



Universidad de Valladolid



ESCUELA DE INGENIERÍAS
INDUSTRIALES

UNIVERSIDAD DE VALLADOLID

ESCUELA DE INGENIERÍAS INDUSTRIALES

Grado en Ingeniería de Organización Industrial

**“Análisis de la influencia del variador
en el diagnóstico de los fallos de motores
mediante técnicas estadísticas”**

Autor:

Rubio del Rey, Fernando

Tutores:

**Fernández Temprano, Miguel Alejandro
García Escudero, Luis Ángel**

*Departamento de Estadística e
Investigación Operativa*

Valladolid, julio 2016.

Índice de contenidos

Resumen	vii
Palabras clave	vii
Abstract	ix
Keywords	ix
1. Introducción y objetivos	1
1.1. Justificación	1
1.2. Objetivos	2
2. Motor de inducción y fallos característicos	5
2.1. Introducción	5
2.2. Características constructivas. Funcionamiento	5
2.3. Fallos en el motor de inducción	8
2.3.1. Causas de fallos en el estator	8
2.3.2. Causas de fallos en el rotor	9
2.3.3. Causas de fallos en rodamientos y eje	10
2.4. Rotura de una barra del rotor	11
2.4.1. Diagnóstico del estado del motor: MCSA	11
2.4.2. Dominio de la frecuencia: FT, PSD y FFT	14
3. Experimentación	17
3.1. Introducción y objetivo de la experimentación	17
3.2. Equipo empleado	17
3.3. Descripción de la experimentación	22
3.4. Características medidas y notación	24
3.4.1. Identificadores del ensayo	24
3.4.2. Variables registradas	26
4. Técnicas estadísticas empleadas	29
4.1. El Análisis de Componentes Principales	29
4.1.1. Obtención de las componentes principales	29
4.1.2. Representación de los datos	30
4.1.3. Aplicaciones del ACP	32

4.2. El Análisis Discriminante Lineal	33
4.2.1. Clasificación en dos poblaciones	33
4.2.2. Clasificación en más de dos poblaciones.....	35
4.2.3. Métodos de selección de las variables	36
4.2.4. Validación del modelo	38
4.3. El Modelo Aditivo Generalizado	40
4.3.1. Modelo Aditivo	40
4.3.2. Modelo semiparamétrico	42
4.3.3. Modelo Aditivo Generalizado	42
5. Análisis descriptivo de los datos	43
5.1. Análisis descriptivo inicial	44
5.2. Correlación de variables.....	50
5.3. Normalidad de las variables	53
5.4. Análisis de la Varianza.....	55
5.5. Análisis de Componentes Principales	58
6. Análisis Discriminante	63
6.1. Análisis Discriminantes preliminares	63
6.2. Análisis Discriminantes finales	69
6.2.1. Análisis Discriminantes representativos: caso M1_UT.....	70
6.2.2. Resumen y conclusiones	74
6.3. Validación del Análisis Discriminante.....	80
6.3.1. Validación discriminación C1-C234	81
6.3.2. Validación discriminación C1-C2-C3-C4	83
6.3.3. Conclusiones sobre la validación	84
7. Análisis del efecto del tipo y del deslizamiento.....	85
7.1. Modelo Aditivo Generalizado para cada variable	86
7.1.1. GAM sobre variables tipo pico.....	86
7.1.2. GAM sobre variables tipo amplitud	88
7.2. Resumen y conclusiones.....	90
8. Análisis Discriminante con residuales del GAM	93
8.1. Análisis Discriminantes GAM preliminares	94
8.2. Análisis Discriminantes GAM finales	96

8.2.1. Análisis Discriminantes GAM representativos: caso M1_UT ...	97
8.2.2. Resumen y conclusiones.....	99
8.3. Validación del Análisis Discriminante GAM	103
8.3.1. Validación discriminación C1-C234.....	103
8.3.2. Validación discriminación C1-C2-C3-C4.....	104
8.3.3. Conclusiones sobre la validación	105
8.4. Comparación entre análisis discriminantes: variables originales o residuales del GAM.....	106
9. Influencia de la condición en el efecto del deslizamiento.....	111
9.1. Modelos Aditivos Generalizados de relevancia.....	112
9.2. Resumen y conclusiones	118
10. Conclusiones	121
A. Banco de ensayos.....	125
A.1. Motores comerciales.....	125
A.2. Variadores de frecuencia	126
A.3. Freno y Unidad de Control	127
A.4. Sensores	127
A.5. Sistema de adquisición.....	128
A.6. Sensor de velocidad	129
B. Análisis complementarios.....	131
B.1. Análisis Discriminantes complementarios	131
B.1.1. Discriminación M1_TE.....	131
B.1.2. Discriminación M1_LN	135
B.1.3. Discriminación M2_UT.....	138
B.1.4. Discriminación M2_AB	142
B.1.5. Discriminación M2_SM	145
B.2. GAM complementarios.....	151
B.3. Análisis Discriminantes complementarios con residuales del GAM.....	162
B.3.1. Discriminación M1_TE.....	162
B.3.2. Discriminación M1_LN	164
B.3.3. Discriminación M2_UT.....	166
B.3.4. Discriminación M2_AB	169
B.3.5. Discriminación M2_SM	171

B.4. GAM complementarios según la condición	173
C. Código de programación en R	179
C.1. Código método de validación <i>bootstrap</i>	179
C.2. Código GAM	181
Bibliografía	183

Resumen

El objetivo más importante del presente trabajo es establecer una metodología de diagnóstico de fallos en motores de inducción de jaula de ardilla mediante técnicas estadísticas. Para ello, se empleó una base de datos ya existente sobre una experimentación que consistió en registrar la señal de la corriente de alimentación de dos motores en varios estados de fallo, introducidos de forma intencionada. Se distinguieron cuatro estados según la severidad del fallo: motor sano, fallo leve, fallo medio y motor con ruptura total de una barra del rotor. Además, se emplearon varios tipos de alimentación y niveles de carga, por lo que se pretende determinar su influencia en el diagnóstico.

Las principales técnicas estadísticas abordadas en este trabajo fueron el Análisis Discriminante Lineal, para comprobar su eficacia en la predicción del estado del motor, y el Modelo Aditivo, para estudiar la posible influencia del deslizamiento en el valor de las variables registradas. También se comparó la idoneidad de los residuales derivados de los Modelos Aditivos Generalizados planteados frente a las variables originales, como variables explicativas.

Palabras clave

Motor de inducción
Análisis discriminante
Modelo aditivo
Diagnóstico de fallos
Análisis del espectro
Análisis estadístico
Mantenimiento predictivo
Variador de frecuencia

Abstract

The priority of this paper is to establish a methodology for fault diagnosis in squirrel-cage induction motors using statistical techniques. To this end, it was used an existing database of an experimentation which consisted of recording the current signal of two motors in several different fault states, introduced intentionally. Four states were distinguished according to the severity of the fault: healthy motor, small failure, medium failure and fully broken bar motor. In addition, different types of supplies and load levels were employed, so it was necessary to determine their influence.

The main statistical techniques used in this work were the Linear Discriminant Analysis, to test its effectiveness in predicting the motor condition, and the Additive Model to study the possible influence of the slip in the value of the recorded variables. The suitability of the residuals derived from Generalized Additive Models against the aptitude of the original variables was also compared, as explanatory variables.

Keywords

Induction motor
Discriminant analysis
Additive model
Fault diagnosis
Spectral analysis
Statistical analysis
Preventive maintenance
Frequency inverter

Capítulo 1

Introducción y objetivos

1.1. Justificación

A medida que la tecnología ha ido evolucionando, los motores eléctricos se han convertido en elementos fundamentales de toda actividad industrial. Ponen en marcha el sistema productivo, posibilitando el desarrollo de todo tipo de aplicaciones. Resulta de vital importancia optimizar su rendimiento y reducir al máximo la aparición de defectos que provoquen un comportamiento incorrecto.

Los motores de inducción de jaula de ardilla son los más empleados y presentan un índice de averías muy reducido. Pese a ello, el continuo funcionamiento del motor hace inevitable su deterioro a largo plazo. El mecanismo comienza a presentar deficiencias físicas que revierten en el rendimiento en un determinado momento de su ciclo de vida, poniendo en peligro la eficiencia del sistema en el que esté operando. Para evitarlo, el mantenimiento preventivo tradicional, consistente en el cambio programado de piezas susceptibles de sufrir una avería, ha ido evolucionando hacia el concepto de mantenimiento predictivo. Este radica en la supervisión y control de ciertos parámetros de funcionamiento, que determinarán el momento adecuado de intervención cuando es probable la futura aparición de la avería.

Surge así la necesidad de promover la investigación y el desarrollo de herramientas capaces de detectar fallos, cuando estos se encuentren en estado incipiente. Entre algunas de las técnicas de detección más comúnmente empleadas encontramos el análisis de las vibraciones, el análisis de la temperatura y la técnica MCSA (*Motor Current Signature Analysis*). Esta última trata de analizar la corriente del estator y presenta como ventaja principal frente a otros métodos su carácter no invasivo, sin necesidad de que el sistema deba cesar su funcionamiento.

2 INTRODUCCIÓN Y OBJETIVOS

Existe gran variedad de averías posibles en los motores de inducción. Una de ellas es la rotura de barras del rotor, la cual provoca la asimetría del mismo. Como consecuencia, el campo magnético del motor se ve afectado y finalmente el fallo se manifiesta a través de la corriente de alimentación del motor, en forma de bandas alrededor del armónico fundamental, así como alrededor de ciertos armónicos enteros superiores.

Para descubrir la aparición de dichas bandas, se debe analizar la respuesta en frecuencia de la corriente de alimentación. El denominado *Power Spectral Density* (PSD) o Densidad de Potencia Espectral, basado en la transformada de Fourier, es una técnica ideal para el propósito deseado, ya que destaca la amplitud de las bandas frente al armónico principal y los enteros superiores de interés.

El papel de los variadores de frecuencia en la medición de la respuesta en frecuencia descrita es también muy a tener en cuenta, ya que introducen un ruido elevado que dificulta la identificación de las bandas. Los variadores de frecuencia son dispositivos utilizados para suministrar la energía de forma adecuada a un motor de corriente alterna. Permiten controlar su velocidad rotacional gracias al control de la frecuencia de alimentación al motor, contribuyendo a la reducción del consumo de electricidad y, consecuentemente, del coste de operación de los sistemas accionados por motores eléctricos. El uso de estos elementos está cada vez más extendido y existe una gran variedad de tipos de variadores dependiendo de la aplicación a la cual el motor esté destinado.

1.2. Objetivos

Se ha mencionado una técnica para la detección de la rotura de barras del rotor y la importancia de factores como el variador de frecuencia en la propia detección. Ahora bien, la pretensión final de todo sistema de mantenimiento es disponer de una metodología capaz de predecir de forma fiable el estado del motor.

En este contexto, el presente Trabajo de Fin de Grado tiene como objetivo principal desarrollar, implementar y validar una metodología capaz de diagnosticar fallos en motores de inducción a partir de técnicas estadísticas de predicción. Asimismo, se pretende que la metodología tenga la flexibilidad y robustez suficiente como para poder ser empleada independientemente del motor y del tipo de alimentación implicados.

Para abordar el propósito descrito, se empleará una base de datos ya existente proporcionada por el Departamento de Ingeniería Eléctrica de la Escuela de Ingenierías Industriales de la Universidad de Valladolid. La base de

datos consiste en mediciones sobre la señal de dos motores de inducción sobre los cuales se introducía, intencionadamente y de forma gradual, una avería en una barra del rotor. Así pues, se dispone de ensayos en cuatro estados de fallo: motor sano, fallo leve, fallo medio y barra rota. Además, fueron empleados diversos variadores de frecuencia, aparte de la propia alimentación de red, con la intención de establecer una regla general aplicable a todos los casos.

El tratamiento de la información y el desarrollo de los análisis pertinentes requerirán el uso de dos herramientas *software* extensamente conocidas en el ámbito de la Estadística: *StatGraphics* y **R**.

Capítulo 2

Motor de inducción y fallos característicos

2.1. Introducción

Los motores eléctricos son máquinas eléctricas que transforman en energía mecánica la energía eléctrica que reciben a través de sus bornes. Se trata de elementos indispensables en toda actividad industrial.

En la actualidad, mientras que los motores de corriente continua (CC) y los motores síncronos de corriente alterna (CA) tienen aplicaciones muy específicas, los motores asíncronos de CA o motores asíncronos de inducción, tanto monofásicos como trifásicos, son los motores eléctricos de uso más generalizado. Aproximadamente el 80% de los motores eléctricos industriales son de este tipo, convirtiéndose en los principales consumidores de energía eléctrica de la industria (Rosa, 2015).

Entre las razones de su éxito, encontramos su facilidad de utilización, su escasa necesidad de mantenimiento y su reducido coste de fabricación, características que les hacen especialmente atractivos.

2.2. Características constructivas. Funcionamiento

Los motores objeto de nuestro estudio son motores asíncronos trifásicos de inducción, un tipo de motor eléctrico de CA. Como todo motor eléctrico (véase la Fig. 2.1.a), están formados, desde un punto de vista constructivo, por un estator y un rotor (véase la Fig. 2.1.b). El rotor es la parte móvil que gira en el interior del estator, que constituye la parte fija.

El motor dispone de un circuito magnético formado por láminas apiladas en forma de anillo en el estator y en forma de cilindro en el rotor. En unas

ranuras dispuestas en la superficie interior del circuito magnético laminado del estator, se sitúa un bobinado inductor (véase la Fig. 2.2). El rotor se coloca en el interior del estator y se le dota de un entrehierro constante lo más pequeño posible, permitiendo su giro con libertad.



Figura 2.1. A la izquierda (a), motor eléctrico exteriormente; a la derecha (b), dos conjuntos rotor-estator de motor eléctrico.

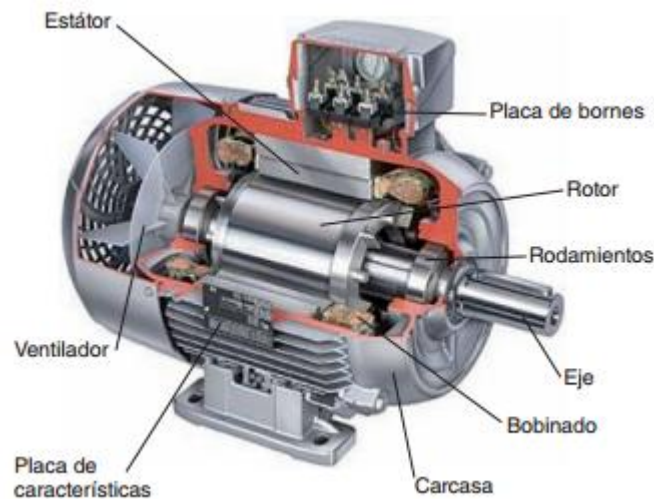


Figura 2.2. Sección de motor eléctrico.

Existen dos tipos principales de rotor según su construcción: rotor bobinado y rotor en cortocircuito (véase la Fig. 2.3). En el rotor bobinado, el circuito eléctrico (bobinado inducido) se coloca sobre ranuras con las que se ha dotado previamente al circuito magnético laminado del rotor. La variante de rotor en cortocircuito o rotor de jaula de ardilla, por su parte, incorpora conductores de gran sección adosados a los anillos laterales.

A nivel industrial, la variante de rotor en cortocircuito es la más utilizada. El motor, de este modo, presenta una construcción simple, tamaño y peso reducido, bajo coste, robustez y escasa necesidad de mantenimiento. El rotor en cortocircuito es comúnmente conocido como rotor de jaula de ardilla, debido a su similitud con esta.

Muchos de los elementos que constituyen el motor de jaula de ardilla son susceptibles de fallo. Entre otros, destacamos los rodamientos, el eje, el

circuito magnético, el circuito eléctrico del estator, el circuito eléctrico del rotor, el bornero de conexiones, etc. La Fig. 2.4 muestra esquemáticamente los elementos del motor. En el apartado 2.3 se expondrán algunos fallos comunes provocados por deficiencias constructivas de algún elemento del motor.

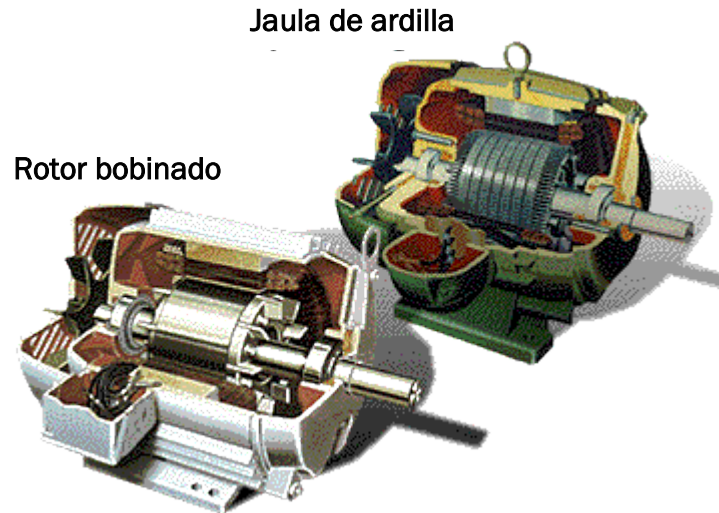


Figura 2.3. Sección motor de rotor bobinado y motor de jaula de ardilla.

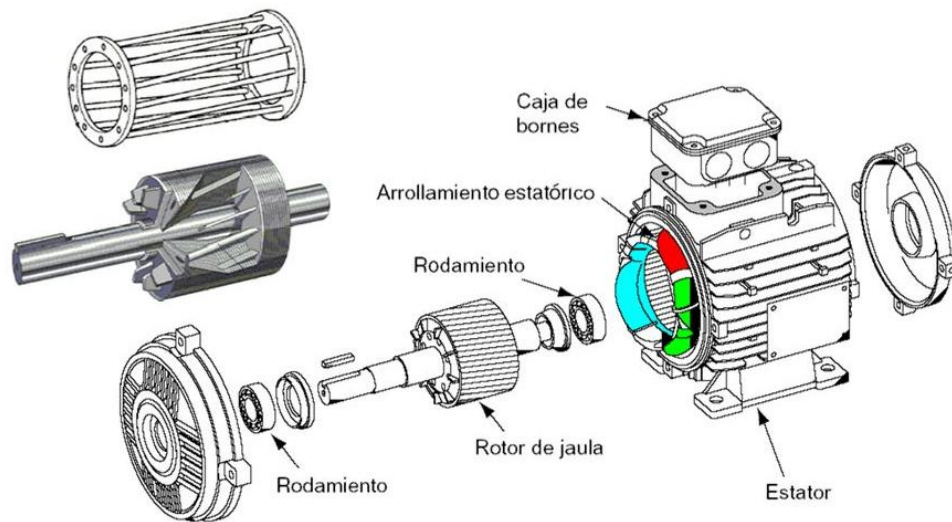


Figura 2.4. Despiece de motor de jaula de ardilla.

El motor funciona siguiendo el principio de inducción mutua de Faraday. Cuando se aplica corriente alterna trifásica al bobinado situado en el estator, se genera un campo magnético giratorio en el entrehierro. Este gira a la denominada velocidad de sincronismo n_{sinc} , condicionada por la frecuencia de alimentación f y por el número de pares de polos de la máquina p , según la ecuación

$$n_{sinc} = \frac{60f}{p},$$

con n_{sinc} en rpm y f en Hz.

El campo magnético induce un sistema de corrientes sobre el rotor. Como consecuencia de la acción de estas y del campo magnético giratorio, se produce a su vez otro campo magnético que sigue el movimiento del campo estático. De este modo, se crea un par motor responsable de que el rotor gire. La velocidad de rotación será inferior a la velocidad de sincronismo y a través de su diferencia se define un parámetro adimensional denominado deslizamiento como

$$s = \frac{n_{sinc} - n}{n_{sinc}}, \quad (2.1)$$

con n_{sinc} y n en la misma unidad de medida.

2.3. Fallos en el motor de inducción

Como consecuencia de continuos esfuerzos a los que están sometidos los motores de inducción, la industria se enfrenta a fallos inesperados y a la reducción del tiempo de vida de esos motores, a pesar de su robustez y fiabilidad (Duque, García, Moríñigo, Gardel y Pérez, 2015). Los fallos pueden agruparse en tres clases, según el elemento afectado del motor: fallos en el estator, en el rotor o en rodamientos y eje.

El fallo objeto de estudio en el presente TFG es la rotura de una de las barras del rotor (fallo en el rotor). No obstante, en este apartado se explicarán y enumerarán de forma muy resumida los fallos más comunes de los motores de inducción de jaula de ardilla. En el apartado 2.4, por su parte, se profundizará en el caso específico de la rotura de barras del rotor.

2.3.1. Causas de fallos en el estator

El bobinado trifásico del estator tiene un aislamiento eléctrico que es el componente más propenso a sufrir un fallo debido a diversos esfuerzos, entre los cuales destacan los térmicos, los eléctricos, los mecánicos y los medioambientales (Gardel, 2013).

- Esfuerzos térmicos

La clase térmica del aislamiento restringe la temperatura límite de funcionamiento del motor. Si esta se supera, el motor sufre un proceso de envejecimiento a través del cual pierde poco a poco sus

propiedades dieléctricas y puede derivar incluso en un cortocircuito. Existe gran cantidad de factores que pueden provocar un incremento de la temperatura, desde funcionamientos incorrectos del motor (sobrecargas, variaciones de tensión, desequilibrio de fases) hasta una deficiente ventilación del mismo.

- Esfuerzos eléctricos

El dieléctrico puede fallar debido a fenómenos de carácter eléctrico, entre los que mencionamos: descargas parciales, efecto corona, transitorios de tensión y efecto “*tracking*”.

- Esfuerzos mecánicos

Las oscilaciones a las que son sometidas las bobinadas ubicadas en las ranuras del estator pueden dañar el aislamiento de la bobina por rozamiento.

- Esfuerzos medioambientales

La humedad, la contaminación o partículas suspendidas en el aire contribuyen a dañar el aislamiento por rozamiento.

2.3.2. Causas de fallos en el rotor

El rotor sufre el mismo tipo de esfuerzos que el estator. Los problemas suelen estar relacionados con el sobrecalentamiento del rotor y con las fuerzas centrífugas que deben soportar las barras en sus extremos y las conexiones con el anillo en cortocircuito, especialmente en situaciones de arranque, parada o variaciones bruscas de carga.

- Esfuerzos térmicos

Para evitar calentamientos excesivos del rotor, se deben respetar las características de funcionamiento transitorio indicadas por la clase de servicio del motor.

- Esfuerzos dinámicos

Los motores están diseñados para trabajar en un rango de velocidades determinado, que no debe ser superado, para evitar pares a los que el motor no sea capaz de hacer frente. Por otro lado, los esfuerzos cíclicos pueden originar fallos por fatiga.

- Esfuerzos magnéticos

El motor ideal presenta su rotor centrado en el entrehierro. Sin embargo, en el motor real el rotor no se encuentra perfectamente centrado con respecto al entrehierro, sino que existe un cierto nivel de excentricidad. De este modo, aparece una fuerza resultante de origen magnético que, a la larga, puede originar que rotor y estator entren en contacto como consecuencia de la reducción del entrehierro.

- Esfuerzos mecánicos

Existen muchos fallos que pueden ser asignados a la categoría de fallos mecánicos. Entre los más comunes, se encuentran: porosidad de la fundición, laminación suelta, piezas fatigadas o rotas, pérdida del entrehierro, eje del rotor torcido, fallos de los cojinetes, defecto de alineación, materiales incorrectos y resonancia de los dientes.

- Esfuerzos medioambientales

Del mismo modo que para el estator, la humedad y ciertos agentes químicos pueden deteriorar la jaula. Además, la existencia de partículas en el entorno de funcionamiento puede atascar el sistema de ventilación.

2.3.3. Causas de fallos en rodamientos y eje

Los rodamientos son elementos que requieren su sustitución sistemática después de un determinado tiempo de vida, debido a la fatiga del material. Entre las averías más habituales relacionadas con los rodamientos, se citan las siguientes:

- Deficiente montaje
- Carga excesiva
- Lubricación incorrecta, por defecto o exceso
- Contaminación externa
- Diseño inadecuado
- Exposición a vibraciones, con el motor parado
- Paso de corriente eléctrica

Respecto a las causas de fallos en el eje, cabe destacar la corrosión, las fracturas, las sobrecargas y la fatiga, siendo esta última la causa más frecuente.

2.4. Rotura de una barra del rotor

El desarrollo del presente TFG centra su atención en un tipo concreto de fallo de los motores de inducción de jaula de ardillas: el agrietamiento o rotura de una de las barras del rotor. Se trata de una avería especialmente común en motores de grandes potencias (Soto y Torre, 2007).

Cuando el motor arranca, una corriente elevada circula por las barras debido a que la velocidad del rotor es muy inferior a la velocidad de sincronismo. Las altas corrientes causan un calentamiento y una expansión de las barras con relación al rotor, pero de forma heterogénea afectando más a unas que a otras. Esto desemboca en un agrietamiento de las juntas a través de las cuales las barras se unen al anillo de cortocircuito. La aparición de una grieta incrementa la resistencia de la barra y, consecuentemente, su temperatura, por lo que poco a poco la grieta se va haciendo de mayor tamaño (Rosa, 2015). Además, las fisuras o fracturas pueden propagarse a barras vecinas, debido a que estas se ven obligadas a conducir corriente extra (Duque et al., 2015), o incluso dañar los devanados de estator cuando se producen deformaciones (Verucchi y Acosta, 2007).

La grieta se inicia normalmente en los puntos de la barra próximos a la unión con el anillo de cortocircuito o incluso en este. Se trata de la zona que soporta mayores tensiones mecánicas.

2.4.1. Diagnóstico del estado del motor: MCSA

La rotura de las barras del rotor de un motor de inducción es un problema grave para la industria, ya que la solución exige la reparación del motor, para lo cual es indispensable la interrupción de la producción. Por ello, la preocupación creciente sobre este fenómeno ha promovido la investigación sobre la implantación de un sistema de mantenimiento predictivo basado en técnicas de diagnóstico que permitan determinar el estado del motor. De este modo, la supervisión de la máquina permite repararla justo cuando el fallo o potencial fallo se manifiesta (Villada, Parra y Ocampo; 2007).

La técnica denominada *Motor Current Signature Analysis* (MCSA) es, en la actualidad, una de las técnicas preferidas en el ámbito del mantenimiento preventivo de motores de inducción, debido a las ventajas que ofrece con respecto a otros métodos por su carácter no invasivo. Se centra en el análisis espectral de las corrientes del estator. Aunque también permite diagnosticar averías eléctricas o mecánicas, la aplicación fundamental del MCSA es la detección de la rotura de barras del rotor.

En la realidad, la disposición de los devanados en las ranuras no es perfectamente sinusoidal, la fabricación del motor origina imperfecciones o irregularidades y la fuente de alimentación presenta componentes armónicas. Todo ello contribuye a que las máquinas de inducción contengan generalmente armónicos que pueden ser detectados en su respuesta en frecuencia a través del MCSA (Verucchi y Acosta, 2007).

Ahora bien, los armónicos se hacen más evidentes cuando el motor tiene rota alguna de las barras del rotor. De momento, vamos a exponer el comportamiento de la respuesta en frecuencia asumiendo el caso de una tensión de alimentación sinusoidal, pura y equilibrada. Como consecuencia de la asimetría de la jaula ante la rotura de una de sus barras, se inducen voltajes y corrientes en el devanado del estator, a las frecuencias

$$f = (1 \pm 2s)f_1,$$

donde f_1 es la frecuencia de alimentación del motor en Hz y s es el deslizamiento (Duque y Pérez, 2005). A la frecuencia $(1 - 2s)f_1$ se la conoce como banda lateral inferior del armónico (*lower-sideband harmonic*, LSH), mientras que a la frecuencia $(1 + 2s)$ se la conoce como banda lateral superior del armónico (*upper-sideband harmonic*, USH).

Para comprender mejor el significado del fenómeno explicado, la Fig. 2.5 ilustra un ejemplo representativo. La Fig. 2.5.a muestra el espectro en frecuencias de las corrientes de una fase de un motor considerado sano; la Fig. 2.5.b muestra el espectro del mismo motor con varias barras rotas. Como se puede apreciar, la diferencia de las bandas con respecto al armónico fundamental ha disminuido considerablemente, de 46 dB a 26 dB, lo cual permite distinguir entre un motor sano y uno averiado. Este será el fundamento de nuestros estudios.

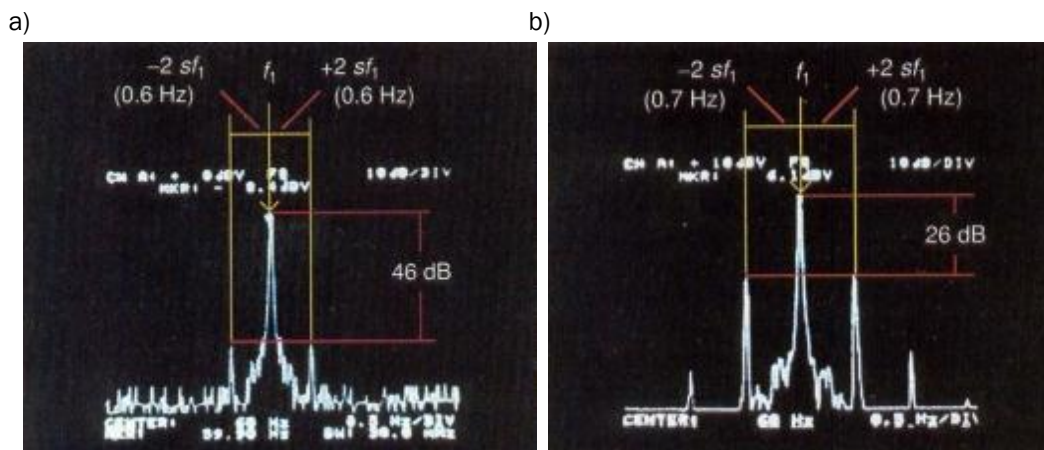


Figura 2.5. A la izquierda (a), espectro de la corriente de un motor sano; a la derecha (b), espectro de la corriente del mismo motor, pero esta vez averiado como consecuencia de la rotura de varias de las barras de su rotor.

Por otro lado, tiene especial interés para el desarrollo del trabajo que nos ocupa la respuesta en frecuencia de los motores alimentados por inversores de frecuencia. Estos introducirán armónicos adicionales, que generarán nuevos entrehierros o modificarán la amplitud de los existentes. Además, introducen mucho ruido a lo largo de todo el espectro. En conclusión, además del espectro característico del motor, serán introducidos nuevos armónicos relacionados con el grado del fallo, la carga o el tipo de alimentación. Por lo tanto, el espectro de la corriente es afectado por muchos factores.

Hasta el momento se ha hecho referencia al caso de un voltaje de alimentación sinusoidal, puro y equilibrado. Cuando un motor con jaula asimétrica es alimentado por un voltaje de alimentación no sinusoidal, el proceso es muy similar al descrito. En esta ocasión aparece un par de bandas alrededor del armónico a frecuencias de

$$f = (k \pm 2ns)f_1,$$

donde k es el orden del armónico y n es cualquier entero positivo (García, Duque, Morínigo y Pérez; 2011). Así pues, por un lado, a múltiplos enteros de la frecuencia fundamental aparecen armónicos, aunque su amplitud disminuye a medida que se alejan de dicha frecuencia fundamental, tal y como ilustra la Fig. 2.6. Por otro lado, alrededor de cada uno de esos armónicos se manifiestan pares de bandas laterales, cuya magnitud también disminuye a medida se alejan del armónico, tal y como refleja la Fig. 2.7.

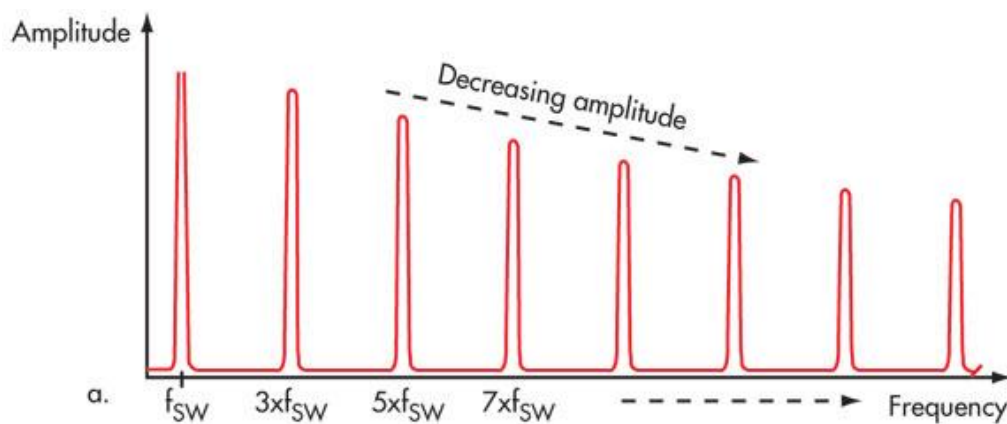


Figura 2.6. Gráfico representativo del decremento característico de la amplitud del armónico a medida que este se aleja del fundamental.

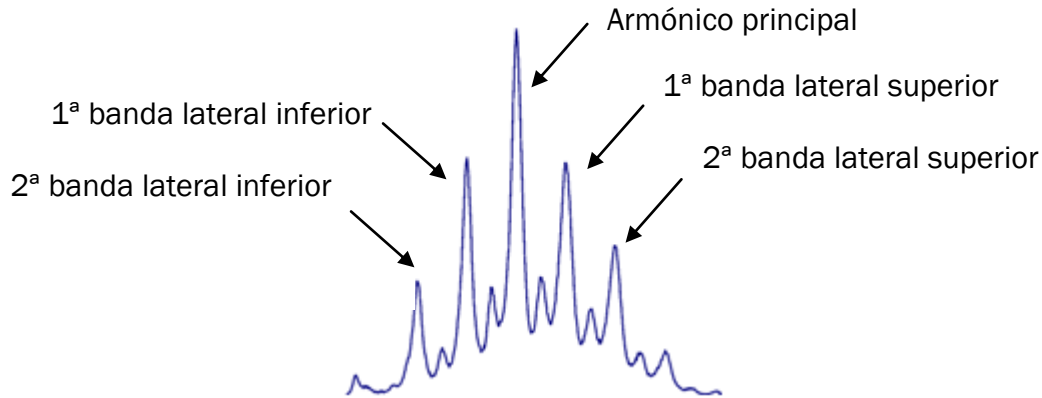


Figura 2.7. Armónico principal y dos pares de bandas laterales a su alrededor.

En nuestro estudio particular, interesa analizar la respuesta en frecuencia en torno a los armónicos fundamental, quinto y séptimo, que son los más representativos en el ámbito de la corriente estatórica. El tercer armónico no es de interés. Como toda la experimentación se realizó a 50 Hz de frecuencia de alimentación, el quinto y el séptimo armónico se presentan a 250 y a 350 Hz respectivamente. Cada uno de ellos está acompañado de sus correspondientes pares de bandas laterales, aunque sólo se centra la atención en el primer par de bandas, el más significativo.

2.4.2. Dominio de la frecuencia: FT, PSD y FFT

El MCSA se basa en el análisis espectral de la corriente del estator, una técnica de medición en línea que permite al motor seguir operando de forma continua (Fernández, García, Alonso, Cano y Solares; 2000). Para obtener este es necesario registrar las señales en el dominio temporal y hacer la transformación conveniente al dominio de la frecuencia. La Transformada de Fourier (FT) permite llevar a cabo esta tarea. La transformación de la respuesta temporal a la respuesta en frecuencia se define a través de la ecuación

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi ft} dt,$$

donde $x(t)$ es la representación en el tiempo de la señal x , $X(f)$ es la representación en la frecuencia de la señal x y $j = \sqrt{-1}$.

La señal $x(t)$ debe ser estacionaria para el correcto desarrollo de la transformación. Una señal es estacionaria siempre que sus componentes en el dominio de la frecuencia existan durante todo el tiempo en el cual se define la propia señal.

Por otro lado, el denominado *Power Spectral Density* (PSD) es una función que ofrece información sobre la distribución en potencial de la señal $x(t)$ en el dominio de la frecuencia. Se trata de una técnica cuya aplicación principal es el análisis espectral de señales aleatorias o señales periódicas contaminadas con ruido. Se define a través de la expresión

$$|X(f)|^2 = X(f)X^*(f),$$

donde $|X(f)|^2$ se mide en V^2/Hz .

El procesamiento de los datos para el cálculo de la FT exige un gran número de operaciones matemáticas. En este contexto, se desarrolló un algoritmo que recibió el nombre de Transformada Rápida de Fourier (FFT) mediante el cual el cálculo era mucho más rápido y eficiente. Esta será la técnica empleada en la experimentación objeto de nuestro estudio.

Capítulo 3

Experimentación

3.1. Introducción y objetivo de la experimentación

El experimento fue realizado por profesores del Departamento de Ingeniería Eléctrica de la Escuela de Ingenierías Industriales de la Universidad de Valladolid en el Laboratorio del Departamento situado en la Sede Paseo del Cauce. Se trata de una serie de ensayos a los que son sometidos dos motores de inducción de jaula de ardilla de corriente alterna (CA), sobre los cuales se va introduciendo una anomalía de forma gradual. Esta consiste en la perforación de una de las barras del rotor. Cada ensayo tiene una duración de 10 segundos, aunque esta podría haber sido reducida sensiblemente sin influir en los resultados.

El objetivo de la experimentación consiste en llevar a cabo mediciones acerca del funcionamiento real del motor bajo distintas situaciones dependiendo del motor utilizado, la severidad del defecto introducido intencionadamente, el variador de frecuencia empleado (o la ausencia del mismo) y el nivel de carga del motor. De este modo, los datos obtenidos servirán para establecer una pauta que permita conocer el estado físico del motor mediante el simple análisis de la respuesta obtenida, sin necesidad de parar su actividad.

3.2. Equipo empleado

A continuación se presentan de forma general los elementos y dispositivos empleados durante la experimentación. Para mayor información sobre los parámetros nominales, datos técnicos y características de los mismos véase el Anexo A.

- Motor comercial

Se emplean dos motores de inducción electromagnética de CA con estructura de jaula de ardilla (véase la Fig. 3.1). Ambos son motores comerciales de la marca *Siemens*. Sus rotores son sometidos a modificaciones para reproducir la avería objeto de estudio. A efectos prácticos, la principal diferencia entre ambos motores es la potencia nominal. Mientras que uno la tiene de 750 W (motor M1), el otro alcanza una de 1.100 W (motor M2).

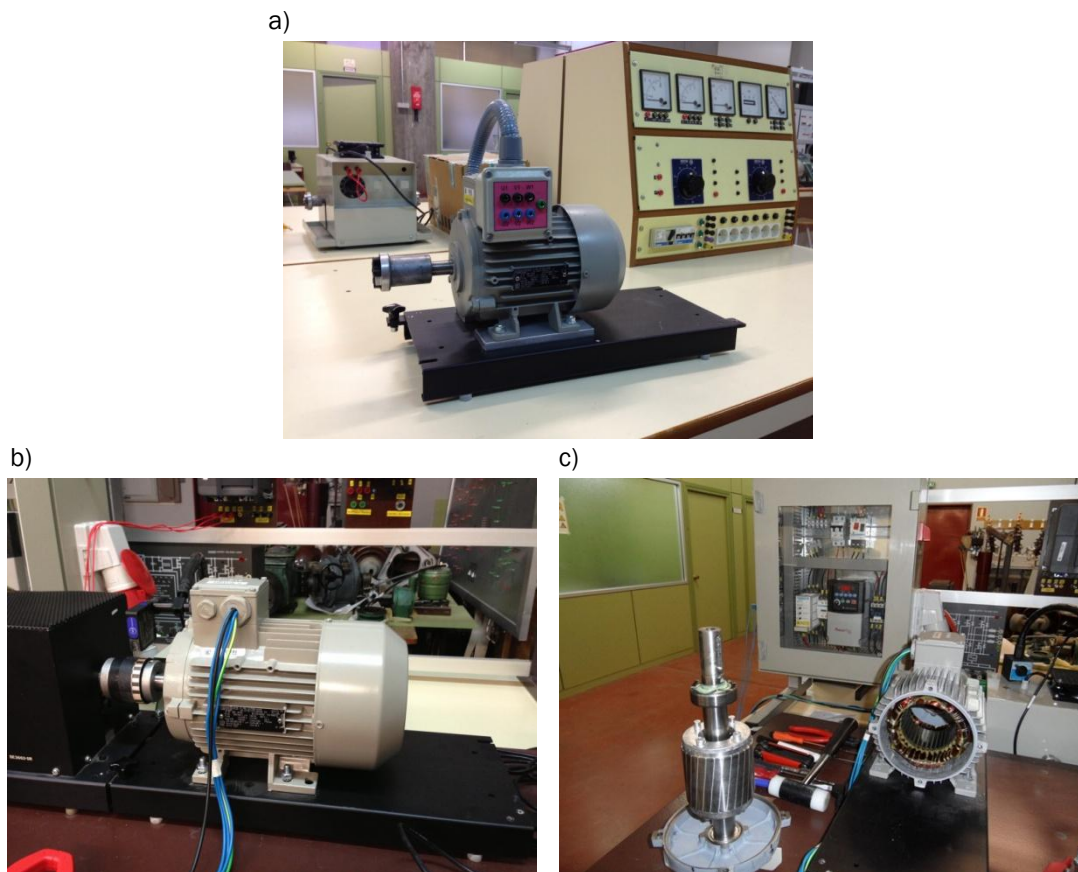


Figura 3.1. Arriba (a), motor de 750 W de *Siemens* (motor M1); abajo, motor de 1.100 W de *Siemens* (motor M2) conectado al freno (b) y desarmado (c).

- Variador de frecuencia

Los variadores de frecuencia o inversores se emplean como fuente de alimentación alternativa a la suministrada directamente por la red, de 50 Hz y 400 V de tensión de línea. A lo largo de la experimentación se utilizan seis variadores comerciales de distintas marcas:

- Variador *PowerFlex 40* de *Allen-Bradley* (véase la Fig. 3.2.a).
- Variador *MicroMaster 420* de *Siemens* (véase la Fig. 3.2.b).
- Variador *Altivar 66* de *Telemecanique* (véase la Fig. 3.2.c).
- Variador de *Lucas-Nülle*.

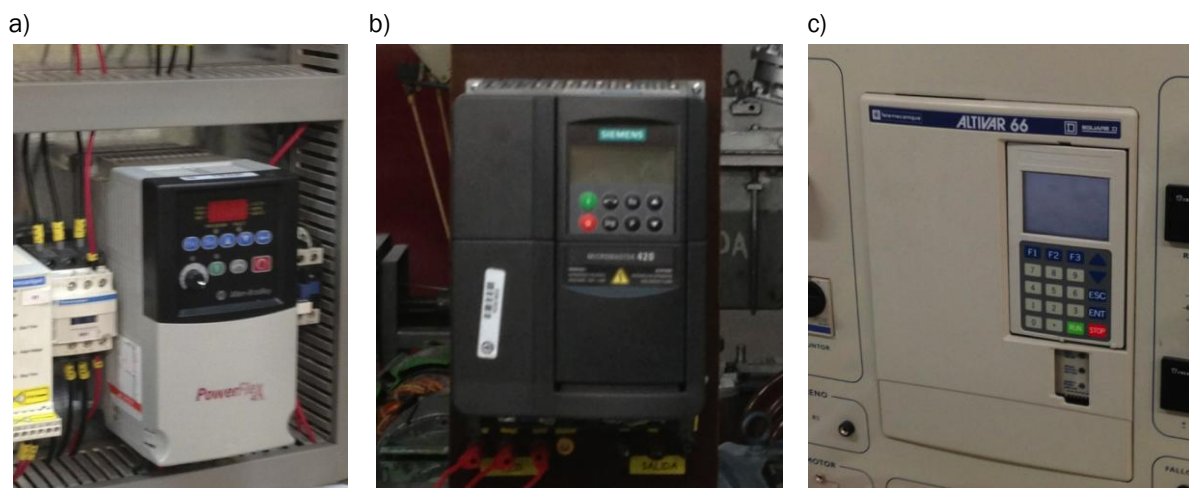


Figura 3.2. Variadores de frecuencia comerciales: *PowerFlex 40* de *Allen-Bradley* (a), *Micromaster 420* de *Siemens* (b) y *Altivar 66* de *Telemecanique* (c).

- Freno y Unidad de Control

Con la intención de cargar el motor y simular un par resistente constante, se utiliza un freno de polvo magnético de *Lucas-Nülle*, representado en la Fig. 3.3. Va acompañado de una Unidad de Control de la misma marca, de tal forma que se pueda seleccionar el valor del par resistente. Los ensayos de la experimentación que nos concierne se llevan a cabo en condiciones estacionarias de carga del motor, es decir, el par resistente ejercido no varía a lo largo de la prueba.



Figura 3.3. Freno de polvo magnético y Unidad de Control de *Lucas-Nülle*.

- Sensores

La medición de las magnitudes deseadas se lleva a cabo mediante transductores o sensores de diversa tipología. Todos ellos cuentan con empaquetamiento PCB y se montaron sobre dos placas de diseño propio (véase la Fig. 3.4) construidas en el propio laboratorio.

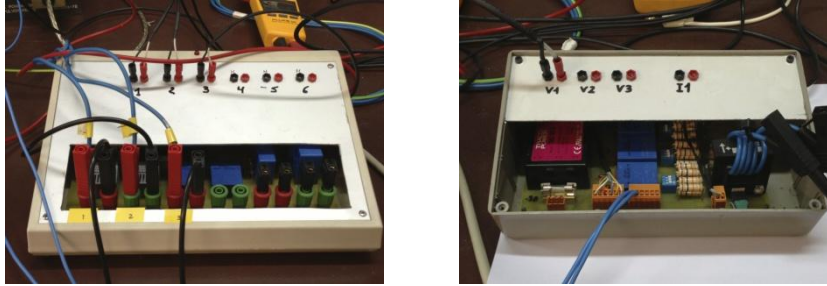


Figura 3.4. Tarjetas dedicadas con los transductores de efecto Hall de tensión y corriente.

Distinguimos entre dos tipos de sensores o transductores: de corriente y de tensión.

- Transductores de corriente

Estos transductores miden la intensidad de corriente que circula por el motor. Para nuestra aplicación, se utilizan dos transductores de *Honeywell* y un tercero de *LEM* (véase la Fig. 3.5), todos basados en el efecto Hall.

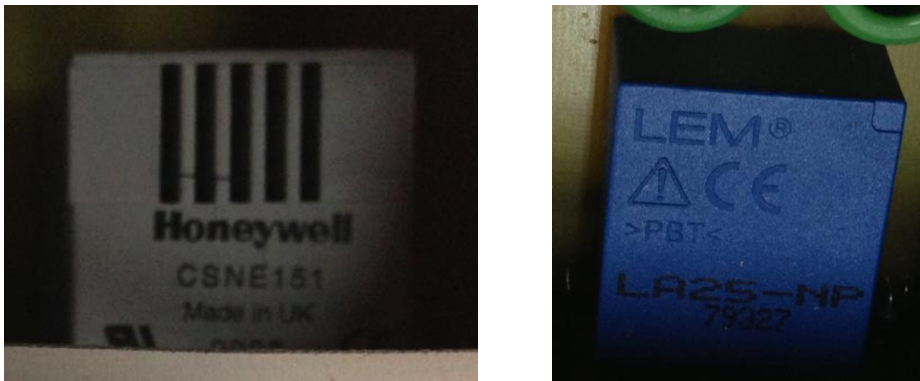


Figura 3.5. Transductores de corriente de *Honeywell* (izquierda) y de *LEM* (derecha).

- Transductor de tensión

La medida de la tensión se realiza mediante un transductor de efecto Hall de LEM (véase la Fig. 3.6), cuyo lado primario se conecta a los terminales de la tensión que se quiere medir. Para la conexión se utiliza también una resistencia externa, de tal forma que pasa una corriente por el primario del sensor. Cuando esta se aproxima a los 10 mA se obtiene la máxima precisión. La resistencia es de un valor tal que a la máxima tensión que se va a medir, la corriente sea precisamente de 10 mA. En ese caso, la precisión será de $\pm 0,8\%$ de la tensión de entrada. El sensor ofrece una corriente secundaria a su salida, proporcional a la tensión de entrada. Se hace circular por otra resistencia externa para efectuar la medida de la tensión.

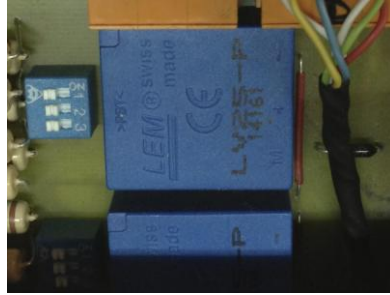


Figura 3.6. Transductor de tensión de LEM.

- Sistema de adquisición de datos

El sistema de adquisición de datos consiste en una plataforma de *National Instruments* (véase la Fig. 3.7), un sistema modular compuesto por un chasis compacto que permite la conexión de hasta cuatro módulos de entrada/salida y un módulo de adquisición con entrada analógica.



Figura 3.7. Sistema de adquisición de datos de *National Instruments*.

- Ordenador portátil

El sistema de adquisición se conecta vía USB 2.0 a un PC portátil, modelo 6717s de HP Compaq (véase la Fig. 3.8). El *software* que realiza la adquisición es Matlab, versión R2010. También calcula el PSD, la transformada de Hilbert y ciertos estadísticos en el dominio temporal. Más adelante se explicará detalladamente el significado y la razón de estos cálculos.



Figura 3.8. Ordenador portátil conectado al sistema de adquisición de datos.

- Sensor de velocidad

La velocidad de giro del motor se mide mediante el contraste entre blanco y negro producido por una banda de esos mismos colores instalada en el eje del rotor del motor. Se trata, por tanto, de un sensor de contraste, en concreto de un S60W de la marca *Datalogic* (véase la Fig. 3.9).

Este sensor de contraste entrega una señal entre 0 y 30 V dependiendo del contraste leído. La señal es recibida por un tacómetro digital que muestra la velocidad en rpm en una pantalla en formato numérico. Como hay 18 ranuras que la banda blanca y negra posee a lo largo de su perímetro, la señal indica que el rotor ha realizado una vuelta completa cuando se completan 18 ciclos.

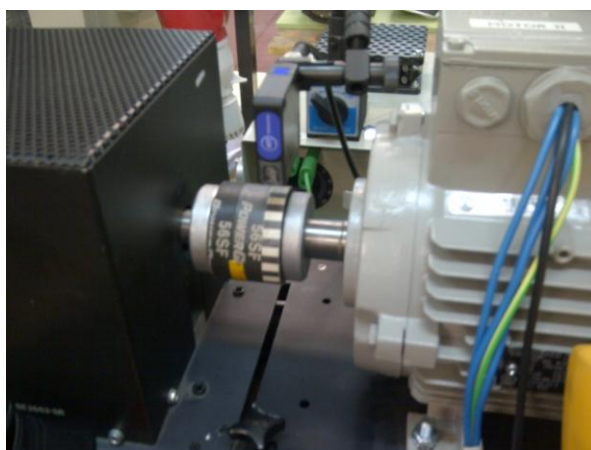


Figura 3.9. Sensor de contraste S60W de *Datalogic*.

3.3. Descripción de la experimentación

El procedimiento es idéntico para ambos motores. Se divide en una serie de etapas.

1ª etapa

En primer lugar, se estudia la respuesta del motor inicialmente sano utilizando distintos tipos de alimentación. Para ello, se realizan unos ensayos con la alimentación de red y otros empleando un variador de frecuencia. El motor 1 opera con variadores de la marca *Telemecanique* y *Lucas-Nülle*, mientras que el motor 2 lo hace con *Allen-Bradley* y *Siemens*. En cualquier caso, tanto si la alimentación es directa como si se emplea un inversor la frecuencia de alimentación es de 50 Hz. Para mayor información, se aconseja

consultar el Anexo A que incluye las características de los variadores empleados.

El motor se pone en funcionamiento y, una vez se encuentra en régimen estacionario, se registran los resultados sobre las variables de interés a través de la tarjeta de adquisición de datos, que se encuentra conectada al ordenador.

2ª etapa

En esta segunda etapa, se introduce al rotor el defecto anteriormente explicado en su menor grado de severidad. Se taladra un agujero de unos 6 mm de profundidad en una de sus barras. Previamente, el motor debe haber sido abierto para la extracción del rotor. Cabe destacar que el procedimiento de taladrado es puramente manual, por lo que resulta complicado perforar justamente hasta la profundidad deseada. No obstante, el intervalo exigido para la consecución de los objetivos es suficientemente amplio como para que la exactitud de la profundidad no sea un factor determinante.

Una vez acoplado correctamente, el motor se pone en marcha y se toman las medidas oportunas análogamente a la etapa anterior, utilizando la alimentación correspondiente en cada caso.

3ª etapa

Esta fase contempla la continuación de la perforación del agujero hecho con anterioridad hasta una profundidad aproximada de unos 12 mm. De nuevo, tiene lugar la medición de las variables clave.

4ª etapa

Por último, se provoca la rotura total de la barra del rotor taladrándola hasta atravesarla por completo. Una vez más, se registran adecuadamente los resultados según la combinación de motor y tipo de alimentación utilizados.

La Tabla 3.1 resume las profundidades del agujero alcanzadas durante el ensayo.

Tabla 3.1. Descripción de los ensayos realizados.

Descripción del estado del motor	Profundidad agujero M1 (mm)	Profundidad agujero M2 (mm)
Motor sano	0	0
Fallo incipiente	6,4	5,7
Barra semi-rotas	11,7	12,1
Barra totalmente rota	17	18

3.4. Características medidas y notación

La velocidad de giro del rotor y la corriente del estator son las señales medidas y procesadas convenientemente para la consecución de los objetivos de la experimentación. Todos los datos obtenidos han sido anotados en una hoja de cálculo. Nos encontramos con hasta 19 columnas que hacen referencia a identificadores del ensayo y a las variables de interés escogidas.

4 identificadores	→	<i>motor</i> {M1, M2}		
		<i>type</i> {M1 _{UT} , M1 _{TE} , M1 _{LN} , M2 _{AB} , M2 _{UT} , M2 _{SM} }		
		<i>condition</i> {1, 2, 3, 4}		
		<i>name</i>		
15 variables	→	<i>current</i>	<i>peak_l.1</i>	<i>amplitude_l.1</i>
		<i>speed</i>	<i>peak_u.1</i>	<i>amplitude_u.1</i>
		<i>slip</i>	<i>peak_l.5</i>	<i>amplitude_l.5</i>
			<i>peak_u.5</i>	<i>amplitude_u.5</i>
			<i>peak_l.7</i>	<i>amplitude_l.7</i>
			<i>peak_u.7</i>	<i>amplitude_u.7</i>

3.4.1. Identificadores del ensayo

Los identificadores recogen toda la información necesaria para reconocer las condiciones en las que es efectuada cada una de las pruebas.

- Motor (*motor*)

Adopta dos valores, M1 o M2, dependiendo de con cuál de los dos motores se trabaje.

- Tipo (*type*)

Indica qué combinación de motor y variador de frecuencia se ha utilizado. Encontramos hasta seis combinaciones posibles, tres para cada motor:

- **M1_UT**: ensayos efectuados con el motor M1 y sin variador de frecuencia, es decir, directamente con la alimentación de red (*utility supply*), de 50 Hz de frecuencia y 400 V de tensión de línea.
- **M1_TE**: el motor M1 se combina con un variador de la marca *Telemecanique*.

- **M1_LN:** se utiliza el motor M1 junto a un variador de la empresa *Lucas-Nülle*.
- **M2_UT:** análogamente al M1_UT, el motor M2 es alimentado por la red, sin la intervención de un variador de frecuencia.
- **M2_AB:** pruebas ejecutadas con el motor M2 y un variador de la compañía *Allen-Bradley*.
- **M2_SM:** el variador usado en combinación con el motor M2 es de la marca *Siemens*.

- Condición (*condition*)

Se registra la información sobre el estado físico del motor. Distinguimos entre cuatro condiciones posibles:

- **Condición 1:** motor sano.
- **Condición 2:** fallo incipiente, motor con perforación de unos 6 mm de profundidad en una de las barras del rotor.
- **Condición 3:** barra semi-rotta, con perforación de unos 12 mm de profundidad.
- **Condición 4:** rotura total de la barra, la cual ha sido perforada por completo.

- Nombre (*name*)

A cada uno de los ensayos se le proporciona un nombre distinto, de tal forma que tenemos una etiqueta única para cada prueba. El nombre incluye el número de ensayo y la frecuencia de alimentación, además de la combinación de características recogidas en los tres identificadores anteriormente explicados. Como ya indicamos en el apartado 3.3, la experimentación se centra en el estudio del comportamiento para una frecuencia de alimentación fija de 50 Hz.

La Tabla 3.2 refleja una cuantificación y clasificación de los datos en base a la información contenida en los identificadores. Como podemos observar, en total se han llevado a cabo 1466 ensayos, de los cuales 758 se corresponden con el motor 1, mientras que los 708 restantes lo hacen con el motor 2.

Tabla 3.2. Descripción general de la experimentación.

MOTOR	Nº ensayos	CONDITION	Nº ensayos	TYPE	Nº ensayos		
M1	758	1	232	M1_UT	83		
				M1_TE	100		
				M1_LN	49		
		2	191	M1_UT	191	M1_UT	67
						M1_TE	65
						M1_LN	59
		3	193	M1_UT	193	M1_UT	70
						M1_TE	63
						M1_LN	60
		4	142	M1_UT	142	M1_UT	60
						M1_TE	40
						M1_LN	42
M2	708	1	317	M2_UT	110		
				M2_AB	108		
				M2_SM	99		
		2	129	M2_UT	129	M2_UT	40
						M2_AB	46
						M2_SM	43
		3	130	M2_UT	130	M2_UT	42
						M2_AB	46
						M2_SM	42
		4	132	M2_UT	132	M2_UT	47
						M2_AB	40
						M2_SM	45
TOTAL	1466	TOTAL	1466	TOTAL	1466		

3.4.2. Variables registradas

Las variables contempladas en el análisis se pueden clasificar en dos grupos: variables relacionadas con la carga del motor y variables relacionadas con el dominio de la frecuencia.

3.4.2.1. Variables de carga del motor

Se trata de características que no son objeto principal de estudio, sino que son meramente informativas de las condiciones de carga a las que se ve sometido el motor en cada ensayo.

- Corriente (*current*)

Intensidad de corriente en amperios (A) consumida por el motor durante el ensayo.

- Velocidad (*speed*)

Velocidad de rotación del motor en revoluciones por minuto (rpm) durante el ensayo. El sensor de velocidad es el encargado de proporcionarnos esta medida. La velocidad puede modificarse mediante la manipulación del par resistente ejercido por el freno, gracias a la Unidad de Control del que está provisto.

- Deslizamiento (*slip*)

Velocidad relativa del rotor con respecto a la velocidad del campo magnético, también conocida como velocidad de sincronismo. Se define según la ecuación (2.1) del Capítulo 2.

3.4.2.2. Variables en el dominio de la frecuencia

El estudio que nos ocupa se centra en el análisis de la respuesta en frecuencia del motor. Sin embargo, las mediciones sobre la corriente del estator se corresponden con su comportamiento temporal. Se debe, por lo tanto, realizar el procesamiento adecuado de la señal para la obtención de los resultados deseados. El *software* Matlab permite llevar a cabo esta función.

El método más usado para obtener el espectro en frecuencias de una señal es la transformada de Fourier (FT); no obstante, en este caso, por razones de reducción de cálculos (véase el apartado 2.4.2), se ha empleado su simplificación: la transformada rápida de Fourier (FFT). A partir de ella, se calcula la función *Power Spectral Density* (PSD) normalizada, la cual permite una mejor identificación de las bandas laterales indicativas de fallo situadas en torno a los armónicos impares no triples. La normalización establece como referencia en 0 dB el nivel alcanzado por el armónico fundamental.

Las bandas laterales son especialmente significativas en los armónicos primero, quinto y séptimo, por lo que son objeto de nuestro estudio. Como la frecuencia de alimentación es de 50 Hz en todos los ensayos, los armónicos primero, quinto y séptimo se presentan a frecuencias de 50, 250 y 350 Hz respectivamente. Aunque alrededor de cada armónico aparecen varias bandas tal y como se indicó en el apartado 2.4.1, la recolección de datos se centra únicamente en la banda inferior (*lower-sideband harmonic*, LSH) y en la banda superior (*upper-sideband harmonic*, USH) más cercana a la

frecuencia del armónico. Así pues, tenemos seis puntos de interés sobre cada uno de los cuales se miden dos características: la amplitud y el pico.

- Amplitud

Diferencia en decibelios (dB) con respecto al armónico fundamental. De este modo, se definen:

$$\begin{array}{ll} \textit{amplitude_l.1} & \textit{amplitude_u.1} \\ \textit{amplitude_l.5} & \textit{amplitude_u.5} \\ \textit{amplitude_l.7} & \textit{amplitude_u.7} \end{array}$$

- Pico

Diferencia en decibelios (dB) entre la amplitud del pico y el valor del ruido de fondo del PSD. Esta medición es seleccionada debido a que se considera muy importante tener en cuenta las diferencias en el nivel de ruido introducidas según el tipo de alimentación. De este modo, se definen:

$$\begin{array}{ll} \textit{peak_l.1} & \textit{peak_u.1} \\ \textit{peak_l.5} & \textit{peak_u.5} \\ \textit{peak_l.7} & \textit{peak_u.7} \end{array}$$

La Fig. 3.10 ilustra los conceptos descritos para las bandas LSH y USH asociadas al armónico fundamental, situado a una frecuencia de 50 Hz. A frecuencias de 250 y 350 Hz se distinguirían análogamente el resto de variables. El gráfico se corresponde con un motor con una barra rota.

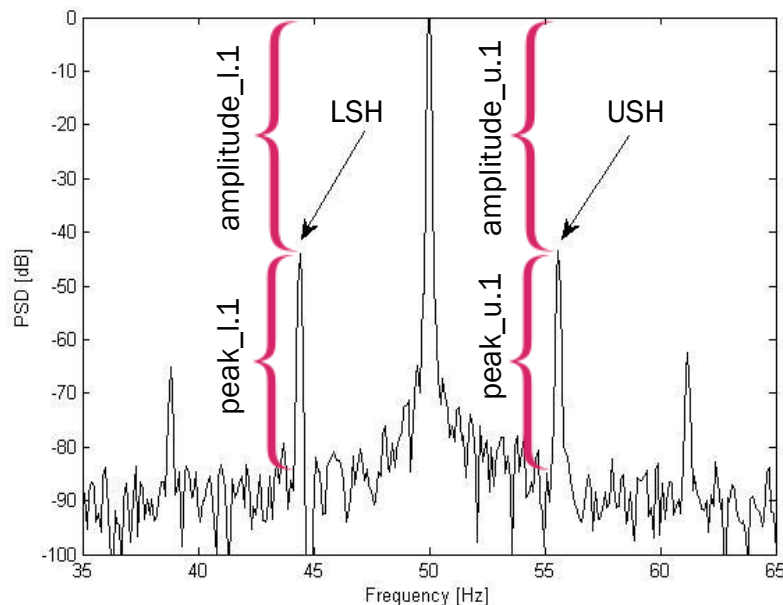


Figura 3.10. Espectro de un motor con una barra rota. LSH y USH en torno al primer armónico. Función PSD.

Capítulo 4

Técnicas estadísticas empleadas

En este Capítulo se expondrán los conocimientos estadísticos necesarios para comprender el desarrollo de los análisis que serán efectuados en Capítulos posteriores.

4.1. El Análisis de Componentes Principales

Para el ser humano es realmente complicado interpretar o “visualizar” conjuntos de datos de dimensión $d > 3$. Como consecuencia de ello, surge la necesidad de tratar de describir con precisión los valores de p variables por medio de un subconjunto $q < p$ de ellas, tratando de perder la mínima cantidad posible de información (Peña, 2002). Mediante el Análisis de Componentes Principales (ACP) se puede abordar este problema. Su principio consiste en generar combinaciones lineales de las variables explicativas registradas, de tal forma que se busquen las proyecciones adecuadas para reducir la dimensionalidad. De hecho, cuando se dispone de variables con elevada dependencia es frecuente que un reducido número de nuevas variables (menos del 20% de las originales) expliquen la mayor parte (más del 80%) de la variabilidad original.

A continuación explicaremos el fundamento teórico de esta interesante y útil técnica estadística multivariante.

4.1.1. Obtención de las componentes principales

Se dispone de n individuos a los que medimos p variables que se almacenan en una matriz de datos:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = (x_1 | \cdots | x_p) \Rightarrow \text{Matriz } n \times p$$

Inicialmente, se quiere obtener una combinación lineal que recoja la máxima variabilidad posible. Dicho de otra forma, su varianza debe ser máxima (Aluja y Morineau, 1999). A esta combinación se la conoce como la primera componente principal. Se busca la dirección dada por el vector u_1 con $\|u_1\| = 1$ más idónea para proyectar los datos en cuestión. Sea $\vec{x}_i = (x_{i1}, \dots, x_{ip})'$ la observación i -ésima para $i = 1, \dots, n$ y su proyección sobre el eje dado por u_1 denotada por $H_i = \text{Pr}_{u_1}(\vec{x}_i)$, se tiene que $d(O, H_i) = |\vec{x}_i' u_1|$. Por tanto, se quiere maximizar la función

$$\sum_{i=1}^n d(O, H_i)^2 = (\vec{x}_1' u_1)^2 + \cdots + (\vec{x}_n' u_1)^2 = \|Xu_1\|^2$$

sujeto a $\|u_1\| = 1$. Usando multiplicadores de Lagrange tenemos que la expresión a optimizar es

$$L(u_1, \lambda_1) = u_1' X' X u_1 - \lambda_1 (u_1' u_1 - 1),$$

cuya solución es

$$X' X u_1 = \lambda_1 u_1.$$

Por lo tanto, u_1 es el autovector de la matriz $X'X$ y λ_1 su correspondiente valor propio, el cual además constituye el mayor autovalor de la matriz $X'X$.

Después de esto, se busca el vector u_2 con $\|u_2\| = 1$ y con $u_1 \perp u_2$ (para no duplicar la información) donde $\{u_1, u_2\}$ ofrezcan el mejor plano donde proyectar los datos. Análogamente a lo anterior, como solución se obtiene que u_2 es el autovector de $X'X$ asociado al segundo autovalor λ_2 .

Si repetimos el proceso sucesivamente, se van obteniendo cada vez más autovectores, tantos como el número de variables de las que se dispone. Nótese que los autovalores asociados son cada vez más reducidos, tal que $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$.

4.1.2. Representación de los datos

Para la representación de la información es necesario calcular determinados factores, tal que $F_\alpha = Xu_\alpha$ nos da las coordenadas de los datos en el eje factorial α y $(F_1(i), \dots, F_q(i))$ nos da las coordenadas de la

observación x_i en el subespacio proyectado ($\mathbb{R}^p \rightsquigarrow \mathbb{R}^q$ con $q \leq p$). Así pues, F_1, \dots, F_p se obtienen como combinaciones lineales de las variables originales con la característica de que $\text{Var}(F_1) \geq \dots \geq \text{Var}(F_p)$ y de que $\text{Cov}(F_j, F_k) = 0$ para $j \neq k$ (Aluja y Morineau, 1999).

Cada eje α calculado recoge una cantidad de información determinada, denominada comúnmente “inercia”, de la siguiente forma:

$$\text{“Inercia recogida en eje } \alpha \text{”} = \|F_\alpha\|^2 = \lambda_\alpha$$

Por consiguiente, dependiendo del número de ejes que se desee considerar, la inercia explicada por los mismos variará. A medida que más ejes se consideren, mayor será la información recogida, pero también aumentará la dimensionalidad, algo no deseable. Por ello, se aprovecha exclusivamente la información de los ejes asociados a los autovalores más altos, desestimando los demás. De este modo, se reduce la dimensionalidad y la mayor parte de la variabilidad es contemplada. Para tener una acotación sobre la eficacia de la estimación de un número determinado de ejes, se precisa la calidad de la representación τ_q , una medida porcentual cuya definición es

$$\tau_q = \frac{\text{“Inercia explicada } q \text{ primeros ejes”}}{\text{“Inercia total”}} = \frac{\lambda_1 + \dots + \lambda_q}{\lambda_1 + \dots + \lambda_p} \cdot 100. \quad (4.1)$$

Hasta ahora se ha hecho referencia a la matriz de datos X como n puntos en \mathbb{R}^p , pero también se puede considerar la matriz X' con p puntos en \mathbb{R}^n . Al realizar un ACP partiendo de la matriz X' obtenemos v_1, \dots, v_n autovectores con norma 1 de $(X')'X' = XX'(n \times n)$ asociados a los autovalores mayores $\mu_1 \geq \mu_2 \geq \dots \geq \mu_p$ y factores calculados $G_\alpha = X'v_\alpha$. Así pues, la relación entre los análisis en \mathbb{R}^p y \mathbb{R}^n es la siguiente:

$$\lambda_i = \mu_i \text{ para } i = 1, \dots, p$$

$$\mu_{p+1} = \dots = \mu_n = 0$$

$$u_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} G_\alpha \qquad v_\alpha = \frac{1}{\sqrt{\mu_\alpha}} F_\alpha$$

La Tabla 4.1 resume las cualidades de los análisis en \mathbb{R}^p y \mathbb{R}^n .

Tabla 4.1. Resumen de las características de los análisis.

	Análisis en \mathbb{R}^p	Análisis en \mathbb{R}^n
Datos	X	X'
Autovectores	u_α	v_α
Autovalores	λ_α	μ_α
Factores calculados	F_α	G_α

4.1.3. Aplicaciones del ACP

El ACP tiene aplicaciones muy diversas dependiendo del tipo de análisis efectuado: análisis en \mathbb{R}^p , análisis en \mathbb{R}^n o ambos de forma simultánea.

1. **Análisis en \mathbb{R}^p :** Las n observaciones se representan usando los factores F_1, \dots, F_q ($q = 2, 3, \dots$). La interpretación del gráfico es que dos observaciones que se encuentren próximas en estos nuevos ejes deben tomar valores parecidos en las p variables originales.
2. **Análisis en \mathbb{R}^n :** Partiendo de datos estandarizados y representando las p variables en los ejes G_1, \dots, G_q se pueden realizar una serie de interpretaciones:

$$\begin{cases} d(x_j, x_k) \simeq 0 \Rightarrow \text{Cor}(x_j, x_k) \simeq 1 \\ d(x_j, x_k) \simeq 2 \Rightarrow \text{Cor}(x_j, x_k) \simeq -1 \\ d(x_j, x_k) \simeq \sqrt{2} \Rightarrow \text{Cor}(x_j, x_k) \simeq 0 \end{cases}$$

Dicho de otra forma, las variables que se sitúan próximas entre sí, están muy correladas positivamente; las que se sitúan en posiciones opuestas, están muy correladas negativamente; y las que son ortogonales entre sí, presentan una correlación prácticamente inexistente.

3. **Análisis en \mathbb{R}^p y análisis en \mathbb{R}^n :** Al superponer los dos análisis en los mismos gráficos es posible ver los valores relativos que toman los n individuos respecto a las p variables analizadas (Aluja y Morineau, 1999).

En el trabajo que nos ocupa, se emplearán análisis en \mathbb{R}^n para visualizar las posibles relaciones entre las variables explicativas (apartado 5.5). Por otro lado, el gráfico derivado del análisis en \mathbb{R}^p servirá como base para efectuar un análisis discriminante sobre el gráfico de dispersión en las dos primeras componentes principales.

4.2. El Análisis Discriminante Lineal

El Análisis Discriminante Lineal aborda la identificación de patrones estadísticos que permitan asignar observaciones a un nivel o clase determinada de un factor cualitativo. Se parte de un conjunto amplio de elementos que pueden proceder de dos o más poblaciones distintas. En cada elemento se registran las mismas variables. Se pretende clasificar un nuevo elemento en una de las poblaciones, partiendo del valor que toman sus correspondientes variables.

Este apartado tratará de explicar el enfoque teórico de esta poderosa herramienta estadística, que será empleada en los Capítulos 6 y 8.

4.2.1. Clasificación en dos poblaciones

Comenzamos con la clasificación de los individuos en el caso de que puedan ser asignados solamente a dos poblaciones, a partir de p variables discriminadoras. Para ello, Fisher enfoca el problema en encontrar la función discriminante, combinación lineal de las variables registradas, que separe las dos clases lo mejor posible (Webb y Copsey, 2011). Así, si se consideran n observaciones, la función se define como

$$\begin{pmatrix} D_1 \\ D_2 \\ \vdots \\ D_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{pmatrix},$$

donde D_i es la puntuación discriminante de la observación i -ésima, X_{ij} es el valor que toma la variable j -ésima de la observación i -ésima, y u_j es el coeficiente de ponderación asociado a la variable j -ésima (Gutiérrez, González, Torres y Gallardo; 1994). Se dispone de la matriz de datos y el objetivo es hallar los coeficientes de ponderación óptimos. También se puede expresar el modelo en función de las desviaciones a la media, con lo que tendríamos:

$$\begin{pmatrix} D_1 - \bar{d}_1 \\ D_2 - \bar{d}_2 \\ \vdots \\ D_n - \bar{d}_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{pmatrix} \Rightarrow d = Xu$$

La variabilidad de la función discriminante se formula como

$$d'd = u'X'Xu,$$

donde $X'X$ expresa la suma de cuadrados total, que puede ser descompuesta en suma de cuadrados entre grupos, F , y suma de cuadrados dentro del mismo grupo, V . Por lo tanto,

$$X'X = F + V \Rightarrow d'd = u'X'Xu = u'(F + V)u = u'Fu + u'Vu.$$

El análisis discriminante lineal busca la proyección que haga que los elementos del mismo grupo tengan proyecciones lo más parecidas posibles entre sí y lo más diferentes posibles del resto de los grupos, con lo que se busca maximizar $u'Fu$ y minimizar $u'Vu$. Es fácil comprobar que esto puede lograrse maximizando el valor de $u'V^{-1}Fu$ con lo que buscaremos los autovectores de la matriz $V^{-1}F$. Estos, ordenados de mayor a menor, delimitan los ejes discriminantes. Por otro lado, las puntuaciones discriminantes pueden entenderse como el resultado de proyectar sobre el eje discriminante cada elemento del espacio p -dimensional de las variables originales (Gutiérrez et al., 1994).

Para abordar la asignación de observaciones a una población u otra, el análisis discriminante se basa en el cálculo de los centroides para cada grupo. Continuando con la clasificación en dos posibles poblaciones, que denotaremos por A y B , se calcula la media de cada una de las variables de forma individual para cada población, tal que

$$\bar{X}_A = \begin{pmatrix} \bar{x}_{1A} \\ \bar{x}_{2A} \\ \vdots \\ \bar{x}_{pA} \end{pmatrix} \quad \bar{X}_B = \begin{pmatrix} \bar{x}_{1B} \\ \bar{x}_{2B} \\ \vdots \\ \bar{x}_{pB} \end{pmatrix}.$$

Así, para cada grupo se determina el centro de gravedad según la fórmula:

$$\bar{D}_A = u_1\bar{x}_{1A} + u_2\bar{x}_{2A} + \dots + u_p\bar{x}_{pA}$$

$$\bar{D}_B = u_1\bar{x}_{1B} + u_2\bar{x}_{2B} + \dots + u_p\bar{x}_{pB}$$

Los centroides son los que finalmente decretan la clasificación de un nuevo individuo. Determinan el punto de corte C que delimita los dos conjuntos. Cuando la probabilidad a priori de que un individuo pertenezca a cualquier clase es la misma, el punto de corte C se define como el punto medio entre los centros de gravedad, es decir,

$$C = \frac{\bar{D}_A + \bar{D}_B}{2}.$$

De este modo, un nuevo individuo i se clasifica en el grupo A si $D_i < C$ y se clasifica en el grupo B si $D_i > C$. Habitualmente se prefiere homogeneizar la expresión, de modo que se define la función discriminante final como

$$D - C = u_1x_1 + u_2x_2 + \dots + u_px_p - C.$$

Ahora, un nuevo individuo i se clasifica en un grupo u otro dependiendo de si $D_i - C < 0$ ó si por el contrario $D_i - C > 0$. La Fig. 4.1 ilustra un ejemplo de análisis discriminante en el que la línea negra determina a qué grupo es asignada cada observación. Como se puede ver, no se acierta con exactitud en todos los casos, pero sí en la inmensa mayoría.

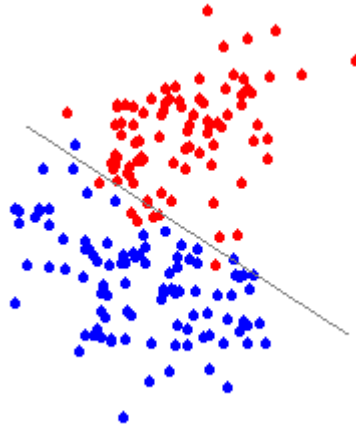


Figura 4.1. Ejemplo de diagrama de dispersión que refleja el análisis discriminante desarrollado.

4.2.2. Clasificación en más de dos poblaciones

Una vez abordada la clasificación en dos poblaciones, procedemos a la generalización para más grupos. En el estudio que nos ocupa, de hecho, interesa distinguir entre los cuatro estados del motor, así que resulta de especial interés.

El número máximo de ejes discriminantes está delimitado por el número de categorías g y por el número de variables p , de la siguiente forma:

$$\text{N}^\circ \text{ máximo de ejes} = \text{mín}(g - 1, p)$$

Lo habitual es que el número de variables exceda con holgura el número de categorías a clasificar, por lo que en general se dispone de $g - 1$ ejes discriminantes (Gutiérrez et al., 1994). Así sucede en nuestro caso particular, donde se podrán definir a lo sumo tres ejes discriminantes.

En esta ocasión, por lo tanto, se dispondrá de varias funciones discriminantes D_i como combinaciones lineales de las p variables explicativas,

$$D_i = u_{i1}x_1 + u_{i2}x_2 + \dots + u_{ip}x_p,$$

donde $i = 1, \dots, g - 1$, y u_1, u_2, \dots, u_{g-1} son los vectores que definen los $g - 1$ ejes, con

$$u_1 = \begin{pmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1p} \end{pmatrix}, u_2 = \begin{pmatrix} u_{21} \\ u_{22} \\ \vdots \\ u_{2p} \end{pmatrix}, \dots, u_{g-1} = \begin{pmatrix} u_{g-1,1} \\ u_{g-1,2} \\ \vdots \\ u_{g-1,p} \end{pmatrix}.$$

4.2.3. Métodos de selección de las variables

En ocasiones, algunas de las variables contempladas para el análisis discriminante no aportan información relevante sobre la población a la que pertenece el individuo. Incluso llegan a dificultar la predicción. Este fenómeno se conoce como sobreajuste u *over-fitting* (Hawkins, 2004). El excesivo número de variables contempladas provoca que la discriminación se ajuste excesivamente al conjunto de datos a partir del cual se ha hecho el análisis.

La Fig. 4.2 ilustra, a modo de ejemplo, el sobreajuste del modelo. Como se puede ver, la discriminación correspondiente a la línea verde tiene un éxito total en la clasificación entre las dos poblaciones. Pese a ello, la trayectoria sinuosa de la línea hace previsible su fracaso ante una nueva muestra de individuos. La línea negra, por su parte, representa una discriminación mucho más acertada. Aunque para este conjunto de observaciones clasifique incorrectamente algunas de ellas, parece que el éxito será similar al abordar la clasificación de una nueva muestra, debido a la robustez del modelo.

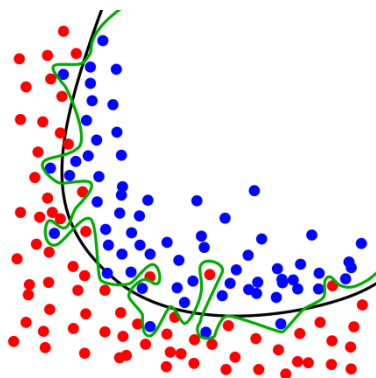


Figura 4.2. Comparación entre sobreajuste del modelo (línea verde) y modelo más adecuado y robusto (línea negra).

Para evitar el sobreajuste u *over-fitting*, se recurre a métodos de selección de variables, de tal forma que sean objeto de estudio solamente las variables con mayor poder discriminante y, por tanto, de mayor relevancia.

Distinguimos tres metodologías de selección: eliminación *Backward*, selección *Forward* y selección *Stepwise* (Faraday, 2015).

- Eliminación *Backward*

Este es el método más simple y puede ser fácilmente implementado sin la necesidad de un *software* especial. El procedimiento es el siguiente:

1. Empieza con todas las variables en el modelo.
2. Descarta la variable con el p-valor más alto superior a un determinado nivel crítico, α_{crit} .
3. Se recalcula el modelo y se vuelve al paso 2 mientras siga existiendo un p-valor más alto que α_{crit} .
4. Para cuando todos los p-valores son inferiores a α_{crit} .

En la eliminación *Backward* el α_{crit} es conocido como “*p-to-remove*”.

- Selección *Forward*

El método *Forward* invierte el funcionamiento del *Backward*.

1. Empieza sin variables en el modelo.
2. Para todas las variables que no se encuentran en el modelo, verifica su p-valor si fueran añadidas al modelo. Elige la de menor p-valor por debajo de α_{crit} .
3. Continúa con el paso 2 hasta que no pueda ser añadida ninguna nueva variable.

En la selección *Forward* el α_{crit} es conocido como “*p-to-enter*”.

- Selección *Stepwise*

Este método es una combinación de la eliminación *Backward* y de la selección *Forward*. Aborda la situación en la cual variables que han sido introducidas o descartadas pronto en el proceso, pueden volver a salir o entrar en el modelo, respectivamente, en cada etapa. En la selección *Stepwise* hay dos α_{crit} que funcionan de forma independiente: son los ya comentados “*p-to-enter*” y “*p-to-remove*”. El procedimiento *Stepwise* es relativamente barato computacionalmente, pero existen varios inconvenientes (Faraday, 2015).

1. Debido a la naturaleza secuencial, ya que en cada etapa solamente puede introducirse o descartarse una variable, es posible perder el modelo “óptimo”.
2. Los p-valores no deben ser tratados de forma demasiado estricta. En muchas ocasiones la validez es dudosa. Además, la eliminación de variables menos significativas tiende a incrementar la significancia del resto de variables, por lo que se exagera su importancia.

3. El procedimiento no está directamente vinculado a objetivos finales de predicción y, por tanto, puede no ayudar a resolver el problema de interés. Con cualquier método de selección es importante tener en mente que la selección del modelo no debe desligarse del propósito subyacente de la investigación. La selección de variables tiende a amplificar la significancia estadística de las variables que están en el modelo. No obstante, las variables que son eliminadas pueden estar correlacionadas con la respuesta. Sería incorrecto decir que no lo están, cuando lo que en realidad sucede es que no aportan ningún efecto explicativo más allá del que proporcionan las variables ya incluidas en el modelo.
4. La selección de variables *Stepwise* tiende a escoger modelos que son más pequeños de lo deseable para propósitos de predicción.

En la práctica, ocurre con cierta frecuencia que los distintos métodos de selección de variables escojan exactamente las mismas variables, en cuyo caso el análisis discriminante será idéntico. El programa *StatGraphics*, con el que serán efectuados los análisis discriminantes de nuestro estudio, proporciona la posibilidad de seleccionar las variables mediante el método *Backward* y el *Stepwise*. El método *Forward*, por lo tanto, no será abordado.

Existe gran variedad de criterios para definir un p-valor crítico que determine la selección de variables. Uno de los más empleados es el criterio del estadístico F , y además es el criterio utilizado por defecto en los análisis discriminantes de *StatGraphics*. Por lo tanto, los análisis discriminantes que se desarrollan en este trabajo emplearán el criterio “*F-to-enter*” y/o el “*F-to-remove*” en la selección de variables. Habitualmente se fija en 4 el valor del estadístico F tanto para entrar como para salir del modelo, así que será el α_{crit} utilizado.

4.2.4. Validación del modelo

Tanto en análisis discriminantes como en cualquier otro tipo de problema de clasificación, la bondad del modelo se mide habitualmente mediante la probabilidad de clasificación errónea o tasa de error verdadero del clasificador. Se trata de la probabilidad de que el modelo no identifique correctamente en su categoría real Y a un nuevo individuo Z . Esta probabilidad se puede expresar como $P(f(Z) \neq Y_z)$, donde f es el clasificador. Normalmente, la probabilidad descrita es imposible de determinar, debido a que depende de las distribuciones de las variables en

cada categoría y estas suelen ser desconocidas. Como consecuencia, se debe recurrir a estimadores para medir la bondad del modelo (Gordo, 2014).

El estimador más sencillo es la tasa de error aparente. Es el que proporciona *StatGraphics* por defecto y su significado es fácil de comprender. La propia muestra de entrenamiento X empleada para calcular la función o funciones discriminantes clasifica incorrectamente un número determinado de individuos. La proporción de estos con respecto al total de individuos a clasificar constituye la tasa de error aparente.

$$err(f, D) = \frac{1}{n} \sum_{i=1}^n (f(X_i) \neq Y_i) \quad (4.2)$$

La simplicidad computacional del estimador lo hace especialmente atractivo. Sin embargo, la tasa ofrecida es, en ocasiones, demasiado optimista. Los individuos clasificados para medir la bondad del modelo han sido a su vez empleados para construir el modelo, por lo que seguramente respondan mejor al mismo que futuros individuos sin relación alguna con las funciones discriminantes.

Para solventar el problema, la solución pasa por dividir el conjunto de datos disponibles en dos muestras: una de entrenamiento y otra de prueba o de test. La muestra de entrenamiento servirá para construir las funciones discriminantes que posteriormente procederán a clasificar la muestra de test, mediante la cual se evaluará la validez del modelo.

Existe gran variedad de métodos de validación, como son el *hold-out* o el *cross-validation*, pero en el estudio que nos ocupa nos centraremos en el método *bootstrap*. A continuación, se explica su fundamento.

Dado un conjunto de n individuos, se toma una muestra aleatoria con reemplazamiento, también de tamaño n , que constituirá la muestra de entrenamiento. El concepto de reemplazamiento significa que, cada vez que se escoge un individuo, este es “devuelto” y puede volver a ser escogido en la siguiente elección al azar. Como el conjunto de datos es muestreado en las condiciones descritas, la probabilidad de que cualquier individuo no sea seleccionado después de n veces es $(1 - 1/n)^n \approx e^{-1} \approx 0,368$, según la probabilidad característica de una distribución binomial. Asimismo, el número esperado de individuos distintos en la muestra se aproxima a $0,632n$. Estos individuos no seleccionados constituirán la muestra de test (Efron y Tibshirani, 1995).

Después de repetir el proceso de muestreo de forma iterativa b veces, se obtiene la tasa de error estimado del *bootstrap*, definida como

$$err_{boot} = \frac{1}{b} \sum_{i=1}^b (0,632\alpha_i + 0,368err(f, D)),$$

donde α_i es la proporción de elementos clasificados de forma incorrecta en la iteración i -ésima del proceso y $err(f, D)$ es el error aparente indicado en la ecuación (4.2).

4.3. El Modelo Aditivo Generalizado

En ocasiones, puede ser de gran utilidad estudiar el efecto que pueda estar ejerciendo algún factor sobre el valor de las mediciones tomadas. En caso de que exista, si se consigue cuantificar y filtrar, las conclusiones desprendidas podrían ser más beneficiosas.

El Modelo Aditivo Generalizado (GAM) es un recurso estadístico que aborda la tarea descrita y será empleado en los Capítulos 7 y 9. Este apartado tratará de exponer de forma general su funcionamiento y significado.

4.3.1. Modelo Aditivo

La relación que puede tener una variable respuesta y con el nivel de determinados factores x_j se mide generalmente a través de un modelo de regresión lineal. Por ejemplo, con dos factores explicativos, la expresión es

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad i = 1, \dots, n, \quad (4.3)$$

donde los términos β_j son los coeficientes de la regresión y ε_i es el término correspondiente al error.

Cuando la asunción de linealidad es insostenible, los componentes como $\beta_1 x_1$ pueden ser sustituidos por polinomios o términos no lineales. La idea de un modelo aditivo es permitir a los componentes del modelo adoptar formas no paramétricas (Bowman y Azzalini, 1997). El modelo (4.3) se reescribiría como

$$y_i = \beta_0 + m_1(x_{i1}) + m_2(x_{i2}) + \varepsilon_i, \quad i = 1, \dots, n, \quad (4.4)$$

donde los términos m_j denotan funciones cuyas formas no tienen restricciones, aparte de la asunción de suavización y condiciones como

$$\sum_{i=1}^n m_j(x_{ij}) = 0 \quad (4.5)$$

para todo j , con el objetivo de hacer única la definición de las funciones.

Ahora bien, en la práctica, las funciones del modelo aditivo (4.4) no son conocidas y se debe recurrir a estimadores que proporcionen una idea de su forma. La asunción de que los efectos son aditivos permite emplear muchos de los métodos unidimensionales de estimación. En primer lugar, el término independiente β_0 puede ser estimado como la media de las respuestas \bar{y} , gracias a las restricciones indicadas en (4.5). La estimación sobre la función m_2 puede ser expresada con notación matricial como $\hat{m}_2 = S_2(y - \bar{y})$. Por otro lado, un reordenamiento del modelo (4.4) como

$$y_i - \beta_0 - m_2(x_{i2}) = m_1(x_{i1}) + \varepsilon_i$$

sugiere que una estimación del componente s_1 puede ser obtenido suavizando los residuales después de haber ajustado \hat{m}_2 , de la forma

$$\hat{m}_1 = S_1(y - \bar{y} - \hat{m}_2),$$

y, análogamente, el estimador \hat{s}_2 puede ser obtenido como

$$\hat{m}_2 = S_2(y - \bar{y} - \hat{m}_1).$$

Mediante esta técnica iterativa, conocida como *backfitting* (Hastie y Tibshirani, 1990), puede extenderse fácilmente a modelos que involucran varios componentes, repitiendo la estimación de cada componente a través de la suavización de los residuos. Así pues, un modelo de la forma

$$y_i = \beta_0 + \sum_{j=1}^p m_j(x_{ij}) + \varepsilon_i$$

puede ser ajustado mediante la construcción iterativa de las estimaciones suavizadas

$$\hat{m}_j = S_k \left(y - \bar{y} - \sum_{j \neq k} \hat{m}_j \right).$$

Para un modelo aditivo, la suma de cuadrados de los residuales está definida por la expresión

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)$$

donde \hat{y}_i denota el valor ajustado, producido por la evaluación del modelo aditivo sobre la observación x_i .

4.3.2. Modelo semiparamétrico

La posibilidad de sistematizar modelos aditivos del tipo (4.4) resulta muy atractiva. Sin embargo, en ocasiones es probable que se pierda la precisión y la potencia del modelo como consecuencia de asignar un término no paramétrico a un componente que, en realidad, responde adecuadamente a un término lineal u otro tipo de término paramétrico. Existen, por lo tanto, muchas situaciones donde un enfoque semiparamétrico será beneficioso. Modelos de este tipo permiten mezclas de componentes lineales y componentes no paramétricos (Bowman y Azzalini, 1997).

El algoritmo *backfitting* puede utilizarse para ajustar modelos de este tipo, como anteriormente. Por ejemplo, un modelo semiparamétrico común es el que incluye un término lineal y otro no paramétrico. De hecho, será el abordado en nuestro estudio, como se verá más adelante en los Capítulos 7 y 9. Se expresa como

$$y = \beta_0 + \beta_1 x_1 + m_2(x_2) + \varepsilon.$$

4.3.3. Modelo Aditivo Generalizado

El origen del GAM proviene de los modelos aditivos. Se trata de un modelo lineal generalizado construido por la suma de funciones suaves (*splines*) de las variables predictoras (Wood, 2006). En general, el modelo tiene una estructura similar a

$$g(\mu_i) = \beta_0 + m_1(x_1) + m_2(x_2) + \dots + m_p(x_p), \quad (4.6)$$

donde

$\mu_i \equiv \mathbb{E}(y_i)$ con $y_i \sim$ alguna distribución de la familia exponencial.

El modelo permite definir las funciones de forma muy general, hasta el punto de poder existir términos de la forma $m_i(x_i, x_j)$. Entre las ventajas más beneficiosas ofrecidas por estos modelos, destacamos la no necesidad de probar la independencia y normalidad de las variables.

En nuestro caso particular, el análisis a efectuar es el más sencillo posible, donde

$$g(x) = x \text{ con } y_i \sim \text{normal.}$$

Los Capítulos 7 y 9 abordarán este tipo de análisis.

Capítulo 5

Análisis descriptivo de los datos

En este Capítulo, se realizará un análisis inicial de los datos con el objetivo de presentar las características principales del problema a abordar. Entre otras técnicas empleadas, se efectuará un Análisis de Componentes Principales, cuyo fundamento teórico se explicó en el Capítulo anterior.

Antes de comenzar, a sugerencia del Departamento de Ingeniería Eléctrica, son eliminados del análisis los datos procedentes de los ensayos cuyo valor de deslizamiento se encuentra fuera de los siguientes rangos:

Ensayos con variador de *Lucas-Nülle* $\rightarrow 0,01 \leq slip \leq 0,09$

Resto de ensayos $\rightarrow 0,01 \leq slip \leq 0,07$

La decisión se toma porque desde un punto de vista operativo el funcionamiento del motor con deslizamientos fuera de esos rangos no es fiable. Tal y como se indica, en el caso de emplear un variador de la marca *Lucas-Nülle* se admite un valor mayor que para el resto de opciones.

La Tabla 5.1 describe las características de los ensayos válidos. En total, han sido descartados 194 ensayos. De ahora en adelante, se trabajará con los 1272 ensayos restantes.

Tabla 5.1. Descripción general de los ensayos válidos.

MOTOR	Nº ensayos	CONDITION	Nº ensayos	TYPE	Nº ensayos		
M1	663	1	218	M1_UT	83		
				M1_TE	100		
				M1_LN	35		
		2	153	2	153	M1_UT	67
						M1_TE	65
						M1_LN	21
		3	169	3	169	M1_UT	70
						M1_TE	63
						M1_LN	36
		4	123	4	123	M1_UT	60
						M1_TE	34
						M1_LN	29
M2	609	1	268	M2_UT	88		
				M2_AB	95		
				M2_SM	85		
		2	117	2	117	M2_UT	37
						M2_AB	41
						M2_SM	39
		3	107	3	107	M2_UT	34
						M2_AB	40
						M2_SM	33
		4	117	4	117	M2_UT	42
						M2_AB	37
						M2_SM	38
TOTAL	1272	TOTAL	1272	TOTAL	1272		

5.1. Análisis descriptivo inicial

Inicialmente es conveniente exponer de forma visual los datos para que su interpretación sea cómoda y obtengamos una idea general del problema que hemos de abordar. Los diagramas de cajas facilitan alcanzar este objetivo, ya que constituyen una herramienta útil para comparar medias y variabilidad entre los diferentes conjuntos de datos. Asimismo, se comprobará si mediante este análisis univariante somos capaces de clasificar los ensayos según el estado físico del motor.

En primer lugar, se deciden representar los diagramas de cajas para cada una de las seis variables denominadas “picos”: peak_l.1, peak_u.1, peak_l.5, peak_u.5, peak_l.7, peak_u.7. Los grupos a distinguir escogidos son la combinación entre motor y condición, incluyendo en cada grupo todas las

fuentes de alimentación posibles, tanto si existe variador de frecuencia como si no.

Como observamos en la Fig. 5.1, en el caso del primer armónico la media aumenta a medida que la severidad del fallo introducido también lo hace. Esto era de esperar, ya que cuanto más dañado esté el motor, mayor será la altura que alcance el pico con respecto al ruido de fondo.

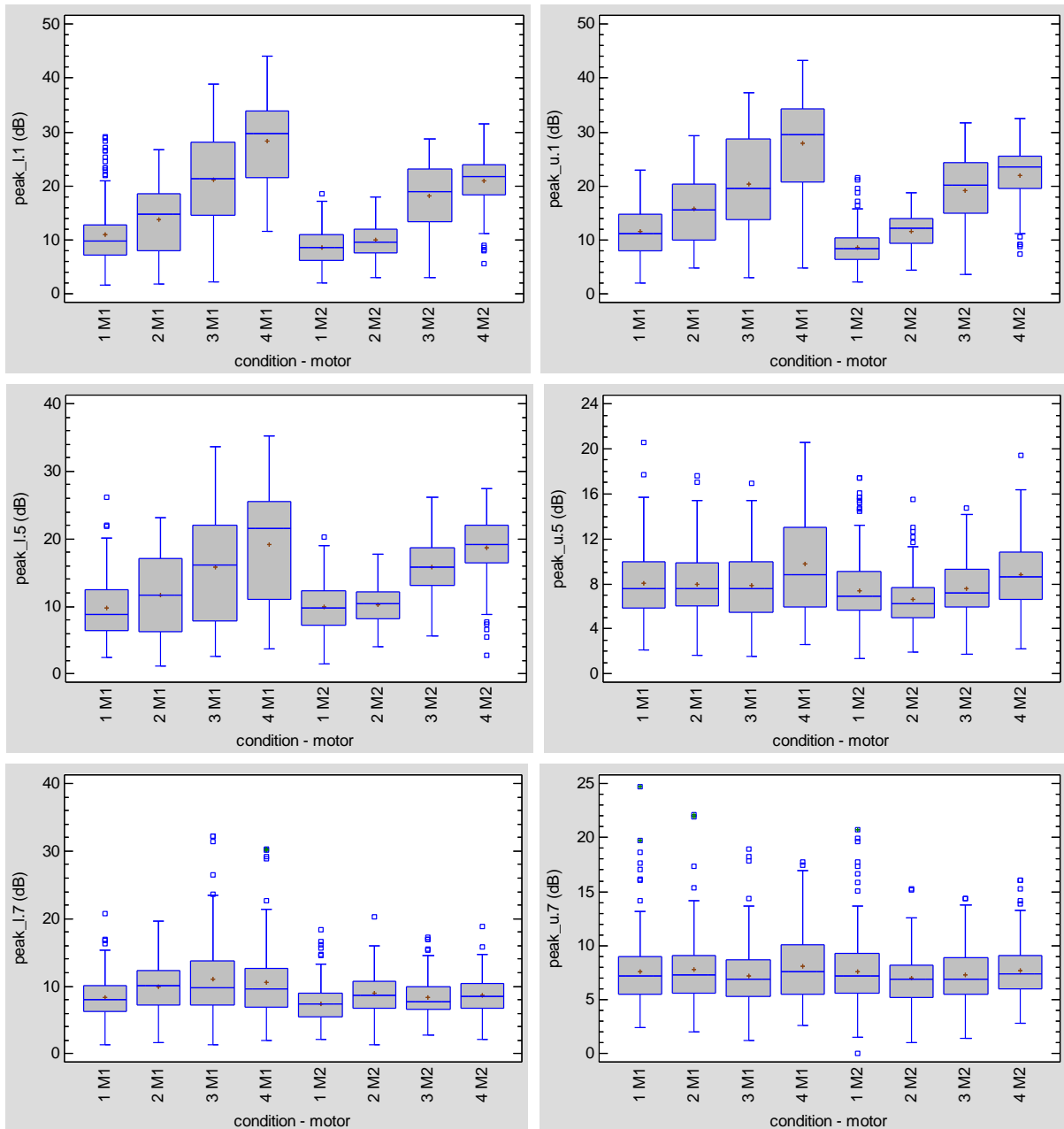


Figura 5.1. Diagrama de cajas de los picos en función de la combinación condición-motor.

En el quinto armónico esta característica persiste para peak_l.5, no así para peak_u.5; en el séptimo, por su parte, se puede decir que desaparece tal característica. En principio, podríamos deducir la disminución de la

sensibilidad al fallo a medida que se trata de un armónico más distante de la frecuencia fundamental. No obstante, se trata meramente de una hipótesis, ya que el análisis descriptivo realizado es demasiado general y aún no se ha filtrado la información adecuadamente.

Por otro lado, los datos de M1 presentan una variabilidad mayor, si bien es cierto que existen numerosos *outliers* en algunos diagramas de ambos motores. La aparición de esos puntos inusuales denota la posible existencia de un factor significativo no contemplado.

En cualquier caso, los diagramas se superponen unos con otros, de tal forma que sería muy osado pronunciarse sobre qué grado de anomalía física tiene el motor analizando de forma individual cualquiera de las variables pico.

Las amplitudes puede que aporten una mayor claridad en las deducciones, de modo que se representan diagramas de cajas para todas ellas: *amplitude_l.1*, *amplitude_u.1*, *amplitude_l.5*, *amplitude_u.5*, *amplitude_l.7* y *amplitude_u.7*. Al igual que antes, se distinguen los grupos según la condición y el motor, resultando los gráficos de la Fig. 5.2.

A diferencia de lo que sucedía en el primer armónico con los picos, la amplitud disminuye con la gravedad del fallo. Esta característica se aprecia bien para el motor M1, pero no tanto para el motor M2. En los diagramas de la amplitud asociada al quinto y al séptimo armónico no se identifica un patrón claramente definido.

De nuevo, la variabilidad del motor M1 es mayor. En cuanto a puntos atípicos, la mayoría de ellos aparecen en el motor M2. De hecho, las amplitudes asociadas a las bandas del quinto y del séptimo armónico en el motor M1 no presentan *outliers*.

Por último, cabe destacar que el solapamiento de los nuevos diagramas, por lo general, es menor que en los diagramas relacionados con los picos. Por ejemplo, en *amplitude_l.1*, las cajas de las condiciones 1 y 2 no se superponen a las de las condiciones 3 y 4. Pese a ello, sigue sin ser prudente clasificar los ensayos según la condición a partir de este análisis univariante.

La inclusión del factor tipo puede esclarecer en parte la aparición de *outliers* y reducir la extensión del diagrama, debido a que se cree que la alimentación utilizada puede influir en los resultados, aunque se use el mismo motor. Como en los gráficos anteriores las evidencias más pronunciadas se reflejaron para el análisis del primer armónico, se decide trabajar sobre este.

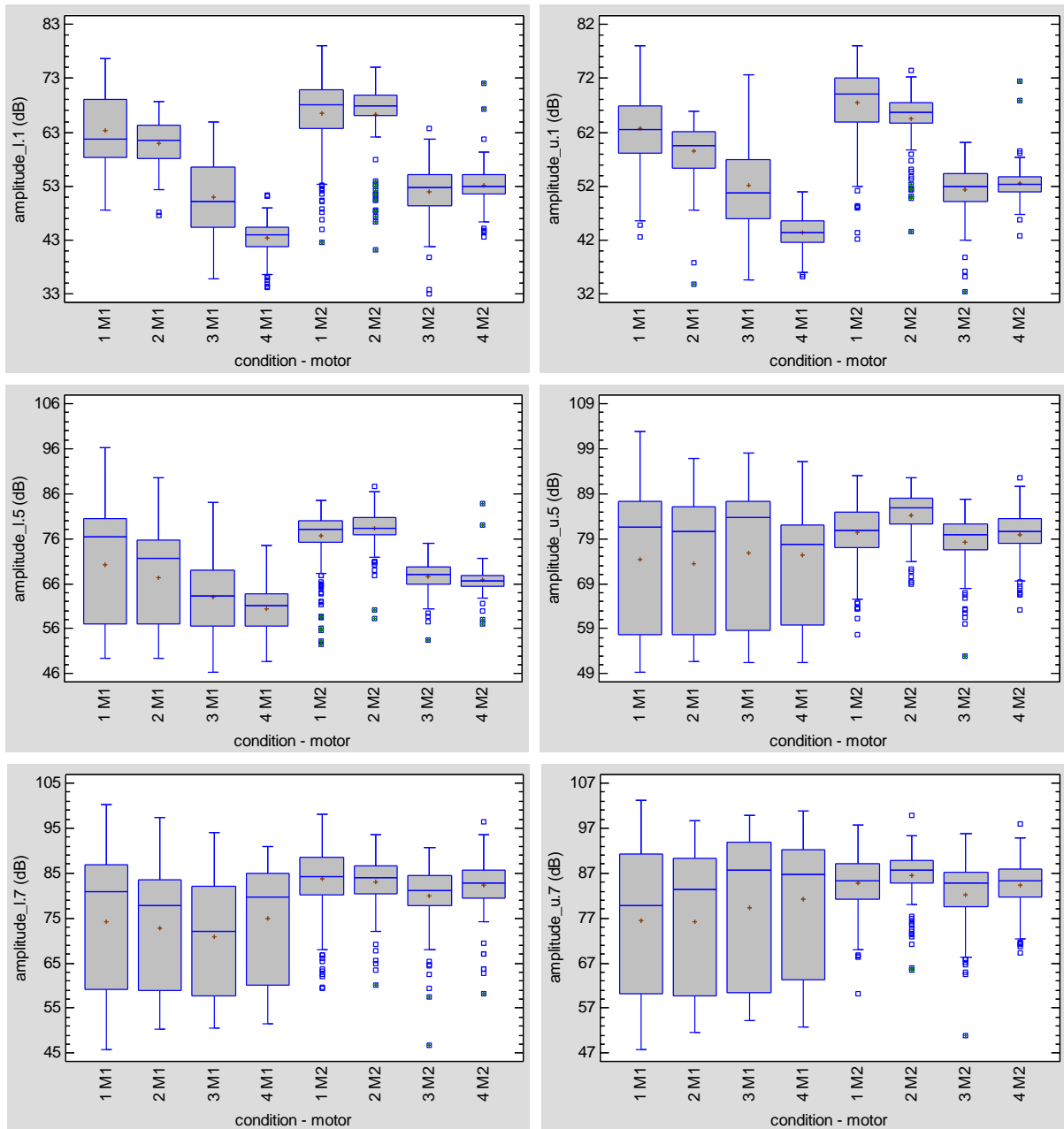


Figura 5.2. Diagrama de cajas de la amplitud en función de la combinación condición-motor.

La Fig. 5.3 refleja los resultados para peak_1.1 y peak_u.1 y confirma la hipótesis planteada: el tipo de alimentación es un factor determinante, ya que disminuye el número de observaciones sin explicación aparente y la variabilidad de los diagramas. Aun así, no es viable la clasificación según la condición a partir del análisis de una única variable.

Por otro lado, de nuevo, aumenta la media a medida que el fallo introducido es más severo, lo cual, como ya hemos indicado antes, tiene sentido.

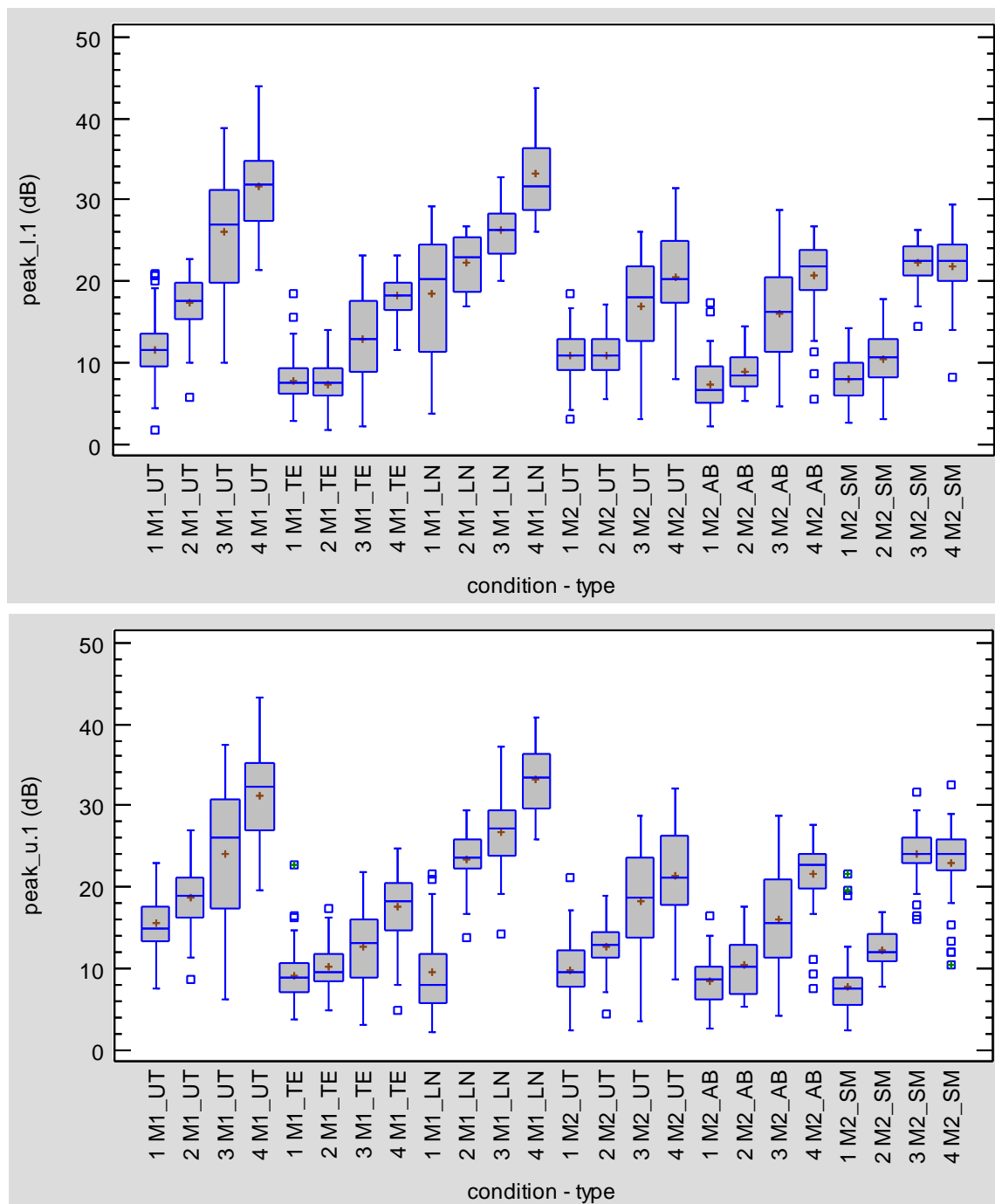


Figura 5.3. Diagramas de caja de los picos asociados al primer armónico en función de la combinación condición-tipo. Arriba LSH, abajo USH.

Los valores de las amplitudes en función del tipo reflejadas en la Fig. 5.4 ofrecen la misma deducción principal: existen diferencias dependiendo del motor y la alimentación proporcionada. Es remarcable la existencia de gran cantidad de puntos inusuales, lo cual hace pensar que haya más factores de interés que puedan influir en la respuesta.

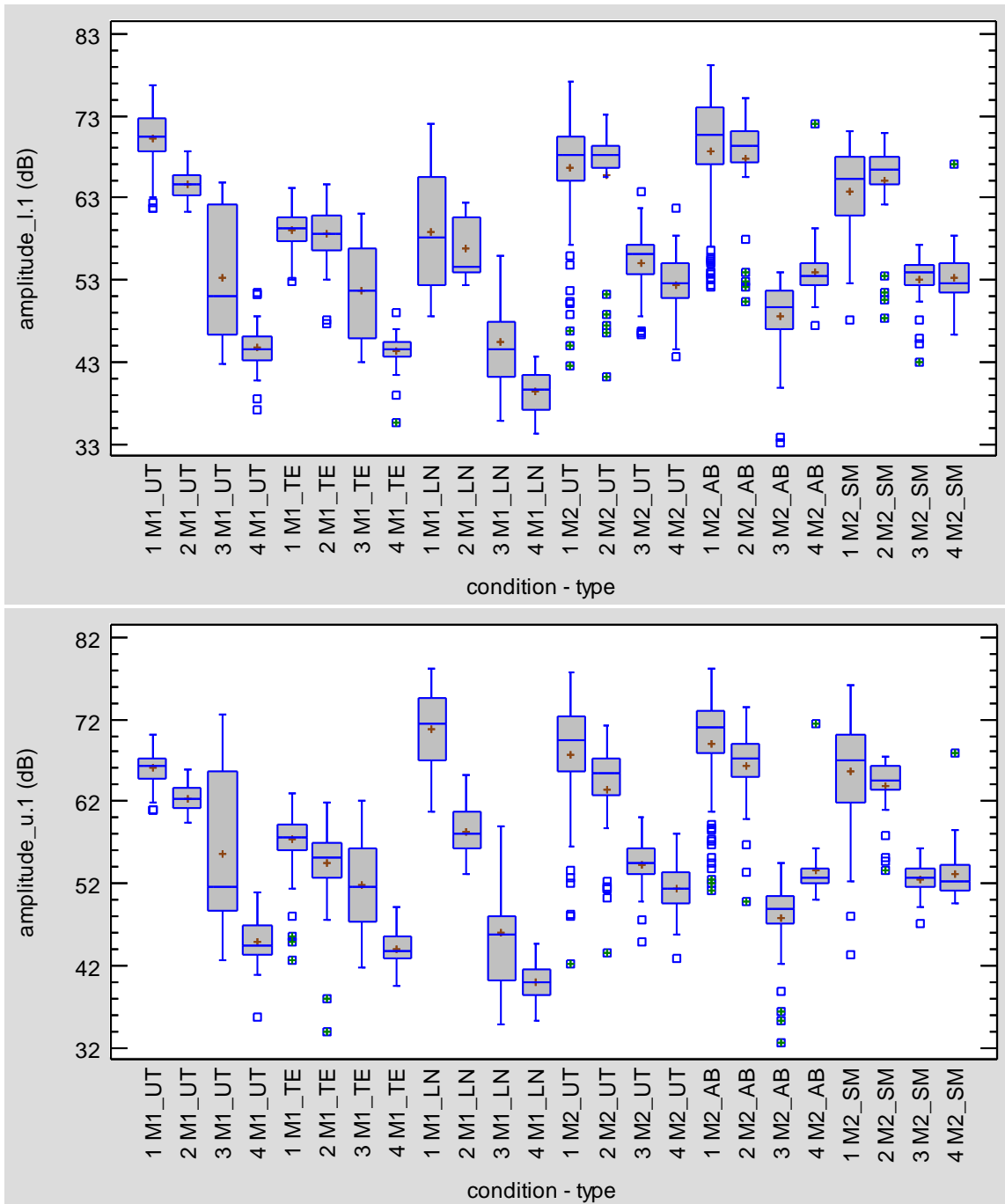


Figura 5.4. Diagramas de caja de las amplitudes asociadas al primer armónico en función de la combinación condición-tipo. Arriba LSH, abajo USH.

El solapamiento generalizado de los diagramas de cajas y bigotes propuestos a lo largo de todo el análisis evidencia la incapacidad de establecer una pauta de distinción entre condiciones mediante esta técnica univariante. El problema deberá ser abordado por otras metodologías más potentes en Capítulos posteriores.

5.2. Correlación de variables

Una cantidad elevada de variables recogidas por un modelo dificulta, en ocasiones, el análisis del mismo y, por tanto, la extracción de conclusiones. Por eso, siempre es conveniente identificar y descartar las variables que no aportan información relevante al modelo.

Teniendo en cuenta la similitud de la naturaleza de algunas de las variables registradas durante la experimentación, se considera oportuno estudiar la matriz de correlaciones, ante la sospecha de que existan fuertes dependencias entre algunas de ellas. El objetivo es reducir al máximo el número de términos contemplados en el estudio.

La Tabla 5.2 muestra los coeficientes de correlación más destacables, con un valor absoluto superior a 0,90. Solamente se hallan dos casos con las condiciones descritas.

Tabla 5.2. Coeficientes de correlación con valor absoluto superior a 0,90. Análisis conjunto de todos los ensayos.

	Coefficiente de correlación
amplitude_l.1 , amplitude_u.1	0,9004
amplitude_u.5 , amplitdute_u.7	0,9254

Llama especialmente la atención que, pese a la conocida relación existente entre velocidad y deslizamiento, indicada a través de la ecuación 2.1 del apartado 2.2, su coeficiente de correlación se sitúe en -0,8707. Aunque se trata de un valor absoluto elevado, no se encuentra entre los más altos.

Al igual que se hizo en el apartado 5.1 haciendo un análisis individual en función del motor del ensayo, se cree oportuno realizar el estudio de las correlaciones entre variables del motor M1 por un lado y entre las variables del motor M2 por otro. La Tabla 5.3 destaca las correlaciones cuyo coeficiente en valor absoluto es superior a 0,90.

Como se puede ver, la relación entre amplitude_u.5 y amplitude_u.7 vislumbrada en la tabla anterior se corresponde fundamentalmente con los ensayos del motor M1, mientras que la relación entre amplitude_l.1 y amplitude_u.1 se debe principalmente a los ensayos del motor M2. Ahora bien, lo más destacable del nuevo estudio desempeñado es que aparece la esperada fuerte relación entre velocidad y deslizamiento para las pruebas del motor M2. Sin embargo, para las pruebas del motor M1 no alcanza el valor absoluto de 0,90, si no que se sitúa en -0,8569.

Tabla 5.3. Coeficientes de correlación con valor absoluto superior a 0,90. Análisis de los conjuntos de ensayos según el motor.

MOTOR		Coeficiente de correlación
M1	amplitud_u.5 , amplitud_u.7	0,9411
M2	speed , slip	-0,9977
	amplitud_l.1 , amplitud_u.1	0,9194

Teniendo en cuenta las conclusiones extraídas en el apartado 5.1, las cuales aseguraban una dependencia del comportamiento en función del tipo de alimentación y del motor utilizado, se efectúan nuevos análisis individuales de las correlaciones para cada conjunto de ensayos con el mismo identificador “tipo”. Los coeficientes cuyo valor absoluto supera 0,90 se presentan en la Tabla 5.4.

Aparecen numerosas correlaciones con un valor reseñable que mediante el análisis individualizado por motores no se habían percibido. Algunas de ellas se repiten para diferentes motores y alimentaciones.

Tabla 5.4. Coeficientes de correlación con valor absoluto superior a 0,90. Análisis de los conjuntos de ensayos según el tipo.

TIPO		Coef. de correlación	TIPO		Coef. de correlación
M1_UT	speed , slip	-1,0000	M2_UT	current , speed	-0,9342
	peak_l.1 , peak_u.1	0,9006		current , slip	0,9327
	peak_l.1 , amplitud_l.1	-0,9452		speed , slip	-0,9979
	peak_u.1 , amplitud_u.1	-0,9154		amplitud_l.1 , amplitud_u.1	0,9081
	amplitud_l.1 , amplitud_u.1	0,9456	M2_AB	current , speed	-0,9786
	amplitud_l.1 , amplitud_l.5	0,9310		current , slip	0,9790
M1_TE	speed , slip	-0,9914		speed , slip	-0,9973
	M1_LN	current , speed	-0,9714	amplitud_l.1 , amplitud_u.1	0,9505
current , slip		0,9716	M2_SM	current , speed	-0,9372
speed , slip		-0,9730		current , slip	0,9375
peak_u.1 , amplitud_u.1		-0,9365		speed , slip	-0,9983

A modo de resumen, la Tabla 5.5 destaca los resultados para los pares de variables cuyo coeficiente de correlación supera el valor absoluto de 0,95 en al menos alguna combinación entre motor y alimentación.

La fuerte dependencia entre la velocidad y el deslizamiento deja de ser una mera suposición y es rotundamente confirmada, puesto que los valores absolutos de los coeficientes en todos los casos superan 0,97. Incluso existe una relación exactamente lineal entre ambas variables cuando se trata de ensayos del tipo M1_UT (coeficiente de -1).

Por otro lado, tanto la relación corriente-velocidad como la relación corriente-deslizamiento son bastante fuertes en varios casos: en cuatro de los seis tipos los coeficientes superan 0,93 en valor absoluto. Con el objetivo de establecer un criterio razonable para continuar reduciendo el número de variables, la Tabla 5.6 muestra el coeficiente medio alcanzado por las pruebas desarrolladas con el mismo motor, así como la media total entre los seis tipos existentes.

Desde un punto de vista global, la correlación media de corriente con velocidad (-0,9055) y con deslizamiento (0,9026) alcanza un valor absoluto de en torno a 0,90, respectivamente. Desde un enfoque menos amplio, para los ensayos llevados a cabo con el motor M2, las correlaciones medias con velocidad (-0,9500) y deslizamiento (0,9497) aumentan considerablemente, situándose aproximadamente en valores absolutos de 0,95. En cuanto al motor M1, resulta llamativo el reducido valor alcanzado por el coeficiente en ensayos del tipo M1_TE, en relación al resto de tipos. La explicación más razonable a este hecho es la antigüedad del variador de *Telemecanique*, mucho menos sofisticado que los demás.

Respecto a las variables *amplitude_l.1* y *amplitude_u.1*, su interdependencia es especialmente pronunciada para los tipos M1_UT (coeficiente de 0,9456) y M2_AB (coeficiente de 0,9505).

Tabla 5.5. Coeficientes de correlación para los pares de variables con al menos un coeficiente cuyo valor absoluto supera 0,95. Análisis de los conjuntos de ensayos según el tipo.

		Coeficientes de correlación					
		M1_UT	M1_TE	M1_LN	M2_UT	M2_AB	M2_SM
current	, speed	-0,8704	-0,7412	-0,9714	-0,9342	-0,9786	-0,9372
current	, slip	0,8704	0,7244	0,9716	0,9327	0,9790	0,9375
speed	, slip	-1,0000	-0,9914	-0,9730	-0,9979	-0,9973	-0,9983
amplitude_l.1	, amplitude_u.1	0,9456	0,8136	0,8759	0,9081	0,9505	0,8733

Tabla 5.6. Coeficientes de correlación medios para los pares de variables con al menos un coeficiente cuyo valor absoluto supera 0,95. Análisis de los conjuntos de ensayos según el tipo.

		Coeficientes de correlación medios		
		MEDIA M1	MEDIA M2	MEDIA TOTAL
current	, speed	-0,8610	-0,9500	-0,9055
current	, slip	0,8555	0,9497	0,9026
speed	, slip	-0,9881	-0,9978	-0,9930
amplitude_l.1	, amplitude_u.1	0,8784	0,9106	0,8945

A partir de los resultados comentados se toman una serie de decisiones, expuestas a continuación:

- Velocidad y deslizamiento presentan una contrastada interdependencia entre sí, lo cual era esperado dada su conocida relación lineal teórica, según la ecuación (2.1). A su vez, muestran una dependencia muy elevada con la corriente en la mayoría de los tipos. Esto se debe a que, para una carga determinada, el motor demanda una cantidad de corriente concreta que hace girar el rotor a la velocidad adecuada. Como en este caso el par introducido por el freno es fijo durante toda la duración del ensayo, la corriente y la velocidad se encuentran íntimamente ligadas.

A consecuencia de ello, se decide escoger una variable de entre corriente, velocidad y deslizamiento y descartar las otras dos. La pérdida de información, de este modo, será relativamente reducida con respecto a los beneficios que suponen la medición de menos características durante la experimentación y la elaboración de modelos más simples gracias a la reducción de variables.

La medición de la intensidad de corriente es la más sencilla y precisa. No obstante, el Departamento de Ingeniería Eléctrica sugiere seleccionar en su lugar la variable deslizamiento, por su carácter adimensional en tanto por uno.

- El resto de correlaciones calculadas no permiten reducir aún más el número de términos del modelo. Su valor depende excesivamente del motor y tipo de alimentación empleados. A pesar de ello, deben tenerse en cuenta los valores más altos alcanzados a la hora de analizar casos combinaciones motor-alimentación concretas.

En conclusión, las matrices de correlaciones planteadas han ofrecido resultados lo suficientemente evidentes como para eliminar del estudio la variables corriente y velocidad. En adelante, ambas variables no serán contempladas.

5.3. Normalidad de las variables

Una vez constituida una base inicial sobre el comportamiento de los datos, es el momento de acondicionarlos para poder desarrollar las técnicas de selección de características adecuadamente.

Conviene que las variables estudiadas sigan una distribución razonablemente normal, porque de este modo los métodos son más eficaces (Gardel, 2013). En el caso de no ser así, se aplicarán las transformaciones necesarias.

Ante la conclusión del apartado 5.1 de que el factor “tipo” puede ser determinante, se analiza por separado cada uno de sus niveles, pero siempre con la intención de establecer transformaciones comunes a todos los tipos, que faciliten el procedimiento a realizar por parte de quien haga uso del contenido de este trabajo. De cada variable se estudia la posibilidad de realizar alguna de las transformaciones comúnmente empleadas: la logarítmica, la raíz cuadrada y el inverso. Comparando los distintos diagramas de probabilidad normal y resultados del test de Shapiro-Wilk, se deciden realizar las siguientes transformaciones:

<i>slip</i>		
<i>amplitude_l.1</i>		
<i>amplitude_u.1</i>		
<i>amplitude_l.5</i>	→ Sin transformación	
<i>amplitude_u.5</i>		
<i>amplitude_l.7</i>		
<i>amplitude_u.7</i>		
<i>peak_l.1</i>		$\log(\text{peak}_l.1)$
<i>peak_u.1</i>		$\log(\text{peak}_u.1)$
<i>peak_l.5</i>	→ Transformación	$\log(\text{peak}_l.5)$
<i>peak_u.5</i>	logarítmica →	$\log(\text{peak}_u.5)$
<i>peak_l.7</i>		$\log(\text{peak}_l.7)$
<i>peak_u.7</i>		$\log(\text{peak}_u.7)$

Es importante señalar que en el caso de *peak_l.5*, *peak_u.5* y *peak_l.7*, la transformación más adecuada era la raíz cuadrada. No obstante, la mejoría introducida era mínima. Además, no resulta coherente aplicar distintas transformaciones a variables de la misma naturaleza. Por todo ello, se toma la decisión de considerar exclusivamente los logaritmos de los picos.

Las modificaciones mencionadas permiten mejorar la distribución, asemejándose más a una normal. Para ilustrar el comportamiento de las variables pico, la Fig. 5.5 refleja las características de la distribución antes y después de la transformación para la variable *peak_u.7* en el caso de ensayos con M1_UT.

Como se puede ver, la transformación es claramente beneficiosa, pues el estadístico W aumenta desde un valor de 0,91 hasta uno de 0,99 posterior a la transformación. También la estructura del histograma mejora bastante, aunque no sea perfectamente ideal.

De la misma forma se podría comprobar la nueva distribución ofrecida por el resto de variables según cada uno de los tipos. No siempre se alcanza una forma aproximadamente normal, pero en general las transformaciones

mejoran la distribución, contribuyendo a la eficacia de las técnicas de clasificación que serán abordadas más adelante.

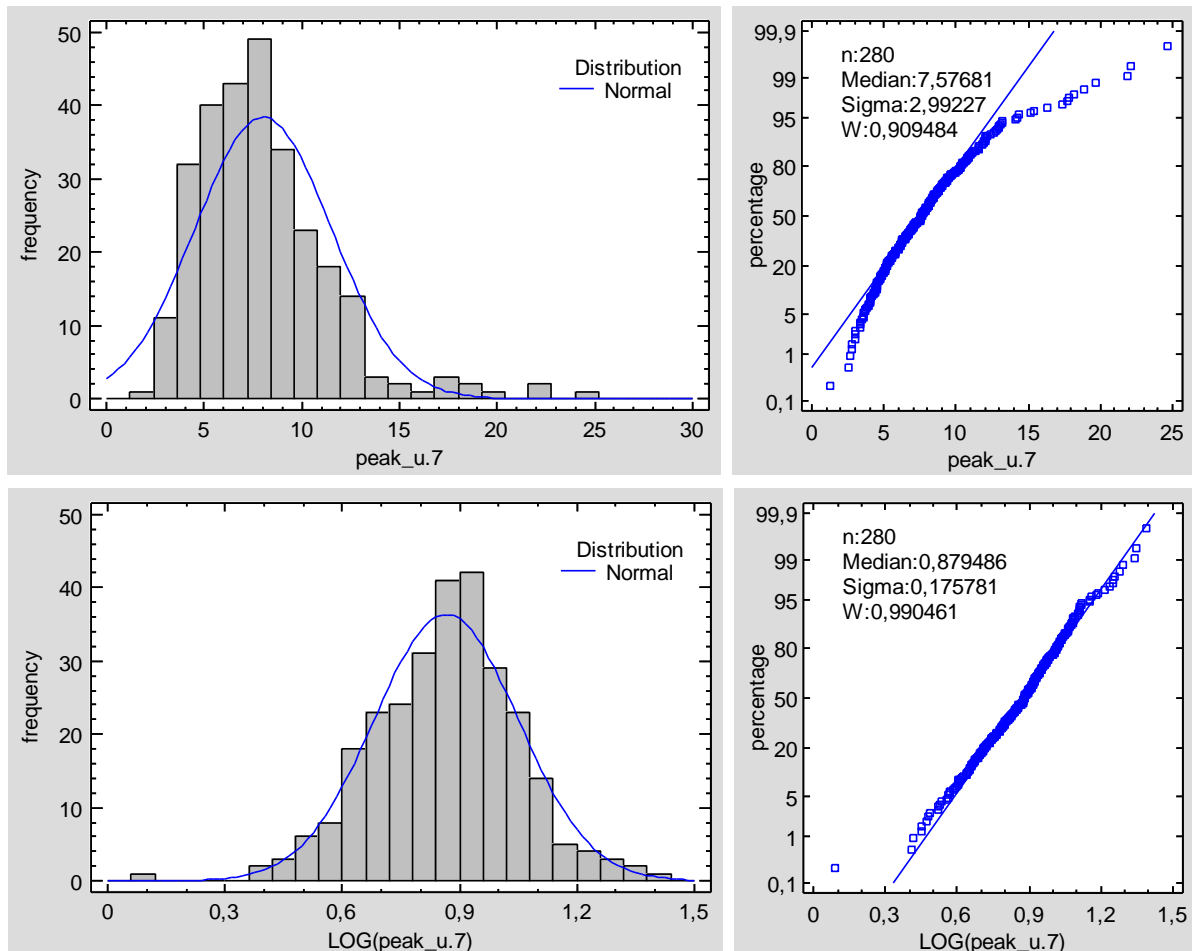


Figura 5.5. Histogramas (izquierda) de peak_u.7 (arriba) y de su transformación logarítmica (abajo) junto con los respectivos diagramas de probabilidad normal (derecha). Ensayos tipo M1_UT.

A partir de ahora, se considerarán las nuevas variables establecidas para continuar con su estudio. Por razones de simplificación de la notación, se seguirá mencionando con el mismo nombre que antes a las variables que hayan requerido transformación. De este modo, cuando hablemos de las variables pico, a lo que realmente nos referiremos será a su transformación logarítmica.

5.4. Análisis de la Varianza

Una vez obtenida una primera aproximación visual y realizadas las transformaciones oportunas, se desea saber si cambios en el estado del motor se reflejan en cambios en alguna de las variables explicativas medidas mediante un Análisis de la Varianza. Aunque el objetivo del proyecto es inverso (determinar si cambios en las variables son indicativos de la condición

del motor), el análisis planteado permitirá conocer si la variable explicativa puede ser interesante. Adicionalmente, se pretende establecer si el factor motor y el factor tipo también influyen. Si hay cambios relevantes según el motor y/o variador empleado no cabe esperar que exista una regla general de clasificación que no dependa de estos factores.

El modelo seguido por la variable de interés y_{ijkl} se expresa como

$$y_{ijkl} = \mu + M_i + T_{j(i)} + C_k + \varepsilon_{ijkl} \text{ con } \begin{cases} i = 1, 2 \\ j = 1, \dots, 3, \\ k = 1, \dots, 4 \end{cases}$$

donde μ es el término general, M_i es el efecto del motor i -ésimo, $T_{j(i)}$ es el efecto del tipo j -ésimo dentro del motor i -ésimo, C_k es el efecto de la condición k -ésima y ε_{ijkl} es el término correspondiente al error. Se trata de un modelo anidado de efectos fijos (Montgomery, 2001). El tipo está anidado en el motor, ya que dependiendo del motor se emplea una alimentación u otra.

Dado que sería excesivo y repetitivo expresar todos los análisis desarrollados estudiando cada variable respuesta, se presentan dos de ellos: uno con una variable tipo pico y otro con una tipo amplitud.

- Variable tipo pico

La Tabla 5.7 muestra el Análisis de la Varianza realizado sobre la variable peak_I.1.

Tabla 5.7. Análisis de la Varianza para LOG(peak_I.1).

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
motor	4,05209	1	4,05209	172,08	0,0000
type(motor)	15,9436	4	3,98589	169,27	0,0000
condition	31,259	3	10,4197	442,5	0,0000
Residual	29,7403	1263	0,0235473		
Total (corrected)	85,7928	1271			

Los efectos del motor, de la alimentación y de la condición son significativos, con un p -valor nulo, lo cual evidencia las diferencias entre sus distintos niveles.

Concretando estas diferencias con un Test de Rango Múltiple mediante el método LSD, se aprecian diferencias de entorno a 0,12 unidades entre los motores M1 y M2, tal y como indica la Tabla 5.8. Del mismo modo, existen diferencias estadísticamente significativas entre cada

una de las cuatro condiciones. La diferencia con respecto a un motor sano se hace mayor a medida que el fallo del motor es más severo, siendo máxima cuando el motor presenta rotura de la barra (diferencia de $-0,39$ unidades).

Tabla 5.8. Comparación múltiple para LOG(peak_l.1) por motor (izquierda) y por condición (derecha). Método LSD al 95%.

motor	Count	LS Mean	LS Sigma	Homogeneous Groups
M2	609	1,10265	0,00645034	X
M1	663	1,22078	0,0064502	X

condition	Count	LS Mean	LS Sigma	Homogeneous Groups
1	486	0,972855	0,00710472	X
2	270	1,04865	0,00948694	X
3	276	1,25978	0,00933311	X
4	240	1,36558	0,00995985	X

Contrast	Sig.	Difference	+/- Limits
M1 - M2	*	0,118124	0,017649

Contrast	Sig.	Difference	+/- Limits
1 - 2	*	-0,0757951	0,0229284
1 - 3	*	-0,286927	0,0228691
1 - 4	*	-0,392724	0,0238582
2 - 3	*	-0,211132	0,0257939
2 - 4	*	-0,316929	0,0267859
3 - 4	*	-0,105796	0,0266326

* denota una diferencia estadísticamente significativa

- Variable tipo amplitud

Análogamente a lo anterior, en esta ocasión se efectúa un Análisis de la Varianza con amplitude_l.1 como variable respuesta. La Tabla 5.9 aporta deducciones similares. De nuevo queda estadísticamente confirmada la existencia de diferencias entre los distintos niveles de motores, alimentaciones y condiciones.

Tabla 5.9. Análisis de la varianza para amplitude_l.1.

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
motor	8163,55	1	8163,55	269,93	0,0000
type(motor)	7605	4	1901,25	62,87	0,0000
condition	62969,6	3	20989,9	694,04	0,0000
Residual	38196,8	1263	30,2429		
Total (corrected)	118219	1271			

Respecto a los Test de Rango Múltiple recogidos en la Tabla 5.10, la diferencia entre los niveles medios de cada motor es de unos 5,3 dB; mientras, la rotura total de la barra aporta una desviación de aproximadamente 16,8 dB con respecto a un motor en buen estado.

Tabla 5.10. Comparación múltiple para amplitude_l.1 por motor (izquierda) y por condición (derecha). Método LSD al 95%.

motor	Count	LS Mean	LS Sigma	Homogeneous Groups	condition	Count	LS Mean	LS Sigma	Homogeneous Groups
M1	663	53,9703	0,231161	X	4	240	47,6444	0,356939	X
M2	609	59,2723	0,231166	X	3	276	51,503	0,334478	X
					2	270	62,9324	0,339991	X
					1	486	64,4053	0,254617	X

Contrast	Sig.	Difference	+/- Limits	Contrast	Sig.	Difference	+/- Limits
M1 - M2	*	-5,30199	0,632499	1 - 2	*	1,47296	0,821704
				1 - 3	*	12,9023	0,819577
				1 - 4	*	16,7609	0,855024
				2 - 3	*	11,4293	0,924395
				2 - 4	*	15,2879	0,959949
				3 - 4	*	3,85863	0,954452

* denota una diferencia estadísticamente significativa

De este análisis se intuye la importancia que tienen el motor y el variador en los picos y la amplitud, lo que sugiere que va a ser como mínimo difícil el establecer un único procedimiento, independiente de motor y variador, que permita discriminar entre los estados del motor.

5.5. Análisis de Componentes Principales

El apartado 5.1 puso de manifiesto la imposibilidad de clasificar los ensayos según la condición del motor a partir del análisis univariante de diagramas de cajas y bigotes. Por ello, el siguiente paso de nuestra investigación aborda un análisis multivariante, con mejores expectativas de éxito.

El Análisis en Componentes Principales (ACP) es idóneo para el propósito deseado, ya que crea nuevas variables, combinación lineal de las anteriores, que recogen la mayor cantidad de variabilidad posible. El objetivo principal es obtener una idea visual sobre las posibles relaciones existentes entre algunas variables y llevar a cabo una clasificación satisfactoria de los ensayos en función del estado del motor.

La Fig. 5.6 presenta el gráfico obtenido en las dos componentes principales. Se pueden extraer una serie de conclusiones observando el comportamiento de las variables.

Las variables relacionadas con la amplitud de las bandas se sitúan en la zona inferior derecha del gráfico. Tres de ellas, amplitude_u.5, amplitude_l.7 y

amplitudu.7, se encuentran muy próximas entre sí. A su vez, lo mismo sucede entre amplitudl.1 y amplitudu.1. Parece haber, así pues, una correlación positiva fuerte entre estas variables.

Por otro lado, las variables pico se sitúan cercanas entre sí, en el cuadrante superior derecho. Están correlacionadas positivamente, destacando la cercanía entre peakl.1 y peaku.1.

Por último, queda por analizar la única variable de carga del motor contemplada en el ACP: el deslizamiento. Se sitúa en la prolongación de la recta correspondiente a la variable peakl.7, lo cual podría inducir al error de afirmar que existe una interdependencia entre ambas. No es así, porque la longitud de las rectas es visiblemente distinta. Lo mismo ocurre entre peakl.1 y peaku.7.

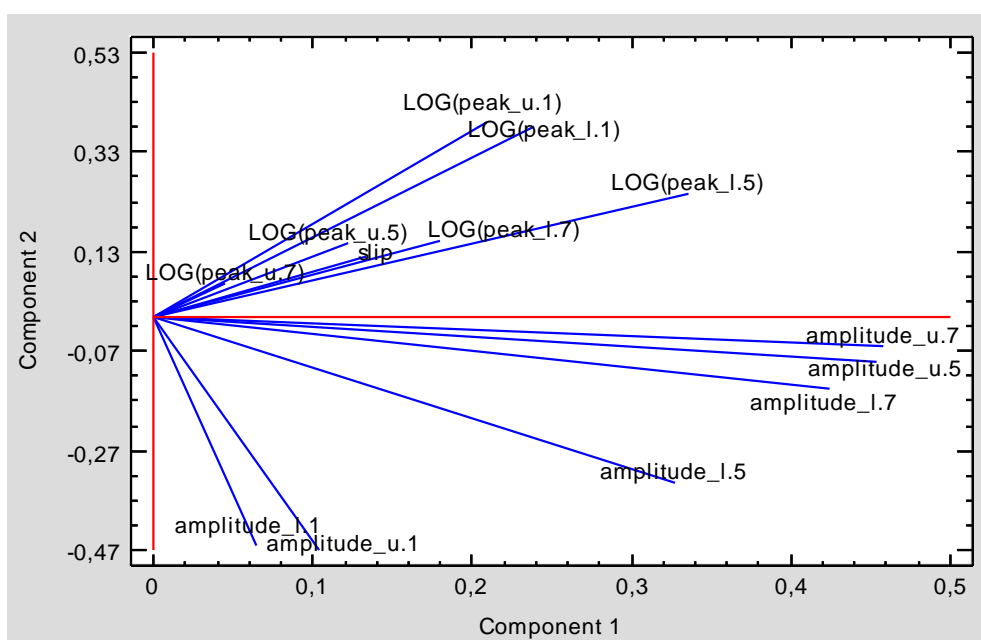


Figura 5.6. Diagrama de las dos primeras Componentes Principales.

La extracción de conclusiones a partir de un Diagrama de Componentes Principales puede ser peligrosa si no es acompañada de una indicación de la potencia de la representación. La Tabla 5.11 proporciona esta información. Tal y como indica, la calidad de la representación, calculada mediante la ecuación (4.1) es del 61,89% utilizando las dos componentes principales, como hemos hecho en el diagrama anterior. Se trata de un valor bastante pobre, ya que en torno a un 40% de la variabilidad no es reflejada por el gráfico.

Tabla 5.11. Análisis de Componentes Principales.

Número de componente	Autovalor	Variabilidad recogida (%)	Acumulado (%)
1	4,2738	32,875	32,875
2	3,77141	29,011	61,886
3	1,14644	8,819	70,705
4	0,955748	7,352	78,057
5	0,885563	6,812	84,869
6	0,77354	5,95	90,819
7	0,393908	3,03	93,849
8	0,328512	2,527	96,376
9	0,242875	1,868	98,245
10	0,0936839	0,721	98,965
11	0,0644748	0,496	99,461
12	0,0381791	0,294	99,755
13	0,0318699	0,245	100

Mediante la representación en tres dimensiones de las tres primeras componentes, la potencia aumenta hasta un 70,70% de variabilidad recogida, de modo que se efectúa el gráfico de la Fig. 5.7. Las conclusiones que ofrece son similares a las anteriores. Se aprecian hasta tres pares de variables con una correlación positiva fuerte entre sí: amplitud_e.l.1 y amplitud_e.u.1; amplitud_e.u.5 y amplitud_e.u.7; peak_e.l.1 y peak_e.u.1.

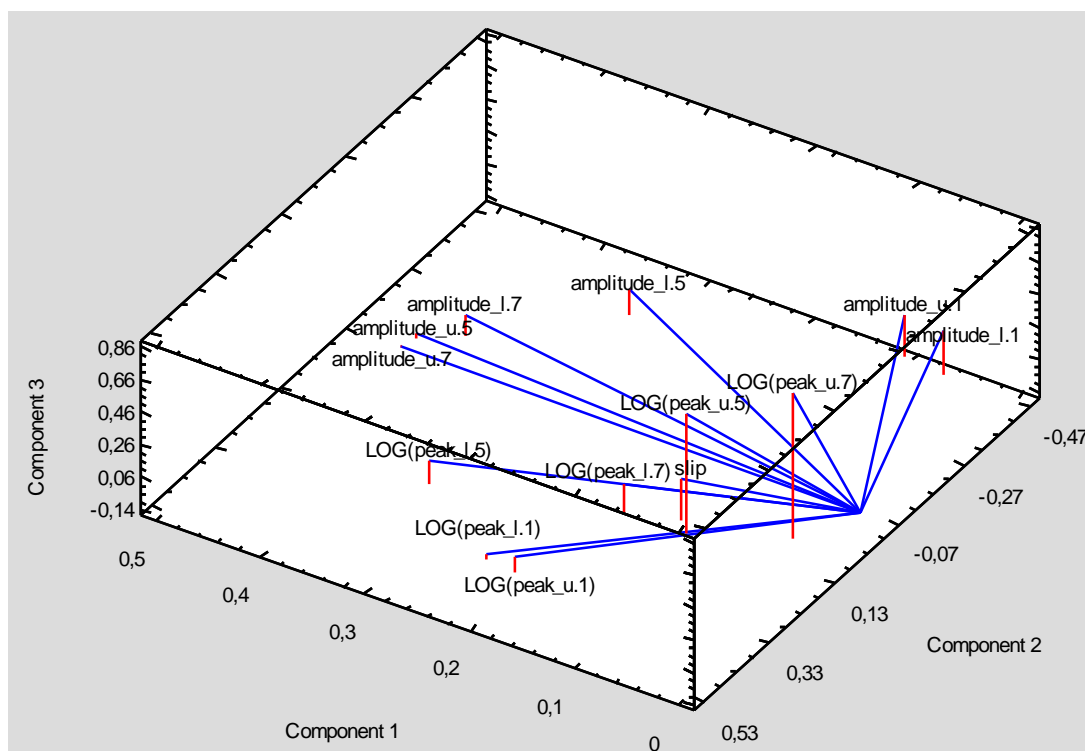
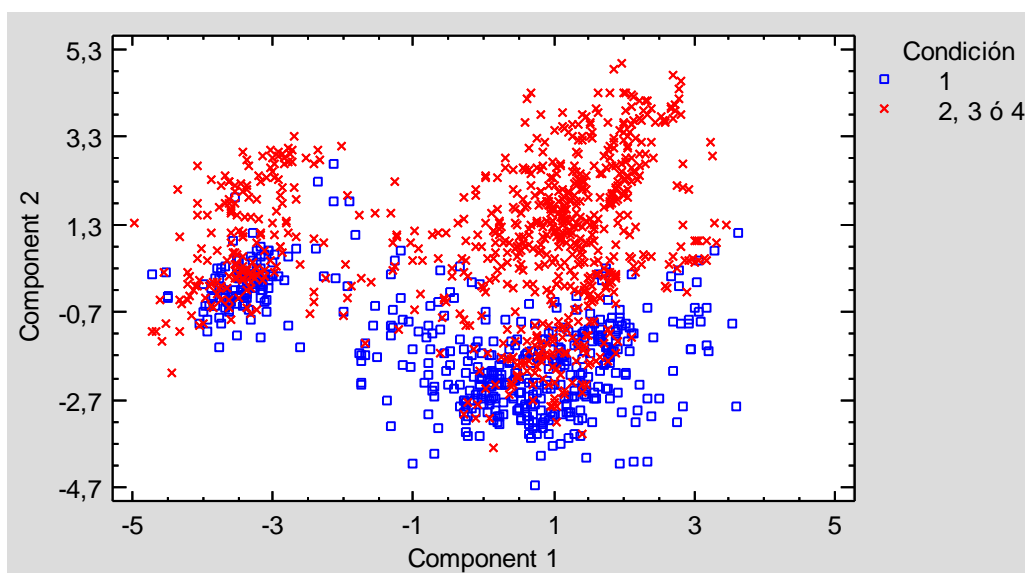


Figura 5.7. Diagrama de las tres primeras Componentes Principales.

La Fig. 5.8 refleja diagramas de dispersión de los ensayos en las dos primeras componentes principales. Con la intención de identificar grupos según la condición del motor empleado, se establecen dos gráficos: el primero (véase la Fig. 5.8.a), destaca en azul los ensayos ejecutados con un motor sano; el segundo (véase la Fig. 5.8.b), distingue entre las cuatro condiciones posibles existentes.

Aunque en la Fig. 5.8.a se aprecia que, por lo general, ensayos con motores sanos se sitúan en la parte inferior del gráfico, los colores se superponen en muchos casos. Además, se visualizan como poco dos grupos de datos: uno grande a la derecha y otro más pequeño a la izquierda.

a)



b)

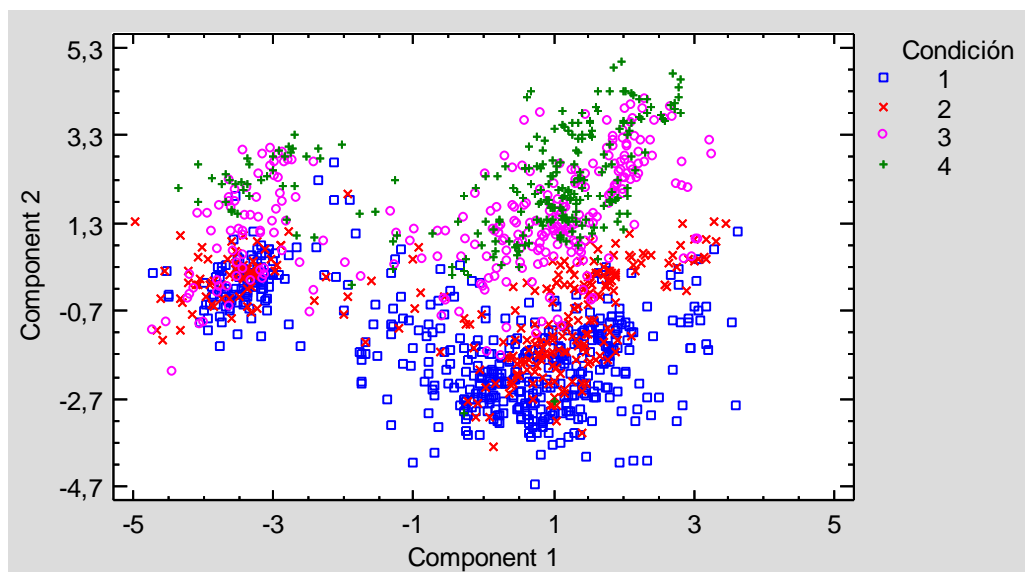


Figura 5.8. Diagramas de dispersión en las dos primeras componentes principales. Arriba (a), distinción entre los ensayos con motor sano y el resto de ensayos; abajo (b), distinción entre las cuatro condiciones posibles del motor.

Respecto a la Fig. 5.8.b, proporciona una confusión aún mayor, de la que apenas pueden ser extraídas conclusiones. Intuitivamente se observa que a medida que aumenta la severidad del fallo, aumenta el valor de la segunda componente principal.

En resumen, resulta muy complicado determinar la condición de un motor a partir de las coordenadas en las dos primeras componentes principales. Se debe procesar mejor la información para obtener resultados más favorables.

Dado que en apartados anteriores se dedujo que podía ser determinante el factor *type* en los resultados, se lleva a cabo el gráfico de la Fig. 5.9, que diferencia los ensayos con una u otra combinación motor-alimentación. Se aprecia que el pequeño grupo situado a la izquierda del diagrama anterior se corresponde en su mayoría con observaciones del tipo M1_TE. También se distingue razonablemente bien el conjunto M1_LN, ocupando la zona más a la derecha del gráfico. El resto de grupos se encuentran bastante superpuestos.

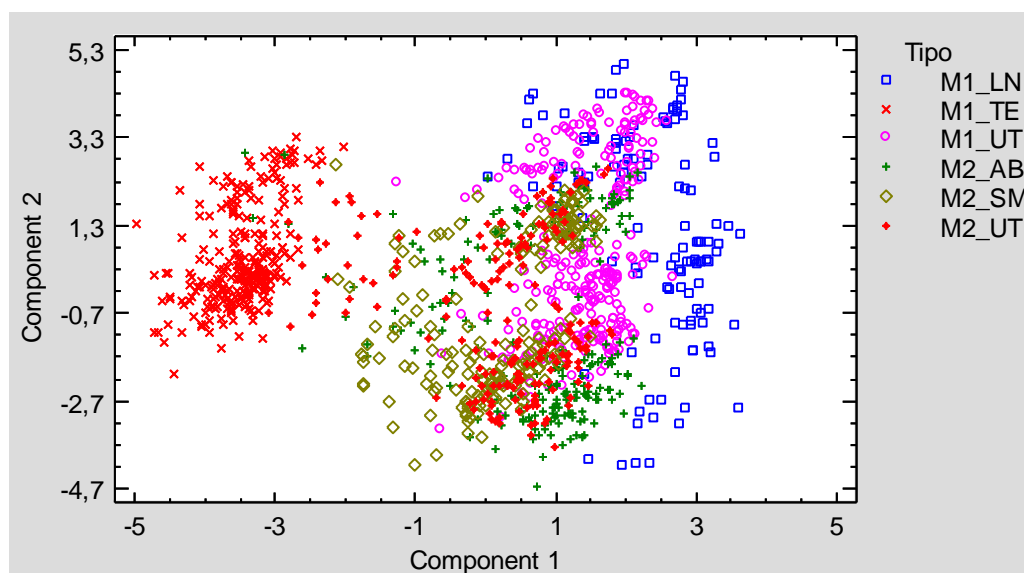


Figura 5.9. Diagrama de dispersión en las dos primeras componentes principales. Distinción según tipo.

Como conclusión del apartado, el ACP no permite clasificar con eficacia los motores según su condición. Las observaciones se encuentran demasiado superpuestas. Se debe recurrir a técnicas más potentes.

Capítulo 6

Análisis Discriminante

Tras verificar a lo largo del Capítulo 5 la incapacidad de un análisis univariante y de la discriminación mediante un ACP para clasificar los ensayos según el estado del motor, el análisis discriminante constituye la herramienta indicada para la finalidad requerida. Puede ofrecer una predicción razonable sobre la condición del motor en función de la respuesta obtenida. El apartado se divide en tres fases:

- En la fase inicial, tendrá lugar una serie de análisis discriminantes preliminares. Se pretende determinar la validez de los datos para alcanzar los propósitos deseados. Por un lado, se identificarán los análisis discriminantes idóneos y, por otro, los incorrectos. Al tratarse de un estudio general, son contempladas todas las variables que no fueron desestimadas en el apartado 5.2 por la existencia de fuertes correlaciones con otras.
- La segunda fase consiste en un exhaustivo estudio de los análisis identificados como válidos en la fase anterior. Serán utilizadas las técnicas paso a paso descritas en el apartado 4.2.3, de modo que no se contemplarán necesariamente todas las variables, sino aquellas que tienen un mayor poder discriminante.
- La fase final contiene una comparación entre el error aparente y el error estimado hallado mediante el método de validación *bootstrap* explicado en el apartado 4.2.4.

6.1. Análisis Discriminantes preliminares

Inicialmente, se contemplan todas las variables, obteniendo la clasificación resumida en la Tabla 6.1.

Tabla 6.1. Clasificación de la condición del motor.

Condición real	Cantidad	Condición predicha			
		1	2	3	4
1	485	329 (67,84%)	129 (26,60%)	24 (4,95%)	3 (0,62%)
2	270	62 (22,96%)	187 (69,26%)	21 (7,78%)	0 (0,00%)
3	276	19 (6,88%)	30 (10,87%)	125 (45,29%)	102 (36,96%)
4	240	2 (0,83%)	1 (0,42%)	63 (26,25%)	174 (72,50%)
Total	1271	412	347	233	279

Porcentaje de casos correctamente clasificados: 64,12%

La calidad de la discriminación no es aceptable: algo menos de dos de cada tres pruebas (64,12%) han sido clasificadas de forma satisfactoria. Destaca especialmente la incapacidad del modelo para identificar los motores con semi-rotura de la barra (condición 3), pues tan sólo se ha acertado en 125 de las 276 pruebas desarrolladas bajo estas condiciones, mientras que en 102 ocasiones se ha interpretado una rotura total de la barra (condición 4).

A continuación, se procede a expresar visualmente la información desprendida del análisis discriminante mediante diagramas de dispersión. En esta sección no se recoge la totalidad de los estudios realizados, sino algunos de especial interés.

El primer gráfico, ilustrado en la Fig. 6.1.a, se centra en la relación de las variables *slip* y *amplitud_u.5*. Se intuye la existencia de al menos dos (tal vez más) conjuntos de datos, situados con cierta separación uno de otro. Si se consigue identificar la fuente de aparición de los grupos, se podrá analizar cada uno de ellos por separado.

Escogiendo los ensayos llevados a cabo con el motor M1, se obtiene el gráfico de la Fig. 6.1.b. Como poco continúan apareciendo dos grupos de datos claramente diferenciados. Sin embargo, haciendo lo propio para el motor M2 (véase la Fig. 6.1.c), el grupo inferior que se apreciaba antes situado entre unos 49 y 59 dB ha desaparecido. Eso sí, la confusión de la clasificación es evidente y por ello se cree que puede haber en realidad varios grupos de datos superpuestos.

El siguiente paso es determinar si las colecciones de datos pueden ser explicadas por el tipo de variador de frecuencia empleado (o la ausencia del mismo). Para ello, se efectúan diversos análisis discriminantes sobre los datos de un único tipo de alimentación cada vez.

Por ejemplo, los ensayos con M1_UT se representan mediante el diagrama de la Fig. 6.1.d. Esta vez sí es evitada la aparición de grupos, lo cual facilita la

clasificación según el estado del motor. Aunque no se hayan reflejado los diagramas para cada uno de los tipos, en ninguno de ellos se forman agrupaciones de datos diferenciadas.

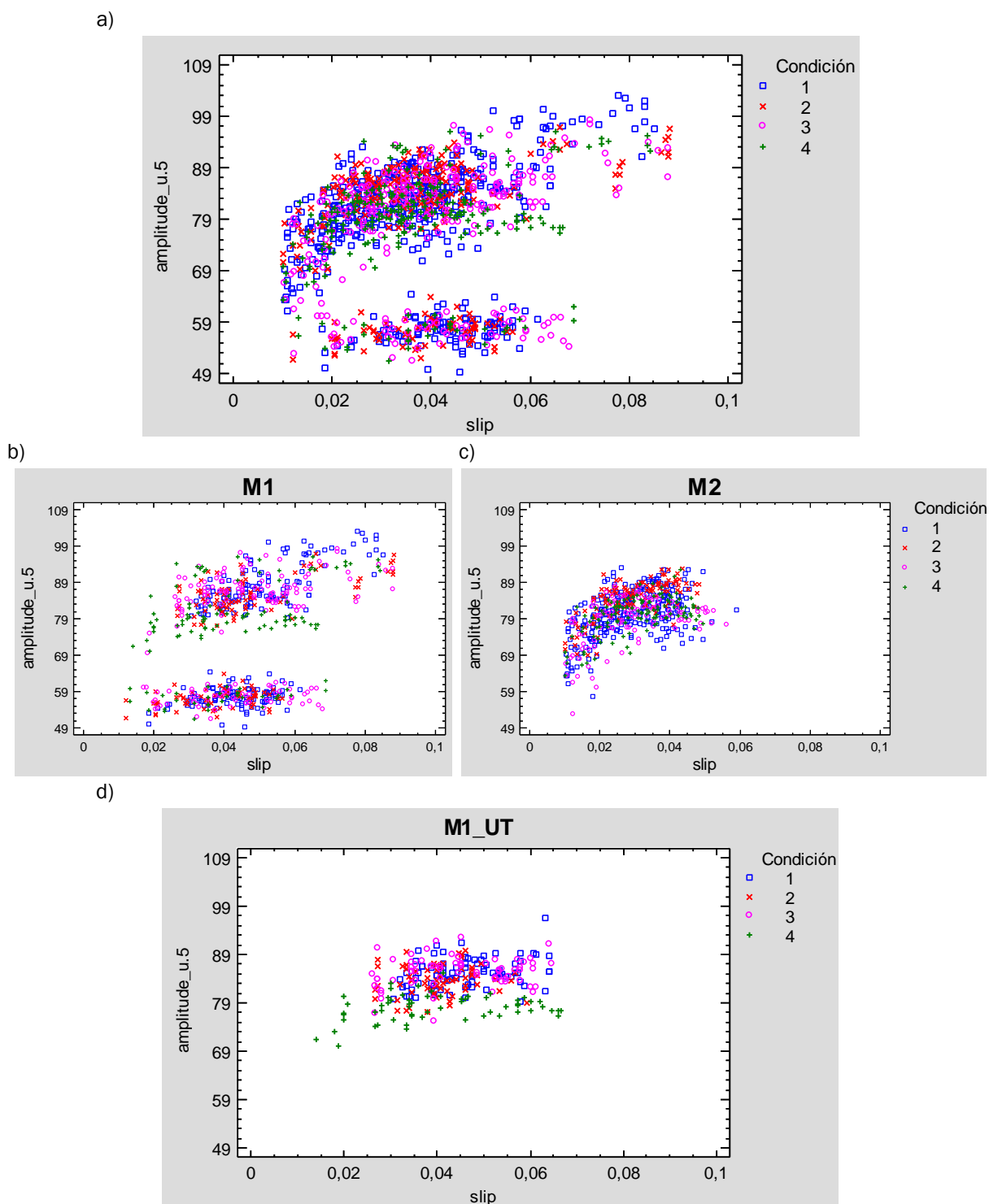


Figura 6.1. Diagramas de dispersión de slip frente a *amplitude_u.5*. Arriba (a), análisis de todos los ensayos; en el centro, análisis de los ensayos con el motor M1 (b) y los ensayos con el motor M2 (c); abajo (d), análisis de los ensayos con M1_UT.

En resumen, se puede deducir lo siguiente:

- El diagrama de dispersión ofrecido por el análisis discriminante de todos los datos no es el apropiado para una correcta clasificación de los ensayos, puesto que se distinguen agrupaciones de datos claramente diferenciadas.
- El factor motor no explica la aparición de estas agrupaciones.
- El análisis individualizado de los datos en función del tipo consigue evitar la formación de grupos diferentes en el diagrama de dispersión, lo cual es fundamental para una clasificación satisfactoria de los ensayos.

Para verificar que el último modelo contemplado es adecuado para los objetivos requeridos, la Tabla 6.2 contiene la relación entre condición real y predicción, al igual que se hizo al inicio de este apartado, cuando se estudiaba de forma general todo el conjunto de datos.

Tabla 6.2. Clasificación de la condición del motor empleando los datos obtenidos para M1_UT.

Condición real	Cantidad	Condición predicha			
		1	2	3	4
1	83	74 (89,16%)	9 (10,84%)	0 (0,00%)	0 (0,00%)
2	67	3 (4,48%)	64 (95,52%)	0 (0,00%)	0 (0,00%)
3	70	0 (0,00%)	1 (1,43%)	64 (91,43%)	5 (7,14%)
4	60	0 (0,00%)	0 (0,00%)	0 (0,00%)	60 (100,00%)
Total	280	77	74	64	65

Porcentaje de casos correctamente clasificados: 93,57%

El porcentaje de casos correctamente clasificados ha aumentado considerablemente con respecto al principio, obteniendo ahora un 93,57% de acierto (antes 64,12%). Además, prácticamente la totalidad de los motores con fallo incipiente (64 de 67 casos) y todos los motores con ruptura completa de la barra (60 de 60 casos) fueron satisfactoriamente identificados.

A modo de resumen, la Tabla 6.3 expresa los porcentajes de éxito obtenidos para la clasificación de todos los ensayos (**A**), de los ensayos según el motor (**B**) y de los ensayos según cada combinación entre motor y alimentación posible (**C**). Como se puede ver, los porcentajes aumentan a medida que el análisis es más específico, excepto en el caso de M1_TE.

La explicación de esta aparentemente extraña deficiencia clasificatoria para M1_TE podría estar relacionada con la antigüedad comparativamente mayor del dispositivo de *Telemecanique* respecto a los demás. En el apartado

5.2 se expuso que la correlación entre corriente y velocidad (o deslizamiento), era notoriamente inferior para este tipo de alimentación. Por ello, se deduce que la pérdida de información al descartar la corriente como variable explicativa es determinante a la hora de establecer una clasificación.

Si se incluye la variable current, el porcentaje aumenta hasta un 76,34% de acierto, lo cual corrobora la hipótesis planteada. Aún así, el porcentaje seguiría siendo inferior al resto de casos. Dado que interesa utilizar las mismas variables, de momento se descarta la posibilidad de introducir de nuevo en el estudio la variable current, ya que la clasificación es aceptable para las demás alimentaciones planteadas.

Tabla 6.3. Resumen del éxito de clasificación dependiendo del conjunto de datos analizado.

		Casos correctamente clasificados (%)
(A)	Totalidad de ensayos	64,12%
(B)	M1	70,74%
	M2	75,99%
(C)	M1_UT	93,57%
	M1_TE	64,89%
	M1_LN	87,60%
	M2_UT	84,58%
	M2_AB	84,04%
	M2_SM	82,99%

Desafortunadamente, la naturaleza de la experimentación no permite conocer si las diferencias que impiden una mejor clasificación del conjunto global se deben al motor o a la alimentación suministrada. Esto es debido a que el factor alimentación no ha sido cruzado con el factor motor, sino que ambos motores han usado variadores de frecuencia distintos. Tan sólo se pueden comparar los resultados obtenidos con la alimentación de red.

Así pues, se decide analizar conjuntamente los ensayos de M1_UT y de M2_UT con la intención de determinar si existen diferencias entre M1 y M2. De no ser así, podría ser posible describir una regla general para identificar el estado de los motores alimentados por la red.

El nuevo análisis discriminante planteado desprende los diagramas de dispersión de la Fig. 6.2. Al igual que antes, se representa amplitud_u.5 frente a slip.

Analizando el gráfico superior izquierdo (véase la Fig. 6.2.a), no se aprecian grupos de observaciones distintos. Se sospecha que puede haber realmente

una superposición de dos colecciones de datos correspondiéndose con el factor motor que impidan una mejor identificación de los ensayos según la condición. Tras efectuar el gráfico que distingue entre los ensayos según el motor empleado (véase la Fig. 6.2.b), se intuye lo siguiente: los ensayos con M1 se sitúan más a la derecha del gráfico (mayor deslizamiento).

Las Fig.s 6.2.c y 6.2.d presentan los análisis individualizados para cada motor. Parece que los diagramas ofrecidos son mejores, menos confusos.

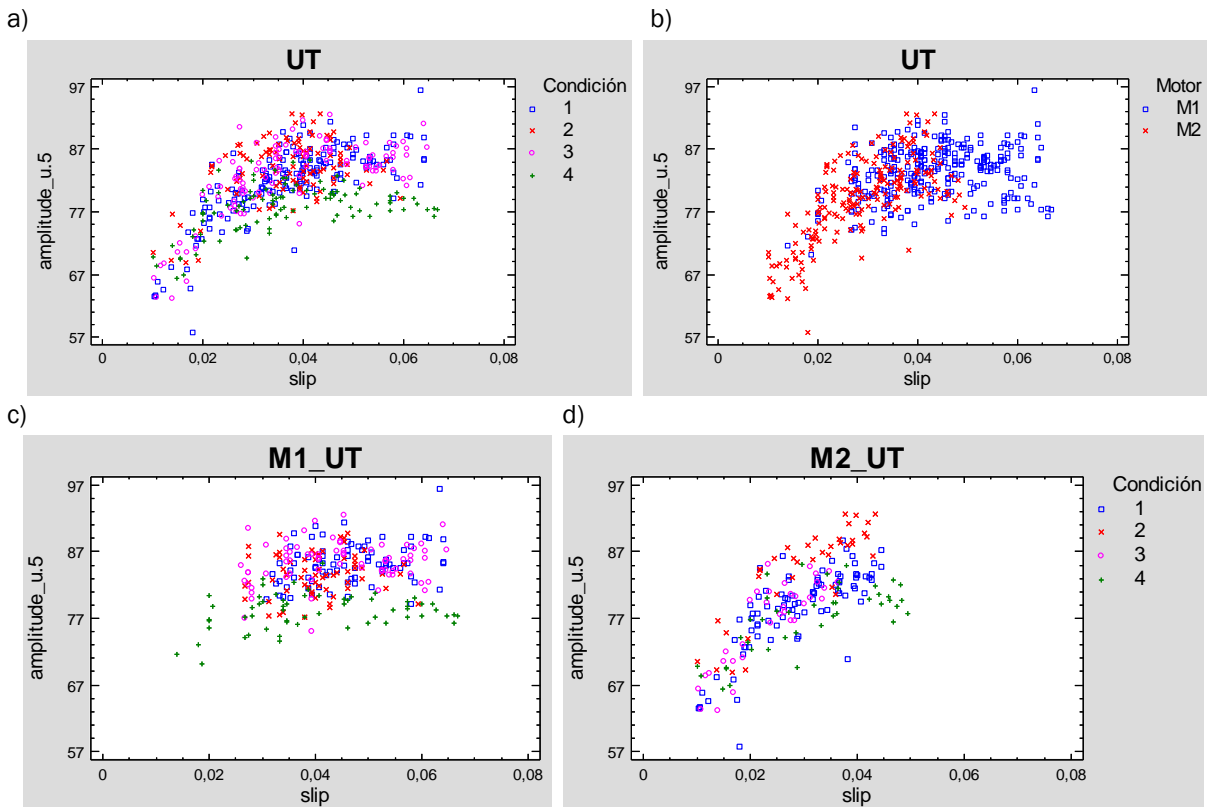


Figura 6.2. Diagramas de dispersión de *amplitude_u.5* frente a *slip*. Arriba, análisis de todos los ensayos con alimentación de red, distinción según la condición (a) y según el motor empleado (b); abajo, análisis de ensayos con M1_UT (c) y de ensayos con M2_UT (d), distinción según la condición.

Como verificación de lo afirmado, la Tabla 6.4 muestra la clasificación para el análisis de todos los ensayos con alimentación de red, utilizando cualquiera de los dos motores. El porcentaje de casos correctamente clasificados es del 79,42%.

La Tabla 6.5 compara el éxito del análisis conjunto de todos los ensayos de alimentación de red con el de los análisis específicos diferenciando según el motor. Aunque un 79,42% es un porcentaje aceptable, los análisis específicos ofrecen mejores resultados; sobre todo para M1_UT, donde el porcentaje era del 93,57%.

Tabla 6.4. Clasificación de la condición del motor empleando los datos obtenidos para alimentación de red.

Condición real	Cantidad	Condición predicha			
		1	2	3	4
1	171	136 (79,53%)	26 (15,20%)	6 (3,51%)	3 (1,75%)
2	104	11 (10,58%)	88 (84,62%)	3 (2,88%)	2 (1,92%)
3	104	1 (0,96%)	7 (6,73%)	75 (72,12%)	21 (20,19%)
4	102	0 (0,00%)	2 (1,96%)	17 (16,67%)	83 (81,37%)
Total	481	148	123	101	109

Porcentaje de casos correctamente clasificados: 79,42%

Tabla 6.5. Resumen del éxito de clasificación dependiendo del conjunto de datos analizado.

Casos correctamente clasificados (%)		
(A)	UT	79,42%
(B)	M1_UT	93,57%
	M2_UT	84,58%

Los análisis discriminantes pormenorizados coinciden con la información dispuesta en el análisis puramente descriptivo del apartado 5.1 y con el numérico del apartado 5.4. En definitiva, los datos deben ser tratados de forma individualizada en función del motor y la alimentación utilizados. Así se hará en la siguiente sección de este Capítulo.

6.2. Análisis Discriminantes finales

En adelante, serán comentados los diagramas de mayor relevancia desprendidos por varios análisis discriminantes realizados en consecuencia con los resultados de los análisis preliminares. La finalidad es la construcción de clasificadores aptos para distinguir entre los estados posibles del motor. Interesa especialmente establecer una distinción entre las siguientes combinaciones de condiciones:

- A. Distinción entre motor sano (1) y el resto de estados (2 U 3 U 4)
- B. Distinción entre motor sano (1), motor con fallo leve (2), motor con fallo medio (3) y motor con ruptura total de la barra (4).

La Tabla 6.6 indica la nomenclatura que se aplicará a partir de este momento para referirnos a cada grupo.

Antes de comenzar es necesario seleccionar las variables trascendentes desde un punto de vista discriminante. Mediante las técnicas paso a paso explicadas en el apartado 4.2.3 se excluyen del modelo las variables

irrelevantes. El criterio estadístico F es el empleado en todos los casos, siendo cuatro el valor crítico del mismo, tanto para entrar como para salir del modelo.

Tabla 6.6. Nomenclatura de los grupos dependiendo de las condiciones (1, 2, 3 ó 4) involucradas en cada combinación (A ó B).

Condición	Combinación	
	A	B
1	C1	C1
2	C234	C2
3		C3
4		C4

Como se considera que el factor *type* influye claramente en la respuesta, se analizarán por separado las pruebas según este identificador, a saber: M1_UT, M1_TE, M1_LN, M2_UT, M2_AB y M2_SM.

6.2.1. Análisis Discriminantes representativos: caso M1_UT

Debido a la repetitividad que supondría explicar todos los análisis y sus diagramas correspondientes, en este apartado sólo se expondrán y comentarán los resultados para el tipo M1_UT, a modo de ejemplo. El procedimiento para el resto de tipos es exactamente el mismo y toda la información relativa a ellos se dispone en el apartado B.1 del Anexo B, por si desea ser consultada.

Se segmenta el estudio en dos partes, según la discriminación: C1-C234 ó C1-C2-C3-C4. A su vez, dentro de cada parte se abordarán dos análisis, aplicando cada uno de ellos una metodología de selección de variables: la *Stepwise* y la *Backward*.

6.2.1.1. Discriminación C1-C234

Selección *Stepwise*

Tras aplicar el método paso a paso *Stepwise*, se reduce a 6 el número de variables del modelo: *slip*, *peak_l.1*, *amplitude_l.1*, *peak_u.5*, *amplitude_l.5* y *amplitude_l.7*. Los coeficientes estandarizados de la función discriminante adquieren los valores de la Tabla 6.7.

Tabla 6.7. Coeficientes estandarizados de la función discriminante. Tipo M1_UT. Distinción entre C1 y C234. Método *Stepwise*.

Función 1	
slip	0,43191
LOG(peak_l.1)	-0,88285
amplitude_l.1	-1,01939
LOG(peak_u.5)	0,198285
amplitude_l.5	1,2609
amplitude_l.7	0,430021

Los mayores coeficientes (en valor absoluto) se corresponden con *amplitude_l.5* y *amplitude_l.1*, en ese orden. Así pues, se trata de las variables más indicadas para representar el diagrama de dispersión, obteniéndose la Fig. 6.3. Se distingue razonablemente bien entre los dos conjuntos de datos estudiados. A priori, se podría decir que la clasificación ofrecerá buenos resultados.

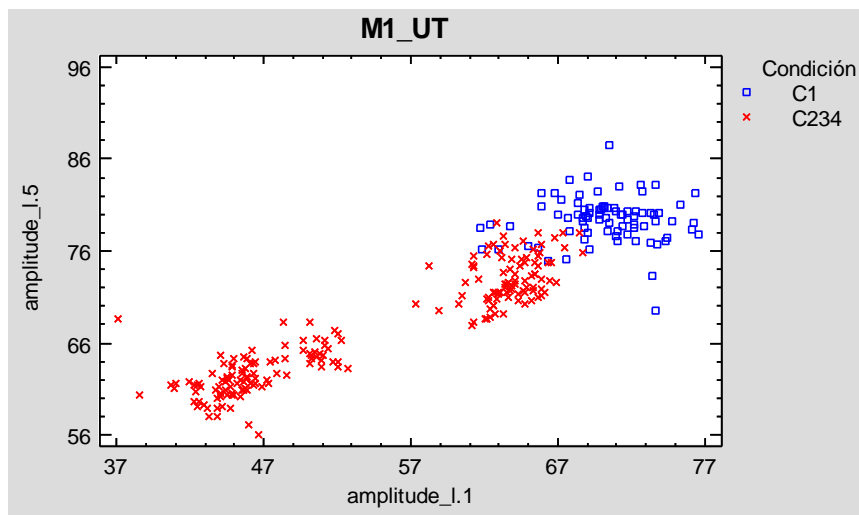


Figura 6.3. Diagrama de dispersión de la relación entre *amplitud_l.5* y *amplitud_l.1*. Tipo M1_UT. Distinción entre C1 y C234. Método *Stepwise*.

La tabla plasmada en la Tabla 6.8 lo corrobora; el porcentaje de casos bien identificados es muy elevado, de en torno al 95%. Únicamente se cometen 15 errores de los 280 casos tratados.

Tabla 6.8. Clasificación entre C1 y C234 para M1_UT. Método *Stepwise*.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	83	79 (95,18%)	4 (4,82%)
C234	197	11 (5,58%)	186 (94,42%)
Total	280	90	190

Porcentaje de casos correctamente clasificados: 94,64%

Selección *Backward*

La selección mediante el método *Backward* introduce dos variables más con respecto al método *Stepwise*: *amplitude_u.5* y *peak_l.7*. La Tabla 6.9 presenta los coeficientes estandarizados.

Tabla 6.9. Coeficientes estandarizados de la función discriminante. Tipo M1_UT. Distinción entre C1 y C234. Método *Backward*.

Función 1	
<i>slip</i>	0,344156
LOG(<i>peak_l.1</i>)	-0,894133
<i>amplitude_l.1</i>	-1,09989
LOG(<i>peak_u.5</i>)	0,34719
<i>amplitude_l.5</i>	1,28162
<i>amplitude_u.5</i>	0,22685
LOG(<i>peak_l.7</i>)	-0,229048
<i>amplitude_l.7</i>	0,296245

Como los coeficientes más destacados coinciden con las variables del método *Stepwise*, su representación es la misma que la Fig. 6.3

La clasificación obtenida se resume en la Tabla 6.10. Supone prácticamente el mismo éxito que la metodología anterior, de un 95%.

Tabla 6.10. Clasificación entre C1 y C234 para M1_UT. Método *Backward*.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	83	79 (95,18%)	4 (4,82%)
C234	197	10 (5,08%)	187 (94,92%)
Total	280	89	191

Porcentaje de casos correctamente clasificados: 95,00%

En principio mediante un *Stepwise* se obtiene un mejor resultado, pues han sido eliminadas dos variables sin que la pérdida de información haya sido significativa.

6.2.1.2. Discriminación C1-C2-C3-C4

Selección *Stepwise*

El método *Stepwise* aplicando la condición como clasificador, contempla 10 de las 13 variables objeto de estudio. Son descartadas *peak_u.1*, *peak_l.5* y *amplitude_u.7*. La Tabla 6.11 muestra los coeficientes estandarizados de la función discriminante. Al haber cuatro grupos posibles, se calculan tres

funciones, a diferencia de lo ocurrido anteriormente, donde se disponía de una única función.

Tabla 6.11. Coeficientes estandarizados de las funciones discriminantes. Tipo M1_UT. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

	Función 1	Función 2	Función 3
slip	0,230705	-0,0519031	-0,341385
LOG(peak_l.1)	0,0183883	0,117084	1,16046
amplitude_l.1	1,04222	-0,276491	1,65114
amplitude_u.1	-0,754609	1,02839	-0,591785
LOG(peak_u.5)	-0,0197049	0,026573	-0,602662
amplitude_l.5	0,918923	-0,41571	-0,36573
amplitude_u.5	0,0261569	0,723874	-0,641064
LOG(peak_l.7)	0,23334	0,275649	0,487954
LOG(peak_u.7)	0,206559	-0,124384	0,171043
amplitude_l.7	0,572204	-0,275444	0,194955

Los coeficientes más elevados en valor absoluto para la primera función se corresponden con *amplitude_l.5* y *amplitude_u.1*. Véase el diagrama de la Fig. 6.4. Llama la atención la división del grupo C3, situándose en dos zonas separadas.

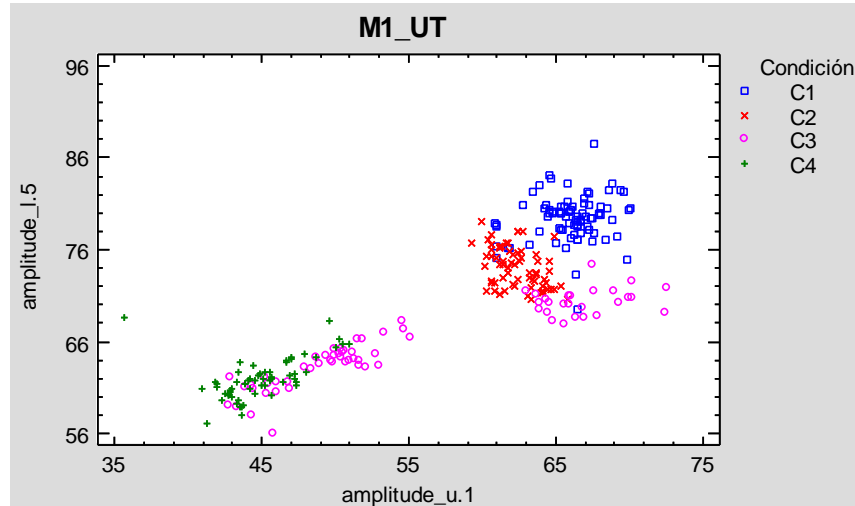


Figura 6.4. Diagrama de dispersión de la relación entre *amplitud_l.5* y *amplitud_u.1*. Tipo M1_UT. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

La Tabla 6.12 de clasificación refleja unos resultados bastante positivos. Algo más del 92% de los ensayos son satisfactoriamente clasificados. La identificación de los motores en su máximo grado de rotura es especialmente buena, del 100%.

Tabla 6.12. Clasificación entre C1, C2, C3 y C4 para M1_UT. Método *Stepwise*.

Grupo real	Cantidad	Grupo predicho			
		C1	C2	C3	C4
C1	83	73 (87,95%)	10 (12,05%)	0 (0,00%)	0 (0,00%)
C2	67	3 (4,48%)	64 (95,52%)	0 (0,00%)	0 (0,00%)
C3	70	0 (0,00%)	1 (1,43%)	61 (87,14%)	8 (11,43%)
C4	60	0 (0,00%)	0 (0,00%)	0 (0,00%)	60 (100,00%)
Total	280	76	75	61	68

Porcentaje de casos correctamente clasificados: 92,14%

En esta ocasión, al disponer de tres funciones discriminantes, es posible hacer una representación de las observaciones utilizando estas como ejes. La Fig. 6.5 lo hace empleando la primera y la segunda función discriminante. Como se puede ver, los grupos se separan bastante bien, característica indicativa de una buena clasificación.

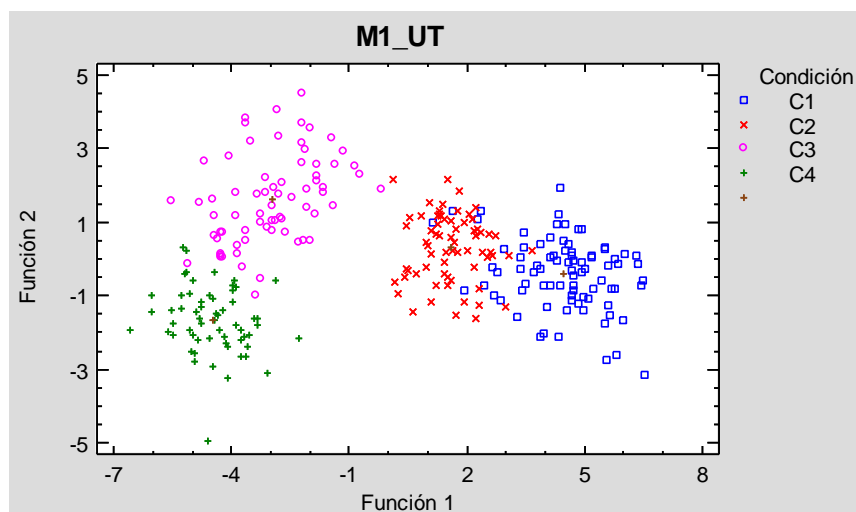


Figura 6.5. Diagrama de las funciones discriminantes 1 y 2. Tipo M1_UT. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

Selección *Backward*

Mediante el método *Backward* las variables escogidas son idénticas a las escogidas mediante el *Stepwise*. Por lo tanto, los resultados coinciden. Se trata de algo habitual, como ya indicamos en el apartado 4.2.3.

6.2.2. Resumen y conclusiones

Para cada combinación motor-alimentación posible (6 tipos) se han formulado dos análisis discriminantes (distinción C1-C234 y distinción C1-C2-C3-C4), ejecutados cada uno de ellos mediante dos metodologías de selección de variables distintas (*Stepwise* y *Backward*). Así pues, han sido

desarrollados y comentados un total de 24 análisis discriminantes entre este Capítulo y el apartado B.1 del Anexo B. En numerosas ocasiones la técnica *Backward* coincidía con la *Stepwise*, lo cual ha permitido reducir el procesamiento de la información.

Un total de 13 variables son objeto de estudio, pero los métodos de selección empleados provocan que no todas ellas sean contempladas en cada análisis. Se descartan así las variables con menor poder discriminante. Ahora bien, la gran cantidad de estudios desempeñados hace inevitable la exclusión de variables diferentes dependiendo de los elementos involucrados en el propio estudio. Es decir, dependiendo del identificador *type*, de los grupos a distinguir y de la metodología de selección, las variables relevantes desde un punto de vista discriminante cambian.

Con el propósito de vislumbrar términos predominantes en la mayoría o todos los análisis, las siguientes tablas sintetizan la información acerca de las variables escogidas en cada caso.

Por un lado, las Tablas 6.13 y 6.14 se centran en la clasificación C1-C234 mediante *Stepwise* y *Backward* respectivamente. Como se puede ver, el número de variables escogidas con *Backward* es siempre mayor o igual que con *Stepwise*. Sin embargo, la eficacia clasificatoria apenas aumenta, si es que realmente lo hace. Por consiguiente, parece que el método *Stepwise* será más favorable desde un punto de vista práctico, dado que el esfuerzo que requiere la recolección de los datos es un factor a tener muy en cuenta.

Tabla 6.13. Variables seleccionadas para cada nivel de *type*, denotadas por una "X". Distinción entre C1 ó C234. Método *Stepwise*.

Variable	M1_UT	M1_TE	M1_LN	M2_UT	M2_AB	M2_SM	FRECUENCIA
slip	X		X		X	X	4
peak_l.1	X		X		X	X	4
peak_u.1						X	1
amplitude_l.1	X	X	X				3
amplitude_u.1		X	X	X	X		4
peak_l.5				X			1
peak_u.5	X						1
amplitude_l.5	X		X	X	X		4
amplitude_u.5				X	X	X	3
peak_l.7					X		1
peak_u.7			X		X		2
amplitude_l.7	X			X			2
amplitude_u.7			X	X			2
TOTAL VARIABLES	6	2	7	6	7	4	

Tabla 6.14. Variables seleccionadas para cada nivel de *type*, denotadas por una "X".
Distinción entre C1 ó C234. Método *Backward*.

Variable	M1_UT	M1_TE	M1_LN	M2_UT	M2_AB	M2_SM	FRECUENCIA
slip	X		X	X	X	X	5
peak_l.1	X	X	X		X	X	5
peak_u.1		X				X	2
amplitude_l.1	X	X	X	X			4
amplitude_u.1			X		X	X	3
peak_l.5		X		X		X	3
peak_u.5	X			X			2
amplitude_l.5	X	X	X	X	X	X	6
amplitude_u.5	X			X	X	X	4
peak_l.7	X				X		2
peak_u.7			X		X		2
amplitude_l.7	X			X			2
amplitude_u.7			X	X			2
TOTAL VARIABLES	8	5	7	8	7	7	

Por otro lado, las Tablas 6.15 y 6.16 se centran en la clasificación C1-C2-C3-C4 mediante *Stepwise* y *Backward* respectivamente. En esta ocasión, ambas tablas coinciden exactamente excepto en el caso de M2_SM, donde el *Backward* ofrece una selección con una variable más que la *Stepwise*. De nuevo este último método es recomendado, por los mismos motivos indicados para la clasificación entre C1 y C234.

Tabla 6.15. Variables seleccionadas para cada nivel de *type*, denotadas por una "X".
Distinción entre C1, C2, C3 ó C4. Método *Stepwise*.

Variables	M1_UT	M1_TE	M1_LN	M2_UT	M2_AB	M2_SM	FRECUENCIA
slip	X		X	X	X		4
peak_l.1	X		X		X	X	4
peak_u.1					X	X	2
amplitude_l.1	X	X	X	X		X	5
amplitude_u.1	X	X	X	X	X		5
peak_l.5				X			1
peak_u.5	X						1
amplitude_l.5	X		X	X	X	X	5
amplitude_u.5	X			X		X	3
peak_l.7	X				X		2
peak_u.7	X						1
amplitude_l.7	X			X	X		3
amplitude_u.7			X	X			2
TOTAL VARIABLES	10	2	6	8	7	5	

Tabla 6.16. Variables seleccionadas para cada nivel de *type*, denotadas por una “X”. Distinción entre C1, C2, C3 ó C4. Método *Backward*.

Variable	M1_UT	M1_TE	M1_LN	M2_UT	M2_AB	M2_SM	FRECUENCIA
slip	X		X	X	X		4
peak_l.1	X		X		X	X	4
peak_u.1					X		1
amplitude_l.1	X	X	X	X		X	5
amplitude_u.1	X	X	X	X	X		5
peak_l.5				X			1
peak_u.5	X						1
amplitude_l.5	X		X	X	X	X	5
amplitude_u.5	X			X		X	3
peak_l.7	X				X	X	3
peak_u.7	X						1
amplitude_l.7	X			X	X	X	4
amplitude_u.7			X	X			2
TOTAL VARIABLES	10	2	6	8	7	6	

Las cuatro tablas expuestas incluyen en su última columna el número total de veces que ha sido seleccionada una variable. Se trata de una información que permite deducir cuáles son las variables más importantes y, por tanto, qué mediciones se deben realizar en detrimento de otras que tal vez puedan ser ignoradas. La Tabla 6.17 condensa esta información para facilitar la extracción de conclusiones.

Como se puede dilucidar, hasta 5 variables aparecen con una frecuencia claramente mayor que las demás, 17 o más veces: *slip*, *peak_l.1*, *amplitude_l.1*, *amplitude_u.1* y *amplitude_l.5*. En una posición intermedia se sitúan *amplitude_u.5* y *amplitude_l.7* con 13 y 11 apariciones respectivamente. Las demás son seleccionadas como mucho en 8 ocasiones, es decir, a lo sumo en uno de cada tres análisis (recordemos que se hacen un total de 24 análisis), lo que supone un bagaje muy pobre.

No obstante, hacer interpretaciones tan específicas no es práctico en la realidad. Conviene establecer recomendaciones genéricas, ya que la propia naturaleza aleatoria de las muestras puede provocar que una variable que realmente sea de especial interés se muestre como no relevante y viceversa. Además, para un hipotético interesado en poner en práctica los análisis sugeridos en este TFG resulta mucho más cómodo llevar a cabo mediciones con algún criterio en común. Por ejemplo, medir exclusivamente las amplitudes (y no los picos) o las variables relacionadas con el primer armónico (y no con el quinto o el séptimo) o las variables correspondientes a la banda inferior de los armónicos (y no la superior), etc.

En consecuencia, la principal singularidad apreciada en la Tabla 6.17 es que las variables asociadas al séptimo armónico son seleccionadas con poca frecuencia. Tan sólo *amplitude_l.7* entra en el estudio en más de 10 ocasiones. Por consiguiente, tal vez sea interesante desestimar la medición de las bandas alrededor de este armónico.

Por otro lado, las variables tipo amplitud parecen más relevantes que las tipo pico. Excepto *peak_l.1*, con una frecuencia de 17 veces, el resto de picos se manifiestan en pocas ocasiones, 8 como mucho. Quizá lo correcto sea centrarse exclusivamente en las variables tipo amplitud.

Aún así, es importante puntualizar que esto se trata de hipótesis no contrastadas; puede que la pérdida de información al eliminar ciertas variables sea determinante en la clasificación según qué tipo de motor y alimentación se utilicen.

Tabla 6.17. Resumen sobre el número de veces que cada variable ha sido seleccionada según el tipo de análisis.

Variable	C1-C234		C1-C2-C3-C4		TOTAL
	<i>Stepwise</i>	<i>Backward</i>	<i>Stepwise</i>	<i>Backward</i>	
<i>slip</i>	4	5	4	4	17
<i>peak_l.1</i>	4	5	4	4	17
<i>peak_u.1</i>	1	2	2	1	6
<i>amplitude_l.1</i>	3	4	5	5	17
<i>amplitude_u.1</i>	4	3	5	5	17
<i>peak_l.5</i>	1	3	1	1	6
<i>peak_u.5</i>	1	2	1	1	5
<i>amplitude_l.5</i>	4	6	5	5	20
<i>amplitude_u.5</i>	3	4	3	3	13
<i>peak_l.7</i>	1	2	2	3	8
<i>peak_u.7</i>	2	2	1	1	6
<i>amplitude_l.7</i>	2	2	3	4	11
<i>amplitude_u.7</i>	2	2	2	2	8

Tras esta primera exposición del probable grado de relevancia de las variables, el siguiente paso es comprobar la eficacia de los distintos análisis en el objetivo final: la clasificación de los ensayos según el estado del motor. La Tabla 6.18 recopila los porcentajes de casos correctamente clasificados alcanzados a lo largo de los análisis planteados en el Capítulo.

Se extraen una serie de conclusiones que sintetizamos a continuación:

- La discriminación empeora a medida que se pretenden identificar más grupos. De este modo, todos los porcentajes obtenidos para

diferenciar entre un motor sano y un motor con fallo de cualquier severidad son mayores que cuando se pretende identificar cada uno de los cuatro estados. Es algo normal, puesto que a mayor número de grupos, mayor será la dificultad de asignación a los mismos. El caso más pronunciado se detecta para M1_LN, que pasa de tener una eficacia del 100% a acertar en el 84,30% de los ensayos.

- La metodología de selección de variables *Backward* clasifica igual o mejor que la *Stepwise*. Sin embargo, esta mejora es muy reducida (como mucho del 3% aproximadamente) y además supone siempre la inclusión de un número mayor de variables (como se vio en la Tabla 6.17). Dado que cuantas menos mediciones sea necesario hacer mayor será la comodidad y rapidez en el proceso de recogida de datos, se puede concluir que la técnica *Stepwise* es más eficiente. Desprende resultados similares reduciendo al máximo la cantidad de medidas a tomar.
- El éxito clasificatorio es razonablemente bueno para todos los motores y alimentaciones excepto para M1_TE. Un 30% de error para la distinción entre motor sano y el resto de estados, y casi un 40% de error para la distinción entre los cuatro estados, son porcentajes inaceptables. Como se ha indicado anteriormente, la explicación más probable a este comportamiento es que el variador de frecuencia de *Telemecanique* es un dispositivo anticuado y manifiestamente menos sofisticado que los demás. Así, la eliminación de la variable *current* en el apartado 5.2 supuso una pérdida determinante de información.

Tabla 6.18. Resumen sobre los porcentajes de casos correctamente clasificados hallados mediante los análisis discriminantes diseñados.

CASOS CORRECTAMENTE CLASIFICADOS (%)				
DISCRIMINACIÓN	C1-C234		C1-C2-C3-C4	
	<i>Stepwise</i>	<i>Backward</i>	<i>Stepwise</i>	<i>Backward</i>
M1_UT	94,64%	95,00%	92,14%	92,14%
M1_TE	70,23%	72,90%	61,45%	61,45%
M1_LN	100,00%	100,00%	84,30%	84,30%
M2_UT	92,54%	92,54%	82,59%	82,59%
M2_AB	89,67%	89,67%	84,98%	84,98%
M2_SM	93,30%	93,81%	78,35%	80,93%

6.3. Validación del Análisis Discriminante

Uno de los principales inconvenientes de abordar la clasificación de observaciones mediante el análisis discriminante de la forma en que se ha realizado hasta ahora, es que los propios datos han servido para calcular las funciones discriminantes que los categorizan. De este modo, el comportamiento de los mismos hacia las reglas clasificatorias suele ser mejor que si se estudia un nuevo conjunto de datos sin el cual se hayan construido las funciones discriminantes, tal y como se expuso en el apartado 4.2.4. Por consiguiente, los porcentajes de casos satisfactoriamente clasificados obtenidos en el apartado anterior no son del todo fiables, sino que tan sólo ofrecen una indicación sobre el error aparente. Este se define según la ecuación (4.2) del apartado 4.2.4, y a efectos prácticos se entiende como

$$e = 1 - c$$

donde

e: error aparente en tanto por uno

c: casos correctamente clasificados en tanto por uno

En esta sección, el método *bootstrap* explicado en el apartado 4.2.4 servirá como técnica de validación de los resultados. Tendrá lugar el cálculo del error estimado por medio de un remuestreo durante 500 iteraciones. Se tomará repetidamente y de forma aleatoria, una muestra con reemplazamiento, de tamaño igual al número de ensayos realizados, que constituirá la muestra de entrenamiento. El resto de observaciones no seleccionadas en cada iteración constituirán las respectivas muestras de test.

Para poder comparar el error estimado con el error aparente se realiza el análisis individualizado según la combinación motor-alimentación, al igual que se hizo en el apartado anterior. Es importante señalar que, mientras que el tamaño de la muestra de entrenamiento es siempre fijo e igual al número total de ensayos, el tamaño de la muestra de test varía en cada iteración como consecuencia del muestreo aleatorio. Es decir, el tamaño de la muestra de test no es fijo y no puede ser conocido hasta después de efectuar el muestreo aleatorio propio del método de validación.

La Tabla 6.19 indica el tamaño de las muestras. El tamaño de la muestra de test es calculado como la media de las 500 iteraciones ejecutadas. Como se ha aprovechado la misma semilla para obtener la muestra de entrenamiento para todo tipo de discriminación (C1-C234 ó C1-C2-C3-C4) y metodología de selección (*Stepwise* o *Backward*), el tamaño medio de la muestra de test es el mismo en todos los casos.

Tabla 6.19. Tamaño de las muestras en función del identificador *type*.

TYPE	Nº TOTAL ENSAYOS	TAMAÑO ENTRENAM.	TAMAÑO MEDIO TEST
M1_UT	280	280	102,61
M1_TE	262	262	96,31
M1_LN	121	121	44,35
M2_UT	201	201	73,63
M2_AB	213	213	78,23
M2_SM	194	194	71,43

En el caso de M1_LN las muestras son especialmente reducidas. Recordemos que en el inicio del Capítulo 5 se tomó la decisión de desestimar gran cantidad de ensayos de este tipo por motivos relacionados con un deslizamiento escaso o excesivo. No obstante, se considera que el tamaño es suficientemente amplio como para ofrecer conclusiones fiables.

A lo largo del Capítulo los análisis se han orientado a remarcar el porcentaje de casos correctamente clasificados, y no el error aparente. Para facilitar la comparativa, aunque el *bootstrap* se haya enfocado para el cálculo del error estimado, en adelante las tablas de comparación de resultados indicarán los aciertos aparente y estimado. El acierto aparente no es más que la proporción de casos correctamente clasificados, mientras que el acierto estimado es la proporción restante a la correspondiente al error estimado. Ambos serán expresados en términos porcentuales.

A continuación, se desarrolla el apartado estructurándolo en dos secciones dependiendo de los grupos de clasificación: C1-C234 ó C1-C2-C3-C4.

6.3.1. Validación discriminación C1-C234

La Tabla 6.20 expresa los aciertos estimados junto con los aciertos aparentes derivados de la discriminación entre los grupos C1 y C234. Para los ensayos del tipo M1_LN o M2_AB la selección *Stepwise* coincidía con la *Backward*, así que se aprovecha un único análisis.

Si se observa detenidamente la tabla, se aprecia que el acierto estimado es siempre inferior al acierto aparente. Se trata de algo que cabía esperar, dado que, como se ha mencionado, la eficacia empeora si el conjunto de datos no ha formado parte de la construcción de los propios clasificadores. En cualquier caso, acierto aparente y acierto estimado no distan demasiado, lo cual es indicativo de que el análisis discriminante entre C1 y C234 ofrece una buena respuesta.

Tabla 6.20. Comparación entre acierto aparente y acierto estimado, en porcentaje. Clasificación en grupos C1 y C234.

ACIERTO SELECCIÓN	APARENTE		ESTIMADO	
	<i>Stepwise</i>	<i>Backward</i>	<i>Stepwise</i>	<i>Backward</i>
M1_UT	94,64%	95,00%	94,32%	94,59%
M1_TE	70,23%	72,90%	70,08%	70,99%
M1_LN	100,00%	100,00%	99,89%	99,89%
M2_UT	92,54%	92,54%	92,39%	88,96%
M2_AB	89,67%	89,67%	87,21%	87,21%
M2_SM	93,30%	93,81%	91,83%	91,68%

Aunque el acierto estimado es un valor fidedigno sobre el éxito clasificatorio de un análisis discriminante, no aporta ninguna indicación sobre las iteraciones desarrolladas para su cálculo. Así pues, con la intención de complementar la información de la tabla anterior, en la Fig. 6.6 se presentan los diagramas de cajas del error cometido en cada una de las 500 iteraciones según el análisis desempeñado. La principal característica percibida es el aumento de la dispersión del diagrama cuanto mayor es el valor medio del error. El número de fallos cometidos en la clasificación de los ensayos M1_TE presenta la mayor variabilidad, mientras que para M1_LN el acierto está prácticamente garantizado.

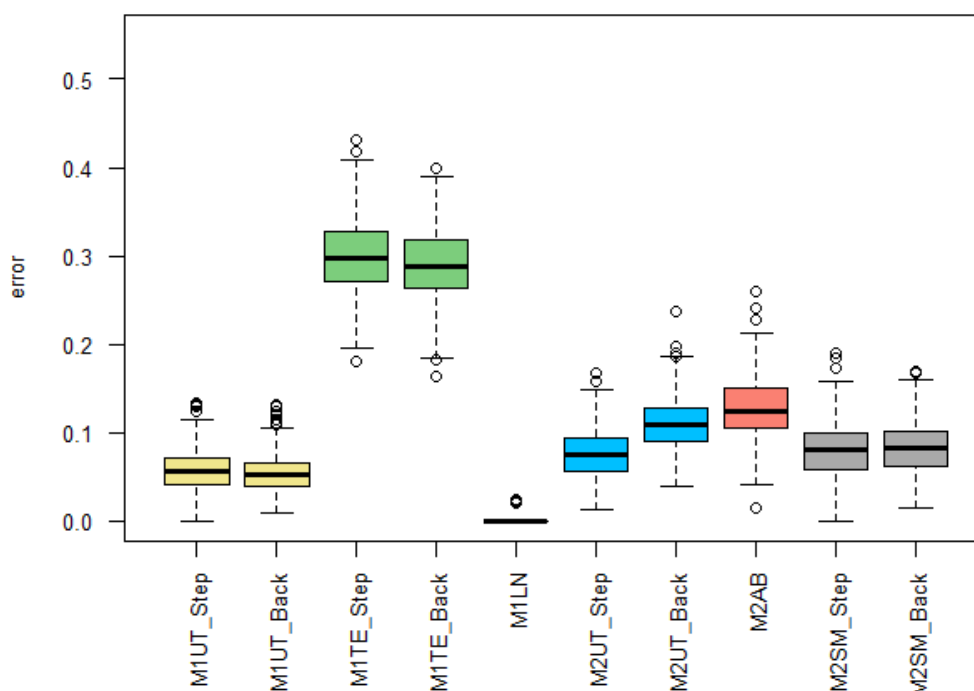


Figura 6.6. Diagramas de cajas del error en función del tipo de alimentación y motor. Distinción entre C1 y C234. Métodos *Stepwise* (Step) y *Backward* (Back). En el caso de M1_LN y M2_AB ambos métodos coinciden, por lo que se muestra un único *plot*.

Por otro lado, se deduce que la metodología de selección de variables, *Stepwise* o *Backward*, no es determinante en la amplitud del diagrama. Además, la media es similar en ambos casos, por lo que se podría decir que cualquiera de los dos métodos promete resultados equivalentes en cuanto a validez se refiere.

6.3.2. Validación discriminación C1-C2-C3-C4

Análogamente a lo anterior, la Tabla 6.21 muestra los aciertos aparentes y estimados derivados de la discriminación entre C1, C2, C3 y C4. La selección *Stepwise* y la *Backward* coincidían en todos los casos excepto para M2_SM, donde consecuentemente los aciertos reflejados son distintos.

De nuevo, en todo momento el acierto estimado supera al aparente. Destacan notoriamente los aciertos estimados para M1_LN y para el método *Backward* de M2_SM, ya que suponen aproximadamente un 4 y un 10% menos que el valor de los aciertos aparentes, respectivamente.

Por lo general, los errores estimados se han incrementado más que en el caso de la discriminación entre C1 y C234, en relación a los errores aparentes. Se trata de algo previsible, dado que el aumento del número de grupos añade una complicación extra en la clasificación. Mediante la técnica *bootstrap* se ha manifestado dicho aumento de dificultad.

Tabla 6.21. Comparación entre acierto aparente y acierto estimado, en porcentaje. Clasificación en grupos C1, C2, C3 y C4.

ACIERTO SELECCIÓN	APARENTE		ESTIMADO	
	<i>Stepwise</i>	<i>Backward</i>	<i>Stepwise</i>	<i>Backward</i>
M1_UT	92,14%	92,14%	92,09%	92,09%
M1_TE	61,45%	61,45%	60,62%	60,62%
M1_LN	84,30%	84,30%	80,53%	80,53%
M2_UT	82,59%	82,59%	79,40%	79,40%
M2_AB	84,98%	84,98%	83,18%	83,18%
M2_SM	78,35%	80,93%	76,93%	70,45%

La Fig. 6.7 representa los diagramas de cajas del error en función del análisis ejecutado. En esta ocasión no se distingue con tanta claridad el aumento de la variabilidad con el aumento del error. Resulta llamativa la pronunciada diferencia entre el *boxplot* de M1_LN obtenido en esta ocasión frente al *boxplot* de M1_LN extraído con anterioridad para la clasificación entre C1 y C234. Tanto la media como la dispersión han aumentado considerablemente.

Adicionalmente, la diferencia entre los dos diagramas de cajas de M2_SM pronostica una mejor clasificación mediante la metodología *Stepwise* de selección de variables, ya que el error medio cometido es inferior y de variabilidad equiparable.

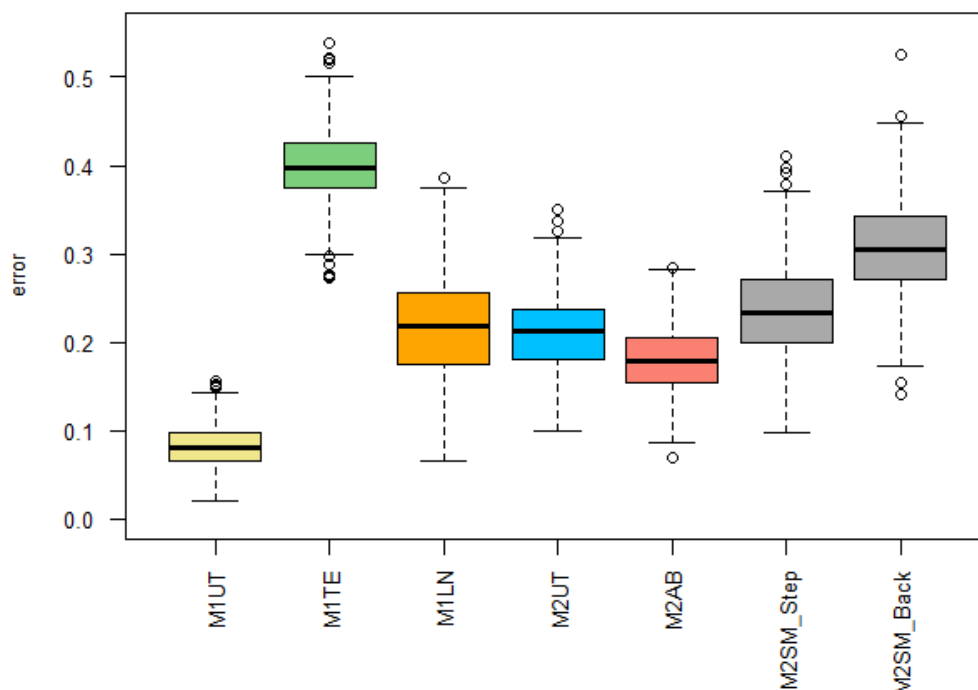


Figura 6.7. Diagrama de cajas del error en función del tipo de alimentación y motor. Distinción entre C1, C2, C3 y C4. Métodos *Stepwise* (Step) y *Backward* (Back). En el caso de M1_UT, M1_TE, M1_LN, M2_UT y M2_AB ambos métodos coinciden, por lo que se muestra un único *plot*.

6.3.3. Conclusiones sobre la validación

La técnica de validación *bootstrap* proporciona una idea fiel sobre el comportamiento real de futuros datos ante un análisis discriminante. Se han manifestado ciertas diferencias entre acierto aparente y acierto estimado, siendo este último siempre inferior.

Para la clasificación entre un motor sano y un motor con anomalía en cualquier grado el acierto estimado es muy similar al aparente. En cuanto a la clasificación entre los cuatro estados, la diferencia entre ambos porcentajes de acierto aumenta, aunque por lo general es aceptable.

En conclusión, se puede afirmar que los análisis discriminantes planteados y la medición de variables que requieren son adecuados. El método *bootstrap* los valida.

Capítulo 7

Análisis del efecto del tipo y del deslizamiento

Como se ha visto a lo largo del desarrollo del presente TFG, la combinación de motor y alimentación empleada es determinante en el comportamiento de la respuesta. Se podría considerar un Modelo Aditivo Generalizado (GAM) en el que cada una de las variables y relacionadas con los armónicos está compuesta por una media general μ , por la variabilidad adicional introducida por el tipo y por el error ε :

$$y \sim \mu + type + \varepsilon$$

Ahora bien, el Departamento de Ingeniería Eléctrica percibió de forma intuitiva durante la recolección de datos que la altura de las bandas registradas aumentaba a medida que el deslizamiento también lo hacía. Por ello, se sospecha que el deslizamiento también podría ser un factor significativo que añade cierta variabilidad a la respuesta. A diferencia del factor *type*, el deslizamiento no adopta niveles fijos, sino que es una variable explicativa de carácter continuo. La nueva hipótesis planteada se refleja añadiendo un término correspondiente a una función del deslizamiento:

$$y \sim \mu + type + s(slip) + \varepsilon$$

Si efectivamente se cumple la suposición formulada, puede que los residuales de las funciones descritas sean de gran interés y proporcionen una mayor eficacia clasificatoria en los análisis discriminantes. Se definen como:

$$r = y - (\hat{\mu} + \widehat{type} + s(\widehat{slip}))$$

r : residual

y : observación

$\hat{\mu}$: estimador de la media

\widehat{type} : estimador del efecto del tipo

$s(\widehat{slp})$: estimador del efecto del deslizamiento

Por lo tanto, en el presente Capítulo se llevarán a cabo las siguientes tareas:

1. Comprobar la existencia del efecto del deslizamiento para cada una de las variables mediante el GAM descrito.
2. Restar el efecto del tipo y del deslizamiento mediante el registro de los residuales. Estos serán las nuevas variables de interés.
3. Comparar de forma visual los diagramas de cajas de los residuales con los de las variables originales, ya procesadas con las transformaciones necesarias sugeridas en el apartado 5.3.

Después de las tareas expuestas, si los resultados son esperanzadores se pondrá en marcha el estudio del comportamiento de los residuales ante exhaustivos análisis discriminantes lineales. El objetivo es determinar si la eficacia clasificatoria mejora en cierta medida con estas nuevas variables de interés y, en caso afirmativo, estimar cuánto lo hace. El Capítulo 8 abordará esta labor.

A pesar de que lo estrictamente correcto sería señalar de algún modo que cada variable responde a un GAM distinto, a lo largo del Capítulo se evitará emplear subíndices en la notación del modelo planteado. De este modo se simplifica la comprensión de la expresión.

7.1. Modelo Aditivo Generalizado para cada variable

Todo el proceso descrito se desarrollará para cada una de las variables: $peak_l.1$, $peak_u.1$, $amplitude_l.1$, $amplitude_u.1$, $peak_l.5$, $peak_u.5$, $amplitude_l.5$, $amplitude_u.5$, $peak_l.7$, $peak_u.7$, $amplitude_l.7$ y $amplitude_u.7$. El objetivo es registrar los residuales para comprobar posteriormente si resultan beneficiosos a la hora de discriminar con respecto a los datos originales.

Ante la similitud del comportamiento de ciertas variables, a continuación se presentan únicamente algunos análisis representativos, divididos en dos secciones, según el tipo de variable: pico o amplitud. El resto de análisis se muestran en el apartado B.2 del Anexo B, para mayor información.

7.1.1. GAM sobre variables tipo pico

Comenzamos con el modelo aditivo planteado para $peak_l.1$, definido como

$$peak_l.1 \sim \mu + type + s(slip) + \varepsilon.$$

La representación obtenida se corresponde con la Fig. 7.1. La influencia del variador en el valor final de la variable es evidente. No existe solapamiento entre la mayoría de los intervalos de confianza estimados para cada tipo y en la mayoría de casos los propios intervalos no contienen al cero.

Por su parte, el efecto debido al deslizamiento parece de menor magnitud, aunque sí que se aprecia. Para valores muy bajos de deslizamiento el pico es menor, mientras que para valores altos el pico aumenta, hasta que llega a un valor cercano 0,07 donde de nuevo se atenúa.

Aunque la magnitud y la forma de la función del deslizamiento difieren según la variable, todos los picos muestran efectos significativos, tanto del tipo como del deslizamiento.

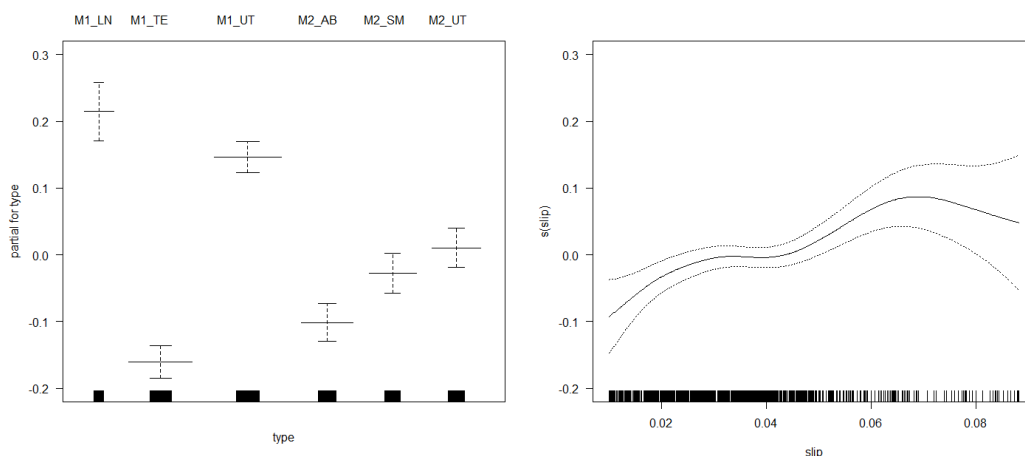


Figura 7.1. Gráficos ofrecidos por el GAM sobre peak_l.1. A la izquierda, efecto del tipo; a la derecha; efecto del deslizamiento.

Volviendo al caso específico de peak_l.1, dado que con los residuales del GAM formulado se descuenta el efecto del tipo y del deslizamiento sobre cada observación, es posible que tengan un mayor interés que las variables originales. El programa **R** permite registrarlos.

La Fig. 7.2 muestra la comparativa entre los diagramas de cajas de los residuales (a la izquierda) y los diagramas de cajas de la variable original (a la derecha), según el grado de rotura del motor. El solapamiento de las cajas se ha reducido considerablemente, lo cual parece indicar una mejor clasificación univariante según el pico de la banda inferior del armónico fundamental.

Este fenómeno se manifiesta también para la variable peak_u.1. Sin embargo, para el resto de picos no. En estos los gráficos adoptan un aspecto similar al de la Fig. 7.3, que representa el caso particular de la variable peak_l.7. Como se puede apreciar, el solapamiento es aproximadamente el

mismo con respecto a las variables originales, así que aquí no se espera que los residuales aporten beneficios.

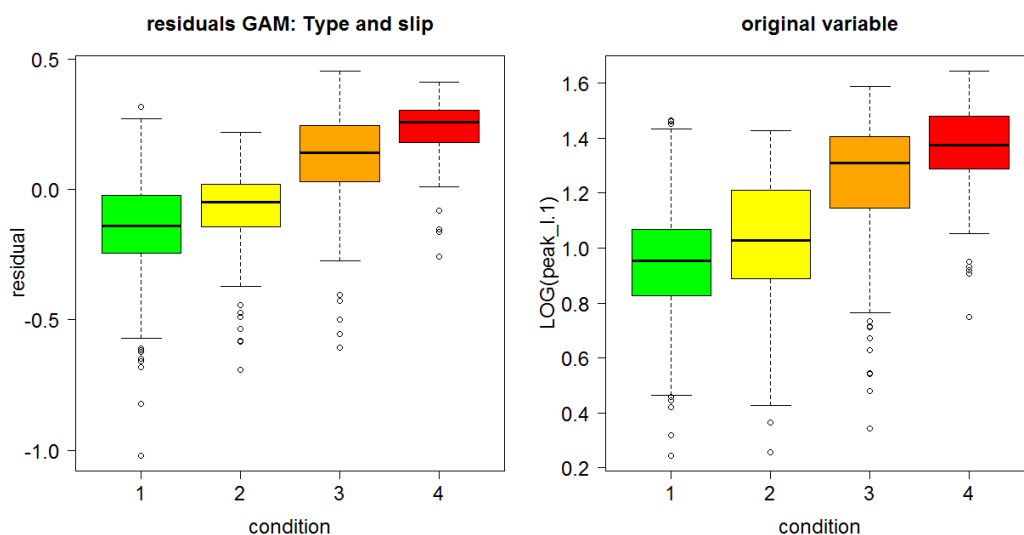


Figura 7.2. Diagramas de cajas relacionados con la variable peak_l.1 en función de la condición del motor. A la izquierda, diagramas de los residuales obtenidos mediante el GAM; a la derecha, diagramas de la variable original.

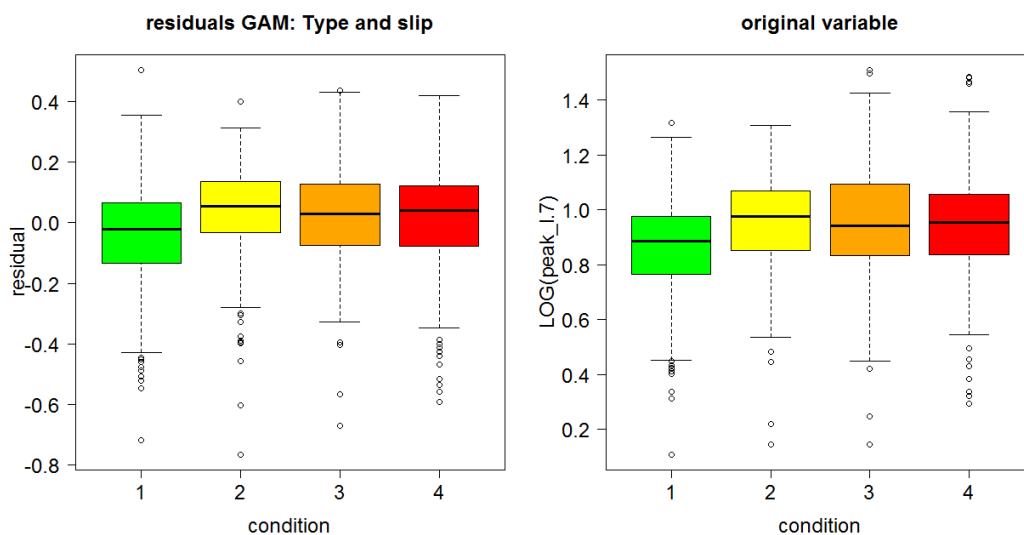


Figura 7.3. Diagramas de cajas relacionados con la variable peak_l.7 en función de la condición del motor. A la izquierda, diagramas de los residuales obtenidos mediante el GAM; a la derecha, diagramas de la variable original.

7.1.2. GAM sobre variables tipo amplitud

El GAM asociado a la variable amplitud_u.1 es representativo de gran parte de las variables tipo amplitud. Se formula como

$$amplitud_u.1 \sim \mu + type + s(slip) + \varepsilon.$$

Como se puede ver en la Fig. 7.4, tanto el efecto del tipo como el del deslizamiento adquieren una forma distinta a los que tenían los gráficos asociados a la variable peak_l.1. Pero, en cualquier caso, la significancia de los efectos es indudable.

En el deslizamiento se aprecia un aumento del intervalo de confianza para valores altos. Si bien también aparecía en el gráfico de peak_l.1, ahora contrasta más con respecto al reducido intervalo de confianza para valores pequeños de deslizamiento. Pero sin duda, el aspecto más reseñable de la función es el efecto negativo a valores bajos y el efecto positivo a valores altos de deslizamiento.

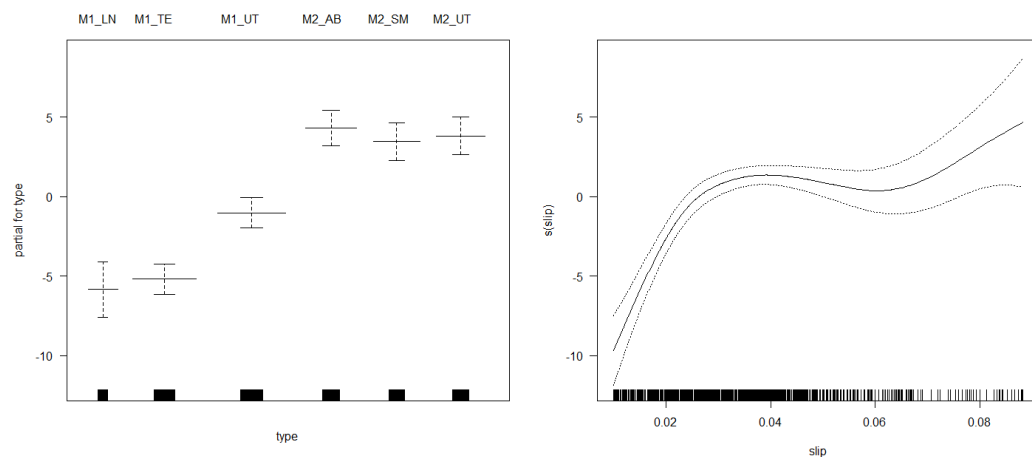


Figura 7.4. Gráficos ofrecidos por el GAM sobre amplitude.1. A la izquierda, efecto del tipo; a la derecha; efecto del deslizamiento.

Los *boxplots* comparativos entre los residuos del GAM y los datos originales de la Fig. 7.5 son de nuevo esperanzadores. Se reduce considerablemente el solapamiento de las cajas.

Esta característica se aprecia también al analizar los residuales asociados a los GAM sobre las variables amplitude.1, amplitude.5 y amplitude.5. Sin embargo, para amplitude.7 y amplitude.7, aunque la amplitud de los diagramas disminuye, el solapamiento vuelve a ser bastante grande. La Fig. 7.6 ilustra la situación para amplitude.7. En principio, no parece que los residuales vayan a mejorar la discriminación en este caso.

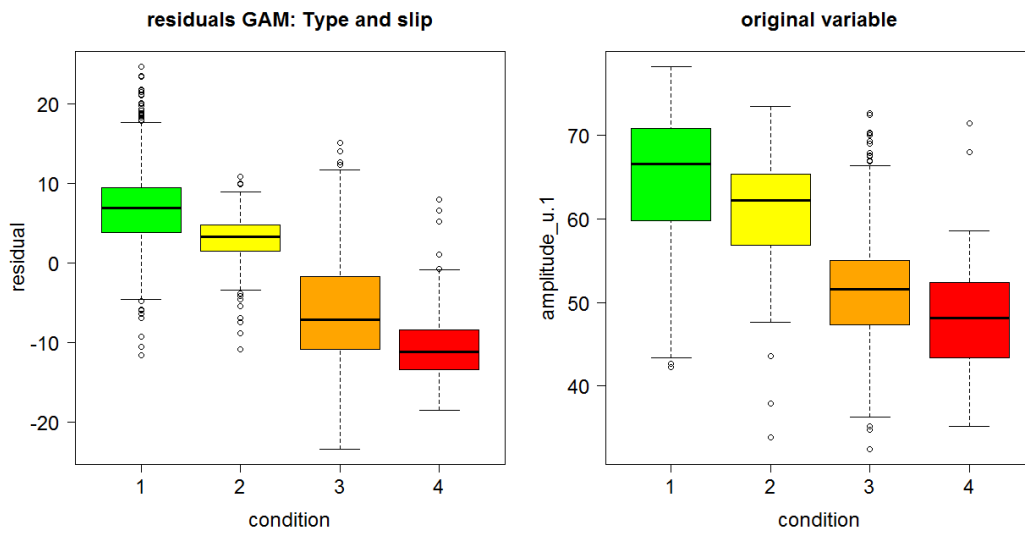


Figura 7.5. Diagramas de cajas relacionados con la variable amplitud_u.1 en función de la condición del motor. A la izquierda, diagramas de los residuales obtenidos mediante el GAM; a la derecha, diagramas de la variable original.

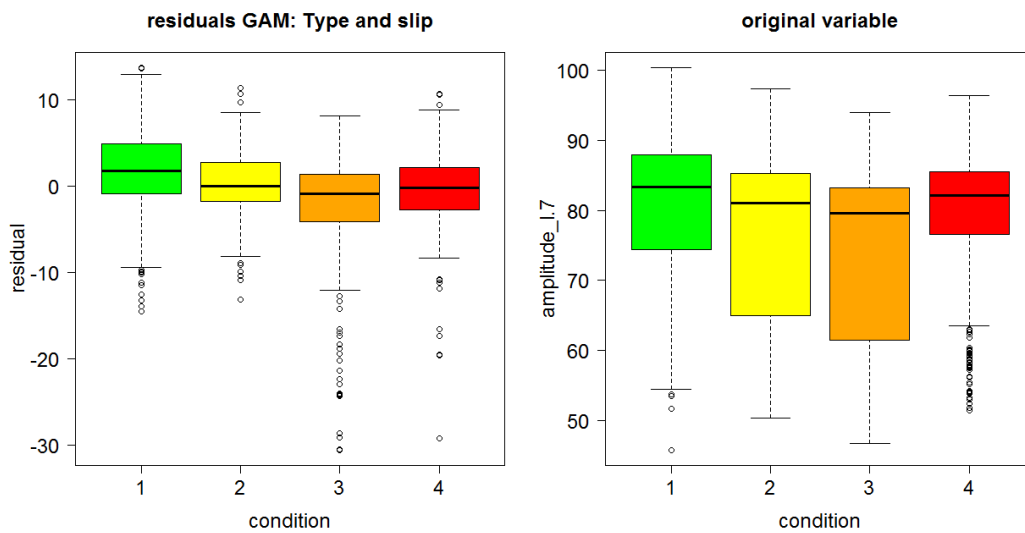


Figura 7.6. Diagramas de cajas relacionados con la variable amplitud_l.7 en función de la condición del motor. A la izquierda, diagramas de los residuales obtenidos mediante el GAM; a la derecha, diagramas de la variable original.

7.2. Resumen y conclusiones

La mayoría de los Modelos Aditivos Generalizados formulados han ofrecido características semejantes que serán sintetizadas a continuación.

- **Efecto del tipo**

El efecto introducido por la combinación motor-alimentación empleada es significativo. En Capítulos anteriores ya se dedujo que esto ocurría,

pues los resultados de los distintos análisis mejoraban a medida que eran más específicos, estudiando de forma individualizada cada tipo.

El signo del efecto introducido dependiendo de si se trata de variables pico o variables amplitud se recoge en la Tabla 7.1. Se trata de una generalización para tener una información aproximada, pero no tiene por qué cumplirse en todos los casos. Por ejemplo, el efecto de M1_LN sobre amplitud_{I.1} realmente es negativo y no positivo como viene indicado en la tabla.

Tabla 7.1. Signo del efecto introducido por el tipo según la naturaleza de las variables: picos o amplitudes.

Tipo de variables	M1_UT	M1_TE	M1_LN	M2_UT	M2_AB	M2_SM
picos	+	-	+	0	0	0
amplitudes	+	-	+	+	+	+

+ denota efecto positivo, - efecto negativo y 0 efecto nulo

- **Efecto del deslizamiento**

El efecto del deslizamiento también es notorio en la mayoría de las variables, aunque lo es de una forma menos significativa que el tipo. Además, en diversos tramos la función incluye al cero en su intervalo de confianza, lo que indica que el efecto para esos valores de deslizamiento es razonablemente nulo.

Por lo general, se puede decir que a medida que aumenta el deslizamiento, también lo hace el efecto tanto sobre picos como amplitudes, aunque la pendiente con la que lo hace varía dependiendo de los casos. De nuevo, se trata de una generalidad que no se cumple en todas las variables.

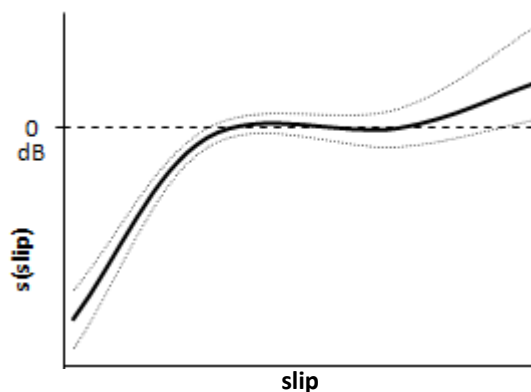


Figura 7.7. Forma característica de la función del efecto deslizamiento sobre las amplitudes.

Cabe destacar que para las variables tipo amplitud la función del efecto del deslizamiento sigue una forma moderadamente similar en todos los

casos. Comienza con efectos de signo negativo que se van reduciendo a medida que aumenta el deslizamiento hasta que el efecto se hace nulo en un valor determinado. Permanece en cero durante un tramo y, a partir de otro valor concreto de deslizamiento, la función empieza a crecer, esta vez con una pendiente menos pronunciada que inicialmente. La incertidumbre aumenta en este último trecho. La Fig. 7.7 ilustra lo descrito.

Las funciones asociadas a los picos, por el contrario, presentan formas muy diversas, sin un patrón definido.

- **Residuales**

Después de llevar a cabo cada uno de los GAM se anotaron y representaron los residuales desprendidos del mismo junto con los datos de la variable en cuestión. De los diagramas de cajas extraídos obtenemos una serie de deducciones.

Los *boxplots* de los residuales asociados a las variables relacionadas con el primer armónico presentan menor solapamiento que los *boxplots* de las variables originales. Esto hace previsible que la clasificación mediante el análisis discriminante mejore al emplear los residuos como nuevas variables de estudio. Lo mismo ocurre para las amplitudes relacionadas con el quinto armónico, pero no así para los picos ni para ninguna de las variables relacionadas con el séptimo armónico. Los residuos de las variables *amplitude_l.7* y *amplitude_u.7* se encuentran menos dispersos; sin embargo, los diagramas según la condición se superponen unos con otros.

De cualquier forma, los residuales no empeoran la información ofrecida por las variables que fueron requeridas para su cálculo. Por ello, puesto que en ciertos casos sí que existe un aparente beneficio, conviene estudiar si efectivamente los residuos son datos más interesantes y eficaces para distinguir entre los cuatro posibles estados del motor.

En el siguiente Capítulo se realizarán análisis discriminantes empleando como nuevas variables los residuales, para verificar o desmentir las deducciones formuladas.

Capítulo 8

Análisis Discriminante con residuales del GAM

Este Capítulo surge como consecuencia de la hipótesis de que los residuales, fruto de los Modelos Aditivos Generalizados planteados en el Capítulo anterior, puedan ser más poderosos en la labor de identificar la condición del motor que las variables contempladas originariamente. Se llevarán a cabo numerosos análisis discriminantes empleando los residuales como nuevas variables explicativas y se compararán los resultados con los del Capítulo 6. Además, dado que los GAM incluían en el modelo el efecto del deslizamiento, no es coherente introducir la variable slip en el estudio.

Para distinguir el nombre de los apartados respecto a los del Capítulo 6, denominaremos como “Análisis Discriminantes GAM” a los nuevos análisis. Asimismo, para referirnos a los residuos emplearemos la siguiente notación:

<i>peak_l.1</i>		<i>r_peak_l.1</i>
<i>peak_u.1</i>		<i>r_peak_u.1</i>
<i>amplitude_l.1</i>		<i>r_amplitude_l.1</i>
<i>amplitude_u.1</i>		<i>r_amplitude_u.1</i>
<i>peak_l.5</i>		<i>r_peak_l.5</i>
<i>peak_u.5</i>		<i>r_peak_u.5</i>
<i>amplitude_l.5</i>	Residuales	<i>r_amplitude_l.5</i>
<i>amplitude_u.5</i>	→	<i>r_amplitude_u.5</i>
<i>peak_l.7</i>		<i>r_peak_l.7</i>
<i>peak_u.7</i>		<i>r_peak_u.7</i>
<i>amplitude_l.7</i>		<i>r_amplitude_l.7</i>
<i>amplitude_u.7</i>		<i>r_amplitude_u.7</i>

El procedimiento a seguir será idéntico al del Capítulo 6. Se establecen tres fases claramente definidas.

- En la fase inicial, tendrá lugar una serie de análisis discriminantes pormenorizados. La finalidad, en este caso, es comprobar si la eficacia de la discriminación es aceptable contemplando todos los datos simultáneamente. De no ser así, se espera que los análisis individualizados según el tipo ofrezcan mayor éxito.
- La segunda fase consistirá en un estudio más exhaustivo de los análisis adecuados según la fase anterior. Vistos los resultados del Capítulo 6, esta vez se empleará solamente la técnica de selección *Stepwise*, que ofrecía mejores resultados. Así se consigue reducir el número de total de estudios a desarrollar.
- Finalmente, la técnica *bootstrap* servirá para certificar la validez de los análisis desarrollados a lo largo de la fase previa.

8.1. Análisis Discriminantes GAM preliminares

El primer análisis a realizar incluye todas las variables para la distinción entre los cuatro estados posibles. Como ya fue indicado, la variable *slip* no forma parte del estudio. La tabla de clasificación obtenida se corresponde con la Tabla 8.1.

La clasificación tiene un éxito del 65,62%. No se trata de un porcentaje aceptable para la aplicación requerida; apenas en dos de cada tres ensayos se identifica el verdadero estado del motor.

Tabla 8.1. Clasificación de la condición del motor. Todos los ensayos.

Condición real	Cantidad	Condición predicha			
		1	2	3	4
1	485	348 (71,75%)	118 (24,33%)	18 (3,71%)	1 (0,21%)
2	270	62 (22,96%)	198 (73,33%)	10 (3,70%)	0 (0,00%)
3	276	28 (10,14%)	26 (9,42%)	105 (38,04%)	117 (42,39%)
4	240	1 (0,42%)	6 (2,50%)	50 (20,83%)	183 (76,25%)
Total	1271	439	348	183	301

Porcentaje de casos correctamente clasificados: 65,62%

Tabla 8.2. Clasificación del motor. Ensayos con motor M1.

Condición real	Cantidad	Condición predicha			
		1	2	3	4
1	218	141 (64,68%)	70 (32,11%)	7 (3,21%)	0 (0,00%)
2	153	25 (16,34%)	123 (80,39%)	5 (3,27%)	0 (0,00%)
3	169	15 (8,88%)	22 (13,02%)	79 (46,75%)	53 (31,36%)
4	123	0 (0,00%)	4 (3,25%)	14 (11,38%)	105 (85,37%)
Total	663	181	219	105	158

Porcentaje de casos correctamente clasificados: 67,57%

La Tabla 8.2 representa la predicción realizada cuando se estiman solo los ensayos en los que interviene el motor M1, para comprobar si de este modo los resultados mejoran. El porcentaje alcanzado es apenas un 2% mejor, no lo suficiente.

La discriminación considerando únicamente los ensayos asociados a un tipo puede ser más interesante, como ocurrió en el apartado 6.1 con los análisis discriminantes a partir de las variables originales. La Tabla 8.3 de clasificación para el tipo M1_UT confirma la hipótesis, con un porcentaje de éxito de más del 90%.

Tabla 8.3. Clasificación del motor. Ensayos con M1_UT.

Condición real	Cantidad	Condición predicha				Total
		1	2	3	4	
1	83	76 (91,57%)	7 (8,43%)	0 (0,00%)	0 (0,00%)	83
2	67	6 (8,96%)	61 (91,04%)	0 (0,00%)	0 (0,00%)	67
3	70	0 (0,00%)	1 (1,43%)	64 (91,43%)	5 (7,14%)	70
4	60	0 (0,00%)	4 (6,67%)	1 (1,67%)	55 (91,67%)	60
Total	280	82	74	65	60	280

Porcentaje de casos correctamente clasificados: 91,43%

Así, en el diagrama de las dos primeras funciones discriminantes de la Fig. 8.1 asociado a este último análisis, se puede distinguir claramente la existencia de zonas diferenciadas según la condición del motor.

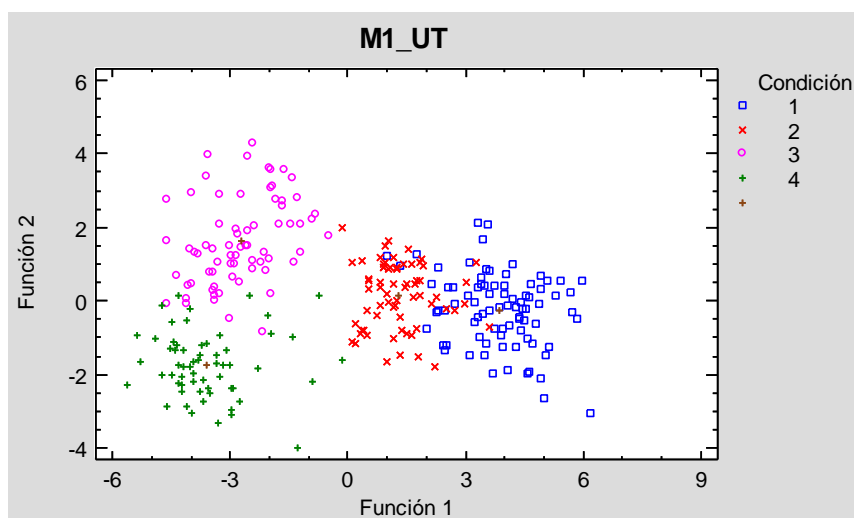


Figura 8.1. Diagrama de las funciones discriminantes 1 y 2. Tipo M1_UT. Distinción entre C1, C2, C3 y C4. Todas las variables contempladas.

Parece evidente que a medida que el análisis es más específico se alcanzan mejor los objetivos requeridos. La confirmación de esta teoría queda

reflejada a través de la Tabla 8.4. En ella, se resumen los porcentajes de acierto hallados en el análisis de todos los datos **(A)** y de los datos por separado según el motor **(B)** y según el tipo **(C)**. Se aprecia un aumento substancial cuando el conjunto de ensayos contemplados es de menor tamaño, excepto cuando se trata de ensayos con M1_TE. Este conjunto de ensayos también daba problemas con las variables originales.

Tabla 8.4. Resumen del éxito de clasificación dependiendo del conjunto de datos analizado.

		Casos correctamente clasificados (%)
(A)	Totalidad de ensayos	65,62%
(B)	M1	67,57%
	M2	76,15%
(C)	M1_UT	91,43%
	M1_TE	62,21%
	M1_LN	86,78%
	M2_UT	84,58%
	M2_AB	86,85%
	M2_SM	82,99%

Vistas las características de los resultados, la solución adecuada para establecer una clasificación óptima es analizar los datos de forma individualizada en función del motor y la alimentación empleados. Así se hará en la siguiente sección de este Capítulo.

8.2. Análisis Discriminantes GAM finales

Del mismo modo que se hizo en el Capítulo 6, serán comentados los estudios de mayor relevancia desprendidos por varios análisis discriminantes realizados sobre cada conjunto de datos en función del tipo. De nuevo, los factores de clasificación deberán ser capaces de establecer las siguientes distinciones:

- A. Distinción entre motor sano (1) y el resto de estados (2 U 3 U 4)
- B. Distinción entre motor sano (1), motor con fallo leve (2), motor con fallo medio (3) y motor con ruptura total de la barra (4).

La nomenclatura empleada coincide con la que se trabajó antes. Véase la Tabla 6.6 del apartado 6.2.

No se incluirán todas las variables en el modelo, sino que se escogerán las de mayor poder discriminante a través de la selección *Stepwise*. El criterio

estadístico F es el empleado en todos los casos, siendo cuatro su valor crítico.

Como se considera que el factor *type* influye en la respuesta, se analizarán por separado las pruebas según este identificador, a saber: M1_UT, M1_TE, M1_LN, M2_UT, M2_AB y M2_SM.

8.2.1. Análisis Discriminantes GAM representativos: caso M1_UT

En el presente apartado se expondrá a modo de ejemplo el procedimiento seguido para M1_UT. Los análisis discriminantes sobre ensayos del resto de tipos se recogen en el Anexo B, en el apartado B.3. Su procedimiento es equivalente al que aquí se presentará y comentará.

El estudio se divide en dos análisis según las categorías a distinguir: C1-C234 ó C1-C2-C3-C4. Recordemos que la única metodología de selección empleada es la *Stepwise*.

8.2.1.1. Discriminación C1-C234

Mediante la técnica *Stepwise* las variables seleccionadas son las indicadas en la Tabla 8.5. Se plasman además los coeficientes estandarizados de la función discriminante.

Tabla 8.5. Coeficientes estandarizados de la función discriminante. Tipo M1_UT. Distinción entre C1 y C234. Método *Stepwise*.

	Función 1
r_peak_l.1	-0,712328
r_amp_u.1	-0,824648
r_peak_u.5	0,512577
r_amp_l.5	1,03208
r_amp_u.5	0,436508
r_peak_l.7	-0,387677

La Fig. 8.2 se corresponde con la representación del diagrama de dispersión en función de las dos variables con mayor poder discriminante, es decir, con mayor coeficiente en valor absoluto. La identificación de los grupos se presupone bastante buena, como podemos ver.

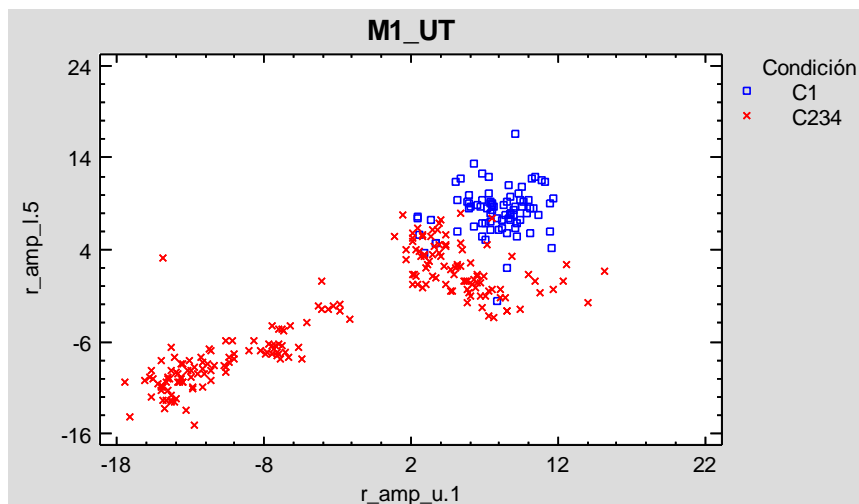


Figura 8.2. Diagrama de dispersión de la relación entre r_amp_l.5 y r_amp_u.1. Tipo M1_UT. Distinción entre C1 y C234. Método *Stepwise*.

La Tabla 8.6 de clasificación refleja un éxito del 92,86%, un porcentaje elevado pese a haber desestimado la mitad de las variables.

Tabla 8.6. Clasificación entre C1 y C234 para M1_UT. Método *Stepwise*.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	83	81 (97,59%)	2 (2,41%)
C234	197	18 (9,14%)	179 (90,86%)
Total	280	99	181

Porcentaje de casos correctamente clasificados: 92,86%

8.2.1.2. Discriminación C1-C2-C3-C4

La distinción entre los cuatro estados del motor, por su parte, requiere la selección de tres variables más: r_amp_l.1, r_peak_u.7 y r_amp_l.7. Los coeficientes estandarizados han sido registrados en la Tabla 8.7.

Tabla 8.7. Coeficientes estandarizados de las funciones discriminantes. Tipo M1_UT. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

	Función 1	Función 2	Función 3
r_peak_l.1	-0,0593178	0,0322758	1,06514
r_amp_l.1	1,05667	-0,330247	1,30482
r_amp_u.1	-0,805753	1,06777	-0,259764
r_peak_u.5	0,182983	0,132662	-0,722606
r_amp_l.5	0,814814	-0,435621	-0,25888
r_amp_u.5	0,0908602	0,779241	-0,754984
r_peak_l.7	0,0783835	0,216722	0,603747
r_peak_u.7	0,21742	-0,117493	0,182518
r_amp_l.7	0,409676	-0,293807	0,287427

Se representan los ensayos en función de $r_amp_l.1$ y $r_amp_l.5$ en la Fig. 8.3. La clasificación parece que vaya a ser buena, aunque algunas observaciones se alejan bastante de la posición ideal, especialmente algunas asociadas al motor con mayor grado de rotura.

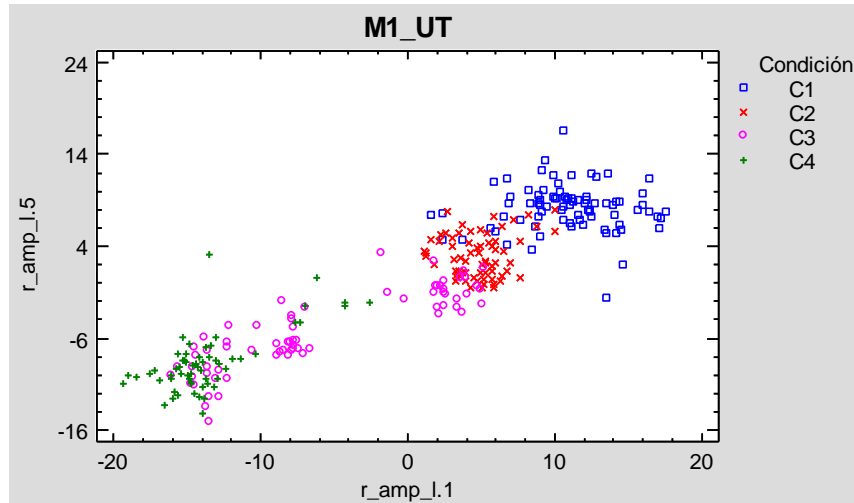


Figura 8.3. Diagrama de dispersión de la relación entre $r_amp_l.5$ y $r_amp_l.1$. Tipo M1_UT. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

Respecto a los resultados numéricos la Tabla 8.8 augura una predicción positiva, con un 90% de acierto, aproximadamente. Efectivamente se cumple el pronóstico favorable.

Tabla 8.8. Clasificación entre C1, C2, C3 y C4 para M1_UT. Método *Stepwise*.

Grupo real	Cantidad	Grupo predicho				Total
		C1	C2	C3	C4	
C1	83	74 (89,16%)	9 (10,84%)	0 (0,00%)	0 (0,00%)	83
C2	67	6 (8,96%)	61 (91,04%)	0 (0,00%)	0 (0,00%)	67
C3	70	0 (0,00%)	0 (0,00%)	64 (91,43%)	6 (8,57%)	70
C4	60	0 (0,00%)	4 (6,67%)	2 (3,33%)	54 (90,00%)	60
Total	280	80	74	66	60	280

Porcentaje de casos correctamente clasificados: 90,36%

8.2.2. Resumen y conclusiones

Durante el transcurso del apartado han sido presentados y comentados para cada combinación entre motor y alimentación posible (6 tipos) dos análisis discriminantes (diferenciación C1-C234 y diferenciación C1-C2-C3-C4), ejecutados mediante la metodología de selección de variables *Stepwise*. Por lo tanto, se dispone de 12 análisis discriminantes enfocados sobre los residuales de los GAM planteados en el apartado 7.2.

Aunque se cuenta con un total de 12 variables de interés, la selección *Stepwise* descarta las irrelevantes, por el hecho de tener un escaso poder discriminante. Ahora bien, es inevitable que dependiendo del tipo de análisis efectuado las variables seleccionadas sean unas u otras, no coincidiendo en muchas ocasiones. Por ello y con el objetivo de identificar los términos predominantes en todos o la mayoría de los análisis, las siguientes tablas sintetizan la información acerca de las variables más frecuentemente escogidas.

La Tabla 8.9 se centra en la clasificación C1-C234, mientras que la Tabla 8.10 se ocupa de la clasificación C1-C2-C3-C4. Por lo general, en la segunda se requiere un número igual o mayor de variables, excepto en el caso de M1_LN. A medida que aumenta el número de niveles del factor clasificador, se podría decir que existen más variables relevantes desde un punto de vista discriminante.

Las tablas expuestas recogen en su última columna el número total de veces que cada variable ha sido seleccionada. A través de la Tabla 8.11 se condensa esta información, que permite vislumbrar los términos requeridos con mayor frecuencia. Como ya se comentó en el apartado 6.2.7, conviene establecer conclusiones generales que sean prácticas en la realidad.

Tabla 8.9. Variables seleccionadas para cada nivel de *type*, denotadas por una "X". Distinción entre C1 ó C234. Método *Stepwise*.

Variable	M1_UT	M1_TE	M1_LN	M2_UT	M2_AB	M2_SM	FRECUENCIA
r_peak_l.1	X		X		X		3
r_peak_u.1						X	1
r_amp_l.1		X	X				2
r_amp_u.1	X	X	X	X	X	X	6
r_peak_l.5			X	X		X	3
r_peak_u.5	X						1
r_amp_l.5	X			X	X	X	4
r_amp_u.5	X	X		X		X	4
r_peak_l.7	X				X		2
r_peak_u.7			X		X		2
r_amp_l.7				X			1
r_amp_u.7			X	X			2
TOTAL VARIABLES	6	3	6	6	5	5	

Tabla 8.10. Variables seleccionadas para cada nivel de *type*, denotadas por una "X". Distinción entre C1, C2, C3 ó C4. Método *Stepwise*.

Variable	M1_UT	M1_TE	M1_LN	M2_UT	M2_AB	M2_SM	FRECUENCIA
r_peak_l.1	X		X		X		3
r_peak_u.1					X	X	2
r_amp_l.1	X	X	X	X		X	5
r_amp_u.1	X	X	X	X	X		5
r_peak_l.5				X		X	2
r_peak_u.5	X						1
r_amp_l.5	X		X	X	X	X	5
r_amp_u.5	X	X		X		X	4
r_peak_l.7	X				X		2
r_peak_u.7	X						1
r_amp_l.7	X			X	X		3
r_amp_u.7				X			1
TOTAL VARIABLES	9	3	4	7	6	5	

Tabla 8.11. Resumen sobre el número de veces que cada variable ha sido seleccionada según el tipo de análisis.

Variable	Selección <i>Stepwise</i>		
	C1-C234	C1-C2-C3-C4	TOTAL
r_peak_l.1	3	3	6
r_peak_u.1	1	2	3
r_amp_l.1	2	5	7
r_amp_u.1	6	5	11
r_peak_l.5	3	2	5
r_peak_u.5	1	1	2
r_amp_l.5	4	5	9
r_amp_u.5	4	4	8
r_peak_l.7	2	2	4
r_peak_u.7	2	1	3
r_amp_l.7	1	3	4
r_amp_u.7	2	1	3

La principal característica apreciada es la escasa frecuencia de aparición en los modelos finales de las variables asociadas al séptimo armónico. El descartar la medición de estas bandas puede ahorrar tiempo y la información perdida se espera que no sea determinante, al menos en la mayoría de los tipos.

Por otro lado, los residuos relacionados con las amplitudes son más relevantes que los relacionados con los picos. En ningún caso alguna variable de esta última naturaleza se ha manifestado con una frecuencia superior a 6,

es decir, aparecen en la mitad o menos análisis efectuados. Parece razonable estudiar la posibilidad de centrarse únicamente en la medición de variables tipo amplitud. De este modo no sería necesario calcular la función *wavelet* correspondiente al ruido de fondo, lo cual conlleva una reducción de la complejidad computacional.

No obstante, una vez más, es importante puntualizar que las deducciones formuladas no son más que meras hipótesis sin contrastar. La pérdida de información al eliminar ciertas variables podría ser determinante según la combinación motor-alimentación empleada.

Una vez estimada la importancia de cada variable, procede comprobar la eficacia de los análisis en las discriminaciones planteadas. La Tabla 8.12 recopila los porcentajes de casos satisfactoriamente clasificados de los mismos.

Tabla 8.12. Resumen sobre los porcentajes de casos correctamente clasificados hallados mediante los análisis discriminantes diseñados.

CASOS CORRECTAMENTE CLASIFICADOS (%)		
DISCRIMINACIÓN	C1-C234	C1-C2-C3-C4
SELECCIÓN	<i>Stepwise</i>	<i>Stepwise</i>
M1_UT	92,86%	90,36%
M1_TE	69,85%	59,54%
M1_LN	100,00%	80,99%
M2_UT	93,03%	81,59%
M2_AB	88,26%	87,79%
M2_SM	92,78%	80,93%

Las conclusiones a extraer son las siguientes:

- La discriminación ofrece mejores resultados cuanto menor es el número de niveles del factor clasificador. Dicho de otra forma, la discriminación C1-C234 es más eficaz que la C1-C2-C3-C4. En principio parece algo obvio, pues al tener más posibilidades de asignación la dificultad aumenta.
- Todos los análisis a excepción de los vinculados al tipo M1_TE desprenden resultados satisfactorios, con un porcentaje de acierto siempre superior al 80%. El fracaso clasificatorio de los ensayos llevados a cabo con el motor M1 y el variador de *Telemecanique* se cree que puede estar relacionado con la antigüedad de este dispositivo, mucho menos sofisticado que los demás. Como ya se indicó en apartados anteriores, la eliminación de la intensidad de

corriente como factor de interés pudo ser determinante en la pérdida de información.

8.3. Validación del Análisis Discriminante GAM

Tal y como se discutió en el apartado 6.3, es necesario validar los análisis discriminantes elaborados, ya que los porcentajes obtenidos ofrecen tan sólo una indicación sobre el error aparente. No es recomendable realizar deducciones a partir de este, sino que es preferible estimar el error a partir de alguna técnica de validación. En nuestro caso, mediante el método *bootstrap* se logra una aproximación más realista del éxito de clasificación.

De nuevo tendrá lugar el cálculo del error estimado por medio de un proceso iterativo. En cada una de las 500 iteraciones se obtiene una muestra aleatoria de tamaño igual al número de ensayos, que constituirá la muestra de entrenamiento. El resto de ensayos formará la muestra de test, en cada iteración.

El tamaño de las muestras difiere en función del número de ensayos disponibles de cada combinación motor-alimentación. En la Tabla 6.19 del apartado 6.3 ya se reflejó esta información, compatible con la resolución de este apartado.

Además, en adelante se expresará la información de nuevo en forma de acierto aparente y acierto estimado para facilitar la comparativa. A continuación, se desarrolla el apartado estructurándolo en dos secciones dependiendo de las poblaciones a distinguir en la clasificación: C1-C234 ó C1-C2-C3-C4.

8.3.1. Validación discriminación C1-C234

En la Tabla 8.13 se observan los aciertos estimados junto a los aparentes desprendidos de la discriminación entre C1 y C234. Como se puede ver, el acierto estimado es siempre menor que el aparente. Esto se debe a que la muestra a partir de la cual el error estimado ha sido calculado no contiene datos que se hayan empleado para la construcción de las funciones discriminantes, mientras que el aparente sí lo hace y consecuentemente responde mejor. En cualquier caso, ambos aciertos no distan demasiado, señal de que el análisis discriminante planteado es fiable.

La representación de los errores mediante diagramas de cajas, permite entender la media y dispersión de las iteraciones. La Fig. 8.4 contiene el gráfico en cuestión. Se percibe un aumento de la amplitud del diagrama a medida que el error medio es mayor. Por ejemplo, el tipo M1_LN, con un error

estimado prácticamente nulo, no presenta dispersión; mientras tanto, el tipo M1_TE, con el mayor error estimado de todos, de casi el 31%, presenta también la mayor amplitud del diagrama.

Tabla 8.13. Comparación entre acierto aparente y acierto estimado, en porcentaje. Clasificación en grupos C1 y C234.

ACIERTO SELECCIÓN	APARENTE <i>Stepwise</i>	ESTIMADO <i>Stepwise</i>
M1_UT	92,86%	92,12%
M1_TE	69,85%	69,39%
M1_LN	100,00%	99,88%
M2_UT	93,03%	92,51%
M2_AB	88,26%	86,63%
M2_SM	92,78%	90,99%

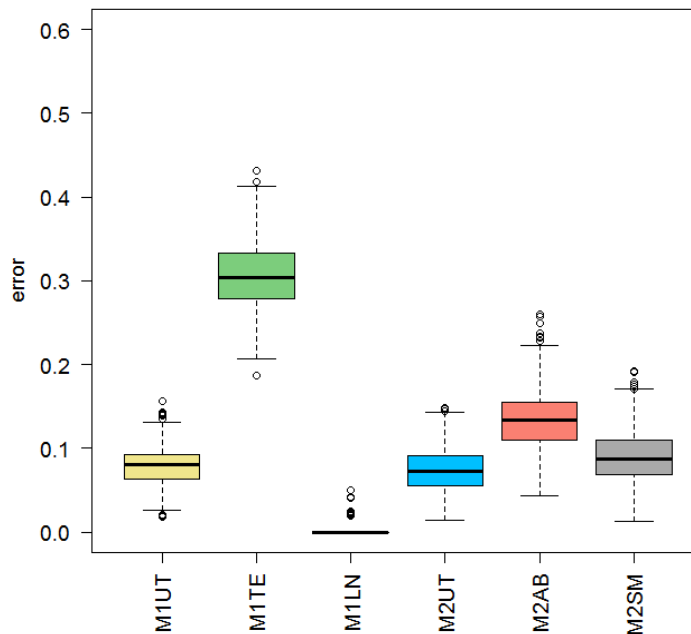


Figura 8.4. Diagramas de cajas del error en función del tipo de alimentación y motor. Distinción entre C1 y C234. Método *Stepwise*.

8.3.2. Validación discriminación C1-C2-C3-C4

Análogamente a lo anterior, se expresan a través de la Tabla 8.14 los aciertos aparentes y estimados obtenidos al diferenciar entre las cuatro condiciones del motor. De nuevo, se manifiestan las mismas características: el acierto estimado es siempre menor, pero no difiere excesivamente del aparente, si bien es cierto que a grandes rasgos la diferencia ha aumentado ligeramente con respecto a la discriminación C1-C234. En definitiva, los análisis diseñados parecen correctos.

Tabla 8.14. Comparación entre acierto aparente y acierto estimado, en porcentaje. Clasificación en grupos C1, C2, C3 y C4.

ACIERTO SELECCIÓN	APARENTE <i>Stepwise</i>	ESTIMADO <i>Stepwise</i>
M1_UT	90,36%	88,88%
M1_TE	59,54%	57,59%
M1_LN	80,99%	78,61%
M2_UT	81,59%	78,67%
M2_AB	87,79%	84,29%
M2_SM	80,93%	76,97%

La Fig. 8.5 ilustra la dispersión del error. Destaca la gran diferencia del diagrama de M1_LN con respecto a antes, en la discriminación C1-C234. Como el acierto estimado ha disminuido considerablemente, del 99,88% al 78,61%, la amplitud del diagrama ha aumentado mucho.

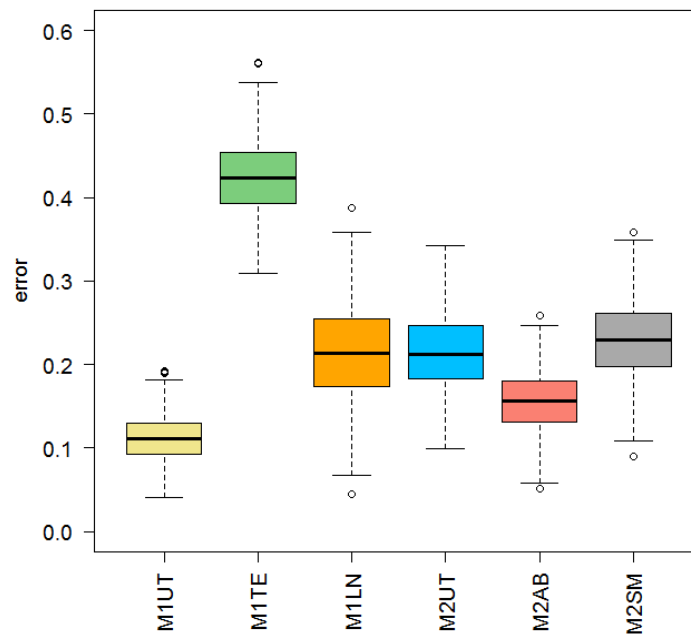


Figura 8.5. Diagrama de cajas del error en función del tipo de alimentación y motor. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

8.3.3. Conclusiones sobre la validación

Gracias al método de validación *bootstrap* se ha podido comprobar la veracidad de los resultados derivados de los análisis discriminantes. Independientemente del número de poblaciones del factor clasificador, el acierto estimado ha presentado en todo momento porcentajes ligeramente superiores al acierto aparente, pero relativamente próximos, según lo

esperado. En definitiva, se corrobora la idoneidad de los análisis discriminantes diseñados.

Ahora bien, se trata de una conclusión desde el punto de vista de la validez y no de la eficacia clasificatoria. En el siguiente apartado se comparará la aptitud de los residuales con respecto a la capacidad de las variables originales en la labor discriminante.

8.4. Comparación entre análisis discriminantes: variables originales o residuales del GAM

Después de formular los GAM adecuados se dedujo que tanto el tipo como el deslizamiento eran factores determinantes que influían en el valor adoptado por las variables explicativas propuestas de forma inicial (Capítulo 7). Así pues, eliminando estos efectos se prevé una mejora en la discriminación según la condición del motor, ya que podrían estar entorpeciendo el análisis encubriendo los efectos derivados exclusivamente de la propia condición. Al considerar los residuales del GAM formulado para cada variable se consigue restar dichos efectos. Se espera que los residuales sean variables de mayor interés y más eficaces para la identificación del estado del motor.

Este apartado pretende exponer los aciertos estimados obtenidos mediante el método *bootstrap* tras efectuar los análisis discriminantes lineales tanto a partir de las variables originales como a partir de los residuales asociados a ellas. El objetivo es establecer una comparación que permita enunciar qué tipo de variables es recomendable registrar para optimizar el rendimiento del análisis discriminante. Toda la información necesaria fue recogida y comentada a lo largo del Capítulo 6 y de apartados anteriores del presente Capítulo.

Sin más dilación, en la Tabla 8.15 se han reunido los aciertos estimados. Nótese que solamente ha sido requerida la información relativa al método de selección *Stepwise*, pese a que para las variables originales también se empleó el *Backward*.

Si se observa con detenimiento, al contemplar en la discriminación entre C1 y C234 los residuales en lugar de las variables inicialmente propuestas, el acierto estimado disminuye ligeramente en todos los tipos, excepto para M2_UT. Eso sí, la diferencia es mínima en este caso (92,39 frente a 92,51%).

Del mismo modo, analizando la discriminación C1-C2-C3-C4 las deducciones son similares; en esta ocasión el tipo M2_AB presenta un aumento aproximado del 2% del acierto estimado al tratar con los residuales

(82,15 frente a 84,29%). Los tipos M1_LN y M2_SM pueden considerarse de eficacia equivalente.

Tabla 8.15. Comparación del acierto estimado entre variables originales y los residuales del GAM correspondiente. Discriminaciones C1-C234 y C1-C2-C3-C4. Método *Stepwise*.

ACIERTO	ESTIMADO			
SELECCIÓN	<i>Stepwise</i>			
DISCRIMINACIÓN	C1-C234		C1-C2-C3-C4	
VARIABLES	Originales	Residuales	Originales	Residuales
M1_UT	94,32%	92,12%	91,80%	88,88%
M1_TE	70,08%	69,39%	60,18%	57,59%
M1_LN	99,89%	99,88%	78,29%	78,61%
M2_UT	92,39%	92,51%	78,80%	78,67%
M2_AB	87,21%	86,63%	82,15%	84,29%
M2_SM	91,83%	90,99%	76,52%	76,97%

Pese a que se esperaba encontrar una mejoría substancial de la potencia clasificatoria al establecer los residuales como nuevos datos de interés, los resultados obtenidos no concuerdan con lo previsto. Por lo general, las variables originales diferencian algo mejor, aunque los residuales no disten demasiado y en algunos casos incluso las superen.

Puesto que los datos sobre las variables originales son claramente más fáciles de extraer desde un punto de vista práctico, se puede concluir su conveniencia frente a los residuales. El esquema planteado a raíz de los GAM ha errado en su intento por hallar un procedimiento más poderoso en la predicción del estado del motor.

Visto el fracaso del nuevo planteamiento abordado, se busca una razón coherente que explique el porqué del mismo. El motivo por el cual los residuales parecían ser más relevantes como variables explicativas fue la disminución del solapamiento de los diagramas de cajas, tal y como se vio, por ejemplo, en la Fig. 7.2 del apartado 7.1.1. Esta característica se apreciaba para gran cantidad de variables, especialmente las asociadas a las bandas del armónico fundamental.

El error pudo residir en asumir que el beneficio aparente reflejado en muchos análisis univariantes individuales (como son los *boxplots*), no sería afectado al proponer un único análisis multivariante (como es el análisis discriminante lineal) que englobase todas las variables. Es decir, asumir que por el hecho de emplear variables que individualmente son beneficiosas, el éxito del análisis multivariante estaría garantizado. Esto en realidad no tiene

por qué ser así y, de hecho, es lo que ocurre en nuestro caso, como se verá a continuación.

Tal y como refleja la Tabla 8.16, al efectuar análisis discriminantes introduciendo una única variable explicativa, la mayoría de los porcentajes de casos correctamente clasificados es mayor al usar los residuales. Así ocurre cuando se consideran todos los ensayos. Por lo tanto, efectivamente los residuos son favorables; pero, como se ha visto, sólo al tratarlos como único término explicativo.

Por otro lado, en la parte derecha de la Tabla 8.16, la característica mencionada en el párrafo anterior no se aprecia con tanta facilidad, al considerar sólo los ensayos del tipo M1_UT. Existen numerosas excepciones y, en cualquier caso, las diferencias han disminuido ampliamente con respecto a antes. Por ejemplo, para peak_l.1 la diferencia era del 20,68%, mientras que ahora es de un 4,64% escaso. Aunque habría que observar si el resto de tipos se comporta de forma semejante, a priori se intuye que los residuales no son tan beneficiosos a nivel específico como a nivel global, siempre hablando desde el punto de vista de un análisis univariante. Esta cualidad habría sido percibida si en lugar de representar diagramas de cajas con todos los ensayos se hubieran representado diagramas de cajas para cada tipo. El solapamiento de estas últimas posiblemente fuese algo más parecido al presentado por los diagramas de las variables originales.

Tabla 8.16. Comparación de la eficacia de los análisis discriminantes incluyendo una única variable explicativa cada vez, bien en su forma original o bien en forma de residual. Discriminación C1-C2-C3-C4.

CASOS CORRECTAMENTE CLASIFICADOS (%)				
DISCRIMINACIÓN	C1-C2-C3-C4			
ENSAYOS	Totalidad		M1_UT	
VARIABLES	Originales	Residuales	Originales	Residuales
peak_l.1	32,15%	52,83%	62,86%	67,50%
peak_u.1	30,97%	55,42%	51,43%	53,93%
amplitude_l.1	32,15%	53,46%	76,43%	74,29%
amplitude_u.1	32,47%	62,34%	73,21%	71,07%
peak_l.5	32,78%	44,03%	47,86%	47,86%
peak_u.5	23,90%	25,79%	34,64%	35,36%
amplitude_l.5	34,43%	51,57%	75,36%	72,50%
amplitude_u.5	35,93%	27,59%	47,50%	47,14%
peak_l.7	25,39%	33,88%	36,79%	38,57%
peak_u.7	23,60%	22,50%	27,14%	26,43%
amplitude_l.7	21,62%	36,32%	31,79%	32,86%
amplitude_u.7	32,47%	25,00%	32,14%	35,36%

Como conclusión final, puesto que los residuales no reportan los deseados beneficios en la eficacia clasificatoria, las variables originales seguirán siendo la referencia a manejar en nuevos estudios. Cabe destacar que aunque el planteamiento haya fracasado, esto no significa que el tipo y el deslizamiento no influyan en la respuesta. Este hecho ha sido contrastado, pese a que a posteriori no se haya podido aprovechar el conocimiento de esta propiedad para operar de forma más óptima a la hora de discriminar entre los ensayos con distinto grado de rotura del rotor.

Capítulo 9

Influencia de la condición en el efecto del deslizamiento

En el Capítulo 7 se mostró el efecto que el tipo y el deslizamiento introducían sobre cada una de variables de estudio, mediante los pertinentes GAM. Eran contemplados todos los ensayos simultáneamente, asumiendo que los efectos serían equiparables independientemente de ciertas características de la prueba. Sin embargo, es conveniente estudiar por otra parte si realmente esta hipótesis es real. En concreto, se estima oportuno vislumbrar si para cada condición del motor posible el efecto del deslizamiento es diferente.

Consecuentemente, el presente Capítulo aborda el estudio planteado. Se formulará un nuevo GAM para cada combinación variable-condición. Aunque el interés se centra en el efecto de deslizamiento, el GAM también debe incluir el efecto del tipo en la ecuación. De lo contrario, el efecto del deslizamiento calculado estaría siendo afectado por otro efecto subyacente provocado por el tipo, del cual se conoce con seguridad su existencia.

La expresión de los GAM será de nuevo de la forma

$$variable \sim \mu + type + s(slip) + \varepsilon$$

aunque esta vez se analice de forma individualizada cada población de ensayos según la condición del motor, como se ha indicado. Como se dispone de 12 variables explicativas y de 4 condiciones, en total se realizarán 48 análisis con sus respectivos gráficos. Con el objetivo de evitar exponer información redundante que provoque la pérdida de interés por parte del lector, a continuación se comentarán los análisis más relevantes y representativos, así como un resumen final con las conclusiones extraídas a nivel global. El resto de análisis pueden ser consultados en el Anexo B.

Como ya se hizo en el Capítulo 7, se evitará el uso de subíndices al exponer la formulación del modelo. Así se simplifica la notación y la comprensión es más sencilla.

9.1. Modelos Aditivos Generalizados de relevancia

En el desarrollo de este apartado se han llevado a cabo diversos GAM para cada una de las variables, tanto si es de tipo pico como si es de tipo amplitud. Sin embargo, los gráficos resultantes de los modelos sobre las amplitudes no han sido esclarecedores y no se ha conseguido vislumbrar un patrón de comportamiento reconocible. Por ello, se ha omitido su exposición en este apartado y se ha desplazado al Anexo B, como mera información descriptiva. En lo sucesivo, las explicaciones pasarán a centrarse exclusivamente en los picos, cuyos modelos sí resultan de interés, como se verá próximamente.

Los cuatro GAM propuestos para la variable $peak_l.1$ son de la forma

$$peak_l.1 \sim \mu + type + s(slip) + \varepsilon$$

La diferencia reside en los ensayos objeto de estudio que son introducidos en el modelo, tal que deben cumplir lo siguiente:

$$\left\{ \begin{array}{l} \text{Modelo I: } \quad condition = 1 \\ \text{Modelo II: } \quad condition = 2 \\ \text{Modelo III: } \quad condition = 3 \\ \text{Modelo IV: } \quad condition = 4 \end{array} \right.$$

La Fig. 9.1 muestra el gráfico desprendido por el Modelo I.

Naturalmente, existe un efecto evidente aportado por el tipo, como se ve en el gráfico de la izquierda. Este hecho ya se dedujo en numerosas ocasiones mediante otros análisis elaborados a lo largo del presente trabajo. Lo que interesa es centrarse en el gráfico de la derecha, referente al deslizamiento.

En el Capítulo 7 se afirmó que existía un efecto aportado por el deslizamiento. Sin embargo, al contemplar sólo los ensayos con motor en estado sano como se está haciendo ahora, no se aprecia la existencia de tal efecto. Las bandas de confianza incluyen al cero a lo largo de los valores de deslizamiento registrados. Por tanto, el efecto vislumbrado en el Capítulo 7 se debe corresponder con los ensayos asociados a otra condición.

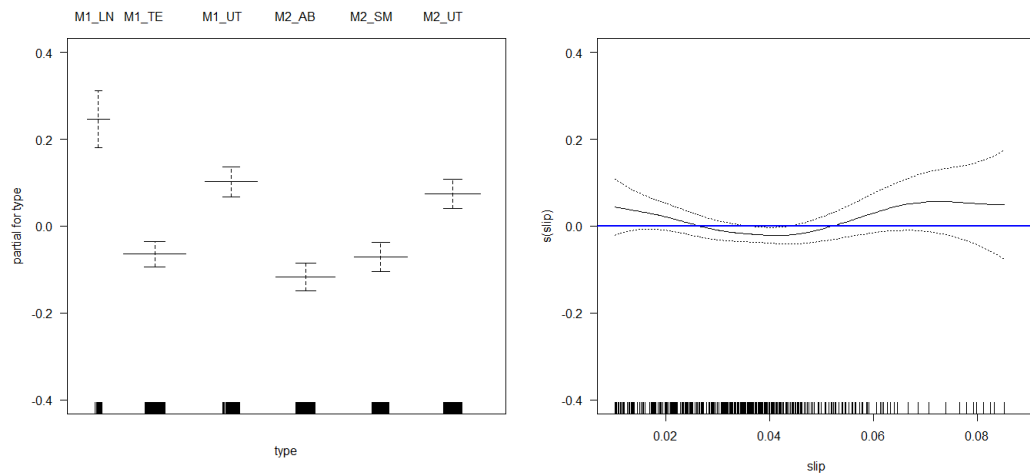


Figura 9.1. Gráficos ofrecidos por el GAM sobre *peak_l.1* al introducir en el modelado sólo los ensayos con motor sano (condición 1). A la izquierda, efecto del tipo; a la derecha; efecto del deslizamiento.

Puesto que, como se ha indicado, el efecto del tipo no es el foco de interés, la Fig. 9.2 presenta únicamente los gráficos del efecto del deslizamiento asociados a los Modelos I, II, III y IV. Ahora bien, es importante señalar que pese a que no se muestre el *plot*, el efecto del tipo sigue estando incluido en el modelado del GAM.

Gracias a la nueva imagen, se facilita la comparación entre los cuatro modelos. Si bien destacamos el efecto nulo en ensayos con motor en estado ideal (condición 1), cuando el motor empieza a estar deteriorado (condición 2), la influencia del deslizamiento sigue oscilando en torno al cero. Aunque esta vez el intervalo de confianza presenta tramos en los que el cero no está comprendido, apenas se aleja de él. Se puede decir que el efecto es de nuevo prácticamente inexistente.

Al estudiar el gráfico correspondiente a la condición 3, la forma de la función cambia. A deslizamientos bajos, se produce una notable contribución negativa al valor final de la variable, en este caso *peak_l.1*. A medida que el deslizamiento aumenta el efecto se va reduciendo hasta ser nulo, pero la función no se detiene y continúa creciendo. De este modo, a deslizamientos altos el efecto introducido es positivo y razonablemente destacable. Nótese que el número de observaciones con deslizamientos superiores a 0,07 es bastante más reducido que para valores inferiores, por lo que el cálculo de la función es menos fiable y consecuentemente el intervalo de confianza aumenta su amplitud.

La interpretación del gráfico relativo al Modelo IV es semejante a la anterior: a valores bajos de deslizamiento el efecto aportado es negativo, mientras que a valores altos es positivo. En esta ocasión, incluso se llega a

apreciar un pequeño tramo final descendiente. Puede que sea fruto del escaso número de ensayos disponibles a deslizamientos altos.

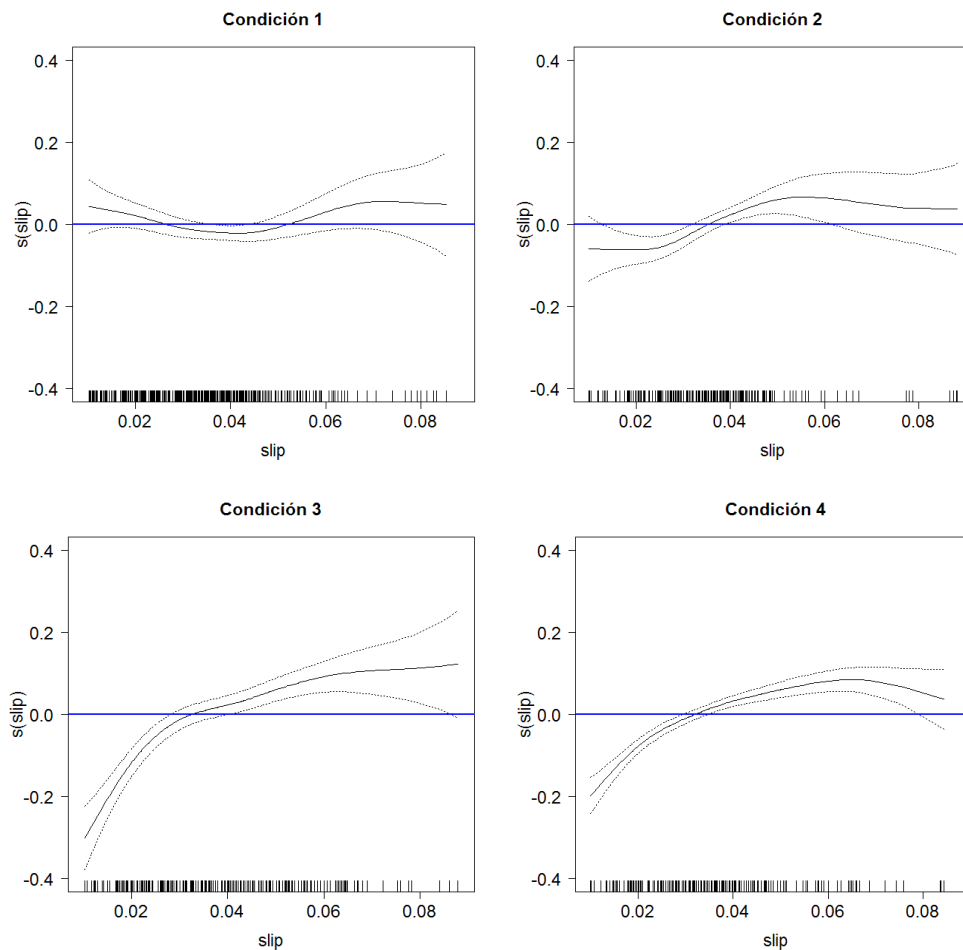


Figura 9.2. Gráficos ofrecidos por los cuatro GAM planteados sobre *peak_l.1* al introducir en el modelado sólo los ensayos con motor en una condición concreta cada vez. Se muestra el efecto del deslizamiento. De izquierda a derecha y de arriba abajo, ensayos bajo condición 1, 2, 3 y 4 respectivamente.

Para la variable *peak_u.1* la situación es muy semejante, como se verá a continuación. Análogamente a lo anterior, mediante la fórmula

$$peak_u.1 \sim \mu + type + s(slip) + \varepsilon$$

se elaboran los cuatro GAM pertinentes y se obtienen los gráficos de la Fig. 9.3. En ellos se muestra de nuevo el efecto del deslizamiento. Los dos primeros siguen razonablemente bien la asíntota horizontal, de modo que el efecto puede considerarse prácticamente nulo. Sin embargo, cuando se trata de ensayos asociados a una condición más desfavorable del motor, condiciones 3 y 4, el efecto es significativo por tramos. Deslizamientos bajos provocan una disminución del valor de la variable de hasta 0,3 unidades;

deslizamientos moderadamente altos, por su parte, también introducen un considerable efecto, pero esta vez positivo.

Como se puede ver se trata de las mismas características apreciadas en el estudio de peak_l.1. Los gráficos presentan a simple vista claras similitudes.

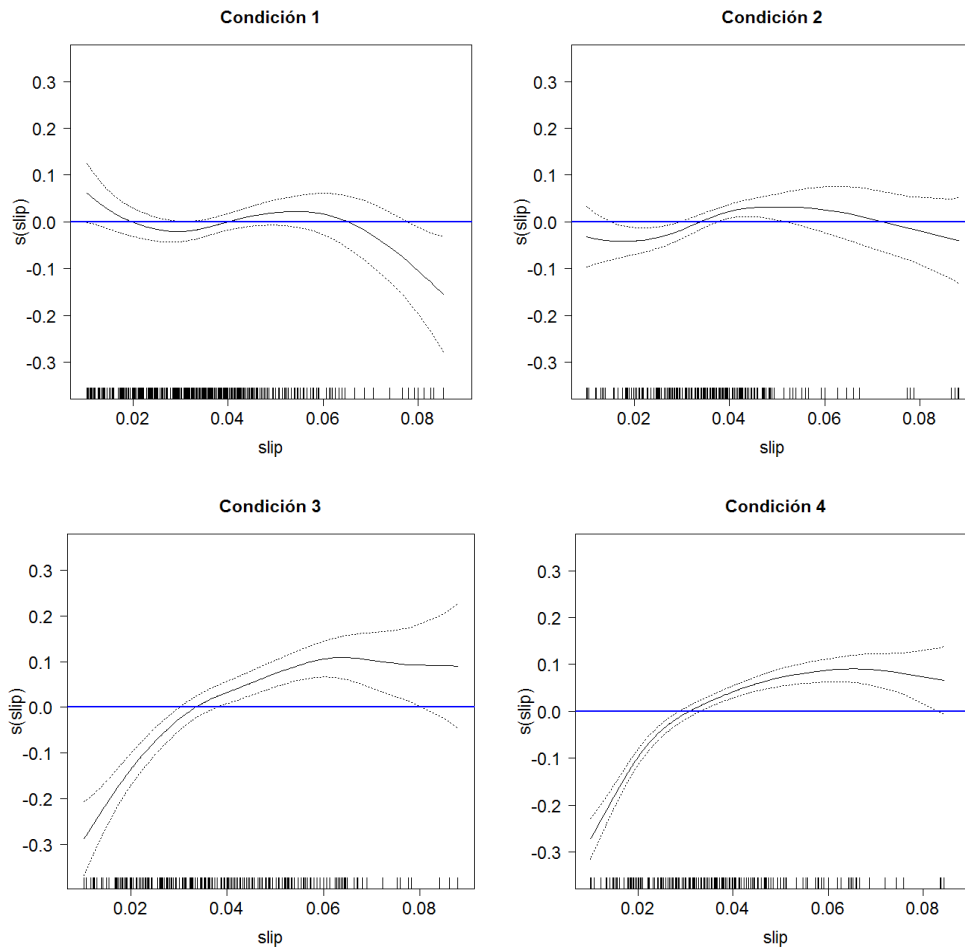


Figura 9.3. Gráficos ofrecidos por los cuatro GAM planteados sobre peak_u.1 al introducir en el modelado sólo los ensayos con motor en una condición concreta cada vez. Se muestra el efecto del deslizamiento. De izquierda a derecha y de arriba abajo, ensayos bajo condición 1, 2, 3 y 4 respectivamente.

Queda comprobada la influencia del estado del motor en el efecto del deslizamiento para las variables peak_l.1 y peak_u.1. Pero veamos qué sucede al tratar con variables asociadas a armónicos distintos del fundamental.

El GAM desarrollado para peak_l.5 es equivalente a los anteriores:

$$peak_l.5 \sim \mu + type + s(slip) + \varepsilon$$

La Fig. 9.4 ilustra, de nuevo, el efecto del deslizamiento en cada uno de los modelos diseñados dependiendo de la condición implicada.

Tal y como se puede ver, algunos de los gráficos desprendidos difieren sensiblemente con respecto a los obtenidos para `peak_l.1` y `peak_u.1`. En concreto, las dos primeras funciones no se sitúan en un valor nulo, por lo que el efecto del deslizamiento no puede afirmarse que sea inexistente. Por otro lado, la tercera función no llega crecer con fuerza para deslizamientos altos, el intervalo de confianza contiene al cero. La única función que se asemeja razonablemente bien a lo deducido anteriormente es la cuarta y última, vinculada a la condición 4 del motor. Para valores bajos de deslizamiento el efecto es negativo y para valores altos es positivo.

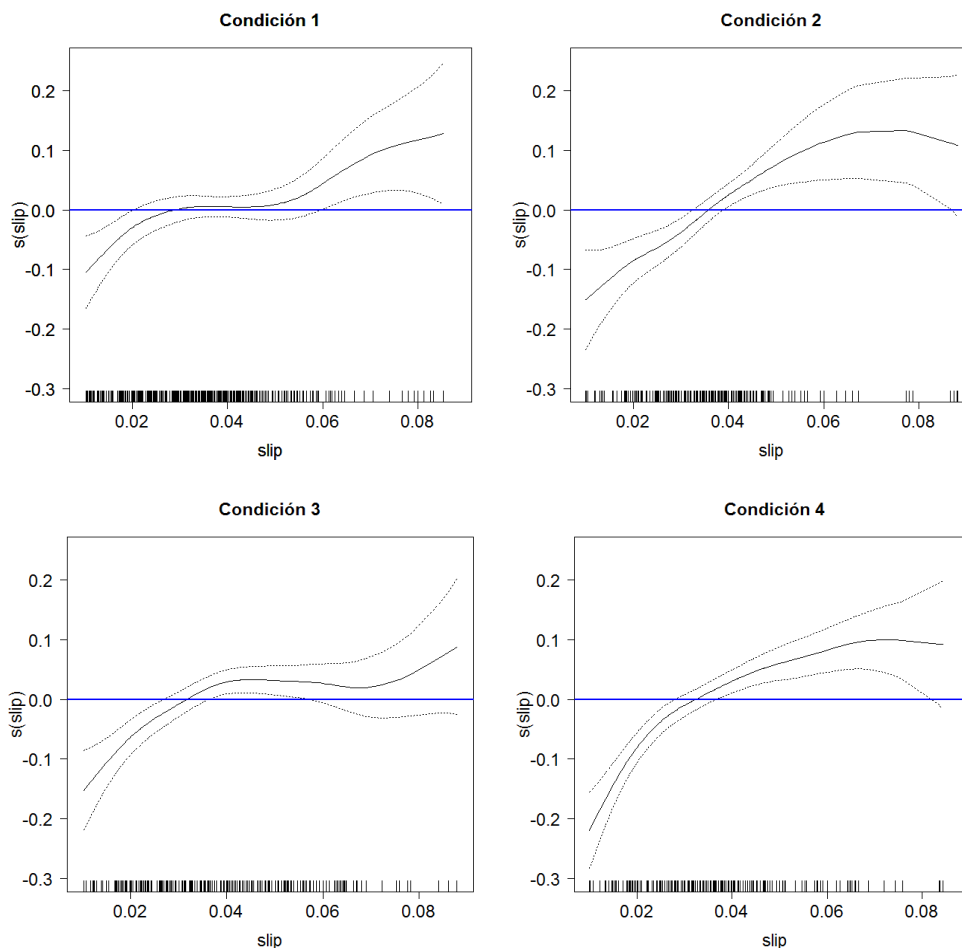


Figura 9.4. Gráficos ofrecidos por los cuatro GAM planteados sobre `peak_l.5` al introducir en el modelado sólo los ensayos con motor en una condición concreta cada vez. Se muestra el efecto del deslizamiento. De izquierda a derecha y de arriba abajo, ensayos bajo condición 1, 2, 3 y 4 respectivamente.

Por su parte, los modelos sobre la variable `peak_u.5` ofrecen conclusiones similares a estos últimos. Para mayor detalle, los gráficos se muestran en el Anexo B.

Las características han cambiado al estudiar variables relacionadas con el quinto armónico. A continuación, se verá si también lo hacen para las relacionadas con el séptimo armónico.

Se plantean los GAM para la variable *peak_l.7*, tal que

$$peak_l.7 \sim \mu + type + s(slip) + \varepsilon$$

Centrando los estudios en el efecto del deslizamiento, se extraen los gráficos de la Fig. 9.5.

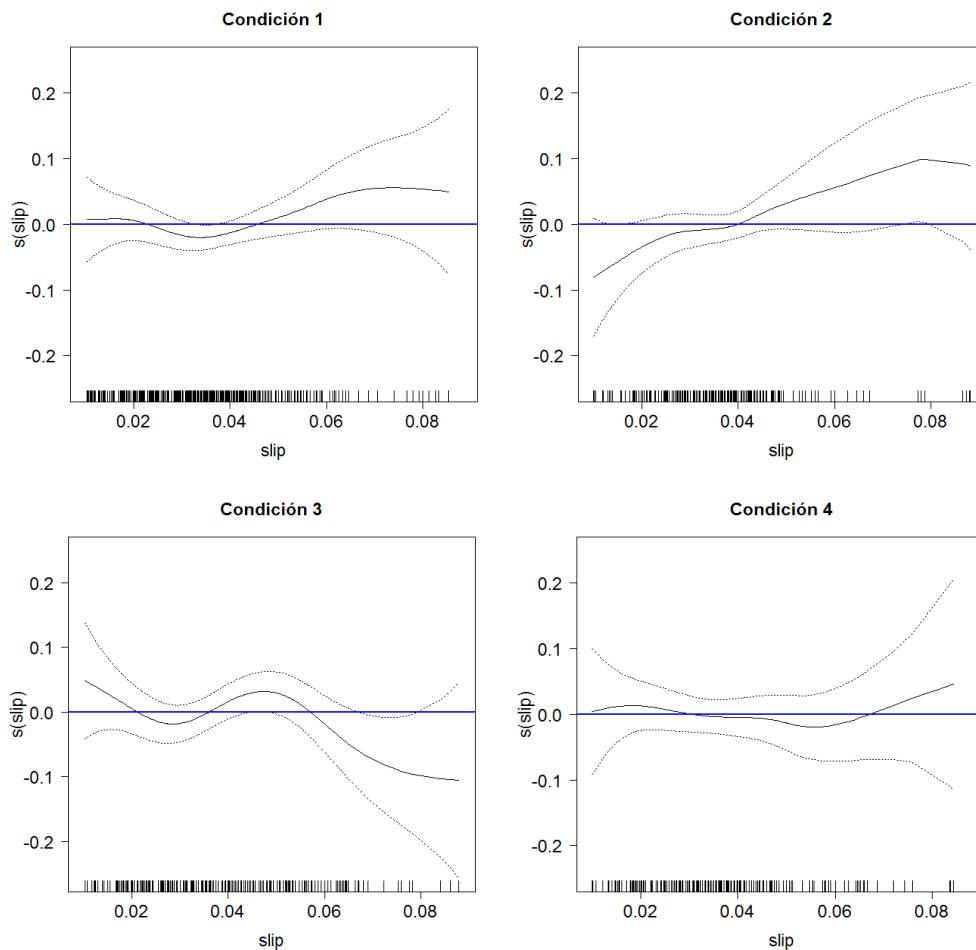


Figura 9.5. Gráficos ofrecidos por los cuatro GAM planteados sobre *peak_l.7* al introducir en el modelado sólo los ensayos con motor en una condición concreta cada vez. Se muestra el efecto del deslizamiento. De izquierda a derecha y de arriba abajo, ensayos bajo condición 1, 2, 3 y 4 respectivamente.

Una vez más, los resultados distan notoriamente de las deducciones indicadas tras los análisis de *peak_l.1* y *peak_u.1*. En esta ocasión, se puede apreciar cómo en todas las funciones las bandas de confianza contienen al cero en prácticamente todo el transcurso de la función. Esto quiere decir que no puede concluirse que la influencia del deslizamiento sea significativa. Se trata de una conclusión que concuerda con los resultados del GAM

estudiando todos los ensayos simultáneamente, en el Capítulo 7. Recordemos que en esa ocasión se estimó que la importancia de las variables disminuía a medida que estaban vinculadas a armónicos mayores.

Los modelos sobre la variable `peak_u.7` presentan las mismas características: un efecto del deslizamiento que puede ser considerado como nulo. El Anexo B recoge los gráficos correspondientes, para mayor información.

9.2. Resumen y conclusiones

El presente Capítulo ha enfocado el interés en el efecto del deslizamiento sobre la respuesta de las variables, aunque la estimación del efecto del tipo se incluye también en el modelo. El objetivo es determinar si existe una dependencia del efecto del deslizamiento con respecto a la condición del motor. Es decir, se pretende comprobar si la función del efecto es la misma independientemente del estado del motor, para lo cual es necesario estudiar por separado cada conjunto de ensayos según la condición.

Tras efectuar los GAM convenientes, se deduce la irrelevancia de las amplitudes para la finalidad comentada. Las funciones del efecto del deslizamiento son muy irregulares y no siguen una pauta reconocible que permita extraer conclusiones concretas. Por ello, los gráficos asociados no son objeto de explicación y se presentan como información adicional en el Anexo B.

Por el contrario, los picos sí muestran un comportamiento identificable, de modo que a continuación se enunciarán las principales características derivadas de su estudio:

- Los GAM propuestos para las variables asociadas al armónico principal, `peak_l.1` y `peak_u.1`, desprenden características comunes. El efecto del deslizamiento es prácticamente inexistente cuando el motor involucrado se encuentra en perfecto estado (condición1) o con un error incipiente (condición 2). Sin embargo, cuando la anomalía es considerable (condición 3) o incluso se produce la rotura total de la barra del rotor (condición4), el efecto introducido es significativo. A bajos niveles de deslizamiento, este contribuye negativamente al valor final de la variable, mientras que a altos niveles lo hace de forma positiva.

Es importante destacar que durante la experimentación fue efectuado un número reducido de ensayos a altos niveles de deslizamiento. Por lo tanto, existe cierta incertidumbre sobre si la

cualidad apreciada es fiable. Los intervalos de confianza así lo indican: su amplitud aumenta claramente a partir de valores de deslizamiento superiores a 0,06. Así pues, en el futuro podría ser interesante estudiar más a fondo el comportamiento a deslizamientos elevados.

- Las variables asociadas al quinto y al séptimo armónico no responden convenientemente a la idea planteada. Las características observadas para peak_l.1 y peak_u.1 se manifiestan de forma mucho más difusa, e incluso se podría decir que no se aprecian. Esto viene a corroborar la conclusión extraída en apartados anteriores, donde se sugería que a medida que nos alejamos del armónico principal, los datos son más confusos e irregulares.

Capítulo 10

Conclusiones

El presente trabajo ha abordado mediante técnicas estadísticas el análisis de la influencia del variador en el diagnóstico de los fallos de motores. A efectos prácticos, el objetivo final es establecer la metodología más indicada para conocer el estado físico real del motor a partir de la recogida de ciertas mediciones no invasivas.

Resulta especialmente interesante obtener una regla general que permita predecir la condición del rotor independientemente del motor y de la alimentación empleados. De este modo, el procedimiento a seguir por el interesado se simplificaría, resultando mucho más útil y eficaz.

Se plantean dos tipos de discriminaciones de interés. El primero consiste en distinguir entre un motor totalmente sano y un motor averiado en cualquier grado (C1-C234). Para aquellos procesos en los cuales sea vital que las prestaciones del motor sean las máximas, esta discriminación es la más adecuada. El segundo trata de identificar cualquier grado de anomalía de motor (C1-C2-C3-C4). Para aquellos procesos en los cuales se requiera conocer con exactitud el estado del motor para promover las medidas necesarias antes de que sea demasiado tarde, esta es la discriminación de oportuna.

La primera conclusión extraída de los análisis desarrollados a lo largo del trabajo es la imposibilidad de extraer la deseada regla general. Los resultados son tan diferentes entre cada conjunto de ensayos que las técnicas estadísticas empleadas, como el Análisis Lineal Discriminante, son totalmente incapaces de hallar una pauta general de comportamiento razonablemente buena. La eficacia se ve claramente perjudicada respecto al análisis individualizado según el tipo y ni siquiera ensayos con el mismo motor pueden ser abordados simultáneamente si el tipo de fuente es distinta.

Después de emplear la técnica del análisis discriminante, formulando los modelos adecuados y utilizando los métodos de selección de variables *Stepwise* y *Backward*, se emplea la técnica de validación *bootstrap* para validar los resultados. Se formulan las siguientes conclusiones:

- Acierto estimado

El análisis discriminante es una técnica adecuada para el fin dispuesto. Ofrece un acierto estimado satisfactorio para todos los modelos, excepto cuando se trata de clasificar ensayos con el motor M1 y el variador de *Telemecanique* (M1_TE). El porcentaje de éxito es manifiestamente inferior en este caso, tal y como se puede comprobar en la Tabla 10.1. La explicación más razonable a este fenómeno es que el inversor de *Telemecanique* es mucho más anticuado y menos sofisticado que los demás. Por ello, el rendimiento de transmisión de la energía es menor y, al desestimar las variables *current* y *speed* (como se hizo tras un análisis de correlaciones), se pierde demasiada información.

- Tipo de discriminación

La discriminación C1-C234 es siempre más efectiva que la discriminación C1-C2-C3-C4. Se trata de algo lógico, pues la asignación a cuatro poblaciones siempre resultará más complicada que a dos. Por ello, se recomienda recurrir a la identificación exacta de la condición del motor únicamente cuando sea estrictamente necesario.

Tabla 10.1. Resumen sobre el acierto estimado hallado mediante los análisis discriminantes diseñados.

ACIERTO	ESTIMADO			
	C1-C234		C1-C2-C3-C4	
	<i>Stepwise</i>	<i>Backward</i>	<i>Stepwise</i>	<i>Backward</i>
DISCRIMINACIÓN				
SELECCIÓN				
M1_UT	94,51%	94,74%	92,09%	92,09%
M1_TE	69,45%	70,80%	60,62%	60,62%
M1_LN	99,98%	99,98%	80,53%	80,53%
M2_UT	92,32%	89,50%	79,40%	79,40%
M2_AB	87,58%	87,58%	83,18%	83,18%
M2_SM	92,31%	92,72%	76,93%	70,45%

- Variables de interés

La selección de variables previa al análisis ha permitido intuir que, cuanto más lejano del principal es el armónico sobre el que se toman mediciones, menor poder discriminante tienen estas. Asimismo, las

amplitudes parecen tener mayor capacidad de discriminación que los picos. Por consiguiente, dicho de otro modo, se cree que con medir las variables tipo amplitud asociadas al primer y tal vez al quinto armónico, la pérdida de información podría ser tan reducida que el análisis discriminante seguiría proporcionando una eficacia satisfactoria.

Este hecho supone una simplificación del procedimiento de medición a llevar a cabo. El número de variables es más reducido y se pueden obtener en menor tiempo, lo cual facilitaría la labor del interesado. Ahora bien, es importante remarcar que las hipótesis formuladas sobre las variables de interés no han sido contrastadas. Es decir, no se ha comprobado si efectivamente trabajando sólo con las variables comentadas, la eficacia del análisis discriminante no se ve afectada. Lo normal es que en la mayoría de casos el acierto sea similar, pero en otros es posible que la pérdida de información sea inaceptable. Consecuentemente, se sugiere que, antes de poner en práctica las hipótesis aquí formuladas, la credibilidad de estas sea verificada a través de tesis más exhaustivas.

Una vez concluida la etapa discriminante, se plantea la posibilidad de que exista un efecto significativo aportado por parte del deslizamiento sobre la respuesta final de las variables medidas. Los GAM diseñados certifican la existencia de dicha influencia, así como la ya conocida influencia del tipo. Los *boxplots* de los residuales desprendidos por los GAM, donde ya se ha restado el efecto tanto del deslizamiento como del tipo, se encuentran menos solapados que los *boxplots* de las variables originales. Así pues, se prevé que los residuales respondan mejor ante un nuevo análisis discriminante.

Sin embargo, después de hacer los estudios necesarios y validarlos convenientemente, se ha podido comprobar que la hipótesis planteada es falsa. El análisis discriminante univariante introduciendo sólo una variable explicativa, es más eficaz al emplear los residuos, como era de esperar. Pero esto no tiene por qué suponer que el análisis discriminante multivariante con todas las variables explicativas de forma simultánea deba ser más favorable. De hecho, es exactamente lo que ocurre. Como consecuencia, sigue siendo recomendable la utilización de las variables habituales como variables explicativas.

Por último, se estudia mediante nuevos GAM el efecto del deslizamiento sobre las variables explicativas por separado según la avería del motor. Lo que se pretende es comprobar si el efecto en cuestión es distinto cuando la barra del rotor está deteriorada. Los resultados muestran una característica de gran interés: los picos asociados al primer armónico se ven aumentados

significativamente a altos deslizamientos a medida que la severidad del fallo es mayor.

La importancia del fenómeno descrito reside en la posibilidad de establecer una predicción sobre el estado del motor observando el comportamiento de la respuesta en frecuencia al someter el motor a grandes deslizamientos. No obstante, por desgracia, la experimentación llevada a cabo no comprendía el análisis exhaustivo del motor a deslizamientos elevados. Fueron tomadas pocas mediciones en este rango de valores, lo cual provoca que el nivel de incertidumbre sobre las deducciones sea considerable.

En definitiva, como conclusión positiva del estudio, el indicio de que a deslizamientos altos los motores cuya barra del rotor se encuentra deteriorada presentan un incremento de la magnitud de los picos asociados al primer armónico, puede abrir una nueva línea de investigación orientada al desarrollo de experimentos del tipo “*accelerated life testing*”. Estos se caracterizan por probar un producto, sometiéndolo a determinadas condiciones (esfuerzos, temperatura, presión, voltaje, etc.) por encima de los parámetros de funcionamiento habituales, en un esfuerzo por descubrir fallos y/o modos potenciales de fallo en un corto período de tiempo (O'Connor y Kleyner, 2012). Dicho de otro modo, se “acelera” la aparición de síntomas indicativos de fallo antes incluso de que este se haya producido, de tal forma que analizando la respuesta del producto a tales pruebas, es posible hacer predicciones sobre su vida de servicio y los intervalos de mantenimiento que será necesario afrontar.

En nuestro caso particular, la idea es someter el motor puntualmente a deslizamientos por encima del rango habitual de funcionamiento, con el objetivo de comprobar si se manifiestan características indicativas de una rotura o potencial rotura de una barra del rotor. De este modo, respuestas correspondientes a un funcionamiento incorrecto que podrían tardar años en aparecer, podrían descubrirse en un periodo mucho más corto de tiempo, de meses o incluso días, haciendo posible la actuación del operario en el momento justo promoviendo las medidas preventivas, correctivas o de mantenimiento necesarias para afrontar la situación antes de que el problema se magnifique. Así, se evita la parada del sistema productivo como consecuencia de una máquina averiada.

Anexo A

Banco de ensayos

Este Anexo proporciona información sobre las características del equipo empleado para llevar a cabo los ensayos objeto de estudio en el presente TFG. Las pruebas se desarrollaron en el Laboratorio del Departamento de Ingeniería Eléctrica de la Escuela de Ingenierías Industriales de la Universidad de Valladolid.

A.1. Motores comerciales

Se emplearon dos motores de la marca *Siemens*, cuyos parámetros nominales se muestran en la Tabla A.1 y en la Tabla A.2.

Tabla A.1. Características del motor M1.

Potencia nominal	0,75 kW
Tensión nominal	3x230/400 V
Conexión del estator	Estrella
Frecuencia nominal 50 Hz	50 Hz
Corriente nominal	1,9 A
Número de pares de polos	2
Número de barras del rotor	28
Velocidad nominal	1395 rpm
Denominación	3 ~ Mot 1LA7083-4AA10
Número de serie	UD 0601/70097477-12
Eficiencia	IE1
Índice de protección	IP55
Tamaño	80M IM B3
Clase térmica	ThcL F
Factor de Servicio	1,1

Tabla A.2. Características del motor M2.

Potencia nominal	1,1 kW
Tensión nominal	3x230/400 V
Conexión del estator	Estrella
Frecuencia nominal 50 Hz	50 Hz
Corriente nominal	2,6 A
Número de pares de polos	2
Número de barras del rotor	28
Velocidad nominal	1415 rpm
Denominación	3 ~ Mot 1LA7090-51
Número de serie	UD 0701/70749029-51
Eficiencia	EFF2
Índice de protección	IP55
Tamaño	90S IM B3
Clase térmica	ThcL F
Factor de Servicio	1,1

A.2. Variadores de frecuencia

Se emplearon los siguientes variadores de frecuencia comerciales:

- Variador *PowerFlex 40* de *Allen-Bradley*
- Variador *MicroMaster 420* de *Siemens*
- Variador *Altivar 66* de *Telemecanique*

Las características de todos ellos se indican en la Tabla A.3, en la Tabla A.4 y en la Tabla A.5 respectivamente.

Tabla A.3. Características del variador *PowerFlex 40* de *Allen-Bradley*.

Potencia nominal	2,2 kW
Tensión y frecuencia de alimentación	3x380 - 480 V (48-63 Hz)
Tensión de salida	0 - 460 V
Corriente nominal	7,5 A
Frecuencia de salida	0 - 400 Hz
Corriente máx. de salida en continuo	6 A
Sobrecarga temporal (60 seg)	9 A
Denominación	22B-D6P0N104
Número de serie	BD6P0N0W8500204
Índice de protección	IP20

Tabla A.4. Características del variador *MicroMaster 420* de *Siemens*.

Potencia nominal	4 kW
Tensión y frecuencia de alimentación	3x380 - 480 V 47-63 Hz
Tensión de salida	0 – Tensión de Entrada V
Frecuencia de salida	0 - 650 Hz
Corriente máx. de salida en continuo	10,2
Denominación	6SE6420-2UD24-0BA1
Número de serie	XAW902-006579
Índice de protección	IP20

Tabla A.5. Características del variador *Altivar 66* de *Telemecanique*.

Potencia nominal	2,2 kW
Tensión y frecuencia de alimentación	3x400 V – 15% a 460 V + 15% 50/60 Hz ±5%
Tensión de salida	0 – Tensión de Entrada V
Corriente nominal	9 A
Frecuencia de salida	0,1 – 400 Hz
Corriente máx. de salida en continuo	5,8 A
Sobrecarga temporal (60 seg)	8 A
Denominación	ATV-66U41N4
Índice de protección	IP20

A.3. Freno y Unidad de Control

El freno de polvo magnético de *Lucas-Nülle* utilizado tiene las características de la Tabla A.6, mientras que la Unidad de Control asociada, también de *Lucas-Nülle*, tiene las características de la Tabla A.7.

Tabla A.6. Características del freno de polvo magnético de *Lucas-Nülle*.

Potencia nominal	1 kW
Tensión y frecuencia de alimentación	230 V 50 Hz
Par resistente	0 – 25 Nm
Velocidad de giro	0 – 6000 rpm
Denominación	SE 2662-5R

Tabla A.7. Características de la Unidad de Control de *Lucas-Nülle*.

Tensión y frecuencia de alimentación	230 V 50 Hz
Denominación	SO 3213-7S

A.4. Sensores

Los datos técnicos de los transductores de corriente de *Honeywell* y de *LEM* se recogen en la Tabla A.8 y en la Tabla A.9 respectivamente.

Tabla A.8. Datos técnicos del transductor de corriente de *Honeywell*.

Denominación	CSNE151
Corriente primaria (valor máximo continuo)	± 36 A
Tensión primaria (AC y DC)	10-500 V
Corriente secundaria nominal	25 mA (eficaces)
Tensión de alimentación	±15 Vdc
Consumo de corriente	10 (@±15V) + Corriente secundaria mA
Ratio de conversión	2500:1000
Precisión	±0,5 %
Tiempo de respuesta	1 µs
Empaquetamiento	PCB

Tabla A.9. Datos técnicos del transductor de corriente de LEM.

Denominación	LA 25-NP
Corriente primaria nominal	25 mA (eficaces)
Rango de medida	0 .. ± 36 A
Corriente secundaria nominal	25 mA (eficaces)
Tensión de alimentación	± 15 V
Consumo de corriente	10 + Corriente secundaria mA
Precisión	$\pm 0,5$ %
Tiempo de respuesta	1 μ s
Ancho de Banda (-1 dB)	DC .. 150 kHz
Empaquetamiento	PCB

Los datos técnicos del transductor de tensión de LEM se recogen en la Tabla A.10.

Tabla A.10. Datos técnicos del transductor de tensión de LEM.

Corriente primaria nominal	10 mA (eficaces)
Tensión primaria (AC y DC)	10-500 V
Corriente secundaria nominal	25 mA (eficaces)
Tensión de alimentación	$\pm 12-15$ V
Consumo de corriente	10 (@ ± 15 V) + Corriente secundaria mA
Ratio de conversión	2500:1000
Precisión a la corriente primaria nominal y a 25 °C de temperatura:	
@ ± 12 V..15 V	$\pm 0,9$ %
@ ± 15 V ($\pm 5\%$)	$\pm 0,8$ %
Tiempo de respuesta a un escalón del 90% de la corriente primaria nominal	40 μ s
Empaquetamiento	PCB

A.5. Sistema de adquisición

La plataforma de *National Instruments* comprende un sistema modular formado por un chasis compacto NI cDAQ-9174 y un módulo de adquisición NI 9215, cuyas características se recogen en la Tabla A.11 y en la Tabla A.12 respectivamente.

Tabla A.11. Características del chasis compacto NI cDAQ-9174.

Denominación	781157-01
Número de serie	14E8524
Alimentación	9-30 V
Potencia máxima consumida	15 W
Número de Contadores/Temporizadores	4
Número de módulos de E/S	4
Resolución	32 bits
Interfaz de Bus	USB 2.0 de Alta Velocidad
Transferencias de Datos de Alto Rendimiento	7
Número de ranuras	4

Tabla A.12. Características del módulo de adquisición NI 9215.

Denominación	779011-01
Número de serie	199269A-1L 14C7DAA
Tipos de Medida	Voltaje
Tipo de Aislamiento	Aislamiento de Canal a Tierra
Compatibilidad con RoHS	Sí
Entrada Analógica	Sí
Canales	4
Canales Diferenciales	4
Resolución	16 bits
Velocidad de Muestreo	100 kS/s
Máx. Voltaje de Entrada Analógica	10 V
Rango de Voltaje Máximo	-10 V a 10 V
Precisión Máxima del Rango de Voltaje	0,003 V
Muestreo Simultáneo	Sí
Conector de E/S	Conectores BNC
Temperatura de Operación	-40 a 70 °C

A.6. Sensor de velocidad

El sensor es un S60W de la marca *Datalogic*, cuyas especificaciones técnicas se indican en la Tabla A.13.

Tensión de alimentación	10 - 30 Vdc
Rizado	2 V _{pp} max.
Consumo de corriente	30 mA
Outputs	PNP o NPN 30 Vdc.
Corriente de salida	100 mA
Voltaje de saturación de salida	≤ 2V
Tiempo de respuesta	100 μs
Frecuencia de conmutación	5 kHz
Indicadores	LED's (amarillo, verde y rojo)
Ajustes	Botón de ajuste
Memorias de datos	EEPROM
Temperatura de operación	-10 a 55 °C
Distancia de funcionamiento	19± 2 mm

Anexo B

Análisis complementarios

Para evitar caer en la monotonía y repetitividad a lo largo del trabajo, se ha decidido desplazar gran parte de los análisis al presente Anexo. Distinguimos cuatro secciones, dependiendo del Capítulo al que correspondan los análisis, a saber: Capítulo 6 (apartado B.1), Capítulo 7 (apartado B.2), Capítulo 8 (apartado B.3) y Capítulo 9 (apartado B.4).

B.1. Análisis Discriminantes complementarios

En el apartado 6.2 sólo se expusieron los análisis efectuados para los ensayos del tipo M1_UT. Aquí se exponen los demás.

B.1.1. Discriminación M1_TE

B.1.1.1. Discriminación C1-C234

Selección *Stepwise*

Los coeficientes de los términos contemplados en el estudio después de la selección mediante el método *Stepwise* se presentan en la Tabla B.1. Tan sólo dos variables han sido contempladas: *amplitude_l.1* y *amplitude_u.1*.

Tabla B.1. Coeficientes estandarizados de la función discriminante. Tipo M1_TE. Distinción entre C1 y C234. Método *Stepwise*.

Función 1	
<i>amplitude_l.1</i>	0,456729
<i>amplitude_u.1</i>	0,609397

El diagrama de dispersión queda de la forma de la Fig. B.1.

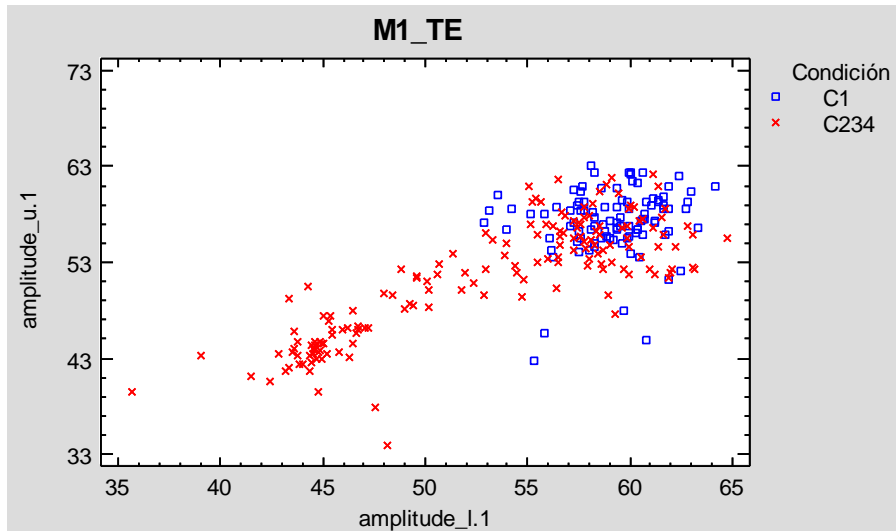


Figura B.1. Diagrama de dispersión de la relación entre *amplitude_l.1* y *amplitude_u.1*. Tipo M1_TE. Distinción entre C1 y C234. Método Stepwise.

El poder clasificatorio del análisis es del 70,23%, según la Tabla B.2. Se trata de un bagaje bastante pobre.

Tabla B.2. Clasificación entre C1 y C234 para M1_TE. Método Stepwise.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	100	95 (95,00%)	5 (5,00%)
C234	162	73 (45,06%)	89 (54,94%)
Total	262	168	94

Porcentaje de casos correctamente clasificados: 70,23%

Selección *Backward*

La técnica *Backward* difiere de la *Stepwise*, recogiendo hasta 5 variables de interés, cuyos coeficientes estandarizados son los de la Tabla B.3.

Tabla B.3. Coeficientes estandarizados de la función discriminante. Tipo M1_TE. Distinción entre C1 y C234. Método *Backward*.

	Función 1
LOG(peak_l.1)	0,394823
LOG(peak_u.1)	-0,375045
<i>amplitude_l.1</i>	0,957383
LOG(peak_l.5)	0,36328
<i>amplitude_l.5</i>	0,370044

Destaca el valor del coeficiente de *amplitude_l.1* sobre los demás. Representando esta variable frente a *peak_l.1* se obtiene la Fig. B.2. No

parece un resultado demasiado esperanzador, muchas observaciones de distinto color se superponen.

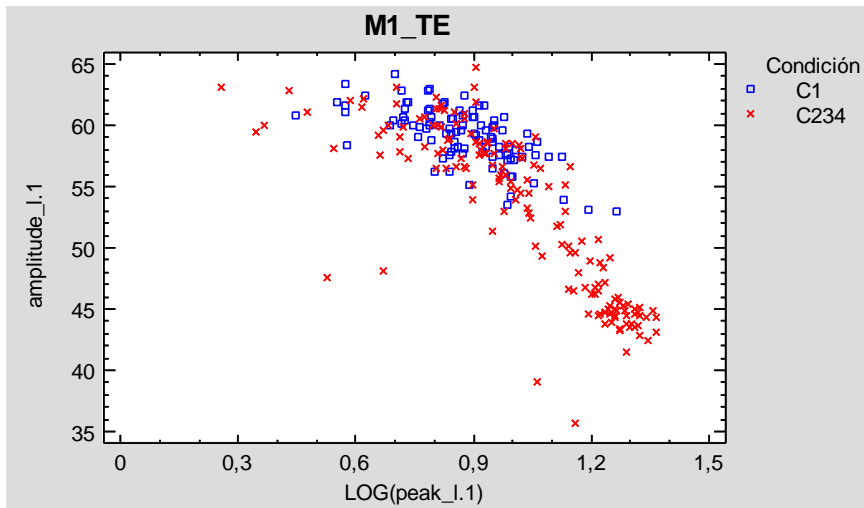


Figura B.2. Diagrama de dispersión de la relación entre amplitud_l.1 y peak_l.1. Tipo M1_TE. Distinción entre C1 y C234. Método *Backward*.

La Tabla B.4 certifica el mal presagio sobre la capacidad clasificatoria. Un porcentaje de casos satisfactoriamente clasificados de 72,90% es un valor escaso.

Tabla B.4. Clasificación entre C1 y C234 para M1_TE. Método *Backward*.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	100	89 (89,00%)	11 (11,00%)
C234	162	60 (37,04%)	102 (62,96%)
Total	262	149	113

Porcentaje de casos correctamente clasificados: 72,90%

En comparación con el método *Stepwise*, mediante el *Backward* solamente se han identificado con éxito 7 casos más, de un total de 262, lo que supone en torno a un 2% más de acierto. Resulta una mejora escasa teniendo en cuenta que se obtiene a costa de contemplar tres variables más. Por consiguiente, se considera que la metodología *Stepwise* sería más conveniente en este caso.

B.1.1.2. Discriminación C1-C2-C3-C4

Selección *Stepwise*

De nuevo son dos las variables consideradas, amplitud_l.1 y amplitud_u.1, como sucedía en la selección *Stepwise* de distinción entre C1 y C234. Aunque haya cuatro condiciones a identificar, como tenemos sólo dos

variables, únicamente se pueden calcular dos funciones de coeficientes estandarizados, como indica la Tabla B.5.

Tabla B.5. Coeficientes estandarizados de las funciones discriminantes. Tipo M1_TE. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

	Función 1	Función 2
amplitude_l.1	0,925331	0,854341
amplitude_u.1	0,115857	-1,25408

La información ofrecida por el diagrama de la Fig. B.3 es bastante confusa, de modo que no se espera una buena clasificación.

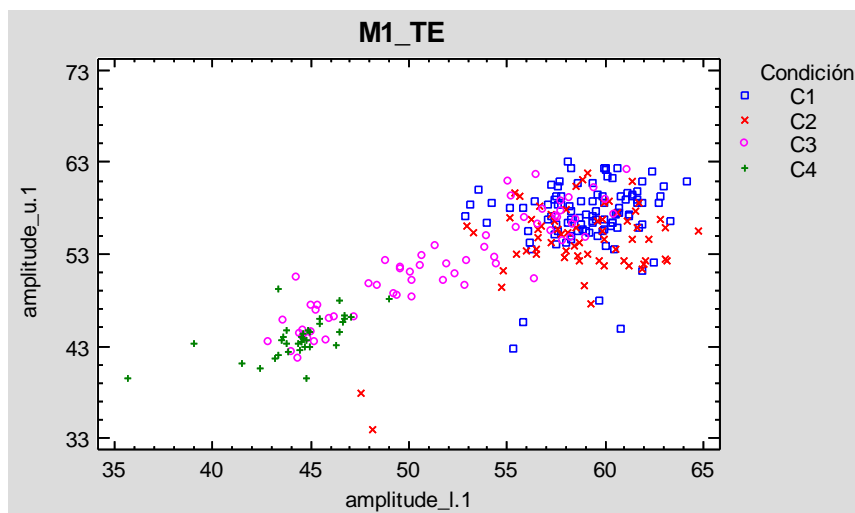


Figura B.3. Diagrama de dispersión de la relación entre amplitude_u.1, slip y speed. Tipo M1_TE. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

La deducción formulada es confirmada a través del análisis numérico de la Tabla B.6. Apenas algo más de un 60% de los ensayos es clasificado de forma adecuada. Además, en el caso de motores con semi-rotura de la barra disminuye considerablemente la eficacia del método clasificatorio, puesto que más de la mitad de los ensayos se asignan a otro tipo de severidad del fallo. Esta situación contrasta con el gran éxito de identificación de los motores con ruptura total de la barra (33 de 34 casos).

Tabla B.6. Clasificación entre C1, C2, C3 y C4 para M1_TE. Método *Stepwise*.

Grupo real	Cantidad	Grupo predicho			
		C1	C2	C3	C4
C1	100	66 (66,00%)	28 (28,00%)	6 (6,00%)	0 (0,00%)
C2	65	22 (33,85%)	38 (58,46%)	3 (4,62%)	2 (3,08%)
C3	63	17 (26,98%)	4 (6,35%)	24 (38,10%)	18 (28,57%)
C4	34	0 (0,00%)	0 (0,00%)	1 (2,94%)	33 (97,06%)
Total	262	105	70	34	53

Porcentaje de casos correctamente clasificados: 61,45%

El diagrama de las dos funciones discriminantes de la Fig. B.4 ilustra de nuevo el desorden y la dificultad de clasificación.

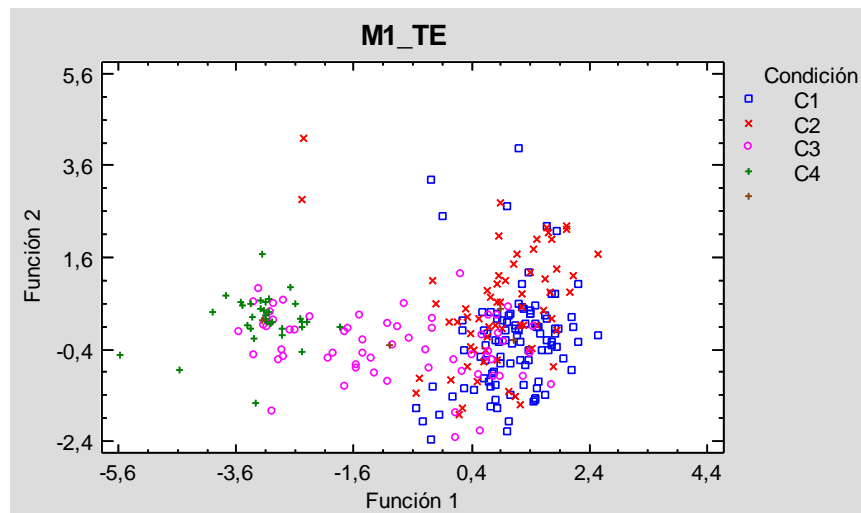


Figura B.4. Diagrama de las funciones discriminantes 1 y 2. Tipo M1_TE. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

Selección *Backward*

La técnica *Backward* recoge las mismas variables que la *Stepwise*, así que los resultados son idénticos.

B.1.2. Discriminación M1_LN

B.1.2.1. Discriminación C1-C234

Selección *Stepwise*

La metodología *Stepwise* desarrollada para M1_LN recoge en su modelo 7 de las 13 variables originales: *slip*, *peak_l.1*, *amplitud_l.1*, *amplitud_u.1*, *amplitud_l.5*, *peak_u.7* y *amplitud_u.7*. Como se puede ver en la Tabla B.7, destacan los coeficientes estandarizados de *amplitud_l.1* y *amplitud_u.1* sobre los demás.

Tabla B.7. Coeficientes estandarizados de la función discriminante. Tipo M1_LN. Distinción entre C1 y C234. Método *Stepwise*.

	Función 1
<i>slip</i>	0,806956
LOG(<i>peak_l.1</i>)	0,776114
<i>amplitud_l.1</i>	2,31185
<i>amplitud_u.1</i>	-2,04334
<i>amplitud_l.5</i>	-0,375447
LOG(<i>peak_u.7</i>)	-0,354888
<i>amplitud_u.7</i>	-0,742418

La representación en dos dimensiones del diagrama de dispersión aporta información suficientemente clara, como muestra la Fig. B.5. Los dos grupos de datos se encuentran visiblemente distinguidos.

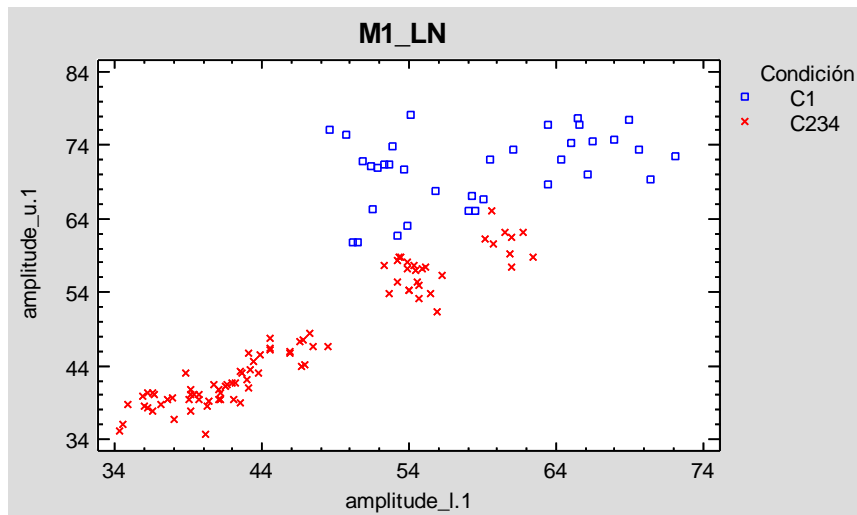


Figura B.5. Diagrama de dispersión de la relación entre amplitud_u.1 y amplitud_l.1. Tipo M1_LN. Distinción entre C1 y C234. Método *Stepwise*.

La separación de los dos grupos permite acertar con exactitud en el 100% de los ensayos, tal y como indica la Tabla B.8.

Tabla B.8. Clasificación entre C1 y C234 para M1_TE. Método *Stepwise*.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	35	35 (100,00%)	0 (0,00%)
C234	86	0 (0,00%)	86 (86,00%)
Total	121	35	86

Porcentaje de casos correctamente clasificados: 100,00%

Selección *Backward*

De nuevo, la selección *Backward* desprende los mismos resultados.

B.1.2.2. Discriminación C1-C2-C3-C4

Selección *Stepwise*

Son seleccionadas las mismas variables que en el apartado anterior a excepción de *peak_u.7*, que no se considera relevante. Los coeficientes estandarizados obtenidos se encuentran reflejados en la Tabla B.9.

Tabla B.9. Coeficientes estandarizados de las funciones discriminantes. Tipo M1_LN. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

	Función 1	Función 2	Función 3
slip	-0,317354	0,794933	-0,0258793
LOG(peak_l.1)	-0,309111	1,03649	0,913026
amplitud_l.1	-0,824713	1,96045	0,289689
amplitud_u.1	1,23376	-0,672779	0,426473
amplitud_l.5	0,386746	0,0692075	-0,608571
amplitud_u.7	0,286256	-0,47363	0,158921

Siguiendo el criterio de mostrar el diagrama de dispersión de las variables con mayor coeficiente (en valor absoluto) de la primera función discriminante, se obtiene la Fig. B.6.

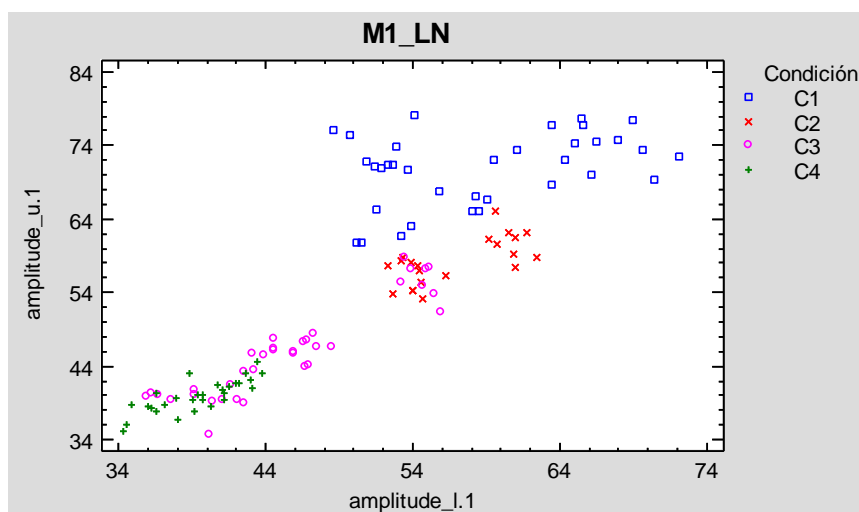


Figura B.6. Diagrama de dispersión de la relación entre *amplitud_u.1* y *amplitud_l.1*. Tipo M1_LN. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

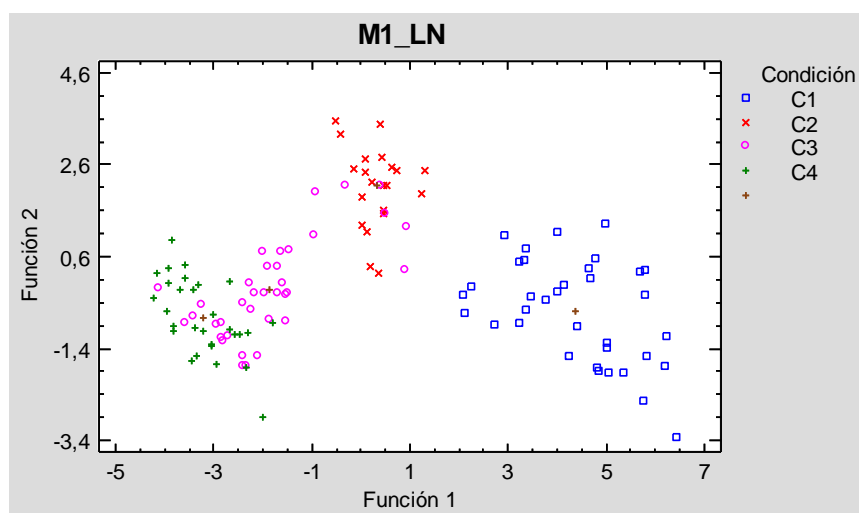
La clasificación realizada se resume en la Tabla B.10. El 84,30% de los ensayos son correctamente identificados, lo cual si nos fijamos es un resultado engañoso. En realidad, la identificación de las condiciones C1, C2 y C4 es muy satisfactoria, con un 100%, 100% y 93,10% de éxito respectivamente; mientras, la de C3 es pésima, apenas la mitad de veces (52,78%) se acierta, aproximadamente.

El diagrama de dispersión sobre las dos primeras funciones discriminantes se muestra en la Fig. B.7.

Tabla B.10. Clasificación entre C1, C2, C3 y C4 para M1_LN. Método *Stepwise*.

Grupo real	Cantidad	Grupo predicho			
		C1	C2	C3	C4
C1	35	35 (100,00%)	0 (0,00%)	0 (0,00%)	0 (0,00%)
C2	21	0 (0,00%)	21 (100,00%)	0 (0,00%)	0 (0,00%)
C3	36	0 (0,00%)	7 (19,44%)	19 (52,78%)	10 (27,78%)
C4	29	0 (0,00%)	0 (0,00%)	2 (6,90%)	27 (93,10%)
Total	121	35	28	21	37

Porcentaje de casos correctamente clasificados: 84,30%

**Figura B.7.** Diagrama de las funciones discriminantes 1 y 2. Tipo M1_LN. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

Selección *Backward*

Una vez más, la técnica *Backward* coincide en la selección de variables.

B.1.3. Discriminación M2_UT

B.1.3.1. Discriminación C1-C234

Selección *Stepwise*

Mediante este método se escogen 6 de las 13 variables posibles. Sus coeficientes estandarizados se expresan en la Tabla B.11.

Tabla B.11. Coeficientes estandarizados de la función discriminante. Tipo M2_UT. Distinción entre C1 y C234. Método Stepwise.

Función 1	
amplitud_u.1	-1,6064
LOG(peak_l.5)	0,484904
amplitud_l.5	0,909604
amplitud_u.5	1,29978
amplitud_l.7	-0,421058
amplitud_u.7	-0,520826

La Fig. B.8 refleja el diagrama de dispersión sobre las dos variables con coeficiente más destacable: amplitud_u.1 y amplitud_u.5. Se intuye una clasificación relativamente buena.

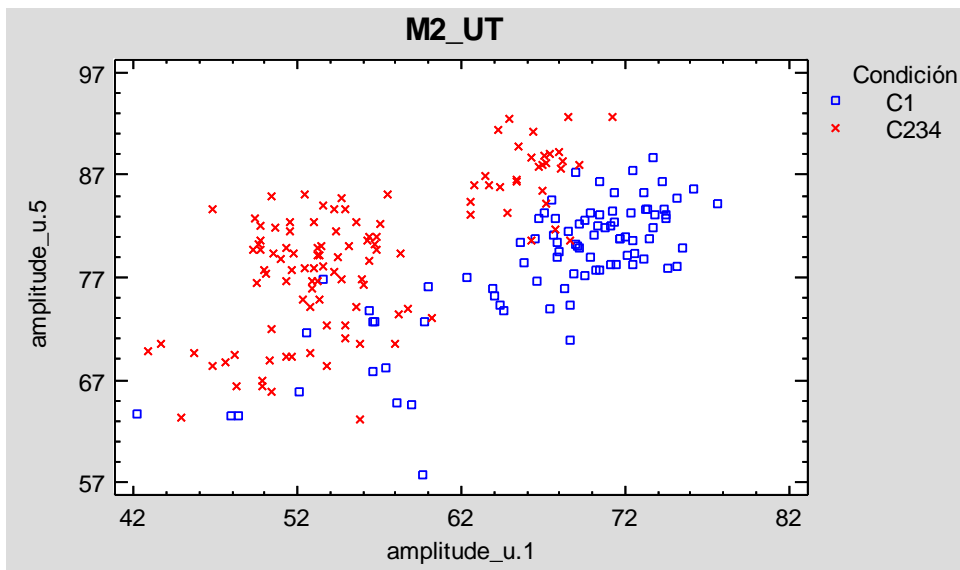


Figura B.8. Diagrama de dispersión de la relación entre amplitud_u.5 y amplitud_u.1. Tipo M2_UT. Distinción entre C1 y C234. Método Stepwise.

La Tabla B.12 de clasificación cumple con las expectativas. El 92,54% de acierto es un porcentaje satisfactorio.

Tabla B.12. Clasificación entre C1 y C234 para M2_UT. Método Stepwise.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	88	77 (87,50%)	11 (12,50%)
C234	113	4 (3,54%)	109 (96,46%)
Total	201	81	120

Porcentaje de casos correctamente clasificados: 92,54%

Selección *Backward*

La selección de variables mediante el método *Backward* difiere considerablemente de la anterior. Se seleccionan 8 variables, indicadas en la Tabla B.13 junto a su coeficiente estandarizado.

Tabla B.13. Coeficientes estandarizados de la función discriminante. Tipo M2_UT. Distinción entre C1 y C234. Método *Backward*.

	Función 1
slip	-0,307931
amplitude_u.1	-1,62025
LOG(peak_l.5)	0,492396
LOG(peak_u.5)	0,204363
amplitude_l.5	0,886954
amplitude_u.5	1,56234
amplitude_l.7	-0,37677
amplitude_u.7	-0,513585

Como los coeficientes más importantes son los mismos que mediante el método *Stepwise*, la representación del diagrama correspondiente ya se hizo en la Fig. B.8.

A continuación, se muestra el resumen de la clasificación realizada en la Tabla B.14. Se asigna la condición correctamente en el 92,54% de los casos.

Tabla B.14. Clasificación entre C1 y C234 para M2_UT. Método *Backward*.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	88	78 (88,64%)	10 (11,36%)
C234	113	5 (4,42%)	108 (95,58%)
Total	201	81	120

Porcentaje de casos correctamente clasificados: 92,54%

Globalmente, el éxito de la clasificación es idéntico al del método *Stepwise*. En principio, se puede decir que la metodología *Stepwise* es más favorable, dado que el número de variables seleccionadas es menor (6 frente a 8 de la *Backward*) y, pese a ello, no se ha perdido información importante.

B.1.3.2. Discriminación C1-C2-C3-C4

Selección *Stepwise*

Tan sólo son descartadas 5 de 13 variables. Se presentan los coeficientes estandarizados de las demás en la Tabla B.15.

Tabla B.15. Coeficientes estandarizados de las funciones discriminantes. Tipo M2_UT. Distinción entre C1, C2, C3 y C4. Método Stepwise.

	Función 1	Función 2	Función 3
slip	0,00199715	0,560521	0,819067
amplitud_e_l.1	-0,701443	-0,556554	-0,0508219
amplitud_e_u.1	-1,02314	0,671411	-0,436847
LOG(peak_l.5)	0,472463	-0,305962	-0,243817
amplitud_e_l.5	0,4668	-1,04642	0,4415
amplitud_e_u.5	0,683218	-1,38479	-0,502496
amplitud_e_l.7	0,335769	1,2838	-0,846237
amplitud_e_u.7	-0,237147	0,340983	1,15869

Las variables amplitud_e_u.1 y amplitud_e_l.1 poseen los coeficientes de mayor valor absoluto para la primera función discriminante. He aquí el diagrama de dispersión correspondiente, en la Fig. B.9.

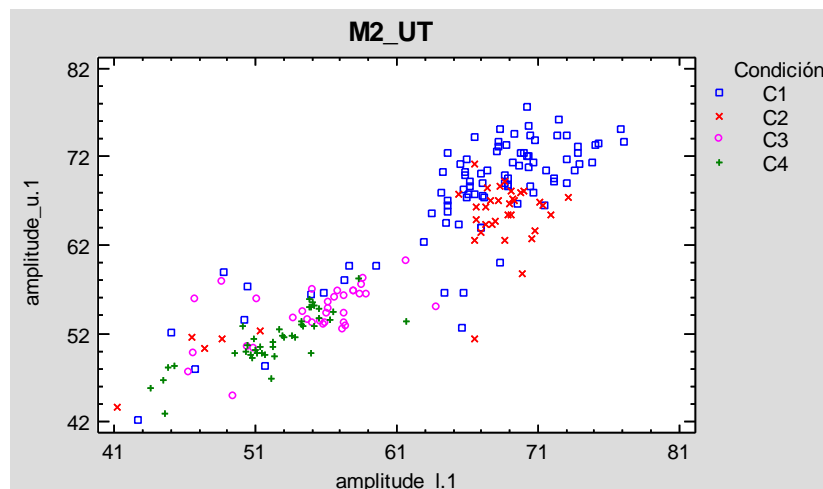


Figura B.9. Diagrama de dispersión de la relación entre amplitud_e_u.1 y amplitud_e_l.1. Tipo M2_UT. Distinción entre C1, C2, C3 y C4. Método Stepwise.

El 82,59% de casos satisfactoriamente clasificados que se resume en la Tabla B.16 es un porcentaje que, sin ser muy elevado, es razonablemente bueno. Las dos primeras condiciones se identifican especialmente bien.

Tabla B.16. Clasificación entre C1, C2, C3 y C4 para M2_UT. Método Stepwise.

Grupo real	Cantidad	Grupo predicho			
		C1	C2	C3	C4
C1	88	78 (88,64%)	5 (5,68%)	4 (4,55%)	1 (1,14%)
C2	37	2 (5,41%)	32 (86,49%)	3 (8,11%)	0 (0,00%)
C3	34	0 (0,00%)	1 (2,94%)	27 (79,41%)	6 (17,65%)
C4	42	0 (0,00%)	0 (0,00%)	13 (30,95%)	29 (69,05%)
Total	201	80	38	47	36

Porcentaje de casos correctamente clasificados: 82,59%

Como consecuencia de esa buena eficacia clasificatoria, la representación de las observaciones en las funciones discriminantes debería ofrecer una separación visual de los grupos. Así se refleja en la Fig. B.10.

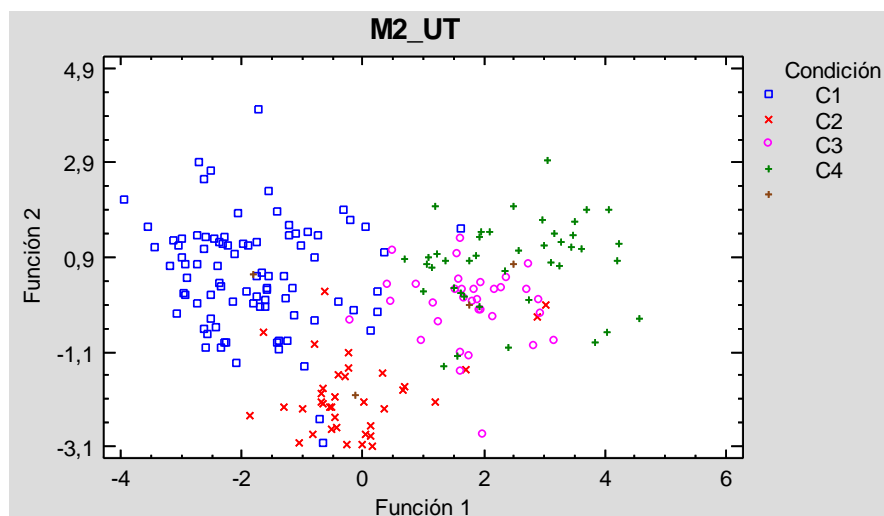


Figura B.10. Diagrama de las funciones discriminantes 1 y 2. Tipo M2_UT. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

Selección *Backward*

Los resultados obtenidos coinciden con la metodología *Stepwise*.

B.1.4. Discriminación M2_AB

B.1.4.1. Discriminación C1-C234

Selección *Stepwise*

El método *Stepwise* contempla las 7 variables reflejas en la Tabla B.17, donde además se indican sus coeficientes estandarizados.

Tabla B.17. Coeficientes estandarizados de la función discriminante. Tipo M2_AB. Distinción entre C1 y C234. Método *Stepwise*.

	Función 1
slip	-0,237386
LOG(peak_l.1)	-0,340911
amplitude_u.1	1,55587
amplitude_l.5	-0,849835
amplitude_u.5	-0,29183
LOG(peak_l.7)	-0,340743
LOG(peak_u.7)	0,207268

Véase la Fig. B.11, indicativa de la clasificación de ensayos en función de las dos variables con mayor poder discriminante. Da la sensación de que la predicción de la condición será buena, porque las observaciones no se encuentran demasiado superpuestas.

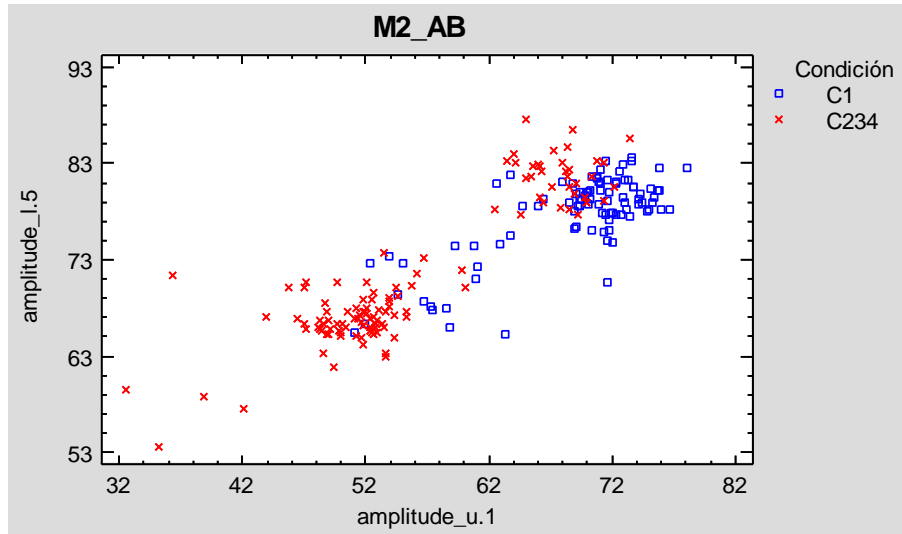


Figura B.11. Diagrama de dispersión de la relación entre amplitud_u.5 y amplitud_u.1. Tipo M2_AB. Distinción entre C1 y C234. Método *Stepwise*.

Efectivamente, el elevado acierto obtenido, de casi el 90%, es muy positivo. Además, es similar para C1 y C234, tal y como refleja la Tabla B.18.

Tabla B.18. Clasificación entre C1 y C234 para M2_UT. Método *Stepwise*.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	95	87 (91,58%)	8 (8,42%)
C234	118	14 (11,86%)	104 (88,14%)
Total	213	101	112

Porcentaje de casos correctamente clasificados: 89,67%

Selección *Backward*

De nuevo, coinciden ambas técnicas, *Stepwise* y *Backward*.

B.1.4.2. Discriminación C1-C2-C3-C4

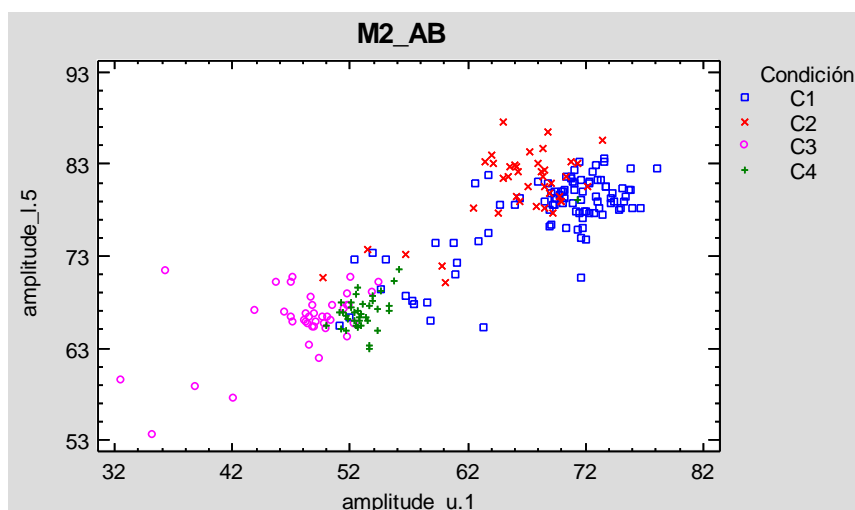
Selección *Stepwise*

Son contempladas 7 de las 13 variables mediante la selección de variables *Stepwise*. Sus coeficientes estandarizados según la función discriminante calculada se indican en la Tabla B.19.

Tabla B.19. Coeficientes estandarizados de las funciones discriminantes. Tipo M2_AB. Distinción entre C1, C2, C3 y C4. Método Stepwise.

	Función 1	Función 2	Función 3
slip	-0,493881	0,698635	-0,534895
LOG(peak_l.1)	-0,307742	-0,0297376	0,311455
LOG(peak_u.1)	0,15143	-0,328561	0,724927
amplitud_u.1	0,899359	-1,07546	0,0829575
amplitud_l.5	0,572551	1,29147	0,603902
LOG(peak_l.7)	-0,303427	-0,0202367	0,329966
amplitud_l.7	-0,628394	-0,866058	-0,0574018

El gráfico de dispersión de amplitud_l.5 frente a amplitud_u.1 se corresponde con la Fig. B.12.

**Figura B.12.** Diagrama de dispersión de la relación entre amplitud_l.5 y amplitud_u.1. Tipo M2_AB. Distinción entre C1, C2, C3 y C4. Método Stepwise.

La predicción ofrecida por el análisis planteado es bastante buena, con un 85% de éxito según refleja la Tabla B.20.

Tabla B.20. Clasificación entre C1, C2, C3 y C4 para M2_AB. Método Stepwise.

Grupo real	Cantidad	Grupo predicho				
		C1	C2	C3	C4	
C1	95	80 (84,21%)	15 (15,79%)	0 (0,00%)	0 (0,00%)	
C2	41	6 (14,63%)	35 (85,37%)	0 (0,00%)	0 (0,00%)	
C3	40	0 (0,00%)	0 (0,00%)	33 (82,50%)	7 (17,50%)	
C4	37	1 (2,70%)	1 (2,70%)	2 (5,41%)	33 (89,19%)	
Total	213		87	51	35	40

Porcentaje de casos correctamente clasificados: 84,98%

En el diagrama de la Fig. B.13 se dibujan las observaciones utilizando como ejes las dos primeras funciones discriminantes. Como signo de buena clasificación, los grupos se distribuyen razonablemente bien en zonas distintas del gráfico.

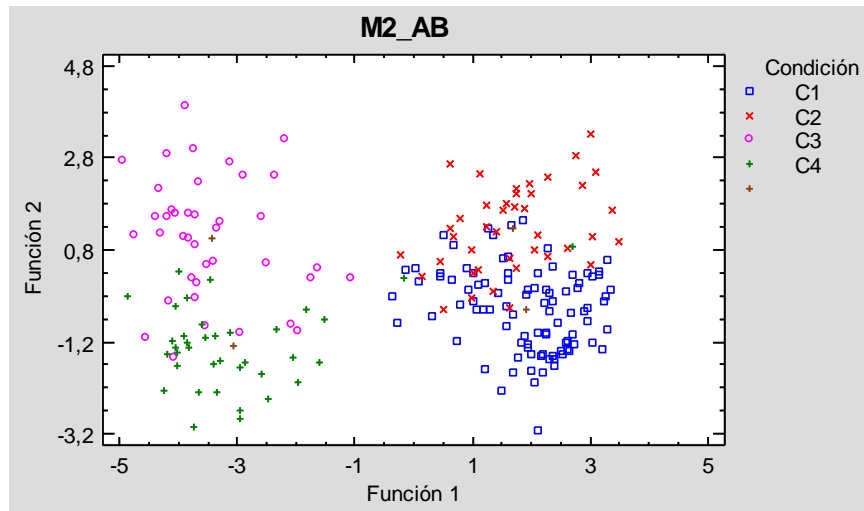


Figura B.13. Diagrama de las funciones discriminantes 1 y 2. Tipo M2_AB. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

Selección *Backward*

En esta ocasión también se obtienen los mismos resultados mediante la técnica *Backward*.

B.1.5. Discriminación M2_SM

B.1.5.1. Discriminación C1-C234

Selección *Stepwise*

Los términos recogidos en el modelo tras la aplicación de la selección tipo *Stepwise* son los de la Tabla B.21, cuyo coeficiente estandarizado se muestra al lado. Se dispone únicamente de 4 variables.

Tabla B.21. Coeficientes estandarizados de la función discriminante. Tipo M2_SM. Distinción entre C1 y C234. Método *Stepwise*.

	Función 1
slip	-0,263052
LOG(peak_l.1)	0,307938
LOG(peak_u.1)	0,740638
amplitude_u.5	0,519858

La representación de los ensayos en función de las variables peak_u.1 y amplitud_u.5 (las de mayor poder discriminante) se corresponde con el diagrama de la Fig. B.14. La distinción entre C1 y C234 es muy significativa.

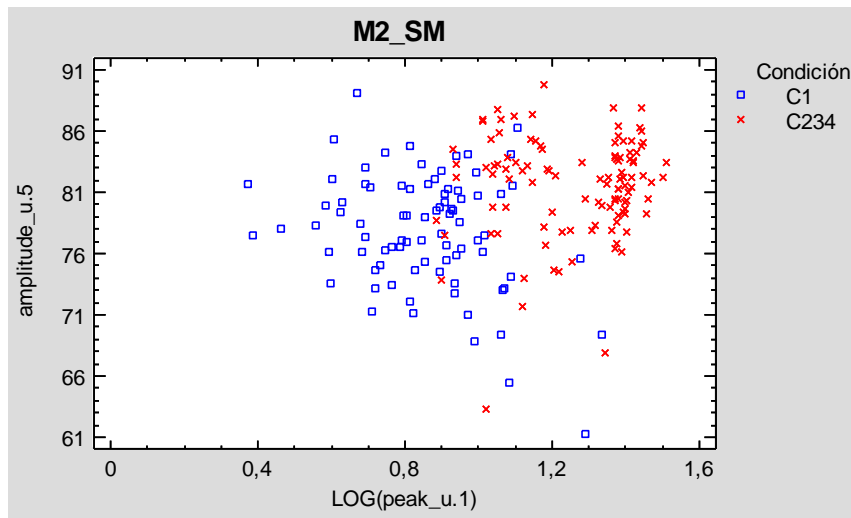


Figura B.14. Diagrama de dispersión de la relación entre amplitud_u.5 y peak_u.1. Tipo M2_SM. Distinción entre C1 y C234. Método *Stepwise*.

El análisis planteado tiene un gran poder discriminante. La Tabla B.22 así lo indica, con un porcentaje global de acierto en la predicción del 93,30%.

Tabla B.22. Clasificación entre C1 y C234 para M2_SM. Método *Stepwise*.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	84	79 (94,05%)	5 (5,95%)
C234	110	8 (7,27%)	102 (92,73%)
Total	194	87	107

Porcentaje de casos correctamente clasificados: 93,30%

Selección *Backward*

Esta vez la selección *Backward* brinda resultados diferentes con respecto a la *Stepwise*. Se añaden tres variables más: amplitud_u.1, peak_l.5 y amplitud_l.5. Véase la Tabla B.23 de coeficientes estandarizados.

Tabla B.23. Coeficientes estandarizados de la función discriminante. Tipo M2_SM. Distinción entre C1 y C234. Método *Backward*.

Función 1	
slip	-0,333459
LOG(peak_l.1)	0,249231
LOG(peak_u.1)	0,495546
amplitud_u.1	-0,499646
LOG(peak_l.5)	0,350783
amplitud_l.5	0,499786
amplitud_u.5	0,480419

La Fig. B.15 ilustra el diagrama de dispersión sobre las variables de mayor coeficiente en valor absoluto (aunque realmente hay cuatro valores prácticamente iguales).

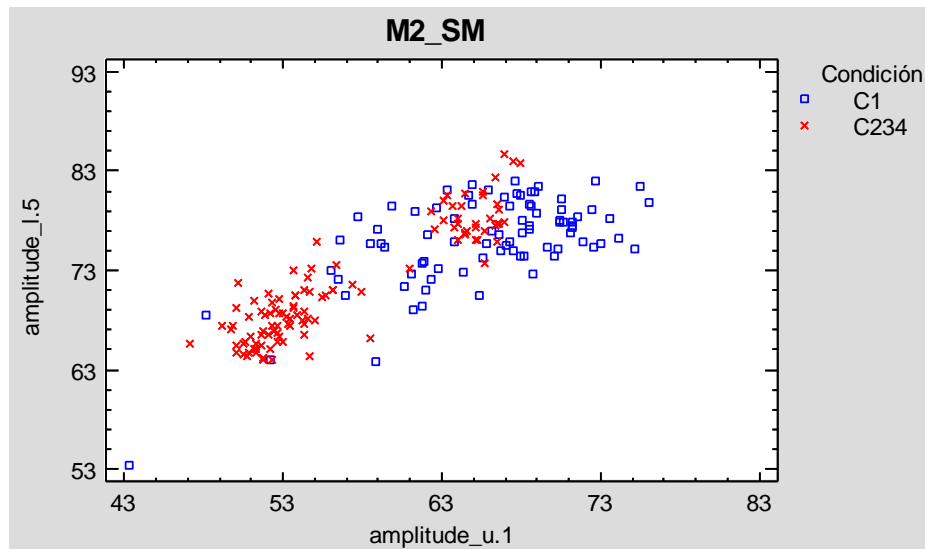


Figura B.15. Diagrama de dispersión de la relación entre amplitud_l.5 y amplitud_u.1. Tipo M2_SM. Distinción entre C1 y C234. Método *Backward*.

El éxito en la clasificación es elevado, del 93,81%. La Tabla B.24 resume la asignación de los ensayos a los grupos clasificadores.

Tabla B.24. Clasificación entre C1 y C234 para M2_SM. Método *Backward*.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	84	82 (97,62%)	2 (2,38%)
C234	110	10 (9,09%)	100 (90,91%)
Total	194	92	102

Porcentaje de casos correctamente clasificados: 93,81%

Comparativamente, la técnica *Backward* ha supuesto el acierto de tan sólo un ensayo más que la *Stepwise*, de un total de 194 ensayos. El esfuerzo que supone registrar el valor de más variables hace que se afirme la conveniencia del método *Stepwise*.

B.1.5.2. Discriminación C1-C2-C3-C4

Selección *Stepwise*

La distinción entre los cuatro estados del motor mediante una selección *Stepwise* requiere la presencia de 5 variables. En la Tabla B.25 se presentan sus respectivos coeficientes estandarizados de las funciones discriminantes halladas.

Tabla B.25. Coeficientes estandarizados de las funciones discriminantes. Tipo M2_SM. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

	Función 1	Función 2	Función 3
LOG(peak_l.1)	0,332184	-0,118225	-0,311201
LOG(peak_u.1)	0,566937	0,477051	-0,307751
amplitude_l.1	-0,418309	-0,0574878	-1,4249
amplitude_l.5	-0,310909	0,725743	0,695749
amplitude_u.5	0,55309	0,365373	0,583089

Como peak_u.1 y amplitude_u.5 poseen el mayor poder de discriminación según la primera función, se expresa el diagrama de dispersión asociado en la Fig. B.16. A priori parece que puede haber alguna dificultad en la clasificación, por la mezcla de los ensayos de diferentes condiciones.

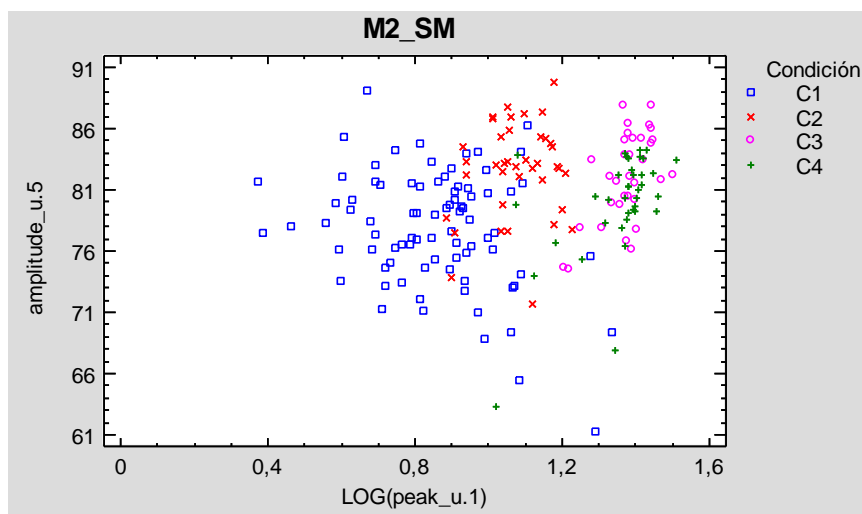


Figura B.16. Diagrama de dispersión de la relación entre amplitude_u.5 y peak_u.1. Tipo M2_SM. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

La Tabla B.26 es un resumen de las predicciones. Mientras que los estados C1 y C2 tienen un porcentaje aceptable, el éxito en los estados C3 y C4 no es el deseado.

Tabla B.26. Clasificación entre C1, C2, C3 y C4 para M2_SM. Método Stepwise.

Grupo real	Cantidad	Grupo predicho			
		C1	C2	C3	C4
C1	84	71 (84,52%)	12 (14,29%)	0 (0,00%)	1 (1,19%)
C2	39	4 (10,26%)	33 (84,62%)	1 (2,56%)	1 (2,56%)
C3	33	0 (0,00%)	0 (0,00%)	24 (72,73%)	9 (27,27%)
C4	38	1 (2,63%)	2 (5,26%)	11 (28,95%)	24 (63,16%)
Total	194	76	47	36	35

Porcentaje de casos correctamente clasificados: 78,35%

Véase el diagrama en las dos primeras funciones discriminantes ilustrado en la Fig. B.17. Los grupos C3 y C4 se sitúan prácticamente en la misma zona, lo cual dificulta la clasificación, como se ha visto antes.

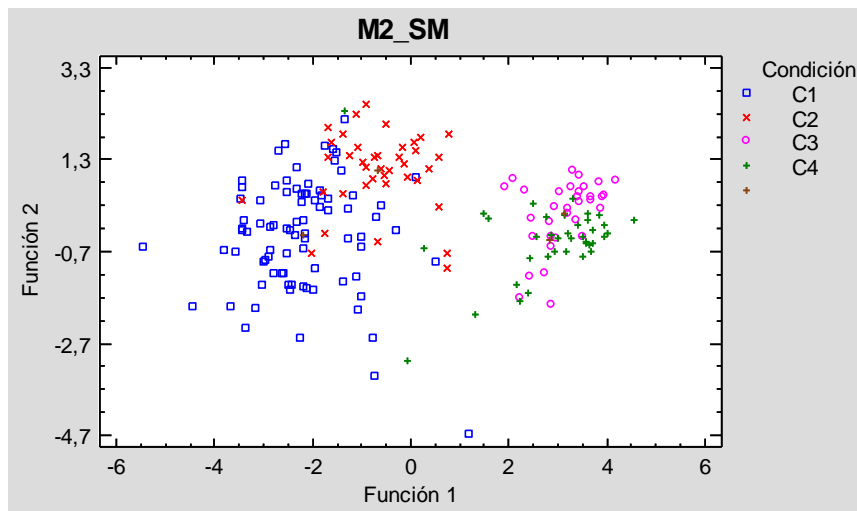


Figura B.17. Diagrama de las funciones discriminantes 1 y 2. Tipo M2_SM. Distinción entre C1, C2, C3 y C4. Método Stepwise.

Selección Backward

La metodología *Backward* comprende 6 variables en el análisis. La Tabla B.27 señala los coeficientes estandarizados obtenidos.

La representación de *amplitud_e.l.1* frente a *peak_u.1* es de la forma de la Fig. B.18. El diagrama muestra claras deficiencias en la distinción entre C3 y C4, que se encuentran situados en la misma zona.

Tabla B.27. Coeficientes estandarizados de las funciones discriminantes. Tipo M2_SM. Distinción entre C1, C2, C3 y C4. Método *Backward*.

	Función 1	Función 2	Función 3
LOG(peak_u.1)	0,573467	0,478852	0,248206
amplitude_l.1	-0,743965	0,0477044	0,814617
amplitude_l.5	-0,371297	0,752672	-0,477767
amplitude_u.5	0,442259	0,383667	-0,501907
LOG(peak_l.7)	0,264121	0,0348356	0,712896
amplitude_l.7	0,522068	-0,180361	0,064475

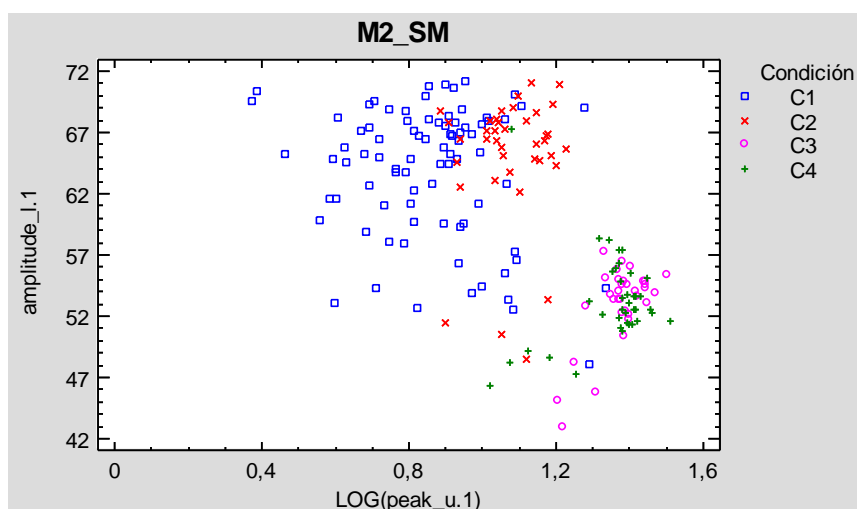


Figura B.18. Diagrama de dispersión de la relación entre amplitude_l.1 y peak_u.1. Tipo M2_SM. Distinción entre C1, C2, C3 y C4. Método *Backward*.

La Tabla B.28 concuerda con lo afirmado. La clasificación empeora con los grupos C3 y C4; para C1 y C2 los porcentajes de éxito son bastante buenos, superando con creces el 80%.

Tabla B.28. Clasificación entre C1, C2, C3 y C4 para M2_SM. Método *Backward*.

Grupo real	Cantidad	Grupo predicho				
		C1	C2	C3	C4	
C1	84	71 (84,52%)	11 (13,10%)	0 (0,00%)	2 (2,38%)	
C2	39	4 (10,26%)	35 (89,74%)	0 (0,00%)	0 (0,00%)	
C3	33	0 (0,00%)	1 (3,03%)	24 (72,73%)	8 (24,24%)	
C4	38	1 (2,63%)	2 (5,26%)	8 (21,05%)	27 (71,05%)	
Total	194		76	49	32	37

Porcentaje de casos correctamente clasificados: 80,93%

Los ensayos representados en las dos primeras funciones discriminantes desprenden el diagrama de la Fig. B.19. De nuevo, se observa una concentración de los grupos C3 y C4 en la misma zona, como era de esperar.

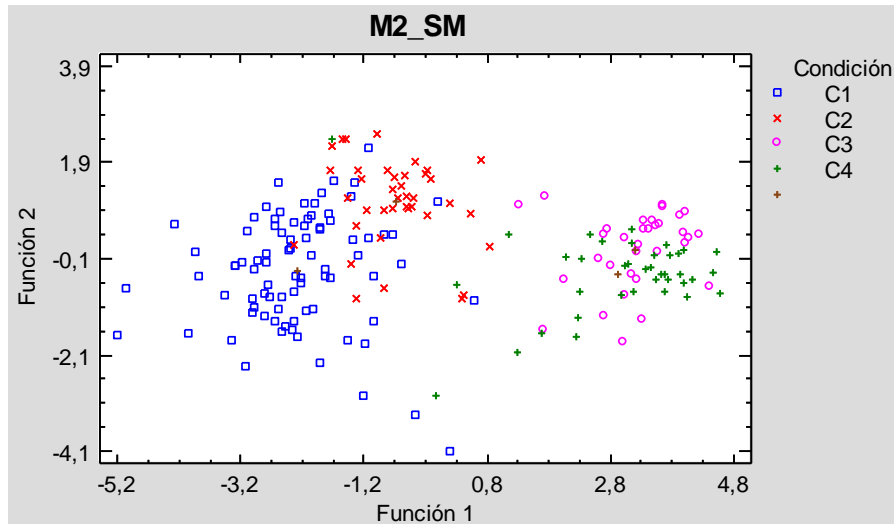


Figura B.19. Diagrama de las funciones discriminantes 1 y 2. Tipo M2_SM. Distinción entre C1, C2, C3 y C4. Método *Backward*.

La metodología *Backward* ha introducido una variable más que la *Stepwise* en el modelo y, sin embargo, la mejora clasificatoria ha sido escasa, de en torno a un 2%. Se puede decir que ambos sistemas responden adecuadamente aunque, en la práctica, se recomienda la técnica *Stepwise*, por el ahorro en la medición de una variable menos.

B.2. GAM complementarios

En el Capítulo 7 se expusieron únicamente los GAM para las variables *peak_l.1* y *amplitud_u.1*, a modo de ejemplo, además de un par de diagramas de cajas sobre las variables *peak_l.7* y *amplitud_l.7*. Aquí se muestran el resto de análisis.

PEAK_L.1

Los resultados fueron comentados en el apartado 7.1.

PEAK_U.1

El modelo contemplado para *peak_u.1* es el siguiente:

$$peak_u.1 \sim \mu + type + s(slip) + \varepsilon$$

La Fig. B.20 ilustra los resultados del GAM. La alimentación es determinante tal y como era de esperar. Los tipos M1_LN y M1_UT presentan picos más pronunciados, mientras que para M1_TE se registran los picos más pequeños.

Respecto al deslizamiento, se ve con mayor claridad las mismas características que para el GAM de *peak_l.1* visto en el apartado 7.1: a

medida que el deslizamiento va aumentando, el pico se amplifica, hasta llegar a un valor ligeramente inferior a 0,07 a partir del cual el efecto comienza a disminuir.

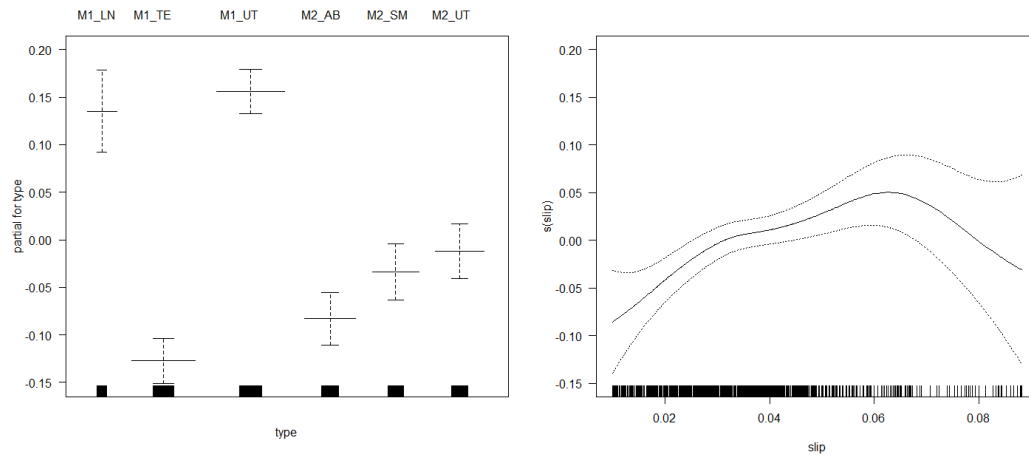


Figura B.20. Gráficos ofrecidos por el GAM sobre peak_u.1. A la izquierda, efecto del tipo; a la derecha; efecto del deslizamiento.

Los residuales representados en función de la condición se reflejan a la izquierda de la Fig. B.21, mientras que la variable original lo hace a la derecha. Los resultados mejoran con este último modelo planteado, pues el solapamiento de las cajas ha disminuido.

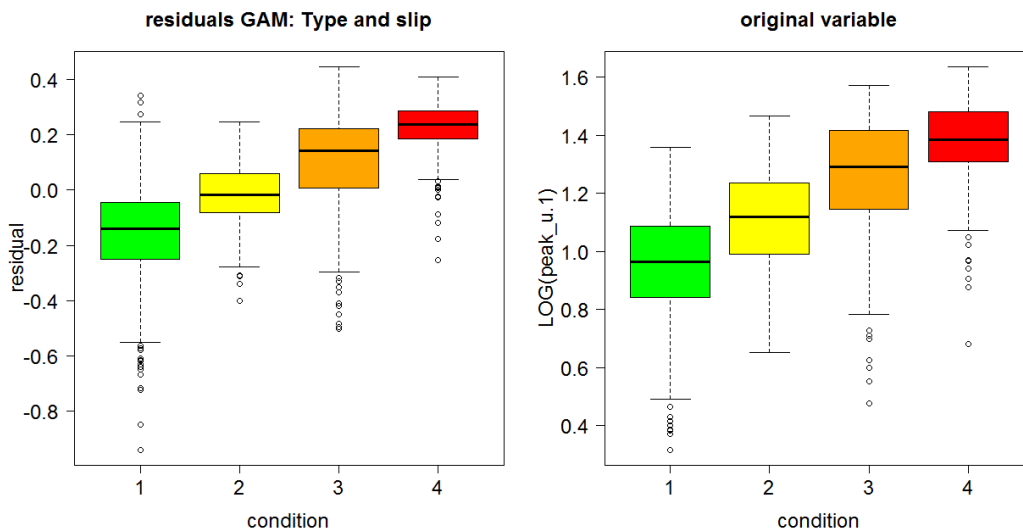


Figura B.21. Diagramas de cajas relacionados con la variable peak_u.1 en función de la condición del motor. A la izquierda, diagramas de los residuales obtenidos mediante el GAM; a la derecha, diagramas de la variable original.

AMPLITUDE_L.1

El GAM asociado a la variable *amplitude_l.1* se formula como

$$amplitude_l.1 \sim \mu + type + s(slíp) + \varepsilon.$$

El efecto del tipo y del deslizamiento, representado en la Fig. B.22, es innegablemente existente, aunque la forma en que lo hace es distinta a *peak_u.1*. El intervalo de confianza aumenta a valores altos de deslizamiento. Aunque para *peak_u.1* también sucedía, ahora contrasta más con el reducido intervalo de confianza a niveles de bajo deslizamiento.

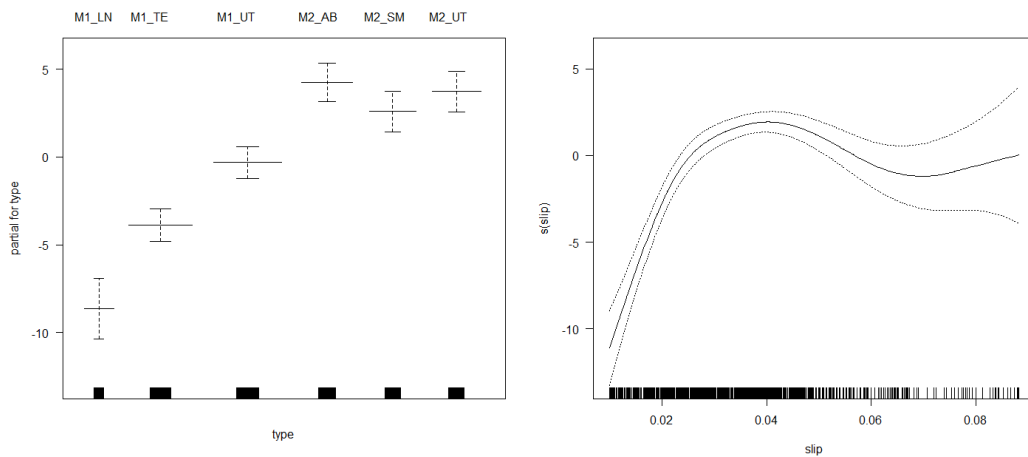


Figura B.22. Gráficos ofrecidos por el GAM sobre *amplitude_l.1*. A la izquierda, efecto del tipo; a la derecha; efecto del deslizamiento.

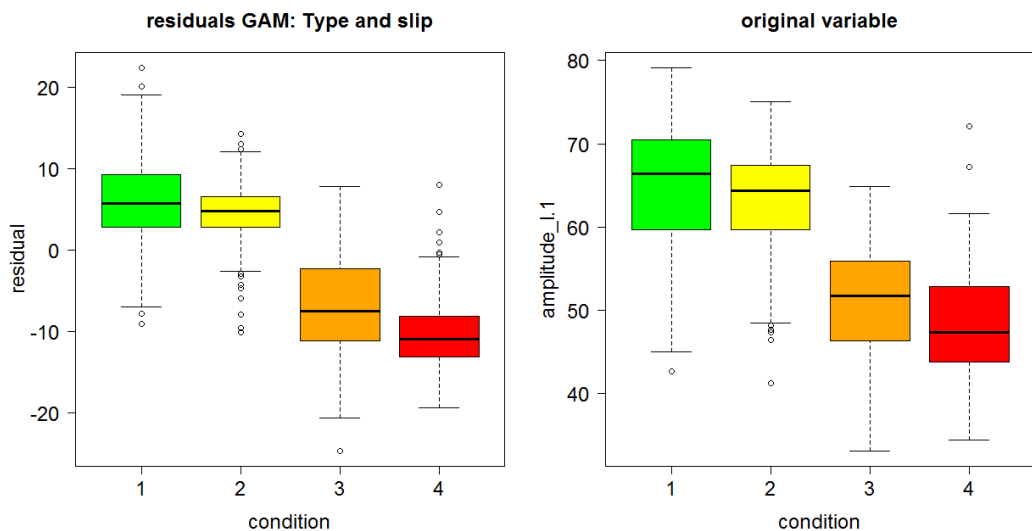


Figura B.23. Diagramas de cajas relacionados con la variable *amplitude_l.1* en función de la condición del motor. A la izquierda, diagramas de los residuales obtenidos mediante el GAM; a la derecha, diagramas de la variable original.

Se registran los residuales correspondientes y se representan junto con la variable original en los diagramas de cajas de la Fig. B.23. También en esta ocasión se produce menor superposición de las observaciones según la condición del motor, lo cual es favorable con respecto a las variables originales.

AMPLITUDE_U.1

Véase el apartado 7.1.

PEAK_L.5

Se contempla el siguiente GAM asociado a `peak_l.5`:

$$peak_l.5 \sim \mu + type + s(slip) + \varepsilon$$

Los gráficos desprendidos se reflejan en la Fig. B.24. Destaca entre otras cosas el prácticamente inexistente efecto de los tipos M2_SM y M2_UT sobre la respuesta. En cuanto al deslizamiento, un deslizamiento bajo reduce el valor de `peak_l.5`, pero cuando alcanza un determinado valor en torno a 0,03 deja de aportar ninguna disminución o aumento de la respuesta.

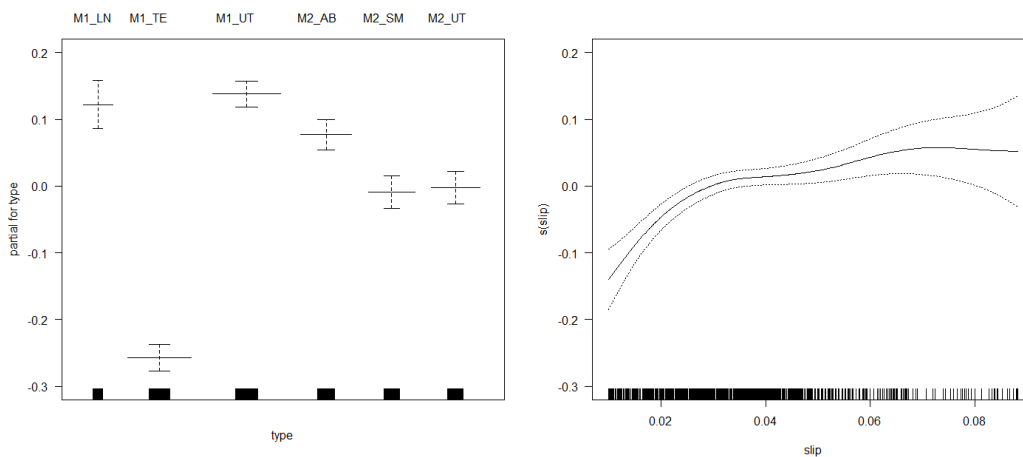


Figura B.24. Gráficos ofrecidos por el GAM sobre `peak_l.5`. A la izquierda, efecto del tipo; a la derecha; efecto del deslizamiento.

Los residuos son registrados y comparados con la variable original mediante diagramas de cajas, reflejados en la Fig. B.25. En esta ocasión no se aprecia con seguridad que los residuos vayan a ser más beneficiosos desde un punto de vista clasificatorio. La dispersión de las cajas entre sí es similar.

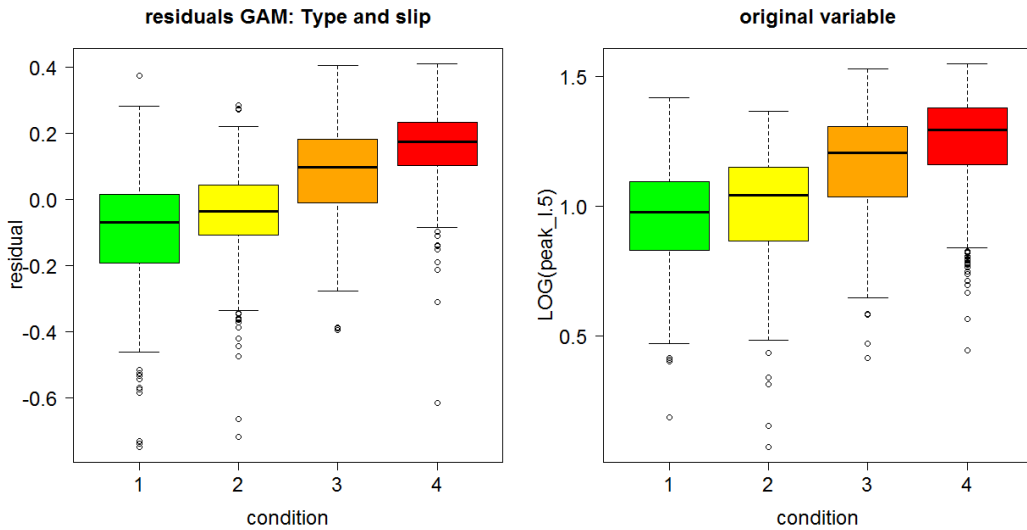


Figura B.25. Diagramas de cajas relacionados con la variable peak_l.5 en función de la condición del motor. A la izquierda, diagramas de los residuales obtenidos mediante el GAM; a la derecha, diagramas de la variable original.

PEAK_U.5

Como siempre, el modelo adopta la forma

$$peak_u.5 \sim \mu + type + s(slip) + \varepsilon$$

Los diagramas ofrecidos por el análisis se presentan en la Fig. B.26. El efecto introducido por el factor tipo es más reducido que en casos anteriores, aunque continúa sin ser despreciable. En cuanto al introducido por el deslizamiento, al principio puede considerarse prácticamente nulo, pero a medida que va aumentando influye cada vez más.

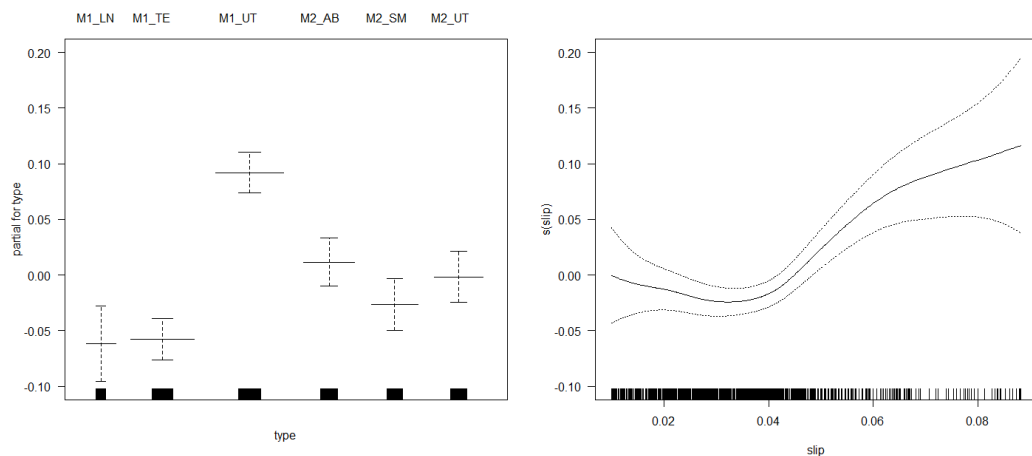


Figura B.26. Gráficos ofrecidos por el GAM sobre peak_u.5. A la izquierda, efecto del tipo; a la derecha; efecto del deslizamiento.

Los residuales son claramente ineficaces, no aportan mayor información que la variable original. La Fig. B.27 así lo ilustra.

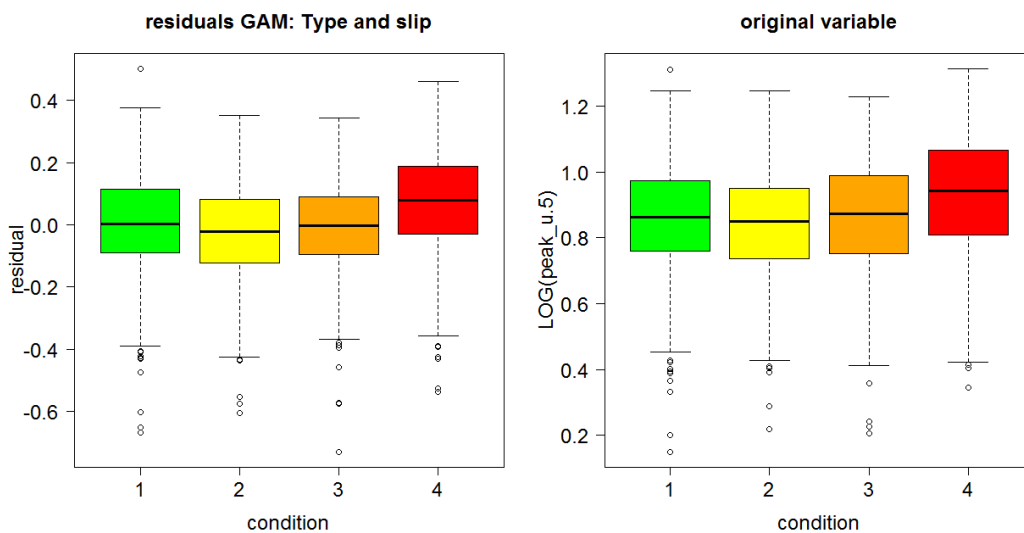


Figura B.27. Diagramas de cajas relacionados con la variable peak_u.5 en función de la condición del motor. A la izquierda, diagramas de los residuales obtenidos mediante el GAM; a la derecha, diagramas de la variable original.

AMPLITUDE_L.5

Se formula el modelo

$$amplitude_l.5 \sim \mu + type + s(slip) + \varepsilon$$

Este GAM ofrece unos gráficos que se representan en la imagen de la Fig. B.28. El efecto negativo introducido por M1_TE es muy grande. Respecto al efecto del deslizamiento es creciente con pendiente creciente por tramos.

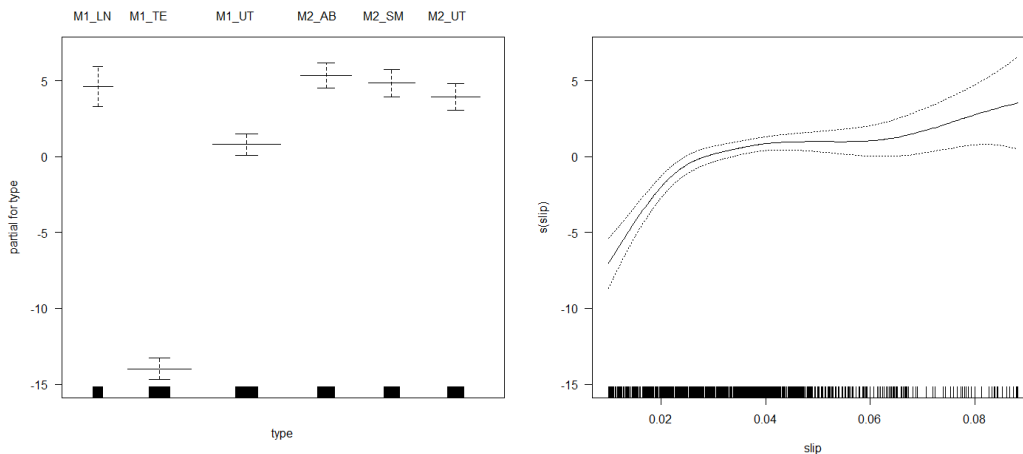


Figura B.28. Gráficos ofrecidos por el GAM sobre amplitud_l.5. A la izquierda, efecto del tipo; a la derecha; efecto del deslizamiento.

Los *boxplots* de los residuales presentan menor solapamiento que los de la variable original, según lo indica la Fig. B.29.

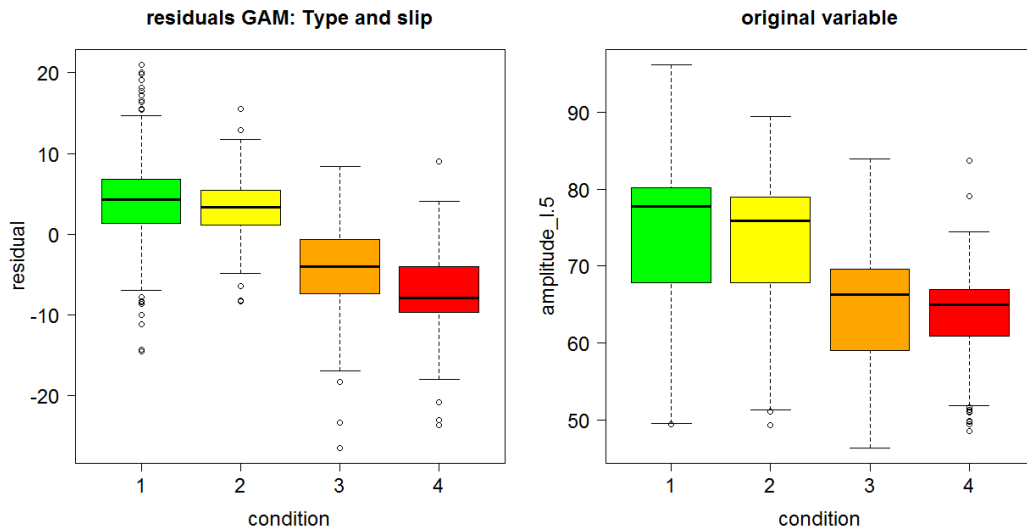


Figura B.29. Diagramas de cajas relacionados con la variable *amplitud_u.5* en función de la condición del motor. A la izquierda, diagramas de los residuales obtenidos mediante el GAM; a la derecha, diagramas de la variable original.

AMPLITUDE_U.5

Se plantea el modelo

$$amplitude_u.5 \sim \mu + type + s(slip) + \varepsilon$$

En esta ocasión, el efecto del tipo se hace aún más pronunciado. En M1_TE llega a ser inferior a 20 dB. El efecto del deslizamiento, por otro lado, sigue la misma forma que en casos anteriores. Véase la Fig. B.30.

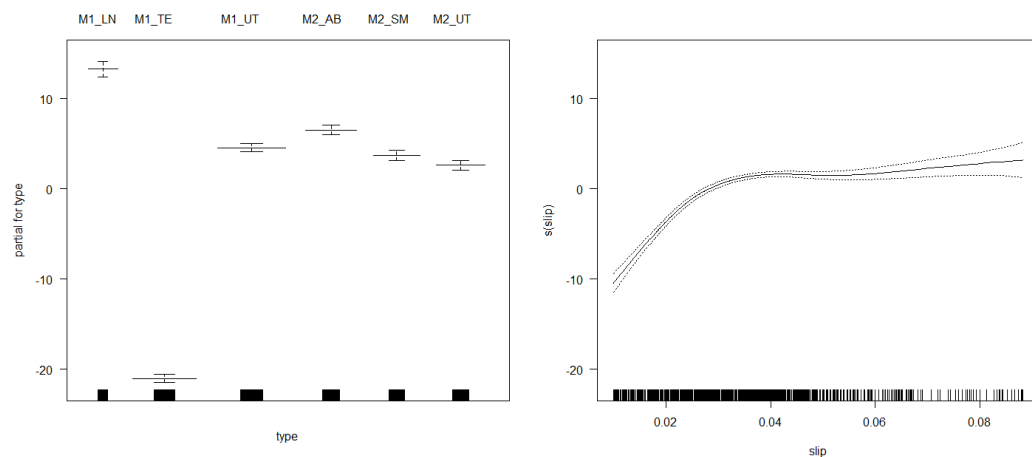


Figura B.30. Gráficos ofrecidos por el GAM sobre *amplitud_u.5*. A la izquierda, efecto del tipo; a la derecha; efecto del deslizamiento.

Del mismo modo, aunque los diagramas de cajas de los residuales se encuentran bastante solapados entre sí, la variabilidad es sensiblemente inferior a la de la variable original. En la Fig. B.31 se puede ver esta cualidad.

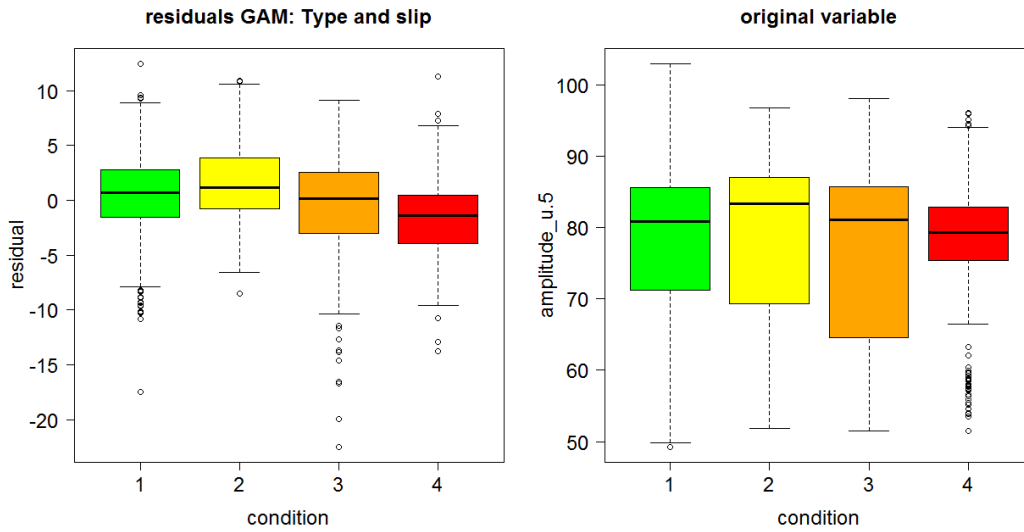


Figura B.31. Diagramas de cajas relacionados con la variable *amplitude_u.5* en función de la condición del motor. A la izquierda, diagramas de los residuales obtenidos mediante el GAM; a la derecha, diagramas de la variable original.

PEAK_L.7

El GAM planteado sigue la fórmula

$$peak_l.7 \sim \mu + type + s(slip) + \varepsilon$$

Según se aprecia en los gráficos de la Fig. B.32, el efecto de la alimentación sigue siendo significativo, pero el efecto del deslizamiento se disipa considerablemente respecto a casos anteriores.

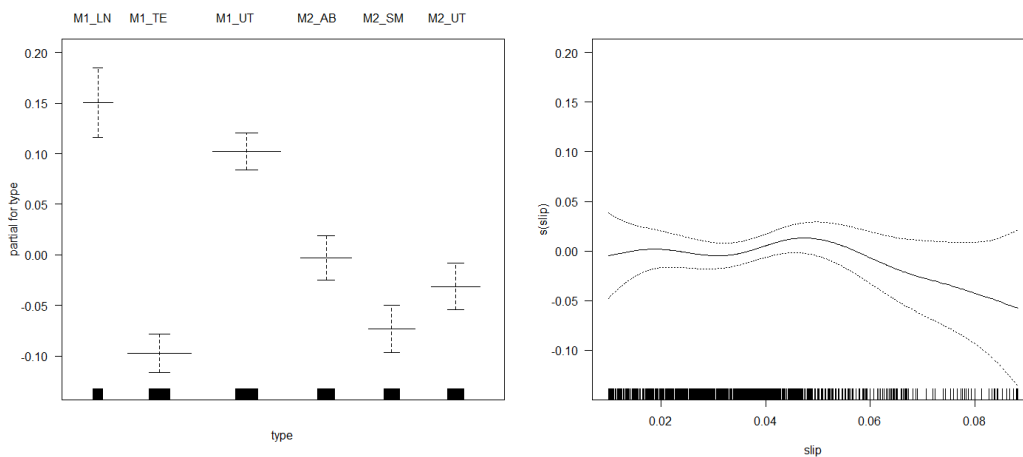


Figura B.32. Gráficos ofrecidos por el GAM sobre *peak_l.7*. A la izquierda, efecto del tipo; a la derecha; efecto del deslizamiento.

Los diagramas de los residuales se comportan de una forma similar a los de la variable inicial, de modo que no parecen aportar información relevante. La Fig. 7.3 recogida en el Capítulo 7 lo ilustra.

PEAK_U.7

La influencia del tipo y del deslizamiento sobre la variable se refleja mediante el modelo

$$peak_u.7 \sim \mu + type + s(slip) + \varepsilon$$

El gráfico izquierdo de la Fig. B.33 refleja que el efecto de la combinación motor-alimentación en la variable peak_u.7 es casi inexistente. Todos los intervalos de confianza establecidos para cada tipo contienen al cero o, como son los casos de M1_UT y de M1_TE, no lo logran por poco.

En cuanto al deslizamiento mostrado en la parte derecha, aunque intuitivamente parezca muy significativo, lo cierto es que la magnitud del efecto es reducida y el intervalo de confianza contiene al cero prácticamente en todos los tramos de la función.

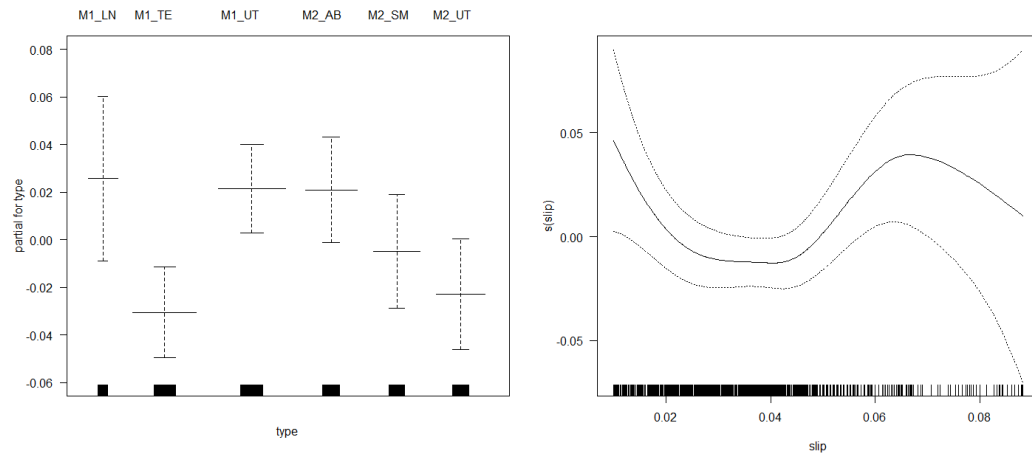


Figura B.33. Gráficos ofrecidos por el GAM sobre peak_u.7. A la izquierda, efecto del tipo; a la derecha; efecto del deslizamiento.

Por su parte, los residuos, cuya dispersión y solapamiento se comparan en la Fig. B.34 con los datos originales, parecen no ser más beneficiosos.

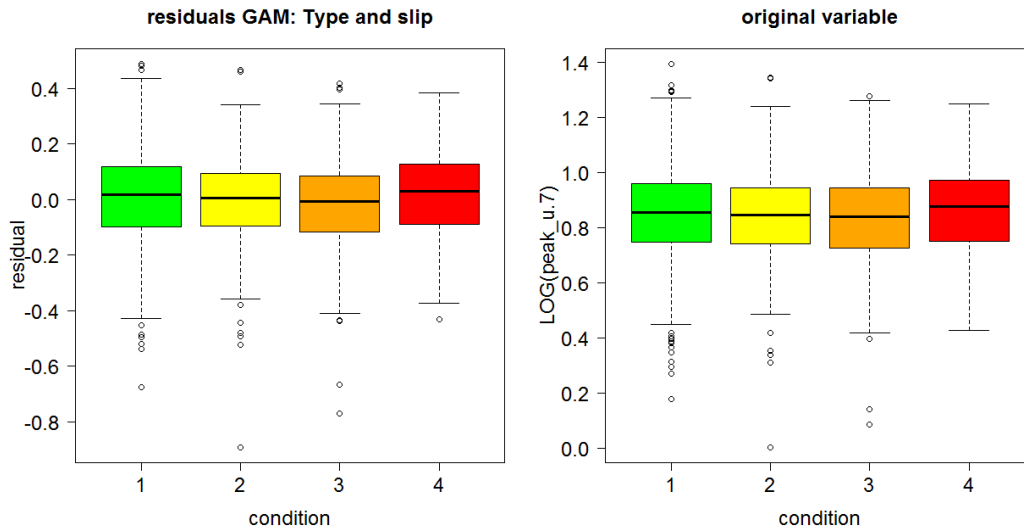


Figura B.34. Diagramas de cajas relacionados con la variable peak_u.7 en función de la condición del motor. A la izquierda, diagramas de los residuales obtenidos mediante el GAM; a la derecha, diagramas de la variable original.

AMPLITUDE_L.7

El modelo planteado es el siguiente:

$$amplitude_l.7 \sim \mu + type + s(slip) + \varepsilon$$

El GAM desprende los resultados percibidos en la Fig. B.35. Se verifica que el efecto asociado al tipo es indudablemente significativo, ningún intervalo contiene al cero. Respecto a la función del deslizamiento, destaca con respecto a casos anteriores el pronunciado aumento del efecto a partir de valores de 0,05 aproximadamente.

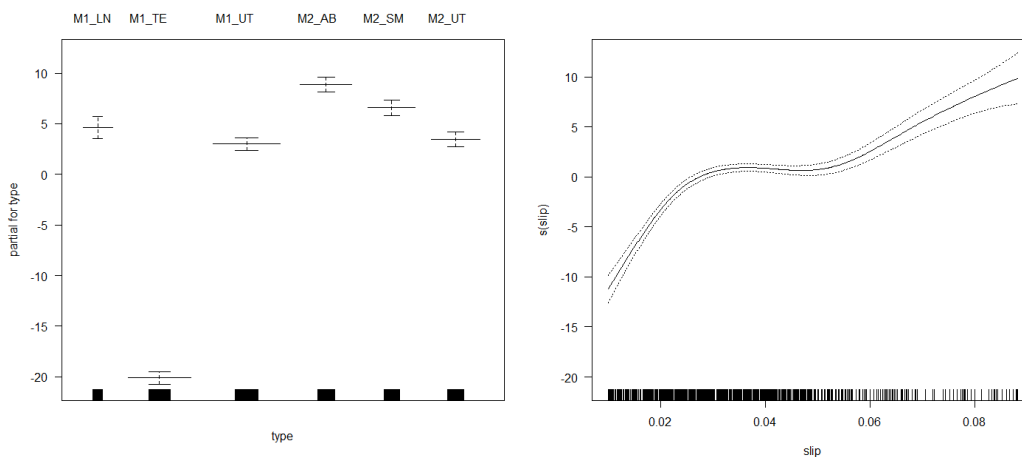


Figura B.35. Gráficos ofrecidos por el GAM sobre amplitude_l.7. A la izquierda, efecto del tipo; a la derecha; efecto del deslizamiento.

Los diagramas de cajas residuales y variables originales fueron presentados en la Fig. 7.6 del Capítulo 7. La extensión de los residuales es más reducida, pero la superposición de unas cajas con otras se produce en

ambos casos, por lo que puede que los residuales registrados no aporten mejores resultados.

AMPLITUDE_U.7

Por último, se modela el análisis

$$amplitude_u.7 \sim \mu + type + s(slip) + \varepsilon$$

Este GAM presenta gráficos muy parecidos a los obtenidos con *amplitude_l.7*, tal y como muestra la Fig. B.36. La diferencia más reseñable es que esta vez la pendiente de la función del deslizamiento para valores altos es más suave.

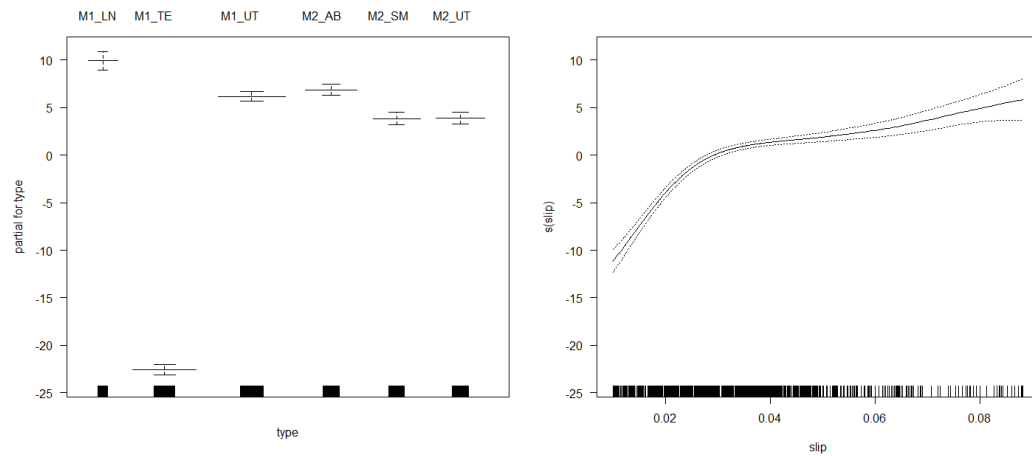


Figura B.36. Gráficos ofrecidos por el GAM sobre *amplitude_u.7*. A la izquierda, efecto del tipo; a la derecha; efecto del deslizamiento.

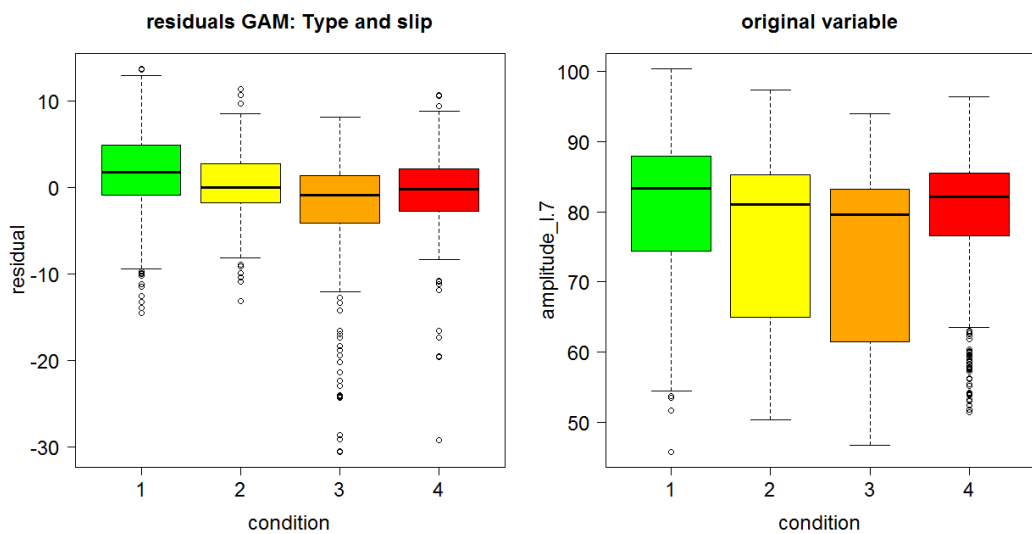


Figura B.37. Diagramas de cajas relacionados con la variable *amplitude_u.7* en función de la condición del motor. A la izquierda, diagramas de los residuales obtenidos mediante el GAM; a la derecha, diagramas de la variable original.

Los residuales se encuentran menos dispersos que las mediciones originales, pero existe un gran solapamiento no favorable. Véase la Fig. B.37.

B.3. Análisis Discriminantes complementarios con residuales del GAM

En el apartado 6.2 sólo se expusieron los análisis efectuados para los ensayos del tipo M1_UT. Aquí se exponen los demás.

B.3.1. Discriminación M1_TE

B.3.1.1. Discriminación C1-C234

Para el tipo M1_TE se seleccionan tan sólo tres variables: $r_{amp_l.1}$, $r_{amp_u.1}$ y $r_{amp_u.5}$. Sus coeficientes estandarizados se pueden ver en la Tabla B.29.

Tabla B.29. Coeficientes estandarizados de la función discriminante. Tipo M1_TE. Distinción entre C1 y C234. Método Stepwise.

Función 1	
$r_{amp_l.1}$	0,401956
$r_{amp_u.1}$	0,659944
$r_{amp_u.5}$	-0,279061

El diagrama de dispersión es muy confuso, tal y como ilustra la Fig. B.38. Además, la información de este diagrama es prácticamente la total, ya que únicamente existe una variable más, con menor poder discriminante.

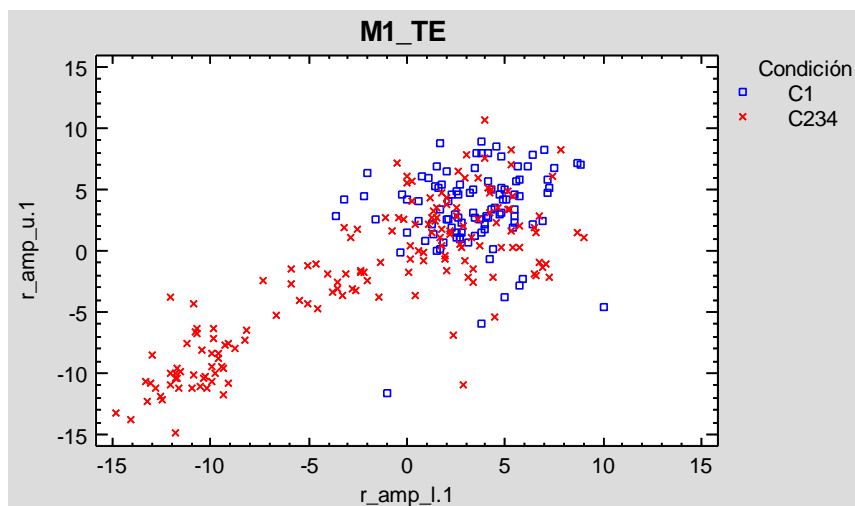


Figura B.38. Diagrama de dispersión de la relación entre $r_{amp_l.5}$ y $r_{amp_u.1}$. Tipo M1_TE. Distinción entre C1 y C234. Método Stepwise.

La Tabla B.30 resume las características de la clasificación y su eficacia. Un porcentaje de en torno al 70% no es suficiente para los objetivos del trabajo. No obstante, nótese que pocos ensayos con un motor sano se identifican erróneamente como ensayos con motor deteriorado; tan sólo el 9%.

Tabla B.30. Clasificación entre C1 y C234 para M1_TE. Método Stepwise.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	100	91 (91,00%)	9 (9,00%)
C234	162	70 (43,21%)	92 (56,79%)
Total	262	161	101

Porcentaje de casos correctamente clasificados: 69,85%

B.3.1.2. Discriminación C1-C2-C3-C4

La discriminación entre los cuatro estados también se vale de las mismas tres variables que antes. La Tabla B.31 expresa los coeficientes hallados.

Tabla B.31. Coeficientes estandarizados de las funciones discriminantes. Tipo M1_TE. Distinción entre C1, C2, C3 y C4. Método Stepwise.

	Función 1	Función 2	Función 3
r_amp_l.1	0,905728	0,885387	0,0950565
r_amp_u.1	0,207647	-1,19601	-0,223697
r_amp_u.5	-0,362505	0,12188	-0,971936

La Fig. B.39 se corresponde con el diagrama de dispersión más representativo. La mezcla de los grupos vaticina una mala clasificación.

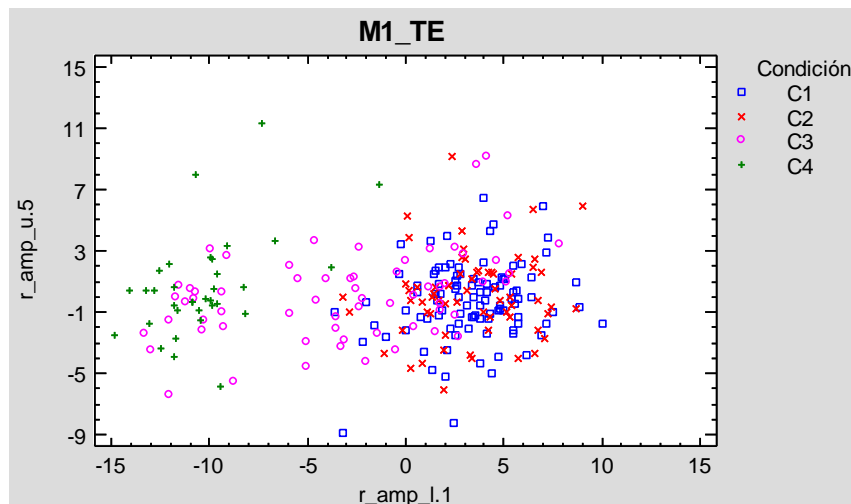


Figura B.39. Diagrama de dispersión de la relación entre r_amp_u.5 y r_amp_l.1. Tipo M1_TE. Distinción entre C1, C2, C3 y C4. Método Stepwise.

Como cabía esperar, se cumple la sospecha, tal y como muestra la Tabla B.32. La eficacia es muy reducida, apenas algo más que de la mitad de casos son satisfactoriamente clasificados. Eso sí, los motores con rotura total de la barra parecen identificarse bastante bien.

Tabla B.32. Clasificación entre C1, C2, C3 y C4 para M1_TE. Método *Stepwise*.

Grupo real	Cantidad	Grupo predicho				
		C1	C2	C3	C4	
C1	100	60 (60,00%)	33 (33,00%)	7 (0,00%)	0 (0,00%)	
C2	65	22 (33,85%)	38 (58,46%)	5 (7,69%)	0 (0,00%)	
C3	63	14 (22,22%)	7 (11,11%)	26(41,27%)	16 (25,40%)	
C4	34	0 (0,00%)	0 (0,00%)	2 (5,88%)	32 (94,12%)	
Total	262		96	78	40	48

Porcentaje de casos correctamente clasificados: 59,54%

B.3.2. Discriminación M1_LN

B.3.2.1. Discriminación C1-C234

Para la distinción entre los grupos C1 y C234 la técnica *Stepwise* selecciona las variables reflejadas en la Tabla B.33. Junto a ellas se ha registrado el valor de los coeficientes de la función discriminante.

Tabla B.33. Coeficientes estandarizados de la función discriminante. Tipo M1_LN. Distinción entre C1 y C234. Método *Stepwise*.

Función 1	
r_peak_l.1	-0,75388
r_amp_l.1	-2,43981
r_amp_u.1	2,16622
r_peak_l.5	-0,325601
r_peak_u.7	0,294391
r_amp_u.7	0,53413

Véase la Fig. B.40. No cabe duda de que la discriminación es muy eficaz. Incluso mentalmente se podría trazar una frontera que separase los dos grupos sin ninguna complicación.

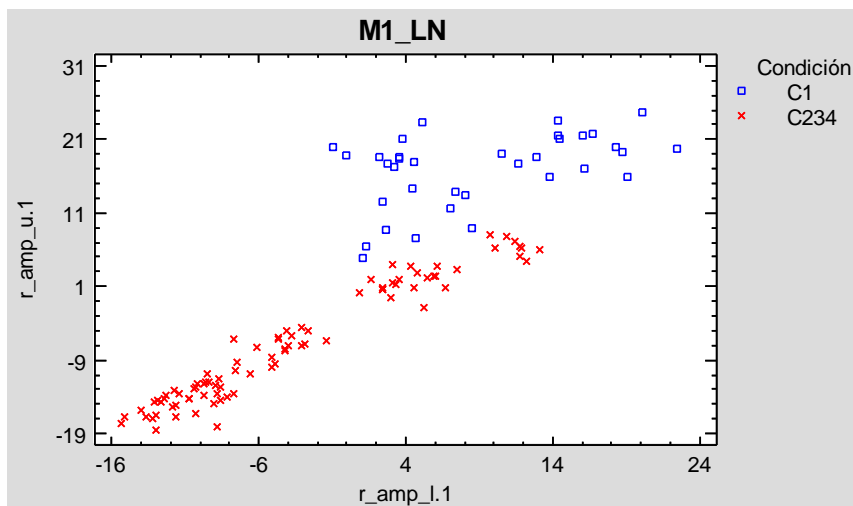


Figura B.40. Diagrama de dispersión de la relación entre r_amp_u.1 y r_amp_l.1. Tipo M1_LN. Distinción entre C1 y C234. Método Stepwise.

Según el resumen de la Tabla B.34, la totalidad de ensayos del tipo M1_LN ha sido correctamente clasificada. El resultado es el ideal.

Tabla B.34. Clasificación entre C1 y C234 para M1_LN. Método Stepwise.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	35	35 (100,00%)	0 (0,00%)
C234	86	0 (0,00%)	86 (100,00%)
Total	121	35	86

Porcentaje de casos correctamente clasificados: 100,00%

B.3.2.2. Discriminación C1-C2-C3-C4

El método de selección Stepwise recoge en el modelo 4 de las 12 variables posibles. Los coeficientes estandarizados asociados se registran en la Tabla B.35.

Tabla B.35. Coeficientes estandarizados de las funciones discriminantes. Tipo M1_LN. Distinción entre C1, C2, C3 y C4. Método Stepwise.

	Función 1	Función 2	Función 3
r_peak_l.1	-0,591898	1,07896	0,864061
r_amp_l.1	-1,17068	2,06753	0,325636
r_amp_u.1	1,37929	-0,704865	0,488096
r_amp_l.5	0,419725	-0,0259305	-0,656375

Al igual que se ha hecho hasta ahora, se representa el diagrama de dispersión más relevante, en este caso el de la Fig. B.41. Algún grupo parece identificarse mejor que otro; nos referimos a C1.

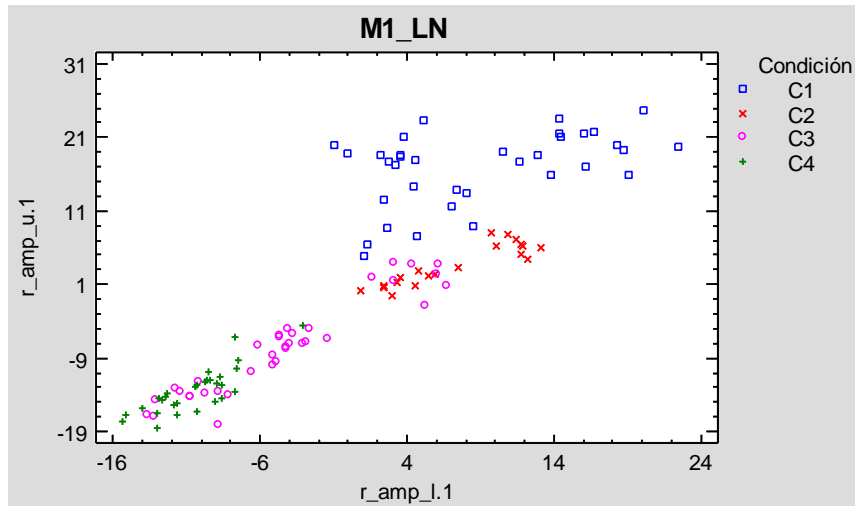


Figura B.41. Diagrama de dispersión de la relación entre $r_{amp_u.1}$ y $r_{amp_l.1}$. Tipo M1_LN. Distinción entre C1, C2, C3 y C4. Método Stepwise.

La Tabla B.36 concuerda con el pronóstico. La clasificación es aceptable, de un nivel situado alrededor del 81%, pero realmente es mayor a la hora de identificar los grupos C1, C2 y C4, mientras que se comete un error muy grande al asignar los ensayos del grupo C3.

Tabla B.36. Clasificación entre C1, C2, C3 y C4 para M1_LN. Método Stepwise.

Grupo real	Cantidad	Grupo predicho			
		C1	C2	C3	C4
C1	35	32 (91,43%)	3 (8,57%)	0 (0,00%)	0 (0,00%)
C2	21	0 (0,00%)	21 (100,00%)	0 (0,00%)	0 (0,00%)
C3	36	0 (0,00%)	7 (19,44%)	19 (52,78%)	10 (27,78%)
C4	29	0 (0,00%)	0 (0,00%)	3 (10,34%)	26 (89,66%)
Total	121	32	31	22	36

Porcentaje de casos correctamente clasificados: 80,99%

B.3.3. Discriminación M2_UT

B.3.3.1. Discriminación C1-C234

El análisis discriminante continúa con las pruebas realizadas con el motor M2 y alimentación de red. Se seleccionan hasta 6 variables de interés, con los coeficientes estandarizados de la Tabla B.37.

Tabla B.37. Coeficientes estandarizados de la función discriminante. Tipo M2_UT. Distinción entre C1 y C234. Método Stepwise.

Función 1	
r_amp_u.1	-1,30313
r_peak_l.5	0,461387
r_amp_l.5	0,764918
r_amp_u.5	0,872015
r_amp_l.7	-0,325023
r_amp_u.7	-0,340268

En la Fig. B.42 se separan razonablemente bien los dos grupos, C1 y C234.

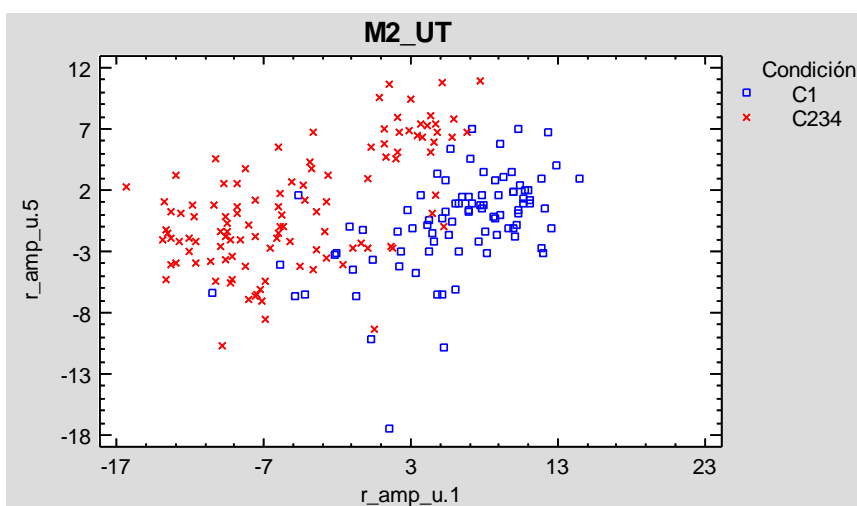


Figura B.42. Diagrama de dispersión de la relación entre r_amp_u.5 y r_amp_u.1. Tipo M2_UT. Distinción entre C1 y C234. Método Stepwise.

La potencia clasificatoria del modelo planteado es correcta. Apenas se comete un error del 7% aproximadamente, según la Tabla B.38.

Tabla B.38. Clasificación entre C1 y C234 para M2_UT. Método Stepwise.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	88	78 (88,64%)	10 (11,36%)
C234	113	4 (3,54%)	109 (96,46%)
Total	201	82	119

Porcentaje de casos correctamente clasificados: 93,03%

B.3.3.2. Discriminación C1-C2-C3-C4

Los coeficientes estandarizados de las variables seleccionadas mediante la técnica Stepwise se han anotado en la Tabla B.39.

Tabla B.39. Coeficientes estandarizados de las funciones discriminantes. Tipo M2_UT. Distinción entre C1, C2, C3 y C4. Método Stepwise.

	Función 1	Función 2	Función 3
r_amp_l.1	0,51905	-0,396512	-0,116562
r_amp_u.1	0,739415	0,523499	-0,322114
r_peak_l.5	-0,349295	-0,306543	-0,142484
r_amp_l.5	-0,314369	-0,860216	0,305791
r_amp_u.5	-0,335997	-0,853092	-0,270605
r_amp_l.7	-0,155762	0,847179	-0,383565
r_amp_u.7	0,215812	0,209843	1,12762

La Fig. B.43 muestra el diagrama de dispersión en función de r_amp_l.1 y r_amp_u.1. No queda muy claro si la clasificación será satisfactoria o no, ya que algunos grupos de observaciones se superponen con otros.

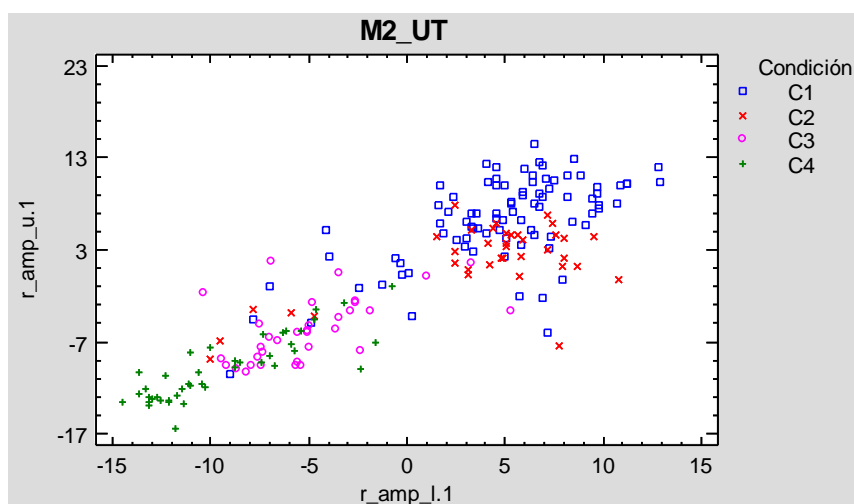


Figura B.43. Diagrama de dispersión de la relación entre r_amp_u.1 y r_amp_l.1. Tipo M2_UT. Distinción entre C1, C2, C3 y C4. Método Stepwise.

A pesar de la aparente superposición mencionada, los resultados numéricos son esperanzadores, alcanzándose un 81,59% de acierto global. Si analizamos mejor la Tabla B.40 se aprecia un mayor éxito con los grupos C1 y C2.

Tabla B.40. Clasificación entre C1, C2, C3 y C4 para M2_UT. Método Stepwise.

Grupo real	Cantidad	Grupo predicho				
		C1	C2	C3	C4	
C1	88	78 (88,64%)	5 (5,68%)	3 (3,41%)	2 (2,27%)	
C2	37	2 (5,41%)	32 (86,49%)	2 (5,41%)	1 (2,70%)	
C3	34	1 (2,94%)	1 (2,94%)	23 (67,65%)	9 (26,47%)	
C4	42	0 (0,00%)	0 (0,00%)	11 (26,19%)	31 (73,81%)	
Total	201		81	38	39	43

Porcentaje de casos correctamente clasificados: 81,59%

B.3.4. Discriminación M2_AB

B.3.4.1. Discriminación C1-C234

La discriminación entre los grupos C1 y C234 de los ensayos M2_AB requiere la selección de 5 variables, cuyos coeficientes estandarizados se indican en la Tabla B.41.

Tabla B.41. Coeficientes estandarizados de la función discriminante. Tipo M2_AB. Distinción entre C1 y C234. Método *Stepwise*.

Función 1	
r_peak_l.1	-0,342875
r_amp_u.1	1,33878
r_amp_l.5	-0,808098
r_peak_l.7	-0,339214
r_peak_u.7	0,229645

Los ensayos se representan sobre las variables con mayor poder discriminante y se obtiene el diagrama de la Fig. B.44. A grandes rasgos, el grupo C1 se sitúa más a la derecha que el C234.

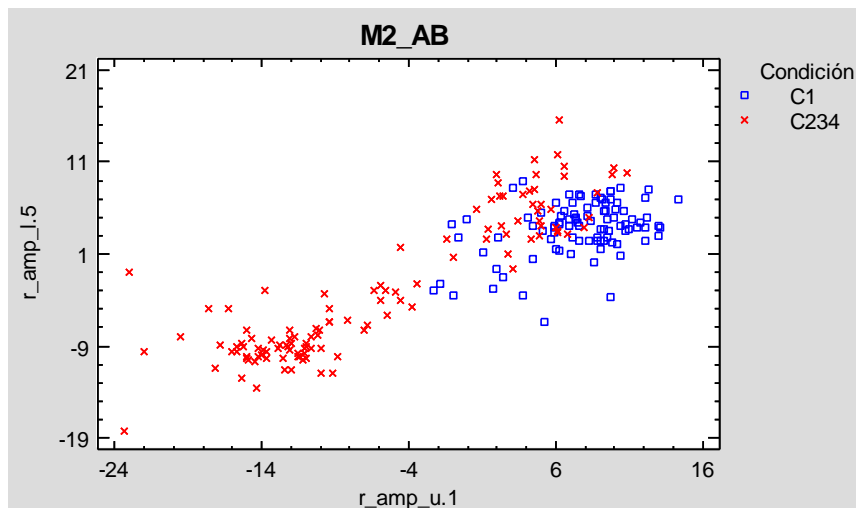


Figura B.44. Diagrama de dispersión de la relación entre $r_amp_l.5$ y $r_amp_u.1$. Tipo M2_AB. Distinción entre C1 y C234. Método *Stepwise*.

Véase la Tabla B.42 de clasificación. La predicción coincide en el 88,26% de las ocasiones con la realidad, lo cual es un porcentaje muy elevado.

Tabla B.42. Coeficientes estandarizados de la función discriminante. Tipo M2_AB. Distinción entre C1 y C234. Método *Stepwise*.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	95	88 (92,63%)	7 (7,37%)
C234	118	18 (15,25%)	100 (84,75%)
Total	213	106	107

Porcentaje de casos correctamente clasificados: 88,26%

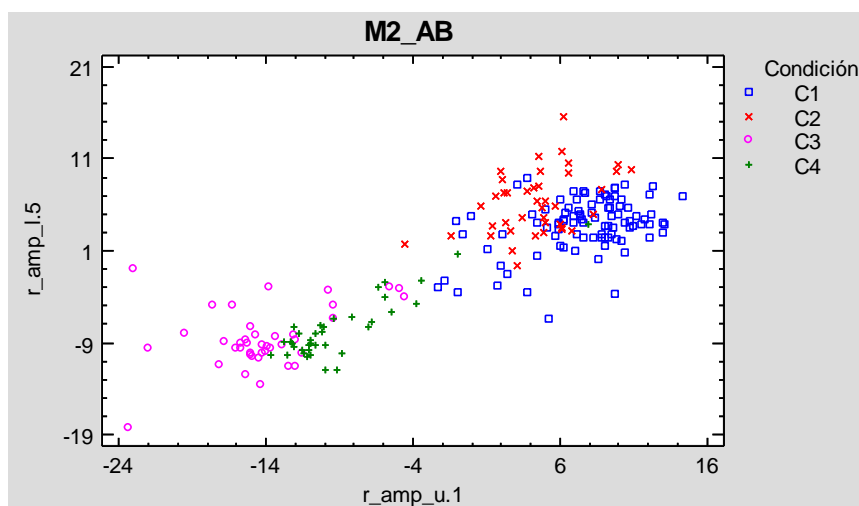
B.3.4.2. Discriminación C1-C2-C3-C4

La Tabla B.43 recoge los coeficientes asociados a las variables seleccionadas mediante el método *Stepwise*. Son necesarias la mitad de las variables iniciales para hacer la discriminación.

Tabla B.43. Coeficientes estandarizados de las funciones discriminantes. Tipo M2_AB. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

	Función 1	Función 2	Función 3
r_peak_l.1	-0,264217	-0,0595196	0,32955
r_peak_u.1	0,173258	-0,215549	0,66424
r_amp_u.1	0,785266	-0,732876	0,0727129
r_amp_l.5	0,421061	1,0701	0,470074
r_peak_l.7	-0,23024	-0,0801077	0,370397
r_amp_l.7	-0,27549	-0,815937	0,052029

Se constituye el diagrama de dispersión en función de las dos variables con mayor valor absoluto del coeficiente estandarizado de la primera función discriminante. La Fig. B.45 lo muestra.

**Figura B.45.** Diagrama de dispersión de la relación entre r_amp_l.5 y r_amp_u.1. Tipo M2_AB. Distinción entre C1, C2, C3 y C4. Método *Stepwise*.

La eficacia de las funciones discriminantes es satisfactoria. Cerca de un 88% de los casos son correctamente clasificados, tal y como resume la Tabla B.44. En comparación con la discriminación entre C1 y C234, apenas se ha perdido eficacia.

Tabla B.44. Clasificación entre C1, C2, C3 y C4 para M2_AB. Método *Stepwise*.

Grupo real	Cantidad	Grupo predicho			
		C1	C2	C3	C4
C1	95	83 (87,37%)	12 (12,63%)	0 (0,00%)	0 (0,00%)
C2	41	5 (12,20%)	36 (87,80%)	0 (0,00%)	0 (0,00%)
C3	40	0 (0,00%)	0 (0,00%)	33 (82,50%)	7 (17,50%)
C4	37	0 (0,00%)	2 (5,41%)	0 (0,00%)	35 (94,59%)
Total	213	88	50	33	42

Porcentaje de casos correctamente clasificados: 87,79%

B.3.5. Discriminación M2_SM

B.3.5.1. Discriminación C1-C234

Por último, queda por exponer el análisis al contemplar los ensayos del tipo M2_SM. Tal y como se puede ver en la Tabla B.45, son 5 las variables recogidas en el modelo para discriminar entre C1 y C234.

Tabla B.45. Coeficientes estandarizados de la función discriminante. Tipo M2_SM. Distinción entre C1 y C234. Método *Stepwise*.

	Función 1
r_peak_u.1	0,507683
r_amp_u.1	-0,611672
r_peak_l.5	0,468965
r_amp_l.5	0,519686
r_amp_u.5	0,46142

El diagrama de la Fig. B.46 es el más representativo. La separación entre las pruebas de distinto tipo es aceptable.

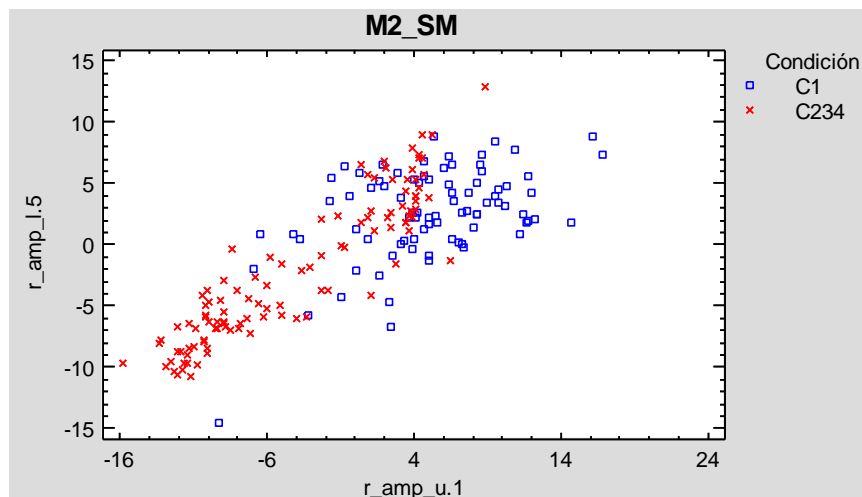


Figura B.46. Diagrama de dispersión de la relación entre $r_{amp_l.5}$ y $r_{amp_u.1}$. Tipo M2_SM. Distinción entre C1 y C234. Método Stepwise.

La discriminación llevada a cabo se resume con la Tabla B.46, donde se señala un porcentaje de casos adecuadamente predichos de casi el 93%. Se trata de un valor muy elevado.

Tabla B.46. Coeficientes estandarizados de la función discriminante. Tipo M2_SM. Distinción entre C1 y C234. Método Stepwise.

Grupo real	Cantidad	Grupo predicho	
		C1	C234
C1	84	79 (94,05%)	5 (5,95%)
C234	110	9 (8,18%)	101 (91,82%)
Total	194	88	106

Porcentaje de casos correctamente clasificados: 92,78%

B.3.5.2. Discriminación C1-C2-C3-C4

Las variables del modelo tras la selección *Stepwise* difieren respecto a la discriminación entre C1 y C234 tan sólo en una variable: en lugar de $r_{amp_u.1}$ se recoge $r_{amp_l.1}$. Los coeficientes estandarizados de las funciones discriminantes se indican en la Tabla B.47.

Tabla B.47. Coeficientes estandarizados de las funciones discriminantes. Tipo M2_SM. Distinción entre C1, C2, C3 y C4. Método Stepwise.

	Función 1	Función 2	Función 3
$r_{peak_u.1}$	0,600694	0,341567	-0,425372
$r_{amp_l.1}$	-0,690276	0,146055	-1,01941
$r_{peak_l.5}$	0,330697	0,322201	-0,0196947
$r_{amp_l.5}$	0,00225441	0,708252	0,760159
$r_{amp_u.5}$	0,347717	0,368238	0,363621

Uno de los diagramas de dispersión desprendidos del análisis es el de la Fig. B.47. Parece que la discriminación pueda ser válida, pero se debe contrastar la información cuantitativa para verificarlo.

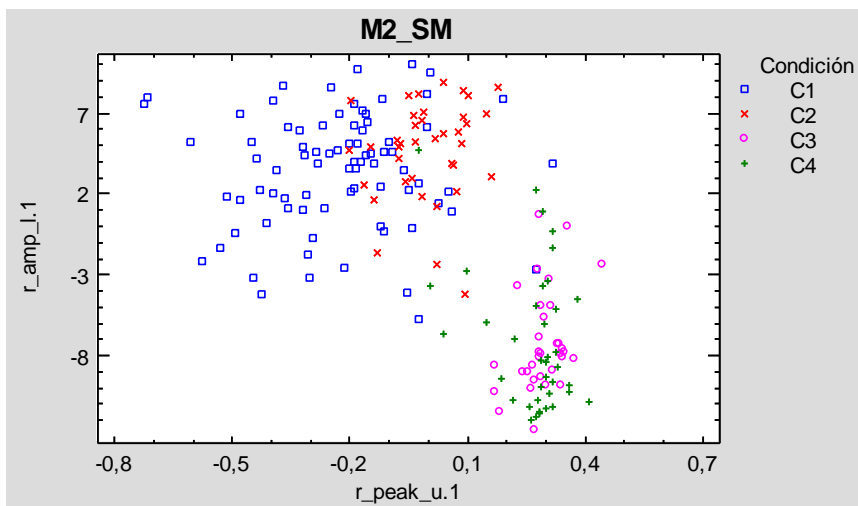


Figura B.47. Diagrama de dispersión de la relación entre r_amp_l.1 y r_peak_u.1. Tipo M2_SM. Distinción entre C1, C2, C3 y C4. Método Stepwise.

La éxito de la predicción es razonablemente bueno, de en torno al 81%. Para mayor detalle véase la Tabla B.48.

Tabla B.48. Clasificación entre C1, C2, C3 y C4 para M2_SM. Método Stepwise.

Grupo real	Cantidad	Grupo predicho			
		C1	C2	C3	C4
C1	84	72 (85,71%)	11 (13,10%)	0 (0,00%)	1 (1,19%)
C2	39	2 (5,13%)	37 (94,87%)	0 (0,00%)	0 (0,00%)
C3	33	0 (0,00%)	0 (0,00%)	24 (72,73%)	9 (27,27%)
C4	38	2 (5,26%)	2 (5,26%)	10 (26,32%)	24 (63,16%)
Total	194		76	50	34

Porcentaje de casos correctamente clasificados: 80,93%

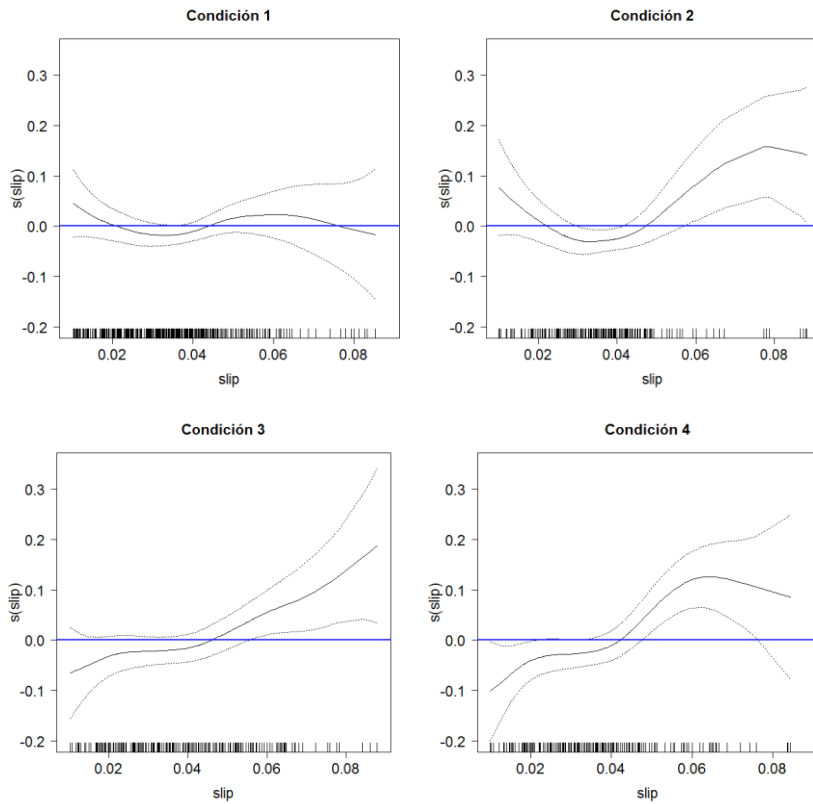
B.4. GAM complementarios según la condición

La influencia de la condición en el efecto del deslizamiento se estudió en el Capítulo 9. Este apartado recoge los diagramas que no fueron representados en dicho Capítulo.

PEAK_L.1, PEAK_U.1, PEAK_L.5 y PEAK_L.7

Estos análisis fueron explicados en el Capítulo 9, en el apartado 9.1.

PEAK_U.5



PEAK_U.7

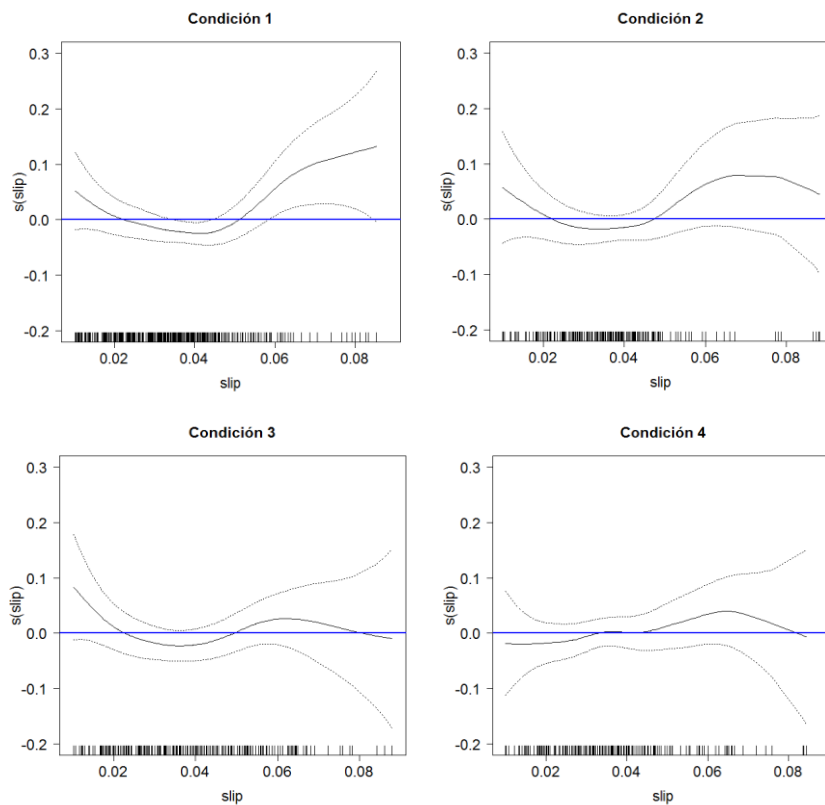
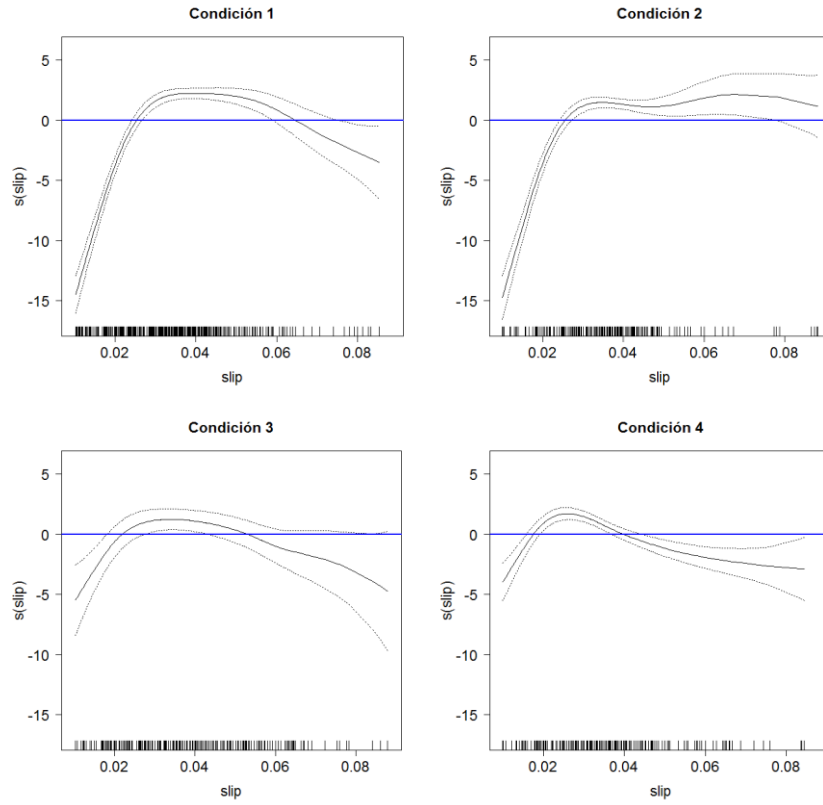


Figura B.48. Efecto del deslizamiento para GAM asociados a peak_u.5 y peak_u.7. De izquierda a derecha y de arriba a abajo, condiciones 1, 2, 3, y 4 respectivamente.

AMPLITUDE_L.1



AMPLITUDE_U.1

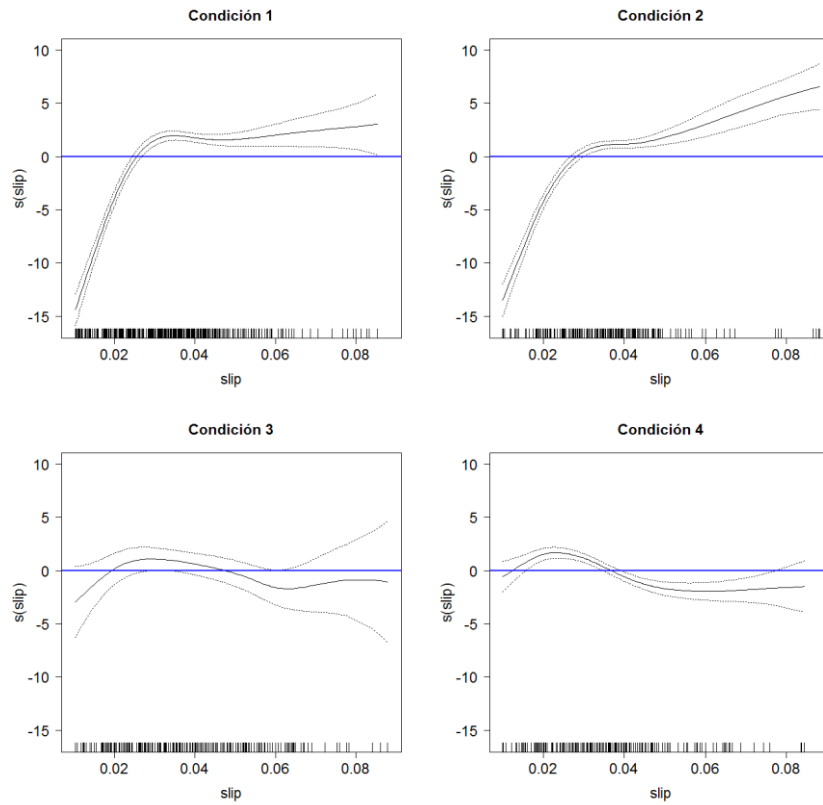
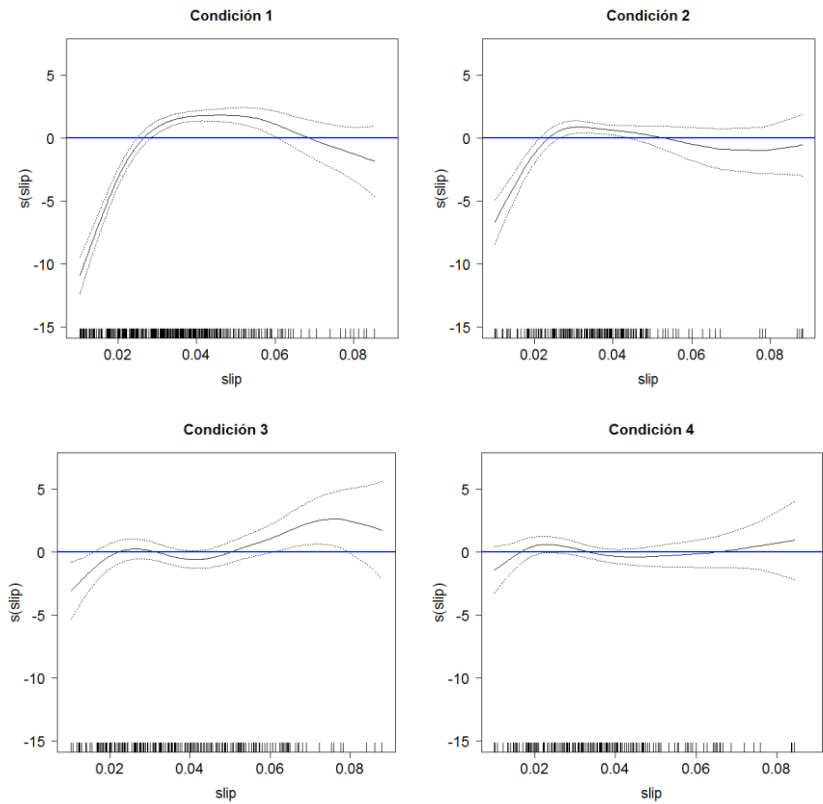


Figura B.49. Efecto del deslizamiento para GAM asociados a amplitud_l.1 y amplitud_u.1. De izquierda a derecha y de arriba a abajo, condiciones 1, 2, 3, y 4 respectivamente.

AMPLITUDE_L.5



AMPLITUDE_U.5

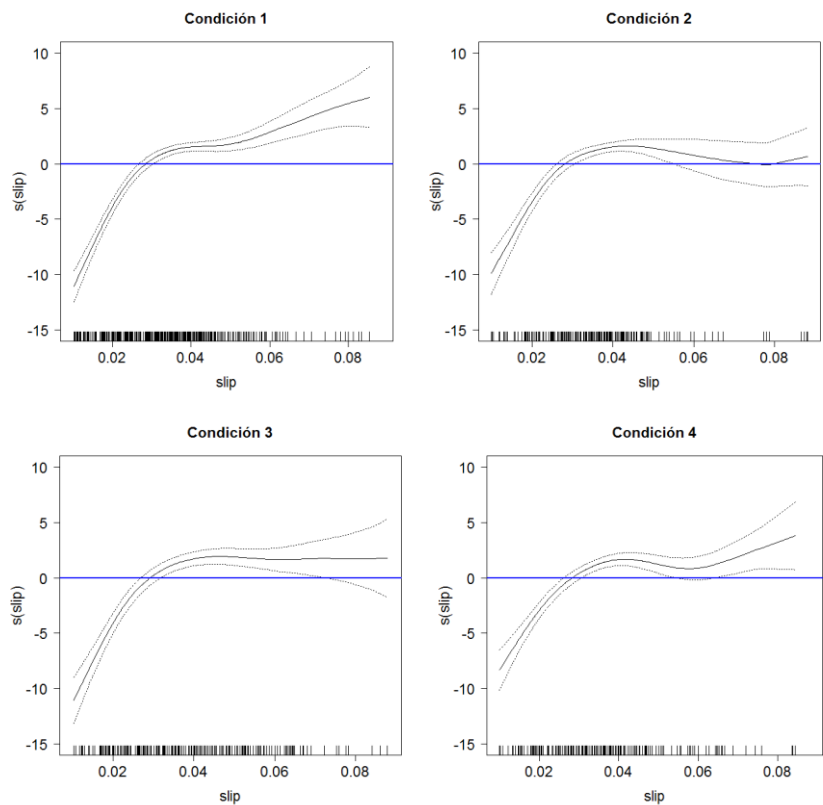
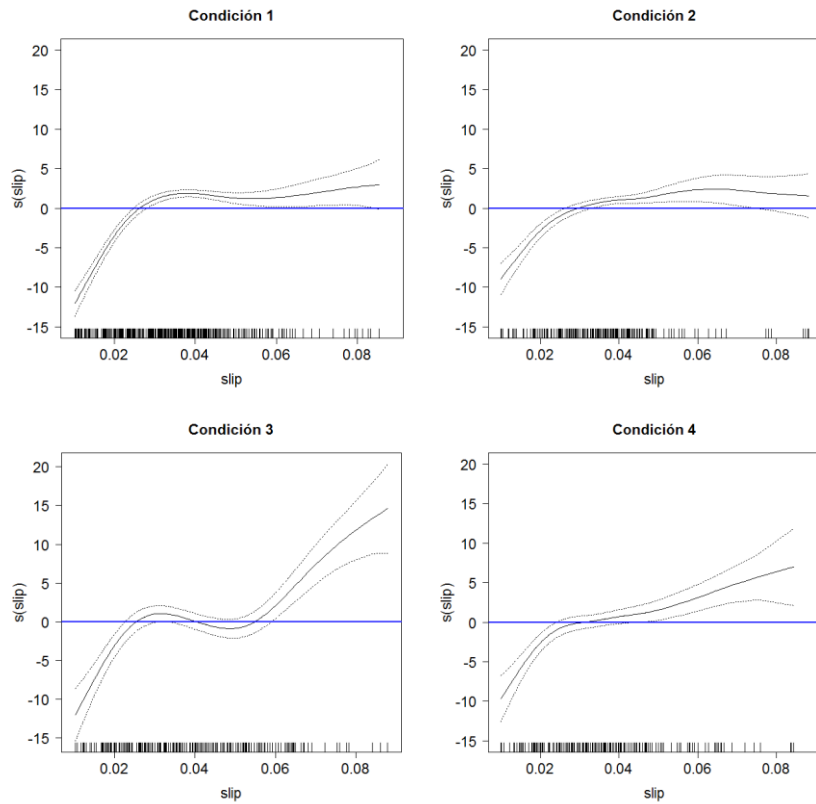


Figura B.50. Efecto del deslizamiento para GAM asociados a amplitud_l.5 y amplitud_u.5. De izquierda a derecha y de arriba a abajo, condiciones 1, 2, 3, y 4 respectivamente.

AMPLITUDE_L.7



AMPLITUDE_U.7

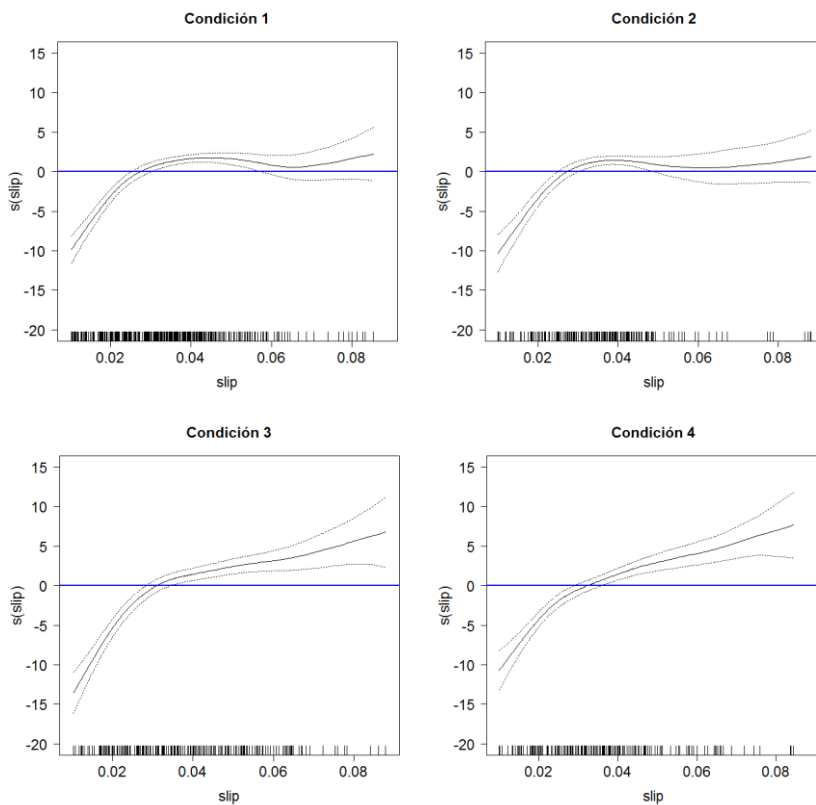


Figura B.51. Efecto del deslizamiento para GAM asociados a amplitud_l.7 y amplitud_u.7. De izquierda a derecha y de arriba a abajo, condiciones 1, 2, 3, y 4 respectivamente.

Anexo C

Código de programación en R

Esta sección complementaria recoge tan sólo parte del código de programación empleado para desarrollar los análisis estadísticos necesarios. La única información que cambia con respecto al código no presentado es el nombre de los archivos a leer, donde fueron guardados previamente los datos con las características apropiadas. Se trata de archivos de valores separados por comas de Microsoft Office Excel (archivos “.csv”).

El nombre asignado a cada archivo indica las características de los datos recogidos. Por ejemplo, el archivo “Stepwise_C1C234_M1UT.csv” contiene los datos de los ensayos realizados con el motor M1 y la alimentación de red, incluyendo únicamente las variables seleccionadas mediante el método *Stepwise* en la discriminación entre C1 y C234.

Para la correcta ejecución del programa en el estudio de Modelos Aditivos Generalizados, ha sido necesaria la instalación del paquete *gam* de R. También se ha empleado el paquete *MASS* para el cálculo del error estimado de los Análisis Discriminantes Lineales, el cual no es necesario instalar puesto que **R** lo proporciona por defecto.

C.1. Código método de validación *bootstrap*

```
##### Bootstrap1, variables originales #####
#C1C234
#M1_UT
#Stepwise

library("MASS")
X <- read.csv2("Stepwise_C1C234_M1UT.csv", header=TRUE, dec=",")
X1<-X[X[,2]!="M1_LN",]
```

```

X1<-X1[(X1[,7]>=0.01)&(X1[,7]<=0.07),]
X2<-X[X[,2]=="M1_LN",]
X2<-X2[(X2[,7]>=0.01)&(X2[,7]<=0.09),]
X<-rbind(X1,X2)

X[,1]<-as.factor(X[,1])
X[,2]<-as.factor(X[,2])
X[,3]<-as.factor(X[,3])
X[,4]<-as.factor(X[,4])
X[,5]<-as.factor(X[,5])

names(X)
#[1] "motor"          "type"          "condition"     "C1_C234"
#[5] "C1_C2_C3_C4"    "name"          "slip"          "slip.1"
#[9] "LOG.peak_1.1." "amplitude_1.1" "LOG.peak_u.5." "amplitude_1.5"
#[13] "amplitude_1.7"

n_iter<-500;
n_aciertos_1<-c(rep(0,n_iter));
error_1<-c(rep(0,n_iter));
tasa_1<-0;
boot_1<-0;
total_aciertos_1<-0;
media_aciertos_1<-0;
tam_test_1<-c(rep(0,n_iter));

for(k in 1:n_iter)
{
  set.seed(k)
  testindex<-sample(1:nrow(X),nrow(X),replace=TRUE)
  trainset<-X[testindex,c(8:ncol(X))]
  testset<-X[-testindex,c(8:ncol(X))]
  grouptrain<-X[testindex,4]
  grouptest<-X[-testindex,4]
  tam_test_1[k]<-length(grouptest)
  lda<-lda(trainset,grouptrain,prior=c(0.5,0.5))
  res<-predict(lda,testset)

  for(i in 1:length(grouptest))
  {
    if(res[[1]][[i]]==grouptest[i])
    {
      n_aciertos_1[k]=n_aciertos_1[k]+1;
    }
  }

  total_aciertos_1=total_aciertos_1+n_aciertos_1[k];
  error_1[k]=1-(n_aciertos_1[k]/length(grouptest));
}

```

```

media_aciertos_1=total_aciertos_1/n_iter;

for(k in 1:length(error_1))
{
  tasa_1=tasa_1+error_1[k];
}

tasa_1=tasa_1/length(error_1);

for(i in 1:n_iter)
{
  boot_1=boot_1+0.632*error_1[i]+0.368*tasa_1;
}

boot_1=boot_1/n_iter;

```

C.2. Código GAM

```

##### Efecto del tipo y del deslizamiento #####
X <- read.csv2("Total_Datos_R.csv",header=TRUE,dec=",")
X1<-X[X[,2]!="M1_LN",]
X1<-X1[(X1[,7]>=0.01)&(X1[,7]<=0.07),]
X2<-X[X[,2]=="M1_LN",]
X2<-X2[(X2[,7]>=0.01)&(X2[,7]<=0.09),]
X<-rbind(X1,X2)

X[,1]<-as.factor(X[,1])
X[,2]<-as.factor(X[,2])
X[,3]<-as.factor(X[,3])
X[,4]<-as.factor(X[,4])
X[,5]<-as.factor(X[,5])

names(X)
#[1] "motor" "type" "condition" "C1_C234"
#[5] "C1_C2_C3_C4" "name" "slip" "LOG.peak_1.1."
#[9] "LOG.peak_u.1." "amplitude_1.1" "amplitude_u.1" "LOG.peak_1.5."
#[13] "LOG.peak_u.5." "amplitude_1.5" "amplitude_u.5" "LOG.peak_1.7."
#[17] "LOG.peak_u.7." "amplitude_1.7" "amplitude_u.7"

library(gam)

###peak_1.1
gam.object_1 <- gam(LOG.peak_1.1~type+s(slip),data=X)
summary(gam.object_1)
par(mfrow=c(1,2),ps=12)
plot(gam.object_1,se=TRUE,las=1,xlim=c(0.01,0.0882),ylim=c(-0.2,0.3),
      cex.axis=1.4,cex.lab=1.4)

```

Bibliografía

- Aluja, T.; Morineau, A. (1999). *Aprender de los Datos: el Análisis de Componentes Principales*. EUB.
- Bowman, A. W.; Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations*. Nueva York: Oxford University Press Inc.
- Duque, Ó.; Pérez, M (2005). *Motores de inducción: técnicas de mantenimiento predictivo*. Badajoz: Abecedario.
- Duque, Ó.; García, L. A.; Moríñigo, D.; Gardel, P. E.; Pérez, M. (2015). "Analysis of fault signatures for the diagnosis of induction motors fed by voltage source inverters using ANOVA and additive models". *Electric Power Systems Research*, n. 121, pp. 1-13.
- Efron, B.; Tibshirani, R. J. (1995). *An Introduction to the Bootstrap*. Boca Ratón, Londres, Nueva York, Washington D.C.: Chapman & Hall/CRC.
- Faraday, J. J. (2015). *Linear Models with R* (2ª ed.). CRC Press, Taylor & Francis Group.
- Fernández, M.; García, M.; Alonso, G.; Cano, J. M.; Solares, J. (2000). *Técnicas para el mantenimiento y diagnóstico de máquinas eléctricas rotativas*. ABB Service, Marcombo S.A.
- García, L. Á.; Duque, Ó.; Morinigo, Daniel; Pérez, M. (2011). "Robust condition monitoring for early detection of broken rotor bars in induction motors". *Expert Systems with Applications*, n. 38, pp. 2653-2660.
- Gardel, P. E. (2013). *Aportaciones al mantenimiento predictivo de motores de inducción mediante una metodología de diagnóstico basada en el uso combinado de técnicas estadísticas y Redes Neuronales Artificiales*. [Tesis Doctoral]. Universidad de Valladolid.
- Gordo, D. (2014). *Diagnóstico de fallos en motores con técnicas de Machine Learning*. [Trabajo de Fin de Grado]. Universidad de Valladolid.
- Gutiérrez, R.; González, A.; Torres, F.; Gallardo, J. A. (1994). *Técnicas de análisis de datos multivariable. Tratamiento computacional*. McGraw-Hill.

- Hastie, T. J.; Tibshirani, R. J. (1990). *Generalized additive model with R*. Chapman & Hall/CRC.
- Hawkins, D. M. (2004). "The Problem of Overfitting". *Journal of Chemical Information and Computer Sciences*, n. 44, pp. 1-12.
- Montgomery, D. C. (2001). *Design and Analysis of Experiments* (5ª ed.). Nueva York, Chichester, Weinheim, Brisbane, Toronto, Singapur: John Wiley & Sons, Inc.
- O'Connor, P. D. T.; Kleyner, A. (2012). *Practical Reliability Engineering* (5ª ed.). John Wiley & Sons, Ltd.
- Peña, D. (2002). *Análisis de datos multivariantes*. McGraw-Hill/ Interamericana de España.
- Rosa, R. (2015). *Estudio sobre la viabilidad de los estadísticos de orden superior de la corriente de alimentación como indicadores para determinar el estado de un motor de inducción*. [Trabajo de Fin de Grado]. Universidad de Valladolid.
- Soto, N.; Torre, F. (2007). "Diagnóstico de problemas de asimetrías rotóricas en un motor de inducción de gran potencia". *Ingeniería Mecánica*, n. 2, pp. 47-50.
- Verucchi, C. J.; Acosta, G. G. (2007). "Fault Detection and Diagnosis Techniques in Induction Electrical Machines". *IEEE Latin America Transactions*, vol. 5, n. 1, pp. 41-49.
- Villada, F.; Parra, D.; Ocampo, G. (2007). "Estudio del comportamiento de motores de inducción ante fallas estatóricas". *Revista Facultad de Ingeniería*, n. 40, pp. 76-94.
- Webb, A. R.; Copsey, K. D. (2011). *Statistical Pattern Recognition* (3ª ed.). John Wiley & Sons, Ltd.
- Wood, S. N. (2006). *Generalized Additive Models: an introduction with R*. Boca Ratón: Chapman & Hall/ CRC, Taylor & Francis Group.