.994271	0.00925203	0.00626504
	Golpe de bomba pistón abajo	1.49606E-6	0.999829	0.0079351	0.0452058
	Falla en caja de transmisión	3.89105E-8	0.999998	0.0260155	0.0306883
	Cabillas partidas	5.88787E-8	0.99995	0.0225855	-2.11354E-13
	Otro estado no determinado en el presente PFM	3.81061E-6	0.999897	0.0285767	0.00530737

CUADRO 4.26: MSE Y  $R^2$  OBTENIDOS CON FD, *Wavelet*, SOM Y FCM COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
FD, <i>Wavelet</i> , SOM y FCM (8 términos cada variable, (32 términos en total))	Bomba normal	0.0495399	0.335975	0.0437104	0.487836
	Bomba funciona con crudo viscoso	0.0563462	0.392151	0.120125	0.0231712
	Filtración en la válvula viajera	1.52491E-7	0.999992	0.046757	-1 (IND)
	Filtración en la válvula fija	0.00314082	0.750504	0.0384902	0.0309063
	Interferencia de gas	0.00403785	0.944095	0.0092887	0.520029
	Golpe de fluido	0.00431797	0.932114	0.0206655	0.450797
	Golpe de bomba pistón arriba	0.00110547	0.524163	0.00874137	0.277458
	Golpe de bomba pistón abajo	0.004580921	0.54187	0.00584472	0.383315
	Falla en caja de transmisión	0.0129229	0.310057	0.0250702	0.0773847
	Cabillas partidas	0.00116859	0.025888	0.0228349	0.18603
	Otro estado no determinado en el presente PFM	0.0100716	0.779206	0.027917	0.113469

CUADRO 4.27: MSE Y  $R^2$  OBTENIDOS CON FD, *Wavelet*, SOM Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
FD, <i>Wavelet</i> , SOM y API (8 términos las tres primeras variables cada una, un término la última variable; (25 términos en total))	Bomba normal	0.0100716	0.779206	0.027917	0.113469
	Bomba funciona con crudo viscoso	0.0433956	0.538832	0.0836277	0.0881754
	Filtración en la válvula viajera	6.50375E-8	0.999997	0.0467427	0.0106735
	Filtración en la válvula fija	0.0029045	0.663075	0.0362724	0.0769559
	Interferencia de gas	0.0218024	0.686484	0.0108823	0.478623
	Golpe de fluido	0.0233787	0.590476	0.0208629	0.548078
	Golpe de bomba pistón arriba	0.00117153	0.532236	0.00895436	0.0340616
	Golpe de bomba pistón abajo	0.00457615	0.52016	0.00523349	0.435612
	Falla en caja de transmisión	0.000539202	0.970234	0.0269097	0.0189997
	Cabillas partidas	0.000113148	0.986777	0.02233	0.000160367
	Otro estado no determinado en el presente PFM	0.00987454	0.766678	0.0329877	0.0591975

CUADRO 4.28: MSE Y  $R^2$  OBTENIDOS CON FD, *Wavelet*, FCM Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
FD, <i>Wavelet</i> , FCM y API (8 términos las tres primeras variables cada una, un término la última variable; (25 términos en total))	Bomba normal	0.035291	0.514726	0.0224774	0.653487
	Bomba funciona con crudo viscoso	0.04625	0.508795	0.0903159	0.03097
	Filtración en la válvula viajera	0.00730051	0.76054	0.0534029	0.0062441
	Filtración en la válvula fija	0.00149297	0.834241	0.0426615	0.0557399
	Interferencia de gas	0.012613	0.825336	0.0157079	0.205866
	Golpe de fluido	0.0135025	0.7752	0.023296	0.396882
	Golpe de bomba pistón arriba	0.00084628	0.616591	0.00875075	0.195125
	Golpe de bomba pistón abajo	0.00364883	0.623054	0.00746551	0.0839475
	Falla en caja de transmisión	0.013647	0.266915	0.0264248	0.046771
	Cabillas partidas	0.00109643	0.841873	0.0228907	0.0307516
	Otro estado no determinado en el presente PFM	0.0145952	0.658628	0.0294758	0.101834

CUADRO 4.29: MSE Y  $R^2$  OBTENIDOS CON FD, SOM, FCM Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

<b>VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR</b>	<b>Estado</b>	<b>MSE de E/V</b>	<b><math>R^2</math> de E/V</b>	<b>MSE de P</b>	<b><math>R^2</math> de P</b>
FD, SOM, FCM y API (8 términos las tres primeras variables cada una, un término la última variable; (25 términos en total))	Bomba normal	0.0329694	0.541908	0.0190538	0.727402
	Bomba funciona con crudo viscoso	0.0455054	0.51772	0.0879442	0.0728446
	Filtración en la válvula viajera	0.0123586	0.618305	0.0467099	0.0211385
	Filtración en la válvula fija	0.00448099	0.738394	0.0374076	0.158015
	Interferencia de gas	0.000382601	0.99356	0.017327	0.243017
	Golpe de fluido	0.00783278	0.874673	0.0207965	0.470296
	Golpe de bomba pistón arriba	0.00130468	0.389441	0.00861276	0.44699
	Golpe de bomba pistón abajo	1.37915E-6	0.999842	0.00773865	0.0613184
	Falla en caja de transmisión	7.36583E-8	0.999996	0.0264152	0.0447811
	Cabillas partidas	0.00114231	0.783907	0.0228593	0.000326843
	Otro estado no determinado en el presente PFM	0.0141947	0.762464	0.0225576	0.120675

CUADRO 4.30: MSE Y  $R^2$  OBTENIDOS CON *Wavelet*, SOM, FCM Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

<b>Variables de entrada y cantidad de términos a la entrada del SVR</b>	<b>Estado</b>	<b>MSE de E/V</b>	<b><math>R^2</math> de E/V</b>	<b>MSE de P</b>	<b><math>R^2</math> de P</b>
<i>Wavelet</i> , SOM, FCM y API (8 términos las tres primeras variables cada una, un término la última variable; (25 términos en total)	Bomba normal	0.0443828	0.463844	0.0501832	0.294404
	Bomba funciona con crudo viscoso	0.0576915	0.391911	0.0782919	0.0276804
	Filtración en la válvula viajera	1.52466E-7	0.999992	0.0467567	-1.28498E-17
	Filtración en la válvula fija	0.0038835	0.684329	0.0364218	0.0137638
	Interferencia de gas	5.2062E-8	0.999999	0.020591	0.165781
	Golpe de fluido	0.0102602	0.842778	0.0386062	0.0793461
	Golpe de bomba pistón arriba	3.15925E-7	0.999952	0.0092337	0.0595546
	Golpe de bomba pistón abajo	1.34739E-5	0.998444	0.00674788	0.295524
	Falla en caja de transmisión	4.20813E-8	0.999997	0.0260741	0.0250365
	Cabillas partidas	0.00117057	0.00013171	0.0228479	0.138545
	Otro estado no determinado en el presente PFM	8.50759E-6	0.99782	0.024553	0.0735545

CUADRO 4.31: MSE Y  $R^2$  OBTENIDOS CON FD, *Wavelet*, SOM, FCM Y API COMO ENTRADA AL SVR POR CADA ESTADO DE FUNCIONAMIENTO

VARIABLES DE ENTRADA Y CANTIDAD DE TÉRMINOS A LA ENTRADA DEL SVR	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
FD, <i>Wavelet</i> , SOM, FCM y API (8 las cuatro primeras variables cada una, un término la última; 33 términos en total)	Bomba normal	0.0371784	0.479461	0.0227577	0.662923
	Bomba funciona con crudo viscoso	0.0426923	0.550658	0.0916309	0.0457236
	Filtración en la válvula viajera	0.0095725	0.721113	0.0485503	0.00379725
	Filtración en la válvula fija	0.00238046	0.735977	0.0401573	0.0986572
	Interferencia de gas	0.00340194	0.950892	0.00861542	0.589675
	Golpe de fluido	0.00388121	0.93427	0.0226536	0.399061
	Golpe de bomba pistón arriba	0.00107036	0.53641	0.0086674	0.41566
	Golpe de bomba pistón abajo	0.00428269	0.563214	0.0063383	0.309369
	Falla en caja de transmisión	0.012652	0.319028	0.0255932	0.0680133
	Cabillas partidas	0.00115453	0.45819	0.0228946	0.27027
	Otro estado no determinado en el presente PFM	0.0090525	0.800795	0.0276563	0.134509

Como el caso donde el pozo BMC funciona con crudo viscoso es un caso especial del caso donde el pozo BMC funciona normal, se decide probar si la unión de estos casos puede mejorar el MSE y  $R^2$  en este nuevo caso llamado bomba funciona normal (incluyendo crudo viscoso). La fusión de los estados bombeo normal y bombeo de fluido viscoso no arrojó buenos resultados. En el siguiente cuadro se muestra este caso, calculando el MSE y  $R^2$  de entrenamiento/validación (E/V) y prueba (P) por cada tipo de entrada (características extraídas).

CUADRO 4.32: MSE Y  $R^2$  OBTENIDOS CON TODAS LAS CARACTERÍSTICAS POSIBLES COMO ENTRADA AL SVR PARA LA FUSIÓN DE LOS ESTADOS BOMBA NORMAL Y BOMBA FUNCIONA CON CRUDO VISCOZO EN BOMBA FUNCIONA NORMAL (INCLUYENDO CRUDO VISCOZO)

VARIABLES DE ENTRADA AL SVR	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P
FD	0.0530514	0.538265	0.102844	0.0867462
<i>Wavelet</i>	0.00114928	0.98999	0.0647896	0.332753
SOM	0.0859281	0.253712	0.126684	0.00198026
FCM	0.0785903	0.313884	0.141156	0.128577
Grados API (° API)	0.113671	0.0174872	0.110073	0.107047
FD y <i>Wavelet</i>	0.0541664	0.5299	0.113156	0.0620715
FD y SOM	0.0572442	0.505533	0.0988327	0.0851677
FD y FCM	0.0540523	0.529022	0.102903	0.07109
FD y API	0.0488484	0.57365	0.111192	0.0692152
<i>Wavelet</i> y SOM	0.0892682	0.224833	0.136592	0.000103715
<i>Wavelet</i> y FCM	0.0723109	0.370535	0.147329	0.0842701
<i>Wavelet</i> y API	0.077966	0.325597	0.120063	0.000204014
SOM y FCM	0.103265	0.101363	0.12871	0.0493261
SOM y API	0.00270687	0.976635	0.091132	0.142056
FCM y API	0.074184	0.352561	0.146497	0.171226
FD, <i>Wavelet</i> y SOM	0.0531133	0.53768	0.111983	0.0524997
FD, <i>Wavelet</i> y FCM	0.0501715	0.564237	0.116056	0.0385246
FD, <i>Wavelet</i> y API	0.0396892	0.653178	0.125149	0.0422163
FD, SOM y FCM	0.0518281	0.551559	0.109371	0.0454298
FD, SOM y API	0.0547477	0.524252	0.0979963	0.0795333
FD, FCM y API	0.0464811	0.59532	0.107279	0.0558456
<i>Wavelet</i> , SOM y FCM	0.0622379	0.456164	0.146655	0.0249488
<i>Wavelet</i> , SOM y API	0.0802966	0.298447	0.140384	0.00244705
<i>Wavelet</i> , FCM y API	0.0690123	0.397028	0.141082	0.120337
SOM, FCM y API	0.0759738	0.340396	0.151491	0.0739422
FD, <i>Wavelet</i> , SOM y FCM	0.0473397	0.590659	0.125319	0.0186834
FD, <i>Wavelet</i> , SOM y API	0.0464527	0.597351	0.116662	0.0563823
FD, <i>Wavelet</i> , FCM y API	0.0483963	0.580985	0.108649	0.039302
FD, SOM, FCM y API	0.0493028	0.573991	0.10872	0.0382097
<i>Wavelet</i> , SOM, FCM y API	0.0582275	0.492397	0.147239	0.0463399
FD, <i>Wavelet</i> , SOM, FCM y API	0.0454085	0.610219	0.123039	0.0138467

Analizando los MSE y  $R^2$  del cuadro 4.32, Aunque los datos de E/V arrojó un  $R^2$  de 0.98999, los datos de prueba arrojaron un  $R^2$  de 0.332753.

Como se pudo observar desde el cuadro 4.1 al 4.32, existen muchos estados con entradas específicas que tienen un MSE muy bajo,  $R^2$  altos en entrenamiento pero MSE altos y  $R^2$

en pruebas. Estos modelos no fueron seleccionados y por ende se eligieron aquellos con MSE muy bajo y  $R^2$  altos en entrenamiento y validación. Para este caso se utilizó la media armónica (siguiendo el caso similar del  $F_1$  score para clasificación) entre el  $R^2$  de validación y prueba. En el siguiente cuadro se muestra un resumen de los mejores regresores obtenidos por cada estado en función de la entrada con los datos de entrenamiento, validación y pruebas.

CUADRO 4.33: CUADRO RESUMEN DE LOS MEJORES REGRESORES EN FUNCIÓN DE LOS EXTRACTORES DE CARACTERÍSTICAS Y LOS ESTADOS DE FUNCIONAMIENTO DE BMC CON LOS DATOS DE E/V Y P, CALCULANDO LA MEDIA ARMÓNICA (M.A) ENTRE EL  $R^2$  DE V Y P

Regresor (cuadro de referencia)	Estado	MSE de E/V	$R^2$ de E/V	MSE de P	$R^2$ de P	M.A $R^2$
WP y SOM Cuadro 4.10	Interferencia de gas	3.66E-4	0.993536	0.0127	0.6819	0.8087
FD, SOM y FCM Cuadro 4.19	Golpe de fluido	8.34E-4	0.983965	0.0175	0.6116	0.7543
SOM y API Cuadro 4.14	Bomba normal	3.5E-4	0.994638	0.0371	0.5937	0.7436
SOM y API Cuadro 4.14	Otro estado no deter- minado en el presente PFM	8.4E-8	0.999998	0.0225	0.5862	0.7392
FD y WP Cuadro 4.6	Cabillas partidas	1.33E-7	0.999936	0.0201	0.4634	0.6333
FD y API Cuadro 4.8	Filtración en la válvula fija	1.06E-3	0.866458	0.027	0.4659	0.6059

Si se desea obtener mejores resultados se puede hacer uso de funciones matemáticas específicas con la finalidad de mejorar la estimación del diagnóstico. Por ejemplo, si a la salida de los datos de prueba del caso golpe de fluido (cuadro 4.19) se le hace pasar la función  $x(1 + \text{sign}(x))$ , la correlación entre este nuevo vector y el vector de prueba es de 0.9957.

Dado los resultados mostrados en el cuadro 4.33, se demuestra la viabilidad de realizar un sistema diagnóstico para pozos BMC usando las características empleadas en el presente PFM, para posteriormente aplicarla sobre sistemas empotrados [13] y tiempo real [14] para los estados: interferencia de gas, golpe de fluido, bomba normal, cabillas partidas, filtración en la válvula fija y otro estado no determinado en el presente PFM.

# Capítulo 5

## Conclusiones

En este último capítulo se presenta las principales contribuciones en el área de reconocimiento de patrones y procesamiento de señales para sistemas diagnósticos para pozos BMC usando CDF, puestas en manifiesto a lo largo de la presente disertación.

### 5.1. Aportes

Generación de modelos satisfactorios para los estados: interferencia de gas, golpe de fluido, bomba normal, cabillas partidas, filtración en la válvula fija y otro estado no determinado en el presente PFM, con  $R^2$  relativamente altos y MSE bastante bajos.

Inclusión de la variable del grado API y la aplicación de FCM como extractores de características para la generación de modelos regresión para sistemas diagnóstico para pozos BMC usando CDF, nunca antes realizado según el estado del arte realizado y sirviendo de ayuda para la generación de los estados: golpe de fluido (FCM), bomba normal, filtración en la válvula fija y otro estado no determinado en el presente PFM (API).

Finalmente el uso de SVR para realizar sistemas diagnóstico sobre sistemas BMC para los estados anteriormente citados no había sido realizado según el estado del arte realizado, arrojando resultados satisfactorios la implementación de este regresor.

### 5.2. Limitaciones

Como se pudo notar en la figura 4.1 no existe una representación total de CDF en todos los tipos de crudo. Esto limita la aplicación de los modelos generados solo para CP/XP y parte de crudos livianos.

No existe una cantidad representativamente equitativa de CDS/CDF por cada estado de funcionamiento. Esto implica que existen clases sesgadas (*skewed classes*) en el supuesto de que se requiera realizar modelos de clasificación.

Se utilizaron solo 14 CDF para realizar las pruebas de los modelos de regresión generados, resultando muy pocas CDF para poder verificar que los modelos generados son satisfactorios. No se aumenta el tamaño del conjunto de prueba porque va en detrimento del conjunto de E/V.

La cantidad de elementos de la matriz de diagnóstico es de 467 (cantidad de CDF)  $\times$  11 (cantidad de estados) = 5137, contiene 3902 ceros (mas del 75% de la matriz es cero), dificultando de esta manera la resolución del problema de optimización asociado al SVR (matriz dispersa) .

Solo se tienen 38 CDF donde pertenecen únicamente a un estado, las 429 cartas restantes (91.86% de todos los datos) poseen al menos dos estados activos, dificultando de esta manera la generación de modelos que determine el grado de pertenencia de una CDF a cada uno de los estados previamente diagnosticados.

### 5.3. Recomendaciones y Trabajo Futuro

- Buscar mayor cantidad de CDS/CDF para aumentar así el tamaño de los conjuntos de entrenamiento, validación y prueba y de los estados de funcionamiento de pozos BMC y evitar así el sesgo de clases.
- En función del punto anterior buscar cantidad CDS/CDF que solo pertenezcan a un estado de funcionamiento para así generar modelos de clasificación.
- Buscar CDS/CDF que representen en su totalidad los diferentes tipos de crudo (condensado / liviano, mediano, pesado y extra-pesado), con especial énfasis para crudos pesados, medianos y livianos. Esto con la finalidad de generar un modelo con mejores propiedades de generalización.
- Implementar otras técnicas de inferencia de CDF, como la descrita en [12], prestando atención al calculo del factor de amortiguamiento tanto simple como doble, descritos en las aplicaciones de patente americanas: US2013 / 0115104A1.2013 y US2013 / 0115107A1.2013 respectivamente.
- Probar otras técnicas de extracción de características, visualizadas en el Estado del Arte del Capítulo 2, por ejemplo: Momentos Invariantes, propiedades geométricas, análisis de las propiedades frecuenciales y otras técnicas de clustering.

- En concordancia del punto anterior utilizar la representación de detalle de *Wavelet* para detectar particularidades de la señal de carga, con la finalidad de extraer características para la generación de los modelos que obtuvieron un MSE alto (o  $R^2$  bajo). En el presente PFM se utilizó la representación de aproximación (gruesa) de la señal de carga.
- Variar la base *Wavelet* usada en el presente PFM: probar con Daubechies, Symmlets y Coeiflets.
- Realizar clustering al espectro obtenido usando FFT, para así obtener mejor una mejor representación del dominio de la frecuencia.
- Variar los parámetros que rige cada extractor de característica usado en el presente PFM. En efecto: variar la cantidad de centros respecto a FCM y SOM, la cantidad de términos tanto en *Wavelet* como en FD.
- Probar con otros tipos Kernel del SVR y los parámetros inherentes a cada tipo de Kernel, para verificar si se pueden obtener mejores MSE y  $R^2$ . Para este ultimo punto se puede utilizar algoritmos de búsqueda: locales y globales (monótonos y no monótonos).
- Variar los parámetros del SVR:  $C$  y  $\xi$  en un rango mas amplio, para verificar si se pueden obtener mejor MSE y  $R^2$ .
- Probar la modalidad de  $\nu$ -SVR [18], variando los parámetros descritos en los dos puntos anteriores y el parámetro  $\nu$ , con la finalidad de controlar la cantidad de SV.
- Formular el problema de clasificación en lugar de multi-clase en uno multi-etiqueta (*multi-label classification*) o en uno multi-tarea (*multi-task classification*) [23, 68].

# Bibliografía

- [1] U.S Energy Information Administration “Short-Term Energy Outlook February 2015”, United States Department of Energy. [Online]. Available: [http://www.eia.gov/forecasts/steo/report/global\\_oil.cfm](http://www.eia.gov/forecasts/steo/report/global_oil.cfm) .
- [2] BP Energy Outlook 2030, January 2013. [Online]. Available: [http://www.bp.com/content/dam/bp/pdf/statistical-review/EnergyOutlook2030/BP\\_Energy\\_Outlook\\_2030\\_Booklet\\_2013.pdf](http://www.bp.com/content/dam/bp/pdf/statistical-review/EnergyOutlook2030/BP_Energy_Outlook_2030_Booklet_2013.pdf).
- [3] B. Guo et al., Petroleum Production Engineering: A Computer-Assisted Approach. Elsevier Science & Technology Books, 2007.
- [4] Central Intelligence Agency (CIA). Crude oil - proved reserves: The World Factbook [Online]. Available: <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2244rank.html> .
- [5] Petróleos De Venezuela Sociedad Anónima (PDVSA). Informe de Gestión Anual 2012 de Petróleos de Venezuela, S.A [Online]. Available: <http://www.pdvsa.com/interface.sp/database/fichero/free/8010/1625.PDF>.
- [6] G. Takacs, Sucker-Rod Pumping Manual, PennWell Corporation, 2003.
- [7] M. Economides, et al., “Pump-Assisted Lift” in Petroleum Production Systems, Prentice Hall, 1994.
- [8] H. Partidas, “Bombeo Mecánico: Optimización, Diagnóstico y Operación”, ESP Oil Engineering Consultants, Venezuela, 2003.
- [9] W. Gilbert., “An Oil-Well Pump Dynagraph”, in API Drilling and Production Practice, New York, USA, 1936, Document ID: API-36-094.
- [10] S. Gibbs. and A.B. Neely, “Computer diagnosis of down-hole conditions in sucker rod pumping wells”, Journal of Petroleum Technology, Vol. 18, Issue 01, pp. 91-98, 1996, doi: 10.2118/1165-PA.

- [11] J.E Chacin, “A Numerical Approach to the Diagnosis of Sucker Rod Pumping Installations and Its Verification With Downhole Pump Field Measurements”, in SPE Production Operations Symposium, Oklahoma, USA, 1989, doi: 10.2118/18829-MS.
- [12] T.A Everitt and J.W. Jennings, “An Improved Finite-Difference Calculation of Downhole Dynamometer Cards for Sucker-Rod Pumps”, SPE Production Engineering, Vol. 7, pp. 121-127, 1992, doi: 10.2118/18189-PA.
- [13] Emdebian - The Universal Embedded Operating System. [Online]. Available: <http://www.emdebian.org/about/>.
- [14] Xenomai - Real-time framework for Linux. [Online]. Available: <https://xenomai.org/>
- [15] Linux. [Online]. Available: <http://www.linux.org/>
- [16] S. Mallat, “A Wavelet Tour of Signal Processing”, 2nd ed., California, USA, Academic Press, 1999.
- [17] V. Cherkassky and F. Mulier, “Support Vector Machines”, in Learning from Data: Concepts, Theory, and Methods, 2nd ed., New Jersey, USA, 2007, ch, 9, pp. 404-466.
- [18] C-C. Chang and C-J. Lin, LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [19] J. Feng et al., “Fault Diagnosis of Sucker-Rod Pumping System Using Support Vector Machine” in International Conference on Communications and Information Processing (ICCIP), Aveiro, Portugal, 2012, pp 179-189. doi: 10.1007/978-3-642-31968-6\_22.
- [20] K. Li et al., “Using the curve moment and the PSO-SVM method to diagnose downhole conditions of a sucker rod pumping unit”, Petroleum Science, Vol. 10, Issue 1, pp. 73-80, 2013. doi: 10.1007/s12182-013-0252-y.
- [21] W. Wu et al., “A Fault Diagnosis of Suck Rod Pumping System Based on SVM”, Applied Mechanics and Materials: Mechatronics and Computational Mechanics, Vol. 307, pp. 285-289, 2012. doi: 10.4028/www.scientific.net/AMM.307.285.
- [22] K. Li et al., “Fault Diagnosis for down-hole conditions of sucker rod pumping systems based on the FBH-SC method”, Petroleum Science, Vol. 12, Issue 1, pp. 135-147, 2015. doi: 10.1007/s12182-014-0006-5.
- [23] Y. Liu et al., “Weighted Task Regularization for Multitask Learning”, in IEEE 13th International Conference on Data Mining Workshops (ICDMW), Dallas, USA, 2013, pp. 399-406, doi: 10.1109/ICDMW.2013.158.

- [24] I.S. Korovin and M.V. Khisamutdinov, “Hybrid Method of Dynamograms Wavelet Analysis for Oil-Production Equipment State Identification”, *Advanced Materials Research: Manufacturing and Applied Research*, Vol. 909, pp. 252-259, 2014, doi: 10.4028/www.scientific.net/AMR.909.252.
- [25] W. Wi et al., “A Fault Diagnosis of Suck Rod Pumping System Based on Wavelet Packet and RBF Network”, *Advanced Materials Research: Manufacturing Process Technology*, Vol. 189 - 193, pp. 2665-2669, 2011, doi: 10.4028/www.scientific.net/AMR.189-193.2665.
- [26] J.A.M.F de Souza et al., “Using artificial neural networks for pattern recognition of downhole dynamometer card in oil rod pump system”, in *8th International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED)*, Cambridge, UK, 2009, pp. 230-235, ISBN: 978-960-474-051-2.
- [27] M.A.D. Bezerra et al., “Pattern recognition for downhole dynamometer card in oil rod pump system using artificial neural networks”, in *11th International Conference on Enterprise Information Systems (ICEIS)*, Milan, Italy, 2009.
- [28] J.A.M.F. de Souza et al., “Artificial Neural Networks for Pattern Recognition in Oil Rod Pump System Anomalies” in *International Conference on Artificial Intelligence (ICAI)*, Las Vegas Nevada, USA, 2009, Vol. 2.
- [29] Y. Yu et al., “Research on feature extraction of indicator card data for sucker-rod pump working condition diagnosis”, *Hindawi Journal of Control Science and Engineering*, Vol. 2013, Article ID 605749, 2013, doi: 10.1155/2013/605749.
- [30] H. Zong and Y. Zhang., “A Real-time Recognition of Working Patterns to Fault Diagnosis Based on BP Neural Network”, in *The Sixth World Congress on Intelligent Control and Automation (WCICA)*, Dalian, China, 2006, Vol. 2, pp. 5769 – 5772, ISBN: 1-4244-0332-4, doi: 10.1109/WCICA.2006.1714181.
- [31] P. Xu et al., “Application of self-organizing competitive neural network in fault diagnosis of suck rod pumping system”, *Journal of Petroleum Science and Engineering*, Vol. 58, Issues 1–2, p.p 43–48, 2007, doi:10.1016/j.petrol.2006.11.008.
- [32] F.S de Lima et al., “Comparison of Border Descriptors and Pattern Recognition Techniques Applied to Detection and Diagnose of Faults on Sucker-Rod Pumping System”, *Digital Image Processing*, InTech, 2012, doi: 10.5772/33245.
- [33] F.S de Lima et al., “A Qualitative Balance of Shape Descriptors on Dynamometers Cards”, in *16th IEEE Conference on Emerging Technologies & Factory Automation (ETFA)*, Toulouse, France, 2011, p.p 1-8, doi: 10.1109/ETFA.2011.6059085.

- [34] J.D Rogers et al., “Artificial Neural Networks for Identification of Beam Pump Dynamometer Load Cards” in SPE Annual Technical Conference and Exhibition, New Orleans, USA, 1990, p.p 23-26, doi: 10.2118/20651-MS.
- [35] G.D. Reges et al., “Application of Curvature-Based Descriptors for Fault Diagnosis in Sucker Rod Oil Pumping Systems” in 22nd International Conference on Production Research, Iguassu Falls, Brazil, 2013.
- [36] X. Wu et al., “Key Technologies of Sucker Rod Pump Card Diagnosis Based on BP Neural Network”, *Advanced Materials Research: Advanced Manufacturing Systems (ICMSE)*, Vol. 201-203, pp. 433-437, 2011. doi: 10.4028/www.scientific.net/AMR.201-203.433.
- [37] Z.J Wu et al., “Research on Automatic Diagnosis Based on ANN Well Conditions Fault”, in *International Conference on Information Science and Computer Applications (ISCA 2013)*, Changsha, China, 2013, ISBN: 978-90786-77-85-7. Available: [http://www.atlantis-press.com/php/download\\_paper.php?id=9552](http://www.atlantis-press.com/php/download_paper.php?id=9552).
- [38] F.S de Lima et al., “Application of Fourier Descriptors and Pearson Correlation for fault detection in Sucker Rod Pumping System”, in *IEEE Conference on Emerging Technologies & Factory Automation (ETFA 2009)*, Mallorca, Spain, 2009, pp. 1-4, doi: 10.1109/ETFA.2009.5347072.
- [39] F.S de Lima et al., “Detecção de falhas no Sistema de Bombeio Mecânico utilizando Descritores de Fourier e ferramentas estatísticas” in *Simpósio Brasileiro de Automação Inteligente (SBAI 2009)*, Brasília, Brazil, 2009.
- [40] T.F.C Porciuncula et al., “Reconhecimento de modos de operação em Sistemas de Bombeio Mecânico via Descritores de Fourier e PCA”, in *XX Congresso Brasileiro de Automática (CBA 2014)*, Belo Horizonte, Brazil, 2014. Available: <http://www.swge.inf.br/CBA2014/anais/PDF/1569935407.pdf>
- [41] F.S de Lima, “Detecção e Classificação de Modos de Operação do Bombeio Mecânico Via Cartas Dinamométricas” Ph.D. dissertation, Dept. Elect. and Computer Eng., Universidade Federal do Rio Grande do Norte (UFRN), Rio Grande do Norte, Brazil, 2014.
- [42] P. Schirmer et al., “Use of Advanced Pattern-Recognition and Knowledge-Based System in Analyzing Dynamometer Cards”, *SPE Computer Applications*, 1991, pp. 21-24, doi: 10.2118/22305-PA.
- [43] J. Abello et al., “A Hierarchy of Pattern Recognition Algorithms for the Diagnosis of Sucker Rod Pumped Wells”, in *Fifth International Conference on Computing and Information (ICCI)*, Sudbury, Canada, 1993.

- [44] M.A Barreto et al., “Automatic downhole card generation and classification”, in SPE Annual Technical Conference and Exhibition, Denver, USA, 1996, doi: 10.2118/36605-MS.
- [45] R. Dickinson and J. Jennings., “Use of Pattern-Recognition Techniques in Analyzing Downhole Dynamometer Cards”, SPE Production Engineering, 1990, Vol. 5, Issue 02, pp. 187-192, doi: 10.2118/17313-PA.
- [46] F. Liu and A. Patel., “Well Failure Detection for Rod Pump Artificial Lift System Through Pattern Recognition”, in International Petroleum Technology Conference, Beijing, China, 2013, doi: 10.2523/16578-MS.
- [47] L. Alegre et al., “Intelligent Diagnosis of Rod Pumping Problems”, in SPE Annual Technical Conference and Exhibition, Houston, USA, 1993, doi: 10.2118/26516-MS.
- [48] L.A. Arditi., “An Intelligent System for Diagnostic and Control to Mechanical Pumping System”, in SPE Latin America/Caribbean Petroleum Engineering Conference, Port of Spain, Trinidad, 1996, doi: 10.2118/36079-MS.
- [49] S. Sánchez y R. Maggiolo, “Fundamentos de producción por bombeo mecánico”, PDVSA CIED (Centro Internacional de Educación y Desarrollo), Primera Edición, 2002.
- [50] A. Pérez y J.C. Iglesias, “Curso de Levantamiento Artificial por Bombeo Mecánico”, PDVSA INTEVEP, 2013.
- [51] J.P. Sanchez et al., “Beam pumping system optimization through automation”, in Latin American & Caribbean Petroleum Engineering Conference, Buenos Aires, Argentina, 2007, doi: 10.2118/108112-MS.
- [52] American Petroleum Institute (API), API RP 11L: Recommended Practice for Design Calculations for Sucker Rod Pumping Systems (Conventional Units), Fourth Edition, 1st June 1988 (Reaffirmed 2000).
- [53] G.J. Salas, “Dinagramas: Manual de Interpretación”, Ediciones de la Facultad de Ingeniería, La Universidad del Zulia (LUZ), Tercera Edición, 1976.
- [54] C.M. Hirschfeldt, “Interpretación de Cartas Dinamométricas Básicas”, Oil Production: Oil & Gas Consulting, 2012.
- [55] A.K Jain., “Image Analysis and Computer Vision”, in Fundamentals of Digital Image Processing, 1st edition, New Jersey, Prentice Hall, 1989, ch. 9, sec. 9.6, pp. 370-374.
- [56] R.C. Gonzalez and R.E Woods., “Representation and Description”, in Digital Image Processing, 3rd edition, Prentice Hall, 2008, ch. 11, sec. 11.2, pp. 840 - 843.

- [57] G.H. Granlund., “Fourier Preprocessing for Hand Print Character Recognition”, IEEE Transactions on Computers, Vol. C-21, Issue 2, pp. 195 - 201, 1972, doi: 10.1109/TC.1972.5008926.
- [58] E. Persoon and K-S. Fu, “Shape Discrimination Using Fourier Descriptors”, IEEE Transactions on Systems, Man and Cybernetics, Vol. 7, Issue 3, p.p 170 - 170, 1977, doi: 10.1109/TSMC.1977.4309681.
- [59] A. V. Oppenheim and R. Schafer, “The Discrete Fourier Transform” in Discrete-Time Signal Processing, 2nd edition, ch.8, sec. 8.1, pp. 542 - 546, 1998.
- [60] V. Cherkassky and F. Mulier, “Structural Risk Minimization” and “Support Vector Machines”, in Learning from Data, 2nd edition, ch. 4 and 9, 2007.
- [61] S. Theodoridis and K. Koutroumbas, “Linear Classifiers, Clustering Algorithms III and Clustering Algorithms IV” , in Pattern Recognition, 4th edition, ch. 3, 14 and 15, 2008.
- [62] G. Pajares Martin-Sanz, J.M. De La Cruz García, y otros, “Máquinas de Vectores Soporte” en Aprendizaje Automático. Un enfoque práctico, 1era edición, ch. 5, pp.139-168, 2010.
- [63] J. Weston, “Support Vector Machine”. In: Lecture in Artificial Intelligence Class, Department of Computer Science, Columbia University, New York, USA. Available: [http://www.cs.columbia.edu/~kathy/cs4701/documents/jason\\_svm\\_tutorial.pdf](http://www.cs.columbia.edu/~kathy/cs4701/documents/jason_svm_tutorial.pdf).
- [64] C-J. Lin, “Optimization, Support Vector Machines, and Machine Learning”, Talk at DIS, University of Rome and IASI, CNR, September 2005. Available: <http://www.csie.ntu.edu.tw/~cjlin/talks/rome.pdf>.
- [65] A.J. Smola and B. Schölkopf, “A tutorial on support vector regression”, Statistics and Computing, Vol. 14, Issue 3, pp 199-222, 2004, doi: 10.1023/B:STCO.0000035301.49549.88
- [66] C-J. Lin, “Support Vector Machines”, Talk at Machine Learning Summer School, National Taiwan University, Taipei, 2006.
- [67] M. Misiti et al., “Wavelet Toolbox: User’s Guide - MATLAB 2013b”, Mathworks, 2013.
- [68] G. Tsoumakas and I. Katakis, “Multi-label classification: An overview”, International Journal of Data Warehousing and Mining 3, 1–13 (2007).

- [69] W.L. Foley and J.G Svinos, “Expert Adviser Program for Rod Pumping”, *Journal of Petroleum Technology*, Vol. 41, Issue 04, pp. 394-400, 1989, doi: 10.2118/16920-PA.
- [70] M. Vazquez and J. Fernandes, “Rod-Pumping Optimization Through Surface and Downhole Parameter Estimation and Simulation”, in *SPE Mid-Continent Operations Symposium*, Oklahoma, USA, 1999, doi: 10.2118/52214-MS.
- [71] L. Schnitman et al., “Modeling And Implementation Of A System For Sucker Rod Downhole Dynamometer Card Pattern Recognition”, in *SPE Annual Technical Conference and Exhibition*, Denver, USA, 2003, doi: 10.2118/84140-MS.
- [72] H. Sun and L. Xie., “Recognition of a Sucker Rod’s Defect with ANN and SVM”, in *International Joint Conference on Computational Sciences and Optimization (CSO)*, Sanya, China, 2009, pp. 3-6, doi: 10.1109/CSO.2009.359.
- [73] J. N McCoy et al., “Efficient and Accurate Monitoring of Rod Pump Well Performance Using Real Time Data Processing and Visualization”, in *SPE Production and Operations Symposium*, Oklahoma, USA, 2013, doi: 10.2118/164485-MS.
- [74] C. Bravo et al., “State of the Art of Artificial Intelligence and Predictive Analytics in the E&P Industry: A Technology Survey”, *SPE Journal*, Vol. 19, Issue 04, pp. 547-563, doi: 10.2118/150314-PA.
- [75] M. Vazquez., “Using Automation for Optimizing Production Fields”, in *SPE Production Operations Symposium*, Oklahoma, USA, 1995, doi: 10.2118/29534-MS.
- [76] M. Vazquez., “Real Time Expert System (R.T.E.S) for Rod Pumping Optimization”, in *SPE Petroleum Computer Conference*, Houston, USA, 1995, doi: 10.2118/30185-MS.
- [77] J. Fernandes and A. Lastra., “Experiences on Heavy and Extra-Heavy Crude Rod Pumping Automation Applications using Thermal and Dilution Techniques”, in *SPE International Thermal Operations and Heavy Oil Symposium*, Porlamar, Venezuela, 2001, doi: 10.2118/69900-MS.
- [78] K. Nolen and S. Gibbs., “Quantitative Determination of Rod-Pump Leakage With Dynamometer Techniques”, *SPE Production Engineering*, Vol. 5, Issue 03, pp. 225-230, 1990. doi: 10.2118/18185-PA.
- [79] GVL Moises et al., “Sucker Rod Pumping Failures Diagnostic System”, in *SPE Annual Technical Conference and Exhibition*, Florence, Italy, 2010. doi: 10.2118/134975-MS.