

UNIVERSIDAD NACIONAL DEL COMAHUE

Maestría en Estadística Aplicada



**Estrategias de Análisis Multivariado para
la Revisión Técnica de Campañas de
Perforación de Pozos de Petróleo y Gas**

Trabajo de Tesis presentado por:

Alfredo Bernardis

Ingeniero Industrial

Universidad Nacional del Comahue

1973

Neuquén, diciembre 2014

Dedicado a
J. Harrison

Agradecimientos

Al concluir esta obra evoco a mis padres. El tesón (igualdad o vicio?) heredado, que el transcurso del tiempo ha tornado paulatinamente en tozudez, es el principio matriz que me ha conducido. Así que debo agradecer a la gens por dotarme de tenacidad, sucedáneo suficiente para suplir la exigüidad de perspicacia.

A nuestra alma mater, madre nutricia, cuyos ministros aportaron sapiencia y guía, no exentas de rigurosidad, con generosidad y tolerancia en justas proporciones.

A mis mentores: Pablo Reeb y Guillermo Sabino, pacientes y eruditos consejeros y a Sergio Bramardi, preclaro director.

A los maestros Silvia Boché, Nora Baccalá, Nora Moscoloni, Pilar Díaz, sin olvidar a otros tantos pródigos representantes del cuerpo académico.

A Mary, amada compañera y soporte, a quien le fue endosado, sin beneficio en contrapartida, un largo plan de cuotas que ella fue honrando puntualmente. Sus pagos, a veces no periódicos y a veces resignados, fueron realizados en diversas monedas: Tolerancia, comprensión, sustento y apoyo.

A mis hijos, queridos estímulos y a mi familia, a mis condiscípulos, a los amigos de toda clase, allegados y próximos, que asistieron con los materiales más vigorizantes: Aliento y ánimo.

El Kilimanjaro es una montaña cubierta de nieve, de 19.710 pies de altura, y dicen que es la más alta de África. Su nombre es, en massai, “Ngàje Ngài”, “la Casa de Dios”. Cerca de la cima se encuentra el esqueleto seco y helado de un leopardo, y nadie ha podido explicarse nunca qué estaba buscando el leopardo por aquellas alturas.

Ernest Hemingway (21/07/1899, Oak Park – 2/07/1961, Ketchum)
Las nieves del Kilimanjaro – Biblioteca Universal Caralt – Barcelona - 1977

¡Loor al gusano loco!

Wimpi (Arthur García Núñez) (12/8/1906, Salto - 9/9/1956, Buenos Aires)
El Gusano Loco – Editorial Freeman – Bs.As. - 1978

RIP KIRBY



John Prentice (17/10/1920, Whitney – 23/05/1999, Connecticut)
Diario Rio Negro – 05/1970

Índice General

PARTE 1 - INTRODUCCIÓN	1
Capítulo 1.1 Introducción.....	1
1.1.1 Justificación del tema propuesto.....	1
1.1.2 Antecedentes y fundamentación.....	2
1.1.3 Formulación del problema.....	2
1.1.4 Estructuración de este estudio.....	6
1.1.5 El Ingeniero Analista de Perforación.....	6
1.1.6 Objetivos de este trabajo	6
Capítulo 1.2 Perforación y antecedentes	8
1.2.1 Introducción.....	8
1.2.2 Equipo de perforación	8
1.2.3 Agentes que afectan la tasa de penetración.....	11
1.2.4 Ecuación de la tasa de penetración	15
1.2.5 Función del Ingeniero de Perforación	16
1.2.6 Revisión de Antecedentes	17
1.2.7 Modelado de la tasa de penetración.....	23
1.2.8 Colofón	24
PARTE 2 - MATERIALES Y METODOS	25
Capítulo 2.1 Materiales.....	25
2.1.1 Introducción.....	25
2.1.2 Antecedentes relacionados a datos.....	26
2.1.3 Pozo tipo en evaluación	27
2.1.4 Origen de los datos	31
2.1.5 Unidades observacionales y variables relacionadas a la perforación	32
2.1.6 Colofón	37
Capítulo 2.2 Métodos.....	39
2.2.1 Introducción.....	39
2.2.2 Curva de aprendizaje	39
2.2.3 Exploración de la tasa de construcción	40
2.2.4 Exploración de la tasa de penetración.....	41
2.2.5 Agrupamiento de pozos	42

2.2.6	Modelos Lineales Generalizados	46
2.2.7	Regresión por Redes Neuronales Artificiales	50
2.2.8	Regresión por Particiones Recursivas	51
PARTE 3 - RESULTADOS Y DISCUSIÓN		52
Capítulo 3.1 Exploración de datos		52
3.1.1	Introducción	52
3.1.2	Tasa de construcción	53
3.1.3	Tasa de penetración	59
3.1.4	Análisis de componentes principales	63
Capítulo 3.2 Agrupamientos		68
3.2.1	Distancias	68
3.2.2	Métodos particionales	68
3.2.3	Agrupamientos jerárquicos	74
3.2.4	Diagnósticos	81
3.2.5	Conclusión	83
Capítulo 3.3 Modelos de tasa de penetración		85
3.3.1	Modelos Lineales Generalizados	85
3.3.2	Regresión por Redes Neuronales Artificiales	98
3.3.3	Regresión por Particiones Recursivas	99
PARTE 4 - CONCLUSIONES		103
Capítulo 4.1 Resumen y Conclusiones		103
4.1.1	Desarrollo y perforación	103
4.1.2	Exploración de los Datos	103
4.1.3	Agrupamientos	104
4.1.4	Modelado	104
4.1.5	Conclusiones	105
Anexo A - Agrupamientos		108
Anexo B - Red Neuronal Artificial		109
Anexo C - Partición Recursiva		111
Bibliografía		112

Índice de Figuras

FIG. 1.1-1 DESARROLLO DE UN ÁREA	3
FIG. 1.2-1 PROCESO DE PERFORACIÓN	9
FIG. 1.2-2 TRÉPANO TRICONO	11
FIG. 1.2-3 EFECTO DEL SOBREALANCE	12
FIG. 1.2-4 TASA DE PENETRACIÓN. A) CON PESO W. B) CON ROTACIÓN N	13
FIG. 1.2-5 FLUJO POR ORIFICIOS.....	14
FIG. 1.2-6 EFECTO DE LA HIDRÁULICA SOBRE LA TASA DE PENETRACIÓN	15
FIG. 1.2-7 CURVA DE APRENDIZAJE	18
FIG. 1.2-8 TIEMPO TOTAL VS. PROFUNDIDAD.....	19
FIG. 1.2-9 APRENDIZAJE A) PROYECTO NUEVO B) CAMBIA ACTIVIDAD C) DESMEJORA.....	19
FIG. 1.2-10 TERMINOLOGÍA DE LOS TIEMPOS DE PERFORACIÓN	21
FIG. 1.2-11 CICLOS DEL CONOCIMIENTO.....	21
FIG. 1.2-12 PROBABILIDAD DE PROBLEMAS.....	22
FIG. 1.2-13 PROCESO DE BENCHMARKING.....	23
FIG. 2.1-1 ACTIVIDADES DEL ANÁLISIS DE DATOS	26
FIG. 2.1-2 ESQUEMA DE POZO	28
FIG. 2.1-3 EJEMPLO DE ACTIVIDADES DE CONSTRUCCIÓN DE POZOS	30
FIG. 3.1-1 CURVA DE APRENDIZAJE - TODOS LOS POZOS.....	52
FIG. 3.1-2 ROC36-TASA DE CONSTRUCCIÓN DE POZO	53
FIG. 3.1-3 ROC36-TASA DE CONSTRUCCIÓN FACTORIZADA.....	54
FIG. 3.1-4 ROC36-DISPERSIONES CRUZADAS	55
FIG. 3.1-5 ROC36-GRÁFICO DE DISEÑO	56
FIG. 3.1-6 DIAGNÓSTICOS DE SUPUESTOS-ANOVA DE ROC36.....	57
FIG. 3.1-7 ROC36-COMPARACIÓN DE DIFERENCIAS DE MEDIAS POR RIG	58
FIG. 3.1-8 ROC36-NORMALIDAD POR TIPO DE POZO.....	59
FIG. 3.1-9 ROC36-HISTOGRAMAS DE ROC36 POR TIPO DE POZO	59
FIG. 3.1-10 SELECCIÓN DEL TIPO DE DISTRIBUCIÓN DE ROP600	60
FIG. 3.1-11 COMPARACIÓN ROP600-GAMMA	61
FIG. 3.1-12 GRÁFICO ENREJADO DE ROP600 POR RIG Y TIPO	62
FIG. 3.1-13 DISPERSIONES CRUZADAS: ROP600 - PROPIEDADES DEL LODO.....	64
FIG. 3.1-14 DISPERSIONES CRUZADAS: ROP600 - CARACTERÍSTICAS DE PERFORACIÓN	65
FIG. 3.1-15 COMPONENTES PRINCIPALES DE LAS VARIABLES.....	66
FIG. 3.2-1 TABLA DE DISIMILARIDADES	68
FIG. 3.2-2 SELECCIÓN DEL NÚMERO DE GRUPOS.....	69

FIG. 3.2-3 AGRUPAMIENTO CON 7 MEDOIDES	69
FIG. 3.2-4 CUATRO POZOS DEL GRUPO 2	70
FIG. 3.2-5 SIETE GRUPOS A 3 COORDENADAS.....	70
FIG. 3.2-6 COMPARACIÓN DEL GRUPO 2 Y EL RESTO DE GRUPOS.....	71
FIG. 3.2-7 COMPARACIÓN DEL GRUPO 2 Y EL RESTO DE GRUPOS.....	71
FIG. 3.2-8 DISIMILARIDADES ORDENADAS	72
FIG. 3.2-9 ANCHOS DE SILUETA PARA 7 GRUPOS	73
FIG. 3.2-10 MÉTODO COMPLETE.....	75
FIG. 3.2-11 MÉTODO AVERAGE	75
FIG. 3.2-12 MÉTODO SINGLE.....	76
FIG. 3.2-13 MÉTODO WEIGHTED.....	76
FIG. 3.2-14 MÉTODO FLEXIBLE	77
FIG. 3.2-15 MÉTODO DIANA.....	77
FIG. 3.2-16 AGRUPAMIENTO JERÁRQUICO - MÉTODO DIVISIVO	79
FIG. 3.2-17 AGRUPAMIENTO JERÁRQUICO - MÉTODO AVERAGE	80
FIG. 3.2-18 BOXPLOT DE LOG(ROP600) SEGÚN 8 GRUPOS	81
FIG. 3.2-19 DISTRIBUCIÓN DE RESIDUALES	82
FIG. 3.2-20 INTERVALOS DE CONFIANZA DE LAS DIFERENCIAS DE MEDIAS.....	83
FIG. 3.3-1 INTERACCIÓN DE TIPO Y RIG CON HRSF600	85
FIG. 3.3-2 GRAFICO DE DIAGNÓSTICO - MODELO COMPLETO.....	87
FIG. 3.3-3 Q-Q PLOT MEDIO NORMAL - MODELO COMPLETO.....	87
FIG. 3.3-4 INFLUENCIA, APALANCAMIENTO Y DIST. COOK-MODELO COMPLETO	89
FIG. 3.3-5 GRAFICO DE DIAGNÓSTICO - MODELO REDUCIDO.....	90
FIG. 3.3-6 INFLUENCIA, APALANCAMIENTO Y DIST. COOK-MODELO REDUCIDO	90
FIG. 3.3-7 DIAGNÓSTICO DE RESIDUALES DE RESPUESTA.....	91
FIG. 3.3-8 DIAGNÓSTICO DE RESIDUALES DE DEVIANZA.....	92
FIG. 3.3-9 AJUSTES DEL SUBSET - MODELO REDUCIDO.....	93
FIG. 3.3-10 AJUSTES DE CONTROL-MODELO REDUCIDO.....	94
FIG. 3.3-11 ÁRBOL DE REGRESIÓN	99
FIG. 3.3-12 PARÁMETRO DE COMPLEJIDAD DE LA PARTICIÓN	100
FIG. 3.3-13 ÁRBOL PODADO	101

Índice de Tablas

TABLA 1.2-1 VARIABLES DE PERFORACIÓN	17
TABLA 2.1-1 ACTIVIDADES DE POZO	29
TABLA 2.1-2 COMPARACIÓN DE TIEMPOS FASE 300, FASE 600 Y TOTAL	30
TABLA 2.1-3 FUENTES DE DATOS	31
TABLA 2.1-4 PERFORACIÓN: VARIABLES BÁSICAS	35
TABLA 2.1-5 VARIABLES DE LA ACTIVIDAD PERFORACIÓN	38
TABLA 3.2-1 INFORMACIÓN DE LOS GRUPOS	73
TABLA 3.2-2 RELACIÓN DE POZOS MEDOIDES CON RIG Y TIPO	74
TABLA 3.2-3 COEFICIENTES DE CORRELACIÓN COFENÉTICA.....	78
TABLA 3.2-4 GRUPOS FINALES	81
TABLA 3.3-1 COEFICIENTE DE COMPLEJIDAD	100

Resumen

En este trabajo de tesis se aplicaron métodos estadísticos multivariados en el proceso de revisión de campañas de perforación de pozos de petróleo o gas. Estas revisiones están orientadas a detectar las condiciones con que se logran perforar los pozos en menos tiempo y para detectar buenas y malas prácticas, mejores y peores técnicas, productos o modalidades de trabajo.

Como todavía no existe un procedimiento estándar para el análisis de datos de perforación, este estudio propone métodos estadísticos. Es un aporte al crecimiento del campo independiente del conocimiento llamado “Análisis de Perforación” y al desarrollo del rol de “Analista de Perforación”.

Inicialmente se introducen las técnicas de perforación y los procedimientos tradicionalmente usados por la industria en el análisis de la actividad. Luego se estudiaron las campañas de perforación de tres años y medio mediante técnicas de análisis exploratorio de datos, agrupamiento de datos y análisis de modelos lineales generalizados.

Los objetivos propuestos se han logrado satisfactoriamente. En efecto:

Se aplicaron técnicas de análisis de datos multivariados a la revisión de campañas de perforación. Se caracterizaron y clasificaron las perforaciones según su eficiencia y variables relacionadas al proceso.

Se aplicaron técnicas multivariadas a la detección y evaluación de las variables determinantes en los tiempos de perforación.

Se evaluaron modelos estadísticos para analizar el efecto de las variables consideradas y para la predicción de tiempos de perforación.

Se desarrollaron e implementaron rutinas en lenguaje R para cada etapa de la revisión de campañas de perforación. Esta premisa se cumplió a lo largo de todos los temas y capítulos de aplicación.

Con estos procedimientos se amplió el alcance del análisis ya que se introdujo una modalidad inédita en este ámbito.

Abstract

In this thesis, multivariate statistical methods were applied to the process of revision of campaigns of oil or gas drilling. These peer reviews are oriented to detect the conditions whereupon the wells are drilled in less time and to detect good and bad practices, better and worse techniques or products, modalities of work, etc.

As a standard procedure for the data analysis of oil drilling still does not exist, this study proposes statistical methods. This is a contribution to the growth of the independent knowledge field of the called “Drilling Analysis” and to the development of the role of “Drilling Analyst”.

Initially, techniques and procedures traditionally used by the industry in drilling analysis are introduced. Then, the drilling campaigns of three year and a half were studied by means of techniques of exploratory data analysis, data clustering and analysis of generalized linear models.

The proposed targets have been achieved successfully. In fact:

Multivariate data analysis techniques were applied to the review of drilling campaigns.

Wells were described and classified according to drilling efficiency and process variables related to drilling.

Multivariate techniques were applied to the detection and assessment of key variables in drilling time.

Statistical models were evaluated to analyze the effect of the variables considered and for the prediction of drilling time.

Routines (scripts) in R language have been developed and applied at each stage of the review of drilling campaigns. This premise was fulfilled over all subjects and chapters of application.

These procedures extended the reach of the analysis since an original modality in this scope was introduced.

PARTE 1 - INTRODUCCIÓN

Capítulo 1.1 Introducción

1.1.1 Justificación del tema propuesto

En este trabajo de tesis se evaluó la aplicación de métodos estadísticos multivariados a revisiones de campañas de perforación de pozos de petróleo o gas. Las aplicaciones incluyen la caracterización y clasificación de las perforaciones según su eficiencia y variables relacionadas al proceso de perforación de pozos, la aplicación de técnicas multivariadas en la detección y evaluación de las variables determinantes en los tiempos de perforación y la evaluación de modelos estadísticos para la predicción de tiempos de perforación. Estas aplicaciones se llevaron a cabo mediante el desarrollo e implementación de rutinas en distintos paquetes del sistema de soft estadístico en lenguaje R para cada etapa de la revisión de campañas de perforación.

El análisis multivariante se refiere a métodos estadísticos que analizan simultáneamente medidas múltiples de cada individuo u objeto sometido a investigación. Puede significar simplemente el examen de las relaciones entre más de dos variables. Todas las variables deben estar interrelacionadas de tal forma que sus diferentes efectos no puedan ser interpretados separadamente (Hair, y otros, 1999),(Uriel Jiménez, y otros, 2005).

En otra definición encontramos que el análisis de datos multivariantes comprende el estudio estadístico de varias variables medidas en elementos de una población con los siguientes objetivos: (1) Resumir los datos en un pequeño conjunto de nuevas variables, (2) Encontrar grupos de datos, si existen, (3) Clasificar nuevas observaciones en grupos definidos y (4) Relacionar conjuntos de variables(Peña, 2002).

Las revisiones de campañas de perforación (también conocidas como Peer Reviews en la jerga de la industria del petróleo) son práctica usual en la industria y están orientadas a detectar, principalmente, las condiciones con que se logran perforar los pozos en menos tiempo. También sirven para detectar buenas y malas prácticas, mejores y peores técnicas o productos, modalidades de trabajo. Además, se emplean en la comparación de diversas campañas para explicar las causas de resultados dispares.

Habitualmente, las revisiones se realizan en base a análisis estadístico univariado o bivariado. En estadesis, se amplió el alcance de la investigación mediante la aplicación de métodos de análisis multivariado.

1.1.2 Antecedentes y fundamentación

Lagreca y otros (Lagreca, y otros, 2004) reconocen que muchos aspectos del análisis de datos de perforación todavía no están claros y siguen bajo debate, que hay diversos conceptos y métodos de análisis propuestos y que todavía hay ámbito para el desarrollo de métodos o procedimientos estructurados. También se explica que dentro de la industria y específicamente, en la ingeniería de perforación, está creciendo un campo independiente del conocimiento llamado “Análisis de Perforación”. Este crecimiento conlleva al desarrollo del rol de “Analista de Perforación”. Justamente, el presente estudio se alineó con estos conceptos, tratando de generar un aporte novedoso al examen de campañas de perforación de pozos.

Tradicionalmente, en la industria se aplica la técnica de “curva de aprendizaje” en la cual se relacionan tiempos de perforación y profundidad de pozo mediante regresiones exponenciales. Esta técnica es particularmente reveladora en las campañas iniciales del desarrollo, pero con la limitación de que no se evidencia cuáles son las variables que modifican el rendimiento de la perforación.

También es frecuente que las revisiones se realicen en base a un análisis estadístico univariado o bivariado. Básicamente, se realizan histogramas, gráficos o regresiones para comparar las variables y tratar de “ver” el efecto de cada una de ellas. Sin embargo, es reconocido que en la práctica de la perforación participan un gran número de variables, las cuales pueden interactuar entre sí y afectar de muy diversas maneras al proceso conjunto. Esta multiplicidad de variables hace que la problemática de la perforación sea suficientemente compleja como para que estos análisis resulten insuficientes y dejen margen para otro tipo de análisis, como el análisis multivariado.

1.1.3 Formulación del problema.

En la Fig. 1.1-1 se muestra un esquema de la historia de un área o yacimiento de hidrocarburos donde se aprecian las etapas de Exploración y Desarrollo.

La exploración se inicia con **estudios** entre los cuales se encuentran los métodos geofísicos. Estos estudios permiten ubicar zonas donde pudieran encontrarse hidrocarburos.

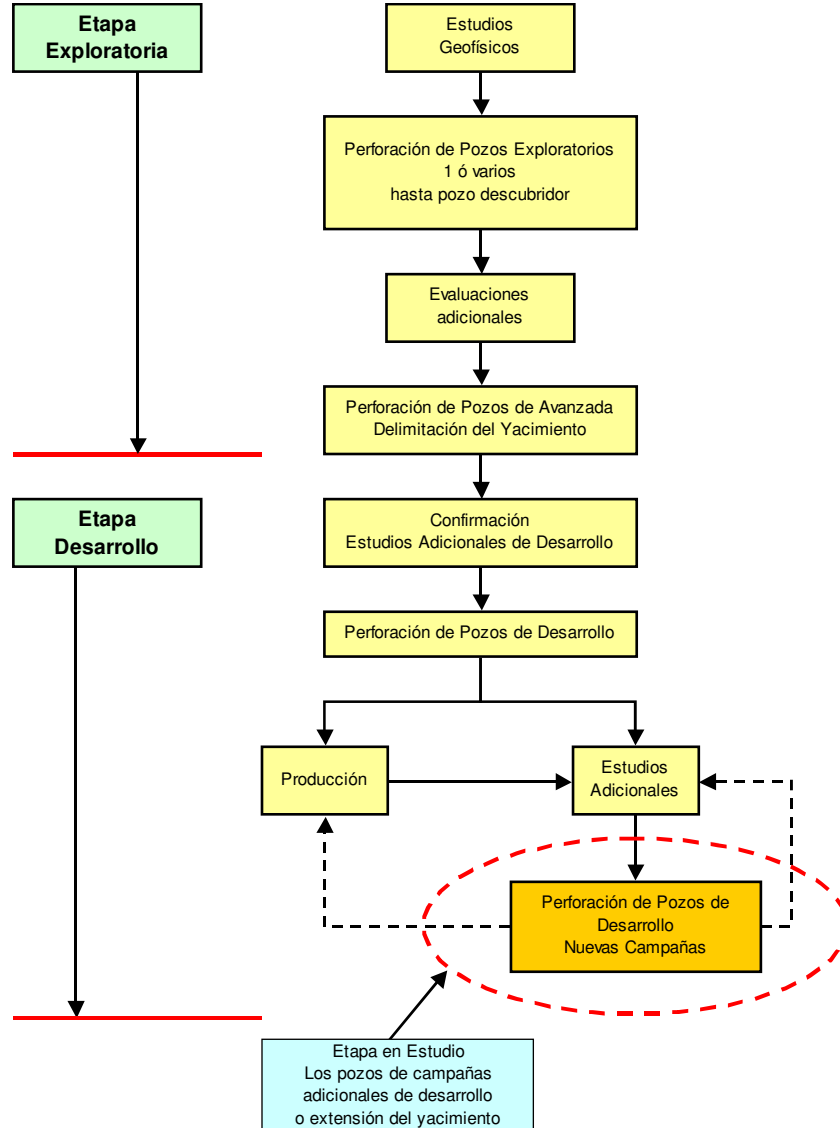


Fig. 1.1-1 Desarrollo de un área

Actividades en las etapas exploratoria y de desarrollo de un área petrolera.
Se resalta dónde se ubican las campañas de perforación objeto de este estudio.
Estas campañas pueden ser sucesivas en distintos momentos del desarrollo.

Seleccionada la zona más favorable, se perfora el primer pozo exploratorio. A veces es necesario perforar varios pozos infructuosos hasta confirmar la existencia de hidrocarburos. Luego del descubrimiento se perforan algunos pozos de avanzada cuya finalidad es delimitar grosso modo la extensión del yacimiento y certificar las reservas.

Posteriormente, se desarrolla el yacimiento con sucesivas campañas de perforación. Luego de cada campaña se realiza una revisión donde se trata de extraer el mayor conocimiento posible de la experiencia previa. El presente estudio se relaciona justamente con esas tareas de revisión.

Las primeras perforaciones se realizan con relativamente escasa información, generalmente en la zona “más segura” del área; es decir, en la zona donde los estudios geológicos y geofísicos previos indiquen la presencia de alguna trampa, existencia de roca madre u otros antecedentes de modo que la posibilidad de encontrar hidrocarburos es mayor. Se perforan los primeros pozos de desarrollo y se ponen en producción. A medida que se perforan pozos y el yacimiento produce hidrocarburos, se obtiene información adicional con la cual se amplía el conocimiento del área. El mayor conocimiento permite una recategorización de las zonas, de manera que zonas cuyo contenido de hidrocarburos era dudoso en algún momento, pueden convertirse en zonas con mayor posibilidad de contener petróleo o gas. La perforación puede avanzar hacia los límites del yacimiento, ampliando los horizontes, tal vez a otras profundidades. Esta es la principal razón por la cual el progreso del desarrollo se realiza en etapas sucesivas, alternando la actividad con la consolidación de conocimiento.

A medida que aumenta el número de pozos perforados, se incrementa la experiencia y el conocimiento de la zona. Las técnicas y condiciones operativas de la perforación de pozos nuevos van cambiando, buscando la eficiencia, es decir el pozo más rápido. El mayor conocimiento de la zona permite seleccionar las técnicas y los procedimientos, productos, herramientas y materiales más adecuadas o eficientes para continuar con la perforación de nuevos pozos. Es decir que la experiencia (cantidad de pozos perforados, cantidad de problemas encontrados y solucionados, cantidad y tipo de equipamientos y técnicas nuevas que se han introducido) es un elemento vital en la mejora de la perforación. El mayor conocimiento se elabora principalmente mediante las revisiones.

Las revisiones se pueden encasillar como un proceso de mejora continua. Se trata de identificar los factores negativos y positivos para modificarlos y lograr los pozos más eficientes, seguros y rápidos. Es decir, se busca lograr el menor tiempo de perforación. Como el costo del pozo está muy relacionado al tiempo de perforación, es casi sinónimo decir que estamos buscando el menor costo de perforación.

Uno de los objetivos de este estudio es delinear procedimientos idóneos para el análisis de los múltiples factores que intervienen en la perforación con el fin de detectar cuáles

son los más importantes, de modo de poder optimizarlos. Otro objetivo es el desarrollo de modelos que permitan hacer predicciones de comportamiento futuro.

La capacidad de realizar predicciones no sólo es una cuestión técnica relacionada a la práctica de la perforación, sino que también tiene que ver con la gestión empresarial. En la planificación de las campañas de perforación intervienen cuestiones económicas y estratégicas, como ser la administración de tiempos y disponibilidades de equipamientos, materiales y recursos humanos. En otras palabras, los tiempos de perforación condicionan los costos relacionados, lo cual implica una previsión presupuestaria y también accionan sobre los cronogramas operativos, lo cual, a su vez, condiciona la cantidad de equipos que deberán operar en la campaña, la cantidad y tipo de personal activo, la cantidad y flujo de materiales y servicios relacionados. En resumen, las predicciones de tiempos de perforación influyen sobre:

- Presupuesto
 - ✓ Evaluación económica del proyecto de perforación
 - ✓ Disponibilidad de dinero
 - ✓ Flujo de dinero
- Cronograma
 - ✓ Plazos de ejecución
 - ✓ Cantidad de equipos activos
- Recursos humanos
 - ✓ Cantidad de personal necesario
 - ✓ Tipo de personal necesario
 - ✓ Tiempos de asignación al proyecto
- Recursos materiales
 - ✓ Previsión de materiales (tuberías, cabezales de pozo, etc.)
 - ✓ Gestión de provisión (compra, fabricación, transporte, etc.)
- Servicios
 - ✓ Previsión de compañías de servicios
 - ✓ Programación de actividades

Si las predicciones son elevadas, se estimarán costos altos y la evaluación del proyecto resultará negativa. Viceversa, predicciones reducidas implican la estimación de presupuestos bajos, situación que hace ver atractivo un proyecto débil, lo cual puede ser desastroso para la empresa.

En la etapa de ejecución, las fallas en la estimación de tiempos afectarán la provisión de dinero, equipos, personal y materiales.

1.1.4 Estructuración de este estudio

Existe ingente cantidad de informes que exponen diversas técnicas de revisión de datos de perforación. Entre ellos se destaca la publicación de Lagreca y otros (Lagreca, y otros, 2004) donde se presenta un marco adecuado para este estudio.

La revisión de artículos previos les reveló algunos aspectos del estado de las técnicas de análisis de perforación en ese momento. Sus comentarios y conclusiones son los siguientes:

1. El uso actual de los datos de perforación todavía no está claro y está bajo debate.
2. La mayoría de los criterios disponibles limitan sus aplicaciones a los estimadores de desempeño de la perforación.
3. Los métodos propuestos varían de acuerdo a las necesidades específicas de cada organización o proceso.
4. Hay una variedad de conceptos y teorías que se han usado para desarrollar los actuales métodos y procedimientos.

Esta situación crea el ambiente apropiado para la definición y objetivos del área del conocimiento de perforación llamada Análisis de Perforación y el desarrollo futuro del rol del Analista de Perforación.

Los ámbitos de aplicación de estos métodos son (1) el análisis de datos históricos y (2) análisis de datos recolectados en tiempo real. Queda claro que los objetivos de esta tesis se encuadran dentro del primer aspecto, es decir, el análisis de datos históricos.

1.1.5 El Ingeniero Analista de Perforación

Lagreca y otros, se alinean con Iyoho y otros (Iyoho, y otros, 2004). Estos autores esbozan las características de una figura casi inexistente hoy día en la industria: el analista de perforación. También puntualizan las cualidades y capacidades que debería tener el mismo y proponen el tipo de formación que se debería impartir a tal especialista. La idea gira alrededor de una capacitación integral pre laboral, es decir, impartida ya en la Universidad.

1.1.6 Objetivos de este trabajo

1.1.6.1. Objetivo General

Desarrollar procedimientos de revisión de campañas de perforación de pozos aplicando técnicas de análisis multivariado.

1.1.6.2. Objetivos Específicos

- Aplicar técnicas de análisis de datos multivariados a la revisión de campañas de perforación.
- Caracterizar y clasificar las perforaciones según su eficiencia y variables relacionadas al proceso de perforación de pozos.
- Aplicar técnicas multivariadas a la detección y evaluación de las variables determinantes en los tiempos de perforación.
- Evaluar modelos estadísticos para la predicción de tiempos de perforación.
- Desarrollar e implementar rutinas en lenguaje R para cada etapa de la revisión de campañas de perforación. Las rutinas (scripts) se pueden ver en el siguiente enlace:

<https://skydrive.live.com/redir?page=view&resid=44F1F16F7978AFE7!927&authkey=!AHL5xzdNhrTheH8>

Capítulo 1.2 Perforación y antecedentes

1.2.1 Introducción

En este capítulo se presenta el conocimiento básico de la tecnología de perforación, con el fin de introducir al lector neófito en esta tecnología, acentuando sobre los aspectos relacionados a la eficiencia de la perforación, que es el tema central de estudio.

1.2.2 Equipo de perforación

La perforación para petróleo y gas requiere grandes inversiones que son realizadas, en principio, por las compañías petroleras o los gobiernos. En ocasiones, los costos de perforación resultan ser tan altos que las compañías se agrupan para compartir el riesgo financiero.

Un pozo se clasifica como exploratorio (wildcat) cuando su propósito es descubrir un nuevo reservorio de petróleo. En contraste, el propósito de un pozo de *desarrollo* es la explotación de un reservorio conocido.

Normalmente, la perforación la realiza un contratista de perforación (compañía de perforación – drilling contractor). Una vez que la gerencia decide perforar el pozo, el grupo de ingenieros de perforación prepara un diseño más detallado y redacta los pliegos de las licitaciones. En éstas, y en los contratos de perforación, se especifica el equipamiento y procedimientos que requerirá el operador, junto con una descripción del pozo.

El ingeniero de perforación recomienda los procedimientos que permitirán perforar el pozo tan segura y económicamente como sea posible. En muchos casos el programa original debe modificarse debido a imprevistos durante la perforación. Estos cambios también son de responsabilidad del ingeniero de perforación.

1.2.2.1. Componentes del Equipo

En casi toda perforación actual se utilizan equipos rotativos. La Fig. 1.2-1 (Bourgoyne, y otros, 1986) es una ilustración del proceso de perforación. El pozo se perfora con un trépano al cual se aplica peso y se lo hace girar. El peso se aplica cargando parte del peso de la columna de perforación.

La mesa rotativa hace girar la columna de perforación (tubería de perforación y portamechas) y al trépano. La fuerza para perforarse aplica al trépano descargando peso de los portamechas.

Los recortes de la perforación se sacan a superficie mediante la circulación de un fluido que baja por el interior de la columna, sale por el trépano y sube por el espacio anular entre la columna y el pozo. Los recortes se separan del fluido de perforación en superficie.

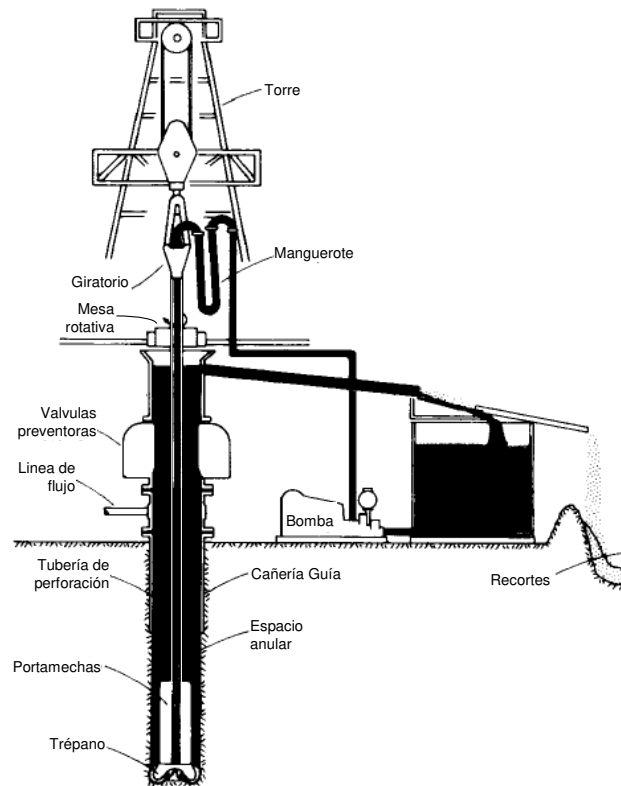


Fig. 1.2-1 Proceso de perforación

Equipo: Torre, giratorio, mesa, válvulas de seguridad, bombas y circuito de lodo.
 Sarta de perforación: Tubería de perforación (drill pipe), portamechas (drill collar), trépano.
 Pozo: Cañería guía, cañería de aislamiento.

Aunque hay muchos tipos y modelos de equipos de perforación, todos tienen el mismo equipamiento básico. Los principales componentes de un equipo rotativo son: (1) el sistema de potencia o energía, (2) el sistema de elevación, (3) el sistema de circulación de fluidos, (5) el sistema de control de pozo (válvula preventora de reventones, blow out preventer o BOP) y (6) el sistema de seguimiento.

1.2.2.2. Fluidos de Perforación

Los fluidos de perforación se usan en el proceso de perforación para (1) limpiar los recortes de roca de debajo del trépano y transportarlos a superficie, (2) ejercer suficiente presión hidrostática sobre las formaciones de subsuperficie para evitar que los fluidos de formación fluyan dentro del pozo, (3) mantener el pozo abierto hasta que se entube y cimente una cañería y (4) enfriar y lubricar la sarta de perforación y el trépano. Además de estas funciones, el lodo NO debe (1) tener propiedades que vayan en detrimento de las técnicas de evaluación de formación planeadas, (2) causar efectos adversos en la formación perforada o (3) causar corrosión en el equipo de perforación o tubulares de superficie.

Los fluidos de perforación más usados son a base de agua. Los lodos base petróleo generalmente son más caros y requieren controles de polución más estrictos que los lodos base agua.

El lodo base agua consiste de una mezcla de sólidos, líquidos y químicos, con el agua como fase continua. Algunos sólidos reaccionan con la fase agua y con los químicos disueltos y, por lo tanto, se clasifican como *sólidos activos*. La mayoría de los sólidos activos presentes son arcillas hidratables.

El Instituto Americano del Petróleo (API) estableció prácticas recomendadas para ensayar fluidos de perforación líquidos. Estas pruebas se desarrollaron para ayudar al ingeniero de lodos a determinar si el fluido de perforación está cumpliendo sus funciones adecuadamente. Corriendo estas pruebas a intervalos regulares, frecuentemente es posible identificar y corregir potenciales problemas de perforación tempranamente y evitar severas pérdidas de tiempo de equipo.

Las pruebas de diagnóstico recomendados por el API incluyen (1) una balanza de lodo para medir la densidad, (2) un embudo Marsh para controlar la consistencia del fluido, (3) un viscosímetro rotacional para determinar la resistencia de gel y propiedades reológicas, (4) un filtro prensa para determinar el caudal de filtrado de lodo y características del revoque de lodo, (5) un pehachímetro para determinar la concentración de H^+ , (7) tamices para determinar el contenido de arena, (8) un destilador de lodo para determinar el contenido de sólidos, petróleo y agua, y (9) un aparato de titulación para análisis químicos.

1.2.2.3. Trépanos de Perforación Rotativos

El proceso de perforación requiere el uso de trépanos. La selección del trépano y las condiciones operativas del mismo es uno de los problemas básicos que enfrenta el

ingeniero de perforación. Hay una gran variedad de trépanos que se fabrican para las diversas situaciones que se encuentran durante la perforación.

Los trépanos de perforación rotativos se clasifican de acuerdo a su diseño en trépanos de arrastre o trépanos de cortadores rodantes.

Trépanos de arrastre:Constan de cuchillas cortadoras integradas al cuerpo del trépano y giran como una unidad con la sarta de perforación.

Trépanos de cortadores rodantes: tienen dos o más conos que contienen los elementos cortadores, los cuales rotan alrededor del eje del cono a medida que el trépano rota en el fondo del pozo. En la Fig. 1.2-2 se ve un trépano de tres conos cortadores rodantes.

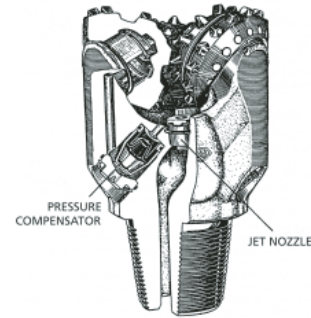


Fig. 1.2-2 Trépano tricono

1.2.3 Agentes que afectan la tasa de penetración

La velocidad de perforación se expresa como tasa de penetración (Rate of Penetration – ROP). La tasa se expresa en unidades de metros por hora o pies por hora o bien en horas o minutos por metro o por pie. La tasa de penetración tiene un efecto directo sobre el costo por metro perforado. Las variables más importantes que afectan la tasa de penetración son: (1) tipo de trépano, (2) características de la formación, (3) propiedades del fluido de perforación, (4) condiciones operativas del trépano, (5) desgaste de los dientes del trépano y (6) hidráulica en el trépano.

Se ha realizado una considerable cantidad de trabajo experimental para estudiar el efecto de estas variables sobre la tasa de penetración. En la mayoría de los trabajos experimentales se estudia el efecto de una variable mientras se mantienen constantes las restantes variables.

Se ha realizado una considerable cantidad de trabajo experimental para estudiar el efecto de estas variables sobre la tasa de penetración. En la mayoría de los trabajos experimentales se estudia el efecto de una variable mientras se mantienen constantes las restantes variables.

1.2.3.1. Tipo de Trépano

El tipo de trépano seleccionado tiene un gran efecto sobre la tasa de penetración. La tasa de penetración inicial de trépanos de cortadores rodantes es más alta cuando se usan dientes largos y el ángulo de excentricidad de conos es grande. Estos trépanos sólo son prácticos en formaciones blandas. En formaciones duras los dientes se destruyen rápidamente y la penetración es menor.

Los trépanos de arrastre se diseñan para lograr una determinada tasa de penetración, la cual depende del número de cuchillas y el ángulo de corte.

1.2.3.2. Características de la Formación

El límite elástico y la resistencia máxima de la formación son las propiedades más importantes que afectan la tasa de penetración junto con la permeabilidad y la composición mineral de la roca. Las rocas que contienen minerales duros y abrasivos pueden causar rápido desgaste de los dientes del trépano. Por otro lado, las rocas que contienen minerales arcillosos gomosos pueden embolar el trépano y hacer que perforo de manera muy poco eficiente.

1.2.3.3. Propiedades de los Fluidos de Perforación

Las propiedades de los fluidos de perforación que afectan la tasa de penetración incluyen (1) densidad, (2) propiedades del flujo, (3) características de filtrado, (4) contenido de sólidos y (5) composición química.

La tasa de penetración tiende a disminuir cuando aumenta la densidad del fluido, la viscosidad y el contenido de sólidos y tiende a aumentar cuando aumenta la tasa de filtrado. La densidad, el contenido de sólidos y las características de filtrado del lodo controlan la presión diferencial a través de la zona triturada debajo del trépano. La viscosidad del fluido controla las pérdidas de fricción parásitas en la sarta de perforación y, por tanto, la energía hidráulica disponible en los chorros (jets) del trépano para la limpieza.

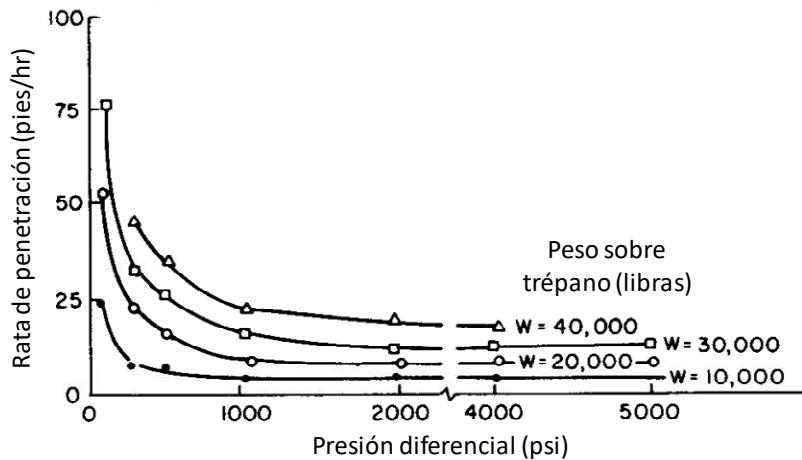


Fig. 1.2-3 Efecto del sobrebalance

Cuando aumenta la densidad del fluido aumenta la presión diferencial sobre el fondo del pozo y la tasa de penetración tiende a disminuir a cualquier peso sobre el trépano.

Un aumento de la densidad del fluido de perforación genera un aumento de la presión de fondo debajo del trépano y, por lo tanto, aumenta la presión diferencial entre la presión del pozo y la presión de los fluidos de formación. Esta presión diferencial frecuentemente

se llama *sobrebalance*. En la Fig. 1.2-3 (Bourgoyne, y otros, 1986) se ven los resultados obtenidos en perforación real.

1.2.3.4. Condiciones Operativas

Estas son el peso sobre el trépano y la velocidad de rotación. La Fig. 1.2-4a, (Bourgoyne, y otros, 1986) es un gráfico de la tasa de penetración versus el peso sobre el trépano (todas las otras variables de perforación constantes). No se logra una apreciable tasa de penetración hasta que se aplica un peso “umbral” sobre el trépano (Punto a). Luego, con moderados aumentos del peso la tasa de penetración aumenta rápidamente (Segmento ab). A pesos algo mayores, frecuentemente se observa una línea curva (Segmento bc). A mayores pesos, subsecuentes incrementos del peso producen solamente pequeñas mejoras en la tasa de penetración (Segmento cd). En algunos casos, se observa que a altos valores de peso sobre el trépano, la tasa de penetración disminuye drásticamente (Segmento de). Esta pobre respuesta se atribuye a una limpieza menos eficiente porque hay una producción muy grande de recortes o a que el elemento cortante se ha clavado completamente en el fondo del pozo. Este efecto se llama “floundering” o estancamiento.

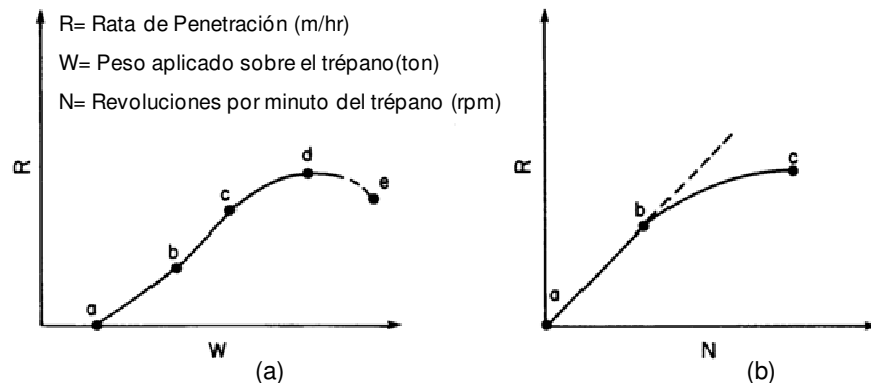


Fig. 1.2-4 Tasa de penetración. a) Con peso W . b) Con rotación N

- a) A partir de un “umbral”, mayor peso sobre el trépano implica mayor ROP, hasta un punto donde decae (posiblemente por remolienda de recortes debajo del trépano).
- b) Mayor velocidad de rotación incrementa linealmente la ROP. Con alta rotación, los incrementos de ROP son menores.

En la Fig. 1.2-4b (Bourgoyne, y otros, 1986) se ve la respuesta típica de la tasa de penetración versus la velocidad de rotación obtenida con todas las otras variables constantes. Usualmente, a bajos valores de rotación la tasa de penetración aumenta linealmente con la velocidad de rotación. A mayores valores de velocidad de rotación la respuesta de la tasa de penetración disminuye. Esta pobre respuesta a altas velocidades de rotación se atribuye a una limpieza de fondo menos eficiente.

1.2.3.5. Desgaste del trépano

A medida que avanza, los dientes sufren un desgaste que va reduciendo la eficiencia del trépano. El desgaste depende de (1) abrasividad de la formación, (2) geometría de los dientes, (3) peso sobre el trépano, (4) velocidad de rotación y (5) la acción de limpieza y enfriamiento del fluido de perforación.

1.2.3.6. Hidráulica en el Trépano

En la Fig. 1.2-5 se ve el trépano en el extremo de la sarta de perforación. El fluido de perforación sale en forma de chorros o jets por los orificios o toberas. Los orificios son restricciones que generan un salto de presión cuyos dos efectos principales son: incrementar la velocidad del fluido y disipar potencia hidráulica sobre el fondo del pozo.

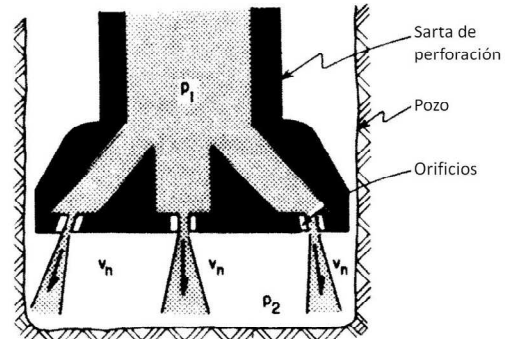


Fig. 1.2-5 Flujo por orificios

El caudal de fluido y su velocidad contribuyen a la limpieza del fondo del pozo, retirando el material cortado por el trépano. Las condiciones de flujo, caracterizadas principalmente por caudal y velocidad, son apropiadas si los recortes que escarba el trépano en el fondo son retirados inmediatamente del fondo. Si el flujo es eficiente y mantiene el fondo permanentemente limpio, el trépano trabaja escarbando sobre el fondo y no sobre material ya cortado. Si la limpieza no es efectiva, el fondo se mantiene cubierto por recortes y el trépano no puede escarbar sobre el fondo. En este caso, el trépano continúa re-moliendo material ya cortado sin avanzar, o con una tasa de penetración menor.

La disipación de potencia sobre el fondo, expresada ya sea en términos de potencia hidráulica o de impacto hidráulico, contribuye a la penetración del trépano. En general, mayor potencia disipada sobre el fondo o mayor impacto aplicado al fondo favorece la tasa de penetración.

Para caracterizar el efecto de la hidráulica en la tasa de penetración se usa la potencia hidráulica en el trépano, la fuerza de impacto del chorro, y la velocidad en la tobera.

La Fig. 1.2-6, (Bourgoyne, y otros, 1986) es un ejemplo hipotético del comportamiento con la hidráulica. A medida que se aplica más peso al trépano y se genera más cantidad de recortes, se llega a un punto fluctuante o estancamiento, donde los recortes no son removidos tan rápidamente como se generan. Si se incrementa el nivel de hidráulica, se

puede aplicarmayor peso sobre el trépano y obtener mayores tasas de penetración antes de llegar al punto fluctuante.

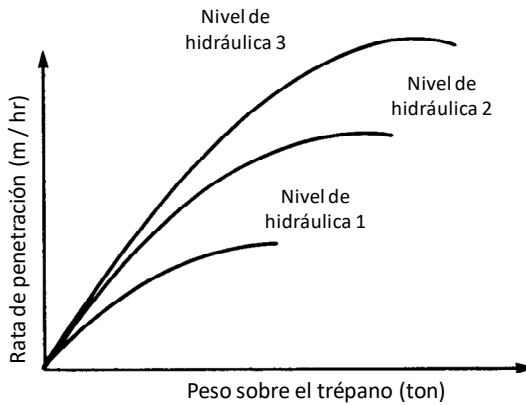


Fig. 1.2-6 Efecto de la hidráulica sobre la tasa de penetración

Mayor nivel de hidráulica (caudal, potencia, impacto sobre el fondo) resulta en mayor ROP a cualquier peso sobre el trépano. Cada nivel de hidráulica llega a un “estancamiento” a mayores pesos.

1.2.4 Ecuación de la tasa de penetración

A continuación se presenta un modelo ingenieril clásico (Bourgoyne, y otros, 1974) y (Bourgoyne, y otros, 1986) para estimar la tasa de penetración. Este modelo se presenta como referencia frente a los modelos estadísticos de este estudio.

Se asume que los efectos del peso sobre el trépano, velocidad de rotación, desgaste del trépano, etc. son independientes unos de otros y que el efecto conjunto toma la forma de un producto:

$$R = (f_1)(f_2)(f_3) \dots (f_n) \quad \text{- Ec. 1.2-1}$$

Donde R = Tasa de penetración y $(f_1)(f_2)(f_3) \dots (f_n)$ representan las relaciones funcionales entre distintas variables y la tasa de penetración.

El modelo de Bourgoyne-Young incluye 8 funciones:

$$f_1 = e^{2.303a_1}$$

$$f_2 = e^{2.303a_2(10.000-D)}$$

$$f_3 = e^{2.303a_3D^{0.69}(g_p-9.0)}$$

$$f_4 = e^{2.303a_4D(g_p-\rho c)}$$

$$f_5 = \left[\left(\left(\frac{W}{d_b} \right) - \left(\frac{W_0}{d_b} \right)_t \right) / \left(4 - \left(\frac{W_0}{d_b} \right)_t \right) \right]^{a_5}$$

$$f_6 = \left(\frac{N}{60} \right)^{a_6}$$

$$f_7 = e^{-a_7 h}$$

$$f_8 = \left(\frac{F_j}{1.000} \right)^{a_8}$$

Donde:

D = Profundidad verdadera

g_p = Gradiente de presión poral

ρ_c = Densidad equivalente de circulación

W = Peso sobre el trépano

d_b = Diámetro de trépano

$(W_0/d_b)_t$ = Peso umbral por pulgada de diámetro al cual el trépano empieza a perforar

N = Velocidad rotacional del trépano (RPM)

h = Desgaste fraccional de diente de trépano

F_j = Fuerza de impacto hidráulico debajo del trépano

a_1 a a_8 = Valores que se deben estimar para las condiciones locales.

La función f_1 depende de a_1 , que representa el efecto de la resistencia de la formación y del tipo de trépano. También podría incluir el efecto de otras variables como el tipo de lodo y contenido de sólidos.

Las funciones f_2 y f_3 representan el efecto de la compactación de la roca. f_2 toma en cuenta el incremento de la resistencia de la roca por la compactación normal con la profundidad y f_3 tiene en cuenta la sobrecompactación debida a zonas con presiones anormalmente altas.

La función f_4 modela el efecto del sobrebalance. Vale 1.0 cuando el sobrebalance es nulo.

Las funciones f_5 y f_6 modelan el efecto del peso sobre el trépano y la velocidad de rotación del mismo.

La función f_7 representa el efecto del desgaste del trépano. Vale 1.0 cuando el trépano es nuevo.

La función f_8 modela el efecto de la hidráulica con la intervención de la fuerza de impacto hidráulica.

1.2.5 Función del Ingeniero de Perforación

La función del Ingeniero de Perforación es diseñar y dirigir la perforación de la manera más eficiente posible. Además de elegir el trépano más apropiado para el trabajo, el ingeniero de perforación debe cuidar que la operación sea lo más eficiente posible. Debe seleccionar el equipo, los fluidos, los demás elementos mecánicos, las compañías de servicio, el personal, y lograr que interactúen todos los elementos de la forma más adecuadamente compatible con la mayor eficiencia. Esta es la instancia donde puede

intervenir el Ingeniero Analista de Perforación y donde esta tesis pretende aportar más conocimiento y herramientas idóneas acordes con las necesidades.

1.2.6 Revisión de Antecedentes

A continuación se presenta una revisión de las publicaciones previas relacionadas a la revisión de datos de perforación. Se observa la evolución de los estilos y métodos de análisis de datos de perforación.

Ya en 1970 se encuentran publicaciones que abordan el tema de la optimización de la perforación mediante el análisis de datos de perforaciones previas, además de investigar los efectos de las variables que intervienen en la gestión de la perforación.

Lummus (Lummus, 1970), (Lummus, 1971) hizo un análisis de cuáles son las variables que afectan la performance de la perforación. En la Tabla 1.2-1(Lummus, 1970) se presentan las variables estudiadas y clasificadas como alterables o inalterables.

Tabla 1.2-1 Variables de perforación

ALTERABLES	INALTERABLES
Lodo	Clima
Tipo	Locación
Contenido de sólidos	Condiciones del equipo
Viscosidad	Flexibilidad del equipo
Pérdida de fluido	Gases corrosivos de pozo
Densidad	Tiempo de carrera
Hidráulica	Propiedades de las rocas
Presión de bomba	Problemas característicos de pozo
Velocidad de jets	Disponibilidad de agua
Caudal de circulación	Formación a perforar
Velocidad anular	Eficiencia de las cuadrillas
Tipo de trépano	Profundidad
Peso sobre el trépano	
Velocidad de rotación	

Ikoku(Ikoku, 1978) introdujo la aplicación de los modelos de curva de aprendizaje, adaptado de la industria manufacturera y de la Investigación de Operaciones. Aplicó la curva de aprendizaje a la predicción del costo por pie de perforación y a la predicción del tiempo total necesario para perforar el pozo.

Brett y Millheim aportaron una evolución importante sobre la misma idea (Brett, y otros, 1986) que fue el desarrollo de una regresión exponencial negativa con tres parámetros: C₁, C₂ y C₃ como la siguiente.

$$t = C_1 e^{(1-n)C_2} + C_3 \quad \text{Ec. 1.2-2}$$

donde

t = tiempo de perforación

n = el número secuencial del pozo

C_1 = el tiempo por encima del óptimo ideal para perforar el primer pozo

C_2 = la velocidad o efectividad de aprendizaje

C_3 = tiempo óptimo idealizado de perforación

Una curva de aprendizaje típica se ve en la Fig. 1.2-7. Estos resultados sugerirían al perforador cuál es el tiempo mínimo que podría alcanzar optimizando su operación, que es C_3 .

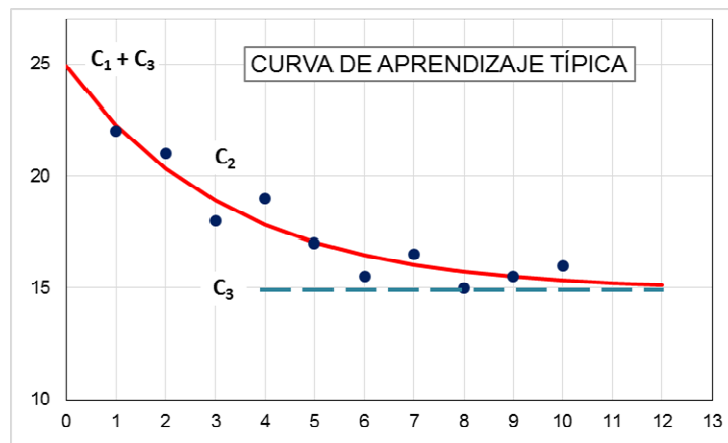


Fig. 1.2-7 Curva de aprendizaje

Los tiempos de perforación de pozos disminuyen a cada pozo sucesivo. C_3 es el tiempo óptimo de perforación, $C_1 + C_3$ es el tiempo inicial ideal, C_2 es la velocidad o efectividad de aprendizaje.

Noerager y otros (Noerager, y otros, 1987) presentaron modificaciones a la curva de aprendizaje. Su estudio se relacionó a la perforación de pozos en el Mar del Norte.

Estos investigadores encontraron una dificultad importante en la **gran dispersión de datos** que se encuentra en las perforaciones típicas. No pudieron atribuir esta dispersión a errores de medición o recolección de datos ni se pudo discriminar el origen o la causa, debido a la estructura de la base de datos. Resignadamente, concluyeron que la gran dispersión se puede deber a muchas pequeñas diferencias entre las operaciones, las cuales, sumadas integran las grandes variaciones encontradas.

En la Fig. 1.2-8 se muestra un ejemplo con el tiempo total de perforación de pozo versus profundidad medida. Se puede ver fácilmente el grado de dispersión presente.

En los años siguientes, aparecieron los primeros intentos de monitoreo de la perforación en tiempo real. Arnold y otros (Arnold, y otros, 1991) presentaron una aplicación práctica de la evaluación de la perforación.

Millheim y otros (Millheim, y otros, 1998) proponen una Gestión de la Curva de Aprendizaje de Perforación, en distintas etapas como en la Fig. 1.2-9. Las etapas a y b representan aprendizajes, pero la gestión continua puede detectar etapas como la c, donde el C_3 se desplaza hacia arriba lo cual significa un desmejoramiento de la performance. Esto requiere acciones correctivas para retornar al buen nivel inicial.

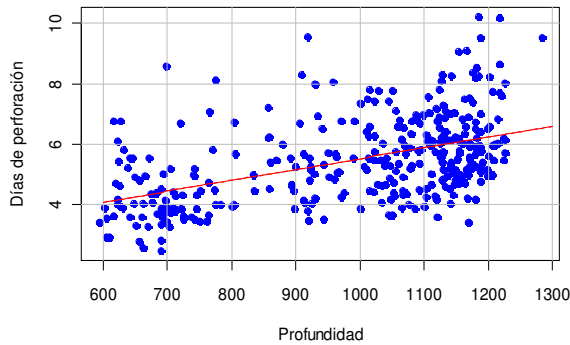


Fig. 1.2-8 Tiempo total vs. Profundidad

Los tiempos totales de perforación aumentan con la profundidad de pozo, pero se observa una gran dispersión.

La propuesta de Kadaster y otros (Kadaster, y otros, 1992) se basó en los conceptos de la Gestión de la Calidad Total. Ellos acuñaron el término Drilling Time Analysis (DTA) o Análisis del Tiempo de Perforación. Además de la aplicación de la Gestión de Calidad, pusieron gran énfasis en la gestión de los datos, proponiendo clasificaciones y estructuras de datos acordes a la actividad de las operaciones de perforación.

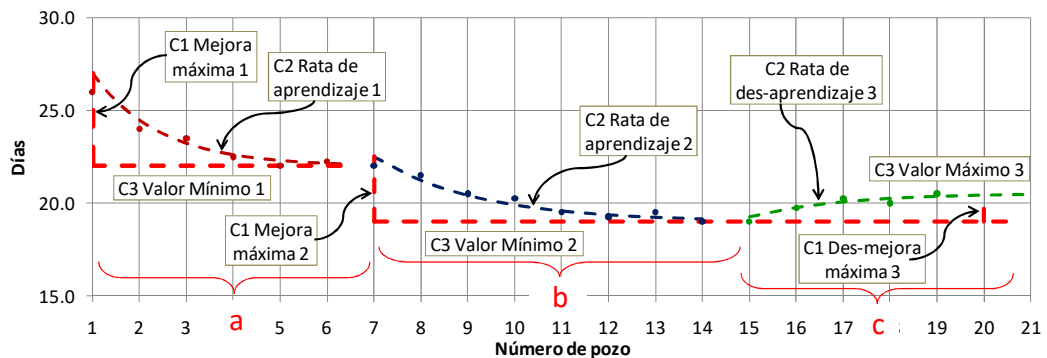


Fig. 1.2-9 Aprendizaje a) Proyecto nuevo b) Cambia actividad c) Desmejora

La curva de aprendizaje aplicada en campañas sucesivas. En las dos primeras campañas (a y b) se mejoran los tiempos. En la campaña c, hay desmejora.

Los métodos tradicionales solamente hacían una comparación del “tiempo planeado” contra el “tiempo real”, sin identificar el origen de los problemas que ocasionaban los mayores tiempos. La Gestión de Calidad permite conocer o saber algo más acerca de la estructura de los tiempos. Ellos identificaron una categoría de tiempos llamada “Tiempos con Problemas” o “Tiempo No-Productivo” (TNP), o “Tiempo Perdido”. La mayor preocupación de la Gestión de Calidad fue promover la obtención de datos con una estructura adecuada, completos y confiables e investigar detalladamente la causa de los tiempos TNP para arbitrar las soluciones.

Peterson y otros publicaron uno de los primeros informes de este tipo (Peterson, y otros, 1995) en el cual presentaron predicciones de la performance de perforación basadas en el análisis de riesgo, realizadas mediante simulaciones Montecarlo. La distribución que mejor se adaptó a sus datos de tiempo históricos fue la **distribución gamma**.

Bond y otros (Bond, y otros, 1996) introdujeron el concepto de Technical Limit o Límite Técnico. El término se refiere a la situación extrema que se lograría en caso que cada operación individual se ejecutara perfectamente, de manera que la suma de todos los tiempos “perfectos” rindiera el mejor tiempo de perforación. Estos tiempos no son meramente los “tiempos libres de problemas”, sino un tiempo de perforación compuesto por operaciones individuales, cada una representando la mejor performance.

Se define un tiempo de pozo teórico que es la suma de todos los tiempos de operación sin fallas, basados en el conocimiento real de las operaciones y la tecnología del diseño. La diferencia entre el tiempo real y el tiempo de pozo teórico es el Tiempo Removible, que incluye los tiempos perdidos convencionales (TNP de Kadaster) y el Tiempo Perdido Invisible, como se ve en la Fig. 1.2-10.

El nudo de la aplicación reside en la disponibilidad de datos adecuados. El seccionamiento provee una cuantificación de los tiempos perdidos invisibles y convencionales.

Irrgang y otros (Irrgang, y otros, 1999) presentaron un proceso de gestión de base de datos y una herramienta computacional para el planeamiento de pozos nuevos. Con las contribuciones de un “Club de Perforación” crearon una gran base de datos con pozos de todo el mundo. Los datos de esta base se pueden capturar con el software que crearon a tal fin.

Iyoho y otros (Iyoho, y otros, 2004) presentaron una revisión y avances en Análisis de Perforación. Recogen varios conceptos ya desarrollados y tratan de darles una estructura

integrada, mediante el análisis de tiempos con problemas, análisis de tiempos sin avance, análisis de tiempos versus profundidad, análisis de costos versus profundidad, evaluación de equipo y cuadrillas, análisis de trépano perforando y tiempos de viajes, gráficos de distribución de tiempos, performance operacional, etc.

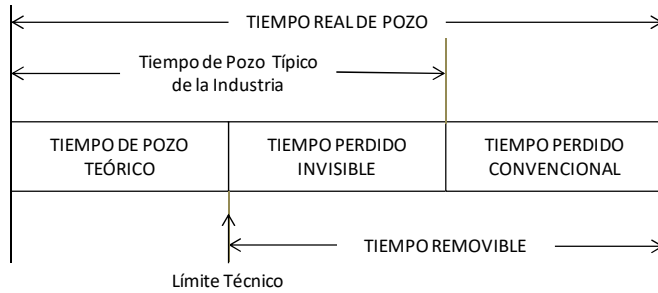


Fig. 1.2-10 Terminología de los tiempos de perforación

La comparación de los tiempos reales de perforación con los tiempos de un “pozo tipo” detecta el tiempo perdido “convencional”. La técnica del “límite técnico” identifica el tiempo perdido “invisible”.

El principal aporte de este artículo es la formulación de una metodología integral de 10 pasos, que integran un proceso cíclico de realimentación entre el planeamiento y la ejecución, que se ve en la Fig. 1.2-11.

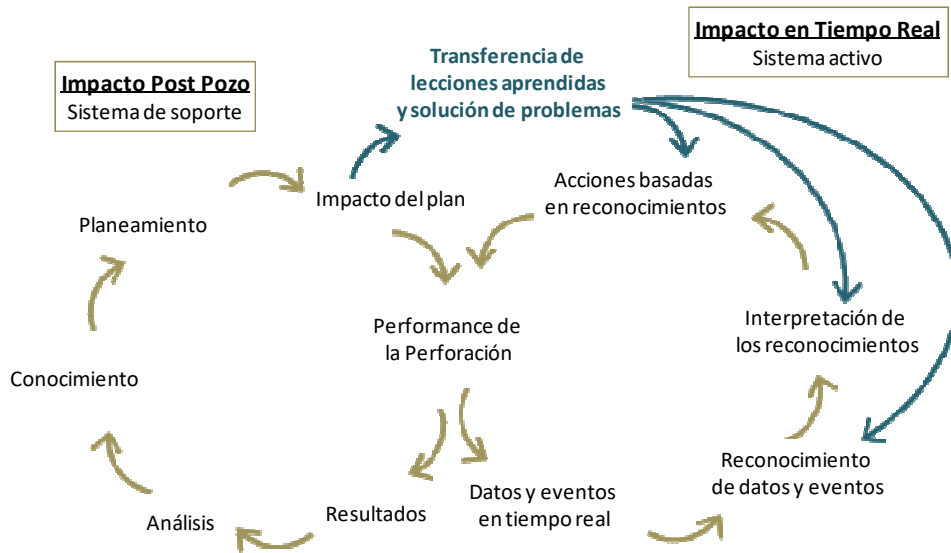


Fig. 1.2-11 Ciclos del conocimiento

El análisis de performance se analiza en tiempo real (ciclo de la derecha) o sobre datos históricos (ciclo de la izquierda). Este último da soporte al primero.

Simultáneamente, Adeleye y otros presentaron (Adeleye, y otros, 2004) la aplicación práctica del artículo anterior. En este artículo señalan el sorprendente hecho (marcado

también por otros artículos previos) de que **la frondosa cantidad de datos generados por la actividad de perforación no se utiliza de manera que se acorte la curva de aprendizaje y de promover la eficiencia de la perforación.**

Lagrecá y otros (Lagrecá, y otros, 2004) propusieron la metodología aquí presentada en el Marco del párrafo 1.1.4. Los nuevos desarrollos siguen las líneas pioneras sumando algunas metodologías y tratando de dar un cuerpo sistemático. Dos de las conclusiones principales son (1) **el uso actual de los datos de perforación todavía no es claro y está bajo debate** y (2) es necesario **definir y dar una descripción de los objetivos del Análisis de Perforación y el desarrollo futuro del rol del Analista de Perforación.**

Kravis y otros (Kravis, y otros, 2004) además de Peterson y otros (Peterson, y otros, 1995) reconocen que las distribuciones de los tiempos ajustan mejor con distribuciones sesgadas (**gamma**, log-normal, exponencial). También han propuesto un gráfico de probabilidad donde presentan la **proporción de tiempos de problemas acumulados en relación a los tiempos totales** (ordenados en orden creciente), contra la **proporción de pozos con problemas en relación al total de pozos analizados**, como se ve en la Fig. 1.2-12. Operaciones maduras tienen una gran proporción de pozos libres de problemas.

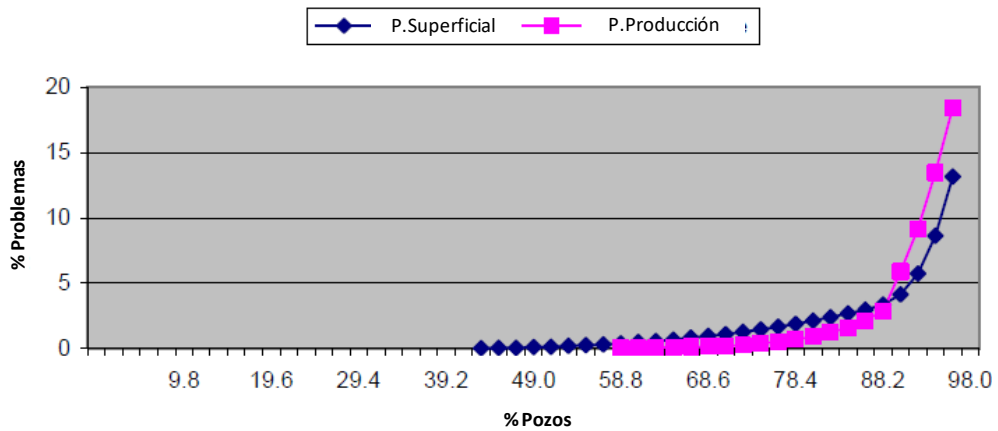


Fig. 1.2-12 Probabilidad de problemas

El gráfico de probabilidad presenta la proporción de tiempos de problemas acumulados en relación a los tiempos totales contra la proporción de pozos con problemas en relación al total de pozos.

Valdez y Sager (Valdez, y otros, 2005) recurren al benchmarking que es una técnica de buscar la excelencia por medio de la comparación con los mejores exponentes de la actividad.

Los pasos a seguir en la implementación del sistema se ven en la Fig. 1.2-13. Lo que se busca es identificar las mejores prácticas que llevan a los mejores resultados y tratar de implementarlas en el caso propio.

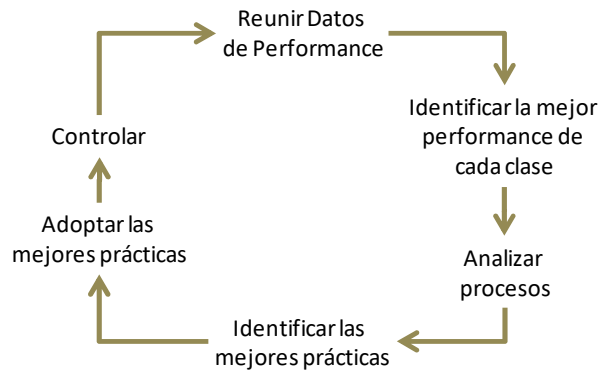


Fig. 1.2-13 Proceso de benchmarking

Aparecieron otros trabajos relacionados a conceptos como la Energía Específica del Límite Técnico (Curry, y otros, 2005), la gestión de datos (Nakagawa, y otros, 2005) y revisiones del Proceso de 10 Pasos(Iyoho, y otros, 2005).

1.2.7 Modelado de la tasa de penetración

Los antecedentes de modelado de la ROP, nos remiten a modelos ingenieriles en los cuales se ha buscado relaciones como la propuesta por Maurer(Maurer, 1962), donde, estableciendo una premisa de limpieza perfecta del fondo, la ROP es proporcional a las revoluciones de la mecha y al cuadrado del peso aplicado, e inversamente proporcional a los cuadrados del diámetro de la mecha y de la resistencia de la roca. El efecto de una limpieza imperfecta afecta los exponentes, reduciéndolos. La estimación de los nuevos exponentes fue analizado por diversos investigadores que propusieron múltiples soluciones. El efecto de la hidráulica fue considerado por Eckel (Eckel, 1968) que relacionó ROP con el grado de hidráulica dado por el número de Reynolds, el cual depende de la viscosidad, la densidad, el diámetro de los orificios y el caudal. Algunas de estas propuestas fueron la base para algunas aplicaciones orientadas a la reducción de los costos de perforación, ya sea en una carrera de trépano o en uno o varios intervalos en particular (Wilson, y otros, 1972). Bourgoyne y Young presentaron una integración de las variables consideradas hasta ese momento en una ecuación de 8 términos($R = (f_1)(f_2)(f_3) \dots (f_n)$ - Ec. 1.2-1), (Bourgoyne, y otros, 1974).

Con el advenimiento del desarrollo computacional, otras aplicaciones más complejas resultaron en simuladores del proceso de perforación completo, desde la selección del equipo hasta las condiciones de perforación (Millheim, y otros, 1983), (Millheim, y otros, 1983-2), (Brett, y otros, 1985), (Al-Betairi, y otros, 1988), (Rampersad, y otros, 1994). Algunos de simuladores resultaron apropiados para generar programas de entrenamiento

(Cooper, y otros, 1995), (Cooper, y otros, 1996). En otros casos se aplicó un análisis retrospectivo, con el objeto de predecir la resistencia de la roca, valores que luego se utilizaban en un simulador (Bratli, y otros, 1997). La modelización ingenieril continuó basándose en los modelos tradicionales de Bourgoyne-Young y de Warren. También se realizaron comparaciones entre modelos ingenieriles y redes neuronales artificiales (Rahimzadeh, y otros, 2010), (Mostofi, y otros, 2010).

También se hicieron aplicaciones de redes neuronales artificiales para predecir el desgaste de trépanos y la ROP (Bilgesu, y otros, 1997). A pesar de estas iniciativas, siguieron apareciendo propuestas de selección personal de variables y observación directa de los posibles efectos (Fear, 1999).

Posteriormente, la iniciativa se inclinó a la optimización en tiempo real, aplicando el concepto de energía mecánica específica (Koederitz, y otros, 2005), (Remmert, y otros, 2007), (Armenta, 2008) o algoritmos específicos (Iqbal, 2008).

1.2.8 Colofón

En general, la técnica de perforación ha identificado los principales factores que afectan la performance y los ingenieros perforadores concentran su pesquisa en tales factores. Sin embargo, en muchos casos los procedimientos de clasificación, ordenación y ponderación de los mismos en cada caso, son artesanales. Debido a esto, los análisis tradicionales serían permeables a técnicas estadísticas.

En su artículo relacionado a la curva de aprendizaje (Noerager, y otros, 1987), presentan situaciones como las de la Fig. 1.2-8 donde la regresión resulta poco convincente debido a la gran dispersión de datos, pero no intentaron una clasificación por factores. La clasificación permite hacer agrupamientos o “conglomerados” más uniformes, reduciendo la dispersión y mejorando la calidad de la regresión.

En cierta manera, se percibe que los análisis propuestos fueron formulados desde el punto de vista del perforador, posición que tiene sus limitaciones. Estas restricciones son las que dan lugar y espacio a la visión del estadístico. La iniciativa de esta tesis es tomar el desafío planteado por Irrgang y otros ampliando el horizonte de acción, mediante la introducción de técnicas estadísticas multivariadas.

PARTE 2 - MATERIALES Y METODOS

Capítulo 2.1 Materiales

2.1.1 Introducción

Generalmente, la obtención y preparación de la materia prima del trabajo, los datos, es uno de los pasos que consume mayor tiempo y esfuerzo(Myatt, 2007).

Los términos obtención y preparación están relacionados a un tema de **calidad de datos**, básicamente cuando se requieren y manipulan grandes cantidades de datos. Generalmente, los principales factores que afectan la calidad de datos son(Vercellis, 2009): Exactitud, integridad, consistencia, vigencia, individualidad, relevancia, interpretabilidad y accesibilidad. En otras palabras, se requiere que los datos sean exactos, que estén todos completos, que sean coherentes entre sí, que correspondan a las fechas en estudio, que no estén repetidos, que sean relevantes, que tengan sentido y que sean accesibles.

Para lograr una buena calidad de datos, Han y Kamber proponen realizar un pre-procesamiento de datos que puede comprender limpieza, integración, transformación y/o reducción de datos(Han, y otros, 2006), en el cual se deben componer las fallas de integridad, eliminar los ruidos presentes y corregir las inconsistencias. Myatt (Myatt, 2007) afirma que alistar los datos para el análisis es la tarea que más tiempo consume en todo el proceso.

Además, Myatt organiza las tareas que comprende el análisis de datos y sus interacciones de acuerdo a la Fig. 2.1-1(Myatt, 2007). Desde resumir datos, hallar relaciones entre ellos y elaborar modelos de predicción, se deben considerar una variedad de tareas como presentaciones gráficas, tablas de resumen, estadísticas descriptivas, búsqueda de relaciones y correlaciones, y análisis de modelos matemáticos, todas las cuales intervienen en un juego interactivo e iterativo

En los anteriores capítulos relacionados a Perforación, quedó claro que la actividad involucra una ingente **cantidad de operaciones, personal, herramientas, técnicas, situaciones, contingencias, ámbitos y escenarios**. Por esta razón los resultados de la perforación dependen de una gran cantidad de factores.

También está claro que, tratándose de una actividad económica, el parámetro principal sea, precisamente, el monetario. Y como, en resumidas cuentas, el negocio económico

está compuesto por tres términos: (1) Costos o inversiones, (2) Ingresos y (3) Beneficios, la gestión de la actividad pasa por maximizar los beneficios, que puede lograrse aumentando los ingresos y/o reduciendo los costos. Es decir, uno de los medios es reducir los costos de perforación, como ya se ha explicado anteriormente.

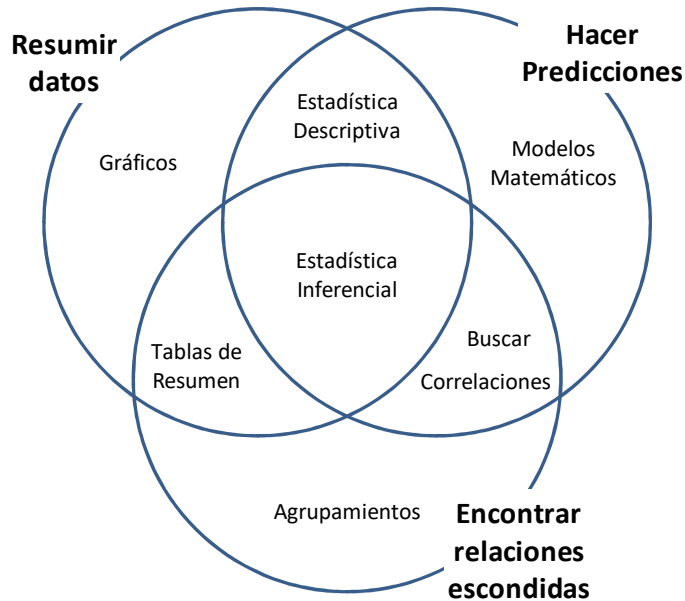


Fig. 2.1-1 Actividades del Análisis de Datos

Esto se ha buscado por diversos medios desde el origen de la técnica de perforación, ya sea por medio de mejoras o cambios en las herramientas, tecnologías, sistemas y también mediante la optimización de las operaciones. También la optimización se ha intentado por diversos medios, ya sean intuitivos u organizados según alguna metodología. De cualquier modo, la optimización requiere la revisión de la actividad precedente, buscar las fallas y las formas de mejorar el trabajo.

2.1.2 Antecedentes relacionados a datos

La propuesta basada en los conceptos de la Gestión Total de la Calidad(Kadaster, y otros, 1992) enfatiza intensamente la gestión de los datos, proponiendo clasificaciones y estructuras de datos acordes a la actividad de las operaciones de perforación. La mayor preocupación de la Gestión de Calidad es promover la estructuración acertada de las bases de datos a fin de **obtener datos adecuados, completos y confiables.**

Peterson y otros(Peterson, y otros, 1995)dan especial importancia a la **disponibilidad de datos históricos confiables y de calidad.**

En 1996, Bond y otros introdujeron el concepto de Technical Limit o Límite Técnico (Bond, y otros, 1996). Ellos también opinan que el nudo de la aplicación reside en la **disponibilidad de datos adecuados**. Estos autores coinciden en que la tarea de revisar y clasificar los datos de perforaciones previas, para seccionar el pozo teórico, **demandan gran cantidad de tiempo y la participación de personal con alto conocimiento de las operaciones**.

En 1998, Millheim y otros hicieron su aporte al tema, (Millheim, y otros, 1998), alentando a implementar una **adecuada estructura de bases de datos e insisten sobre su calidad**. Dicen que con una buena estructuración sería posible analizar la perforación tanto en conjunto como en sus partes constitutivas. Pero, por supuesto nada de eso es posible si la cantidad y calidad de datos no es correcta. A este respecto mencionan tres situaciones afrontadas por todo intento de análisis de perforación: **(1) más cantidad de datos no significa datos útiles, (2) la calidad de datos es muy variable y (3) armar una base de datos adecuada demanda mucho tiempo y es una tarea frustrante**. Estos conceptos serán desarrollados por otros autores, más adelante.

Además, Adeleye y otros señalan el sorprendente hecho (marcado también por otros artículos previos) de que **la frondosa cantidad de datos generados por la actividad de perforación no se utiliza de manera que se acorte la curva de aprendizaje y de promover la eficiencia de la perforación** (Adeleye, y otros, 2004).

Diversos autores se refieren al tipo, cantidad y calidad de datos como uno de los nudos cruciales en el Análisis de Perforación. Todos coinciden en que la gestión de datos es una cuestión primordial, pero compleja que requiere especial atención. En parte, esto se debe a que, en general, (1) no hay normalizada una estructura común de datos y a que también (2) hay problemas en los registros.

2.1.3 Pozo tipo en evaluación

En cada área los pozos se diseñan de acuerdo a las características regionales. En este estudio el pozo tipo es el que se muestra en la Fig. 2.1-2.

2.1.3.1. Etapas y actividades de pozo

En las técnicas tradicionales de análisis es común seccionar los distintos procesos de construcción de pozo en **etapas** y/o en **actividades**. El seccionamiento del pozo en etapas permite analizar en detalle cada una de las partes de la perforación. El seccionamiento en actividades permite analizar el pozo completo o cada etapa según los

grupos de actividades, para detectar mejores o peores prácticas, detectar fallas y analizar problemas.

Pozo 1001		
Etapa	Días	Profund.
1-DTM	2.31	0
2-Perforación de la Guía	3.94	155
3-Perforación de la Aislación	8.07	1117

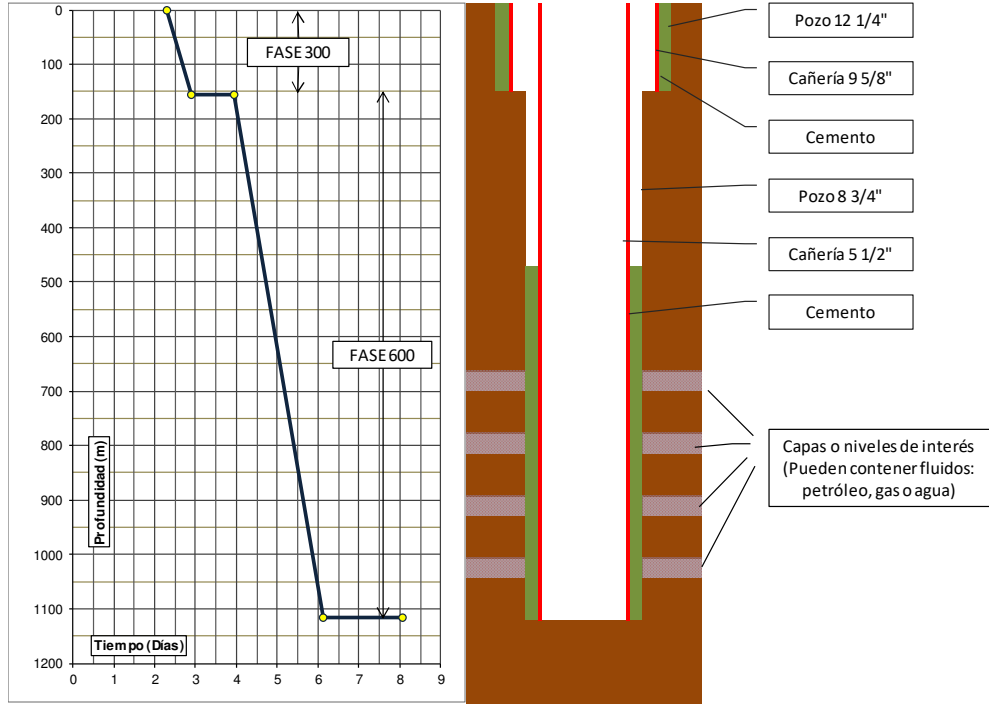


Fig. 2.1-2 Esquema de pozo

A la derecha, la arquitectura del pozo tipo en estudio. A la izquierda, los tiempos y las etapas o fases de la construcción de estos pozos.

La Fig. 2.1-2 es un ejemplo donde se ven los tiempos ocupados por cada etapa. En este caso se ha considerado un seccionamiento de tres etapas: (1) Traslado y montaje del equipo (DTM), (2) Perforación de la guía y (3) Perforación de la aislación. Estas etapas se denominaron Fase 100, Fase 300 y Fase 600. Las fases 300 y 600 se pueden integrar en el concepto de Construcción de Pozo.

En la Tabla 2.1-1 se ven actividades operacionales típicas de las fases de perforación. Las actividades pueden cambiar según los casos o necesidades. Por ejemplo: Los tiempos de la etapa Perforación de la Guía se pueden organizar por las siguientes actividades sucesivas: Perforación Pura, Entubación, Cementación, con el Cabezal y con la BOP, pero también podrían aparecer otras actividades contingentes, como Horas perdidas, Mantenimiento o Pruebas.

2.1.3.2. Tiempos de Perforación

La Fase 300 incluye todos los tiempos de perforación desde superficie hasta la profundidad de la guía, la entubación de la cañería guía y su correspondiente cementación. La Fase 600 comprende todos los tiempos de perforación desde el fondo de la guía hasta profundidad final de pozo, los registros, entubación de la cañería de aislación y su correspondiente cementación. Están incluidos todos los tiempos, como por ejemplo, cambios de trépano, tiempos perdidos, circulación de pozo, armado de BOP y de cabezal de boca de pozo.

Tabla 2.1-1 Actividades de pozo

con la BOP	con el Cabezal
en Cementación	en Control Pozo
en el DTM	en la Entubación
en Horas Perdidas	en Maniobras
en Mantenimiento	en Perfilaje
en Perforación Pura	en Pruebas
en Seguridad	

En la Fig. 2.1-3 se ve un ejemplo donde se han seccionado las actividades de construcción de varios pozos sucesivos. Se nota que la actividad Perforación tiene un peso importante.

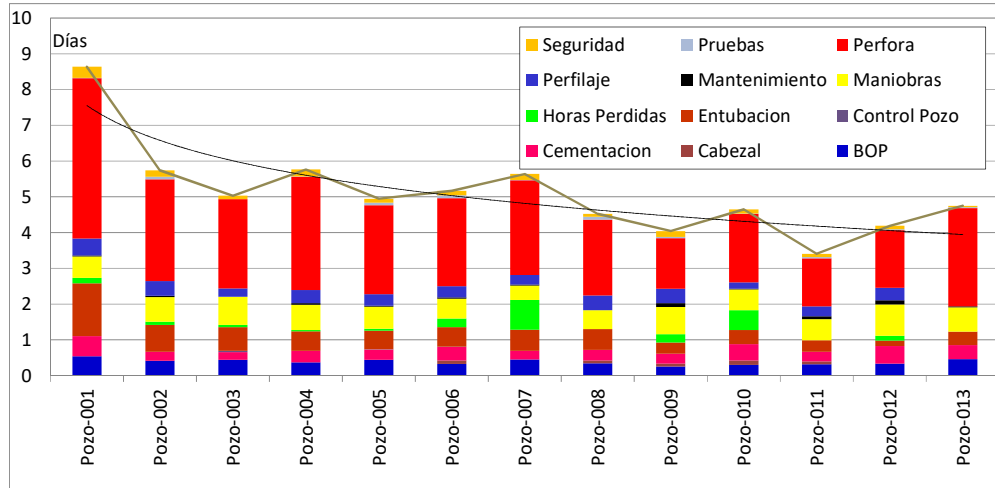


Fig. 2.1-3 Ejemplo de actividades de construcción de pozos

Los tiempos de construcción de pozos particionados según las actividades principales. La perforación pura (en rojo) ocupa los mayores tiempos.

En la Tabla 2.1-2 se ven los tiempos de todos los pozos bajo estudio, clasificados por Fase y por Actividad. Entre todas las operaciones se destaca la cantidad de tiempo ocupado por la perforación pura. En las Fases 300 y 600 el tiempo de perforación pura ocupa el 47.8% del tiempo total, es decir, casi la mitad del tiempo. En la Fase 600 la perforación pura ocupa el 53.1% del tiempo, es decir, más de la mitad del tiempo. Otras tareas u ocupaciones que le siguen en importancia son Maniobras (13.7% del tiempo en Fases 300 y 600 y 15.6% del tiempo en Fase 600), Tiempo Perdido (8.2% del tiempo en Fases 300 y 600 y 7.8% del tiempo en Fase 600).

Tabla 2.1-2 Comparación de tiempos Fase 300, Fase 600 y Total

	TOTAL FASE 300 (hrs)		TOTAL FASE 600 (hrs)		TOTAL FASES 300-600	
BOP	2649	20.5%	337	1.0%	2986	6.6%
Cabezal	43	0.3%	664	2.1%	707	1.6%
Cementacion	1618	12.5%	1407	4.4%	3024	6.7%
Control Pozo	16	0.1%	35	0.1%	51	0.1%
DTM	6	0.0%	12	0.0%	18	0.0%
Entubacion	1126	8.7%	2256	7.0%	3382	7.5%
Horas Perdidas	1206	9.4%	2520	7.8%	3726	8.2%
Maniobras	1171	9.1%	5032	15.6%	6202	13.7%
Mantenimiento	182	1.4%	206	0.6%	388	0.9%
Perfilaje	18	0.1%	2086	6.5%	2105	4.7%
Perforación Pura	4453	34.5%	17169	53.1%	21622	47.8%
Pruebas	17	0.1%	35	0.1%	51	0.1%
Seguridad	388	3.0%	553	1.7%	941	2.1%
Total general	12890	100.0%	32312	100.0%	45202	100.0%

Este estudio comienza con una exploración de la tasa de construcción (Rate Of Construction o ROC) total de la construcción de los pozos, o sea, de las Fases 300 y 600.

Luego, continúa con una exploración de la ROP de la Fase 600.

Finalmente el estudio se concentra en el análisis de la ROP de la Fase 600. La actividad Perforación Pura en la Fase 600 fue seleccionada para este estudio por dos razones. Además de su importancia (en cuanto a tiempos consumidos), los registros de esta actividad contaban con datos abundantes y adecuados.

2.1.4 Origen de los datos

En este estudio se utilizaron los datos de una base de datos de perforación DIMS (Drilling Information Management System) (Landmark, 2000), que es el sistema más difundido en la industria para administrar registros y datos de perforación.

En la Tabla 2.1-3 se presenta una lista de posibles fuentes de datos típicos de la industria. La existencia de diversas fuentes permitiría cotejar datos en casos dudosos y puede proveer los datos en casos de faltas.

Tabla 2.1-3 Fuentes de datos

GRUPO VARIABLES	DIMS	Compañía Perforación	Compañía Lodos	Compañía Trépanos	Compañía Registros	Informes Internos
Grupo Equipo	X	X				
Grupo Parámetros de Perforación	X	X				X
Grupo Fluidos de Perforación	X		X			X
Grupo Pozo	X		X		X	X
Grupo Herramienta	X		X			
Grupo Trépano	X			X		
Grupo Geología					X	X
Problemas	X	X	X	X	X	X

La información del presente estudio se tomó de una base de datos de perforación de un área petrolera. Las observaciones corresponden a 382 perforaciones realizadas entre 2004 y mediados de 2007. Como esta es un Área ya casi completamente desarrollada, se encuentra en la fase final de su explotación. En el Área hay una gran cantidad de pozos y una gran experiencia en perforación. Las observaciones ya estaban registradas en la base de datos y ésta ya tiene una estructura establecida, de manera que no ha sido posible elaborar ningún plan de registros ni estructurar a priori la base de datos.

La información básica se obtuvo en cuatro reportes.

Reporte Perfora: Este reporte consta de casi 35.000 registros. Cada registro incluye los detalles operacionales y tiempos ocupados en la construcción de los 382 pozos perforados en el área durante el tiempo mencionado.

Reporte Rig: Consta de casi 1.500 registros relacionados a la identificación y características de los equipos que operaron en la perforación de los 382 pozos.

Reporte Mud: Compuesto por casi 3.000 registros, abarca la información de las características y propiedades de los fluidos de perforación utilizados.

Reporte Bits: Consta de casi 6.000 registros relacionados a la descripción, características y detalles operativos de los trépanos utilizados.

La primera tarea sobre estos reportes fue verificar la calidad, exactitud, extensión e integridad de los datos. Tal como anticipan varios autores ya citados, esta tarea resultó ser extremadamente demandante. Luego, estos reportes se integraron y redujeron a una sola tabla en donde cada línea, registro u observación corresponde a un pozo (en análisis multivariado corresponde a un *individuo*) y donde cada columna implica alguna propiedad o característica observada o medida durante la perforación. Debido a la detección de circunstancias particulares que desmerecían la presencia de algunos casos en la nómina, como ser paros de actividad debido a prolongados conflictos gremiales, problemas especiales encontrados durante la perforación (presiones excepcionales, pérdidas de circulación, pescas prolongadas), pozo exploratorio, la tabla de trabajo resultó ser una matriz de 343 filas (pozos observados) y 38 columnas (variables), como se ve en la Tabla 2.1-4.

Luego, las variables se consolidaron en las siguientes dimensiones:

- ✓ Performance
- ✓ Equipo perforador
- ✓ Tipo de Pozo
- ✓ Lodo
- ✓ Trépanos
- ✓ Condiciones de Perforación

2.1.5 Unidades observacionales y variables relacionadas a la perforación

En esta investigación se consideraron las variables típicamente usadas en los análisis tradicionales de perforación. Son las variables que se reconocen en la técnica como las que tienen un efecto importante en el comportamiento de la perforación.

Se definió que la Variable Respuesta es el Tiempo ocupado en ejecutar cada tarea o cumplir algún proceso durante la realización del pozo. En diversos análisis este tiempo puede ser, por ejemplo, el tiempo total necesario para llegar a una cierta profundidad, el tiempo neto para perforar un tramo específico, el tiempo para realizar alguna operación determinada o el tiempo perdido en solucionar problemas imprevistos. En el análisis de la perforación de tramos, el tiempo total se considera en forma de tasa de construcción de pozo (Rate Of Construction - ROC). En el análisis de la perforación pura, este tiempo se convierte en velocidad lograda para perforar un tramo específico en forma de tasa de penetración de pozo (Rate Of Penetration - ROP). Tanto el Tiempo (ocupado en perforar cierta longitud de pozo) como las Tasas son expresiones de la Performance de Perforación, que es la característica de interés en esta investigación, porque el tiempo es un indicativo de la performance de la perforación del pozo.

Las variables que caracterizan la tasa de construcción son la longitud perforada en el pozo o profundidad medida verdadera (variable TMD, en metros), el tiempo ocupado en la perforación (variable DiasF36, días) y la relación entre ambas o sea, la tasa de construcción (variable ROC36, m/hr).

Las variables que caracterizan la tasa de penetración son el intervalo o longitud perforada en la Fase 600 (variable Drill600, m), el tiempo ocupado en la perforación (variable HrsF600, hr) y la relación entre ambas o sea, la tasa de penetración (variable ROP600, m/hr). El cuerpo de variables quedó integrado en 297 pozos perforados en Diámetro 8,75”.

❖ **Variables Respuesta**

Una variable respuesta es ROC36.

- ✓ ROC36: Tasa de construcción de pozo. Incluye todos los tiempos de todas las operaciones. Se expresa en metros perforados por hora (m/hr).

Otra variable respuesta es ROP600, pero como ésta es derivada de las variables Drill600 y HrsF600, eventualmente se pueden usar combinaciones de estas últimas.

- ✓ ROP600: Tasa de penetración en la etapa de perforación de la Aislación. Se expresa en metros perforados por hora (m/hr).
- ✓ Drill600: Intervalo de pozo perforado en la etapa de Aislación expresado en metros (m).
- ✓ HrsF600: Tiempo de perforación demandado por la etapa de Aislación expresado en horas (hr).

❖ **Variables Categóricas**

- ✓ Tipo: Se refiere a la trayectoria del pozo, ya sea vertical o dirigido.
- ✓ Rig: Designa al equipo que perforó el pozo.
- ✓ Diametro: Se refiere al diámetro de los trépanos usados en la perforación.

PARTE 2 – MATERIALES Y MÉTODOS

POZO								EQUIPO	
Pozo	Fin	X	Y	Z	Clasificacion	Tipo	Año	Contractor	Rig
Pozo-001	26/08/06	5231.00	8068.00	715.90	INYECTOR	DIRIGIDO	Año3	OILCO	OIL-01
Pozo-002	03/11/06	7376.00	13174.00	632.90	INYECTOR	DIRIGIDO	Año3	OILCO	OIL-01
Pozo-003	14/01/04	8580.00	13642.00	641.30	PRODUCTOR	DIRIGIDO	Año1	PETCO	PET-01

PERFORMANCE									
TMD	MBNM	HrsF600	Drill600	ROC600	ROP600	DiasF136	ROC136	DiasF36	ROC36
1188	472.10	34.50	1012.00	19.18	29.33	7.28	6.80	6.43	7.70
1010	377.10	43.25	894.00	17.79	20.67	6.04	6.97	5.17	8.15
975	333.70	60.75	863.00	12.69	14.21	6.75	6.02	4.40	9.24

LODO								
Compañía	Densidad	Viscosidad	VisPlast	PtoCed	Filtrado	pH	MBT	Potasio
BarCo	1168.13	41.00	11.18	9.81	6.00	8.00	24.51	53327.45
BarCo	1200.00	44.20	12.08	9.08	7.00	8.00	28.00	62488.43
BarCo	1147.60	44.42	12.97	10.05	11.01	8.16	25.82	42048.38

TREPANOS		PERFORACION							
Fabricante	Diametro	Presion	PresionMecha	Peso	RPMMin	RPM	Caudal	VelocidadJets	HHP
PerCo	8.75	1498.07	277.39	4.50	114.63	160.51	426.71	177.10	1.17
PerCo	8.75	1394.75	417.71	4.66	53.88	53.88	435.78	215.20	1.80
PerCo	8.75	1070.10	482.52	5.03	134.00	194.00	480.00	237.60	2.25

Tabla 2.1-4Perforación: Variables básicas

Las variables disponibles, clasificadas en grupos o dimensiones.

❖ **Variables Cuantitativas**

- ✓ Densidad: Peso unitario del fluido de perforación, expresada en gramos por litro (gr/lit).
- ✓ ViscPlast: Es la viscosidad plástica del fluido de perforación, en centipoises (cp).
- ✓ PtoCed: Es el punto cedente del fluido de perforación, expresado en libras por cada 100 pies cuadrados (libras/100 pies²).
- ✓ Filtrado: Volumen de fluido filtrado en una prueba API, expresado en centímetros cúbicos cada 30 minutos (cc/30 min).
- ✓ MBT: Prueba de azul de metileno según la norma API, expresada en kilogramos por metro cúbico (kg/m³).
- ✓ Potasio: Concentración de ion potasio en el fluido de perforación. Se expresa en miligramos por litro (mgr/lit).
- ✓ Presion: Presión de bombeo de fluido de perforación medido en la bomba en superficie, expresada en libras por pulgada cuadrada (psi).
- ✓ Peso: Peso aplicado sobre el trépano expresado en toneladas (ton).
- ✓ RPM: Es la velocidad de rotación del trépano, en revoluciones por minuto (RPM).
- ✓ Caudal: Es el caudal volumétrico bombeado por la bomba en superficie, expresado en galones por minuto (gal/min o gpm).

Un resumen se ve en la Tabla 2.1-5.

2.1.6 Colofón

Se realizarán exploraciones sobre la variable tasa de construcción en las Fases 300 y 600 (ROC36) y sobre la tasa de perforación en la Fase 600 (ROP600).

Específicamente, se analizará la tasa de penetración en la Etapa Perforación de la Aislación con las variables de la Actividad de Perforación pura aplicando técnicas estadísticas multivariadas, que es el propósito de esta investigación.

Tabla 2.1-5 Variables de la actividad Perforación

GRUPO	NOMBRE	TIPO	UNIDADESCATEGORÍA
Performance	HrsF600	Continua	hr
Performance	Drill600	Continua	m
Performance	ROC36	Continua	m/hr
Performance	ROP600	Continua	m/hr
Pozo	Tipo	Categórica	Dirigido; Vertical
Equipo	Rig	Categórica	OIL-01; PET-01; PET-02; PET-03
Trépanos	Diametro	Categórica	8 ½"; 8 ¾"
Lodo	Densidad	Continua	gr/lt
Lodo	ViscPlast	Continua	cp
Lodo	PtoCed	Continua	lb/100 pies ²
Lodo	Filtrado	Continua	cc/30 min
Lodo	MBT	Continua	kg/m ³
Lodo	Potasio	Continua	mgr/lt
Perforación	Presión	Continua	psi
Perforación	Peso	Continua	ton
Perforación	RPM	Continua	1/min
Perforación	Caudal	Continua	gpm

Capítulo 2.2 Métodos

2.2.1 Introducción

La base de datos comprende la actividad de perforación de 3 años y medio en un Área petrolera. Dado que los datos eran pre-existentes, se desarrolló un análisis **observacional** de los tiempos de perforación. Se inició con una **exploración** de los datos. A los métodos tradicionales, en este estudio se incluyeron técnicas multivariadas. Inicialmente se analizó el estado de la técnica de perforación con el fin de saber si se encontraba en un estado “estable” o si estaba sufriendo procesos de mejoras (o desmejoras). Una condición de estabilidad permite que los análisis desarrollados no sean alterados por cambios transitorios.

El primer análisis del estado de aprendizaje en la técnica de perforación se orientó a determinar:

- Si la técnica de la curva de aprendizaje podía ser de utilidad
- Si la perforación había alcanzado un nivel C_3 estable, que sería el límite operativo para las condiciones presentes (Peterson, y otros, 1995). Si se verifica que la técnica tiende a mantenerse estable en un nivel C_3 alcanzado, entonces se puede continuar con el análisis o estudio del estado actual de la técnica de perforación.

Se estudiaron los tiempos totales de perforación de los pozos seleccionados. El tiempo total de perforación corresponde a los tiempos de las Fases 300 y 600.

2.2.2 Curva de aprendizaje

Para dilucidar las primeras dudas planteadas, se analizaron los datos con la curva de aprendizaje. Las formas típicas de la curva de aprendizaje están vinculadas a los modelos no lineales. Se corrió un modelo relacionando el tiempo de perforación de los 343 pozos seleccionados con su número de orden de perforación.

El modelo seleccionado para realizar los ajustes es la función asintótica con parámetro exponencial, cuya expresión es:

$$E(T_{(n)}) = \alpha - \beta e^{-\gamma n} \quad \text{Ec. 2.2-1}$$

donde T representa el tiempo de perforación (en días) y n es número de orden de perforación de cada pozo.

El modelado requiere estimaciones iniciales de los parámetros α , β y γ . La propuesta de Crawley (Crawley, 2007) es graficar los datos y luego poner el modelo en sus límites; es

decir, donde n vale cero o infinito. Para $n=0$, es $e^{-n}=1$, de modo que $y=\alpha-\beta$. Con $n=\infty$, es $e^{-n}=0$, lo que deja $y=\alpha$. Así, el los valores iniciales propuestos resultan: El valor asintótico en la graficación es α inicial y el valor asintótico menos el intercepto se asume como β inicial. El valor de γ inicial se estima con la pendiente previa en la zona de crecimiento de la curva.

2.2.3 Exploración de latasa de construcción

2.2.3.1. Normalidad

Se analizó inicialmente la tasa de construcción total del pozo mediante gráficos de histograma, densidad de la variable y normalidad (qqplot). Para verificar si la variable sigue una distribución normal, se aplicó una prueba de Shapiro-Wilks.

2.2.3.2. Gráficos de caja y enrejados

Inicialmente, la tasa de construcción (variable ROC36) se analizó con gráficos de caja y de enrejados, clasificados por distintos factores. Los factores aplicados son:

- ✓ Año (Año1, Año2, Año3, Año4)
- ✓ Rig (OIL-01, PER-01, PER-02 y PER-03)
- ✓ Tipo de pozo (DIRIGIDO, VERTICAL)
- ✓ Diámetro de Trépano (de 8 3/4" o de 8 1/2").

2.2.3.3. ANOVA de la tasa de construcción

Se realizó un gráfico de diseño para apreciar las medias de la variable respuesta para los distintos niveles de cada factor.

Luego, se verificó el efecto de las variables categóricas Tipo, Rig, Año y Diametro de forma numérica. A continuación se ve el modelo de análisis de varianza para este sistema. En este modelo se descartó el factor Diametro debido a que el gráfico de enrejados reveló la inexistencia de datos en algunos Rigs y Años.

$$C_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \varepsilon_{ijkl} \quad \text{Ec. 2.2-2}$$

$$\varepsilon_{ijkl} \sim N(0, \sigma)$$

Donde:

C_{ijkl}	ROC36 del l -ésimo Pozo
μ	Media general
α_i	Efecto de Año
β_j	Efecto de Tipo de Pozo

γ_k	<i>Efecto de Rig</i>
$\alpha\beta_{ij}$	<i>Interacción Año-Tipo</i>
$\alpha\gamma_{ik}$	<i>Interacción Año-Rig</i>
$\beta\gamma_{jk}$	<i>Interacción Tipo-Rig</i>
$\alpha\beta\gamma_{ijk}$	<i>Interacción Año-Tipo-Rig</i>

y con:

$i =$ Número o índice de Año	1 - 4
$j =$ Número o índice de Tipo	1 - 2
$k =$ Número o índice de Rig	1 - 4

2.2.3.4. Comparación según variables categóricas

Se analizó el efecto de las variables categóricas en detalle con procedimientos de comparaciones múltiples en caso de cuatro factores y con pruebas t en caso de dos factores.

Los procedimientos de comparaciones múltiples permiten comparar todos los pares de los niveles de un factor mientras se mantiene el nivel de significancia nominal como valor seleccionado y produciendo intervalos de confianza ajustados para las diferencias de medias. Uno de estos procedimientos son las *diferencias significativas honestas* propuestas por Tukey(Tukey, 1953).

2.2.4 Exploración de la tasa de penetración

Se realizó una exploración de la variable respuesta tasa de penetración de pozo en la Fase 600 (ROP600).

2.2.4.1. Distribución de la tasa de penetración

Se consideró la posibilidad de asimilar la distribución de ROP600 a una distribución Gamma antes que a una distribución normal.

Además de las comparaciones gráficas, se realizó una prueba de Kolmogorov-Smirnov para probar la bondad de ajuste de una distribución gamma con los datos reales disponibles. En la prueba inicial se utilizaron los parámetros teóricos, tomando:

$$\begin{aligned} \text{Parámetro de forma} &= \left(\frac{\bar{x}}{\bar{s}}\right)^2 \\ \text{Parámetro de tasa} &= \bar{s}^2 / \bar{x} \end{aligned}$$

A fin de mejorar el ajuste de los parámetros, se corrió una aproximación iterativa y además, se realizó una estimación con el paquete `riskDistributions` de R.

Como la variable bajo análisis es una tasa de penetración que no puede tomar valores menores que cero, esta condición invalida las distribuciones de la familia normal o gaussianas y le aporta crédito a la distribución Gamma.

2.2.4.2. Dispersiones cruzadas

Una visión adicional del comportamiento de la ROP600 según los factores, se percibe mediante la graficación enrejada con la cual se pueden observar las relaciones entre variables y grupos de observaciones.

Se revisaron las variables relacionadas a ROP600 de acuerdo a dos dimensiones importantes: Lodo y Perforación. Se inició la exploración por separado con gráficos cruzados.

En caso que la distribución de ROP600 no sea normal, se aplica el coeficiente de correlación por rangos de Spearman, que es un indicador no paramétrico del grado de asociación entre dos variables. Este coeficiente evalúa si la relación entre dos variables es una función monótona. Si una de las variables es función monótona perfecta de la otra, el coeficiente resulta 1 ó -1. El signo indica si esa relación es creciente (+) o decreciente (-). A los efectos visuales, el tamaño de letra del coeficiente es proporcional a su importancia: Coeficientes mayores se presentan con letra más grande.

2.2.4.3. Análisis de componentes principales sobre ROP600

Se realizó un análisis de componentes principales sobre con la variable ROP600 y las variables continuas de perforación y lodo. Las variables consideradas fueron: ROP600, Peso, RPM, Caudal, Filtrado, Densidad, ViscPlast, PtoCed, Presión, Potasio y MBT. Como estas variables tienen distintas unidades, fueron estandarizadas. El interés de este análisis se centra en las relaciones de las variables con la variable respuesta ROP600, observando si los resultados concuerdan con los principios ingenieriles de las técnicas de perforación.

2.2.5 Agrupamiento de pozos

La exploración previa alienta la aplicación de técnicas de agrupamientos sobre la matriz de pozos y variables seleccionadas. Además de descubrir eventuales estratificaciones, el ingeniero de perforación podrá percibir distintos comportamientos. Este conocimiento permitirá gestionar la presupuestación, la operación y los recursos necesarios.

El agrupamiento es una continuidad de la exploración previa, sólo que en este caso se trata de **entender lo que los datos puedan decir** (Kaufman, y otros, 2005). Una de las

formas de estudio es tratar de **agrupar los datos de manera “natural”**(Sneath, y otros, 1973), si es posible o si los datos lo admiten.

Con estos agrupamientos se logró “sintetizar” en cierto grado los datos, lo cual permitió organizar mejor el proceso de análisis de modelos posterior.

2.2.5.1. Introducción

En R existen varios programas para realizar agrupamientos y clasificaciones. Hay algunos programas dentro del paquete stats y también algunos paquetes específicos de agrupamientos como el paquete cluster y el mclust02, entre otros. En este estudio se utilizó exclusivamente el paquete cluster(Maechler, y otros, 2005). Si bien aquí no se han comparado las diversas alternativas disponibles, se encontró que el paquete cluster alberga las funciones y capacidades apropiadas y suficientes para el presente caso, además de contar con adecuada y conveniente documentación.

La primera actividad fue obtener la matriz de disimilaridades. Luego, se realizaron exploraciones con el programa de particiones alrededor de medoides y se exploraron las clasificaciones jerárquicas.

2.2.5.2. Distancias

Con la tabla de casos y variables se preparó la correspondiente tabla de distancias entre objetos, que en este caso son los Pozos. Considerando que entre las variables se contaba con una combinación de variables nominales (Tipo, Rig) y numéricas (ROP600, Densidad, ViscPlast, PtoCed, Filtrado, MBT, Potasio, Presion, Peso, RPM y Caudal), se determinaron las distancias de Gower con el programa daisy del paquete cluster.

A la matriz de distancias obtenida se realizaron agrupamientos particionales y jerárquicos.

Las particiones se realizaron con el programa de particiones alrededor de medoides. Las particiones se realizaron con el programa pam del paquete cluster sobre la tabla de disimilaridades. Una de las virtudes de este algoritmo de agrupamiento es que reporta el valor de “silhouette” que es una medida de la calidad del agrupamiento y con la cual es posible hacer una estimación del número óptimo de grupos para los datos bajo análisis.

2.2.5.3. Agrupamiento particional

Número de grupos óptimo

El método propuesto para la optimización del número de grupos consiste en evaluar la función pam para todo número de grupos, desde $k=2$ hasta $k=\text{óptimo} + 1$ y comparar los

Anchos de Silueta Medio de cada agrupamiento. Se grafican los Anchos de Silueta y se observa si hay un k óptimo, con mayor Ancho de Silueta Medio.

Agrupamientos con programa pam

Se realizaron agrupamientos con el programa pam del paquete cluster en base a la tabla de disimilaridades. Los grupos se grafican sobre los dos primeras coordenadas de un escalado multidimensional aplicado a la matriz de disimilaridades (Pison, y otros, 1999). Eventualmente, es posible realizar representaciones tridimensionales en base a las 3 primeras coordenadas.

2.2.5.4. Visualización por matriz ordenada

En apoyo de la graficación de los grupos, se aplicó el ordenamiento de la matriz de disimilaridades. Este es exclusivamente un método de visualización que facilita la evaluación de la calidad de los agrupamientos.

2.2.5.5. Análisis de los grupos

Se analizó la calidad de los grupos con los anchos medios de siluetas.

Además, considerando que la disimilaridad máxima, la disimilaridad media y el diámetro son indicativos de la compacidad o cohesión del grupo, se pueden comparar con la separación. Esto da una idea de la fortaleza de la estructura grupal del conjunto de datos.

2.2.5.6. Agrupamientos en árboles jerárquicos

Siguiendo los consejos de Kaufman y Rousseeuw de probar y comparar diversos métodos de agrupamiento o de organización, se exploró la alternativa de las clasificaciones jerárquicas.

Proceso de análisis

El paquete cluster posee dos programas de clasificación jerárquica. Uno trata con el proceso aglomerativo (agnes) y el otro con la técnica divisiva (diana). Es decir, el primero empieza con tantos grupos como objetos tiene la matriz de distancias y termina con un solo grupo, mientras que la segunda técnica procede a la inversa.

En este caso se aplicaron ambos métodos y se compararon resultados buscando la mejor solución.

Métodos de clasificación

Se realizaron los seis árboles aglomerativos y el árbol divisivo aplicados a la matriz de disimilaridades de Gower y luego se hicieron comparaciones entre cada uno de los

métodos por medio de la correlación entre la matriz de distancias original y la matriz de distancias cofenéticas.

Selección del agrupamiento

Se analizaron y compararon los resultados de agrupamientos particionales y jerárquicos con el criterio de ingeniería de perforación, llegando a una solución de compromiso “natural”.

2.2.5.7. Diagnósticos

Como la variable que se desea investigar, en definitiva, es la tasa de penetración en la etapa de Aislación (ROP600) se procedió a hacer una evaluación de esta variable con respecto a los grupos conformados. Dado que la variable ROP600 responde a una distribución gamma, en este primer análisis se transforma la variable por su logaritmo, para obtener mayor normalidad. Se obtiene una comparación visual mediante un gráfico de caja, mientras que una comparación formal se logra con un análisis de varianzaa un modelo lineal sobre el logaritmo del promedio de las ROP600 de cada grupo.

A continuación se presenta el modelo.

$$P_{ij} = \mu + G_i + \varepsilon_{ij} \quad \text{Ec. 2.2-3}$$

$$\varepsilon_{ij} \sim N(0, \sigma)$$

Donde: P = log(ROP600); G = Grupo

2.2.5.8. Comparación múltiple

A fin de probar las diferencias de ROP600 entre pares de grupos, se aplicó una prueba de comparación de Tukey. Esta prueba también se conoce como Diferencia Significante Honesta(o Honest Significant Difference – HSD).Primero se corrió un anova a los datos y luego se hizo la comparación.

La Diferencia Significante Honesta de Tukey se basa en la distribución de rango studentizada (de Tukey), que surge de la siguiente manera: Se supone de tener X_1, \dots, X_n que son i.i.d. $N(\mu, \sigma^2)$ y que el rango es $R = \max X_i - \min X_i$. Entonces R/σ^2 tiene la distribución de rango studentizada $q_{n,v}$ donde v es el número de grados de libertad usados para estimar σ .

Los Intervalos de Confianza para la Diferencia de Medias de Tukey son:

$$\hat{\alpha}_i - \hat{\alpha}_j \pm \frac{q_{1,df}}{\sqrt{2}} \text{erstd}(\hat{\alpha}_i - \hat{\alpha}_j) \quad \text{Ec. 2.2-4}$$

El error estándar de las diferencias es $erstd(\hat{\alpha}_i - \hat{\alpha}_j) = \hat{\sigma} \sqrt{1/n_i + 1/n_j}$ y df (grados de libertad) es $df = N - I$ donde N es el número de casos e I = número de parámetros.

Cuando los tamaños de muestras son muy diferentes, esta prueba puede ser muy conservadora ya que este método asume la peor situación, al considerar la máxima diferencia.

2.2.6 Modelos Lineales Generalizados

Tal como se indicara previamente, la posible relación de la tasa de penetración con las modalidades de trabajo detectadas, es motivo de investigación. Aquí se analiza el modelado de la variable tasa de penetración definida como variable (ROP600) o fraccionada en sus componentes, el tiempo de perforación (HrsF600) y la longitud del tramo perforado (Drill600).

La selección de la variable respuesta ROP600 ya fue justificada y se considerará que aparentemente responde a una distribución gamma. La esperanza de una variable gamma es positiva y su varianza es no-constante, porque depende del valor de su esperanza.

Debido a esto, el tratamiento corresponde a modelos con coeficiente de variación constante (McCullagh, y otros, 1989), (Davidian, 2007). Además, como se trata de una tasa impone algunas condiciones al modelado, por ejemplo, no tolera respuestas negativas. La variable en forma de tasa también define la forma del modelo. Por otro lado, el presente modelo es multiplicativo (Parr. 1.2.4), lo que propicia la introducción del enlace log (McCullagh, y otros, 1989).

En esta situación se consideró oportuna la aplicación de un modelo lineal generalizado a la variable tiempo de perforación (HrsF600), con distribución gamma, enlace log y con una variable offset del log de la variable tramo perforado (Drill600) con coeficiente unitario.

2.2.6.1. Ajustes sobre sub conjuntos

Una forma ideal de probar la performance del modelo es aplicándolo sobre un nuevo conjunto de datos. Si los datos disponibles son escasos, se puede usar el método de validación cruzada (Maindonald, y otros, 2003). Como en este estudio se cuenta con una considerable cantidad de casos, se recurrió a la división de la tabla de datos en dos conjuntos (entrenamiento y validación). Hay varias maneras de dividir los datos. Distintos

autores sugieren el uso de muestreo aleatorio simple, del estadístico PRESS, divisiones basadas en los tiempos (por ejemplo en secuencias o series de tiempo) o basadas en algoritmos específicos (Montgomery, y otros, 2004). En este estudio se eligió una muestra mediante la función *sample* de R. En el Análisis de Regresión Múltiple se recomienda contar con un mínimo de 5 observaciones y un óptimo de 20 observaciones por cada variable predictora en el conjunto de desarrollo del modelo, mientras que el grupo menor (de prueba) debe contar con por lo menos 20 observaciones (Hair, y otros, 1999). Ambas recomendaciones se cumplen con los conjuntos seleccionados en este estudio.

Para modelar se procedió a separar un subconjunto compuesto por el 80% de los datos disponibles. Se reservó el 20% remanente para realizar el correspondiente control, a fin de probar el modelo logrado. Como se tenía una tabla de 297 casos totales, el subconjunto de modelado quedó compuesto por 238 casos y el subconjunto de control, por 59 casos. La partición 80/20 es arbitraria, pero aseguró que hubiera abundantes casos en el conjunto de estimación (para lograr mejor ajuste) y que hubiera suficientes casos en el conjunto de control como para tener varios casos en cada grupo de factores.

2.2.6.2. Modelo con todas las variables

Se corrió inicialmente un modelado con todas las variables seleccionadas, donde las variables cualitativas intervienen como variables indicadoras (Montgomery, y otros, 1996),(Faraway, 2006).

Las variables cualitativas en este caso son Tipo (de pozo) y Rig (o equipo de perforación).

La variable indicativa para Tipo puede asumir dos valores: 0 en caso que Tipo sea “Dirigido” o 1 en caso que Tipo sea “Vertical”.

La variable indicativa para Rig es una matriz de contrastes como sigue:

<i>Rig</i>	<i>Var. Indicativa</i>	<i>P1</i>	<i>P2</i>	<i>P3</i>
<i>OIL-01</i>		0	0	0
<i>PET-01</i>		1	0	0
<i>PET-02</i>		0	1	0
<i>PET-03</i>		0	0	1

Donde P1, P2 y P3 son las variables indicadoras que corresponden a los equipos PET-01, PET-02 y PET-03, respectivamente (Montgomery, y otros, 2004). Si un pozo ha sido perforado por el Rig PET-02, P1 y P3 serán iguales a 0 y P2 será igual a 1. Si un pozo ha sido perforado por el Rig OIL-01, P1, P2 y P3 valen 0.

La selección de la distribución para modelar se basa fundamentalmente en el hecho de que la suma de ν variables aleatorias exponenciales independientes e idénticamente distribuidas, con tasa λ tiene distribución gamma (Faraway, 2006). También se dice que una distribución $Gamma(\alpha, \beta)$ es la suma de α distribuciones $Exp(\beta)$ (Vose Software, 2007).

Justamente, esto se relaciona al hecho de que la variable en estudio, la tasa de penetración, proviene de una expresión de forma exponencial.

El modelo resulta como sigue:

Predictor lineal:

$$\eta_i = X_i\beta \quad \text{Ec. 2.2-5}$$

$$H_i = \alpha + \beta T_i + \gamma_1 P1_i + \gamma_2 P2_i + \gamma_3 P3_i + \sum_{k=1}^k \delta_k V_{ki} + 1.0 * \log(D_i) \quad \text{Ec. 2.2-6}$$

con:

H_i	<i>Log de horas de perforación en la Fase 600 (HrsF600) del i-ésimo Pozo</i>
α	<i>Media general</i>
β	<i>Efecto de Tipo de Pozo</i>
T_i	<i>Tipo de Pozo del caso i: Dirigido= 0; Vertical= 1</i>
γ_1	<i>Efecto de Rig PET-01</i>
$P1_i$	<i>Rig del caso i. Si es PET-01: Ausente= 0; Presente= 1</i>
γ_2	<i>Efecto de Rig PET-02</i>
$P2_i$	<i>Rig del caso i. Si es PET-02: Ausente= 0; Presente= 1</i>
γ_3	<i>Efecto de Rig PET-03</i>
$P3_i$	<i>Rig del caso i. Si es PET-03: Ausente= 0; Presente= 1</i>
δ_k	<i>Coficiente de la k-ésima variable de lodo o perforación</i>
V_{ki}	<i>Valor de la variable k-ésima de lodo o perforación del i-ésimo Pozo</i>
D_i	<i>Tramo perforado en la Fase 600 (Drill600) del i-ésimo Pozo</i>

y con:

i	<i>Número o índice de Pozo</i>	$1 - 238$
k	<i>Variable lodo o perforación</i>	$1 - 10$

Función de enlace: Tal como se expresó en la presentación del modelo, el enlace es log.

$$\eta_i = g(\mu_i)$$

$$\eta_i = \log(\mu_i)$$

$$\eta_i = \log(\text{HrsF600}_i)$$

De esta manera, es:

$$\mu_i = \exp(\eta_i)$$

$$\text{HrsF600}_i = \exp(H_i)$$

donde:

HrsF600_i Tiempo de perforación en la Fase 600 (en horas) del i -ésimo Pozo

El número de pozos corresponde a la cantidad de casos conservados en el sub conjunto seleccionado para el modelado, o sea 238 casos. Las variables continuas son las variables operativas de lodo y perforación que son 10 en total.

Luego de los análisis, el modelo lineal generalizado propuesto es el siguiente:

Predictor lineal:

$$\eta_i = X_i\beta \quad \text{Ec. 2.2-7}$$

$$H_i = \alpha + \beta T_i + \gamma_1 P1_i + \gamma_2 P2_i + \gamma_3 P3_i + \delta_{pc} PC_i + \delta_w W_i + \delta_q Q_i + 1.0 * \log(D_i) \quad \text{Ec. 2.2-8}$$

con:

H_i Log de tiempo de perforación en la Fase 600 (HrsF600) del i -ésimo Pozo

α Media general

β Efecto de Tipo de Pozo

T_i Tipo de Pozo del caso i . Si es: Dirigido= 0; Vertical= 1

γ_1 Efecto de Rig PET-01

$P1_i$ Rig del caso i . Si es PET-01: Ausente= 0; Presente= 1

γ_2 Efecto de Rig PET-02

$P2_i$ Rig del caso i . Si es PET-02: Ausente= 0; Presente= 1

γ_3 Efecto de Rig PET-03

$P3_i$ Rig del caso i . Si es PET-03: Ausente= 0; Presente= 1

δ_{pc} Coeficiente de variable de lodo PtoCed

PC_i Valor de la variable de lodo PtoCed del caso i .

δ_w Coeficiente de la variable de perforación Peso

W_i Valor de la variable de perforación Peso del caso i .

δ_q Coeficiente de la variable de perforación Caudal

Q_i Valor de la variable de perforación Caudal del caso i .

D_i Tramo perforado en la Fase 600 (Drill600) del caso i .
 $HrsF600$ Tiempo de perforación en la Fase 600 del caso i

y con:

$i =$ Número o índice de Pozo $1 - 238$
 $k =$ Variable lodo o perforación $1 - 10$

Función de enlace: Tal como se expresó en la presentación del modelo, el enlace es log.

$$\eta_i = g(\mu_i) = \log(\mu_i) = \log(HrsF600_i)$$

De esta manera, es:

$$\mu_i = \exp(\eta_i)$$

$$HrsF600_i = \exp(H_i)$$

donde:

$HrsF600_i$ Tiempo de perforación en la Fase 600 (en horas) del i -ésimo Pozo

2.2.7 Regresión por Redes Neuronales Artificiales

En esta aplicación, se utilizó un modelo tipo Perceptrón Multicapa consistente de una capa de entrada, una capa oculta y una capa de salida.

Las neuronas de entrada corresponden a las variables del conjunto de datos. Los factores Tipo y Rig se presentaron como variables dummy.

El número de neuronas en la capa oculta se seleccionó inicialmente en base a las reglas generales sugeridos y luego se ajustó mediante pruebas.

Debido a la variedad de unidades de medida y rangos de cada variable, las tablas de datos se escalaron para obtener uniformidad.

2.2.7.1. Número de neuronas en la capa oculta

Para la mayoría de los problemas, se considera que podría ser

$$\frac{P}{2N} < n < \frac{2P}{N}$$

Donde P =número de patrones de entrenamiento

N = número de neuronas de entrada

n = número de neuronas en la capa oculta

2.2.7.2. Entrenamiento y de validación de la RNA

El grupo de entrenamiento alojó 238 casos (80%) y el grupo de validación sumó 59 casos (20%). El grupo de entrenamiento se utilizó para obtener los resultados de la red y el de validación se usó para verificar la calidad de los pesos obtenidos.

La capa de salida consta de una sola neurona, que representa la variable respuesta ROP600.

Por ser la salida una variable continua, se eligió como función de activación, la función lineal

2.2.7.3. Aplicación de RNA

Se usó el paquete nnet donde la función nnet se ocupa de analizar redes neuronales artificiales.

Debido a que el algoritmo de aprendizaje busca un valor de error mínimo mediante la aplicación de gradientes descendentes, puede ser que se detenga en un mínimo local de error. Para minimizar este inconveniente, cada corrida se repitió varias veces, para dar oportunidad al algoritmo de hallar el mínimo global.

El muestreo se realizó con la función sample con una semilla para poder repetir el mismo muestreo en varias corridas. Cada corrida sirvió para evaluar el comportamiento de la red frente a distintos números de neuronas en la capa oculta.

La calidad de los resultados se evaluó con el coeficiente de determinación R^2 .

2.2.8 Regresión por Particiones Recursivas

Los métodos de los árboles de decisión se pueden usar tanto en clasificación como en regresión (Maindonald, y otros, 2003). En este estudio se analizan las relaciones entre variables y se realizan estimaciones predictivas. Los árboles de regresión son un compromiso entre los modelos lineales y los modelos no paramétricos (Faraway, 2006). Se aplicó el método de particiones recursivas (Recursive partitioning, paquete rpart de R) a la matriz de datos y se realizó un análisis de los resultados.

PARTE 3 - RESULTADOS Y DISCUSIÓN

Capítulo 3.1 Exploración de datos

3.1.1 Introducción

Curva de aprendizaje

Se inició la exploración con una clásica curva de aprendizaje aplicada a los tiempos de construcción de pozos, según el siguiente modelo.

$$t \sim \alpha + \beta * \exp(-\gamma * n) \quad \text{Ec. 3.1-1}$$

donde

t = tiempo de perforación

n = el número secuencial del pozo

α = tiempo óptimo idealizado de perforación

β = diferencia entre el tiempo óptimo ideal y el tiempo para perforar el primer pozo

γ = la velocidad o efectividad de aprendizaje

Los parámetros obtenidos resultaron ser:

	Estimado	Error Std.	Valor t	Pr(> t)
α	5.738	0.105	54.60	< 2e-16
β	-1.927	0.416	-4.63	5.27E-06
γ	0.023	0.008	2.68	0.00774

El ajuste para las observaciones se puede ver en la Fig. 3.1-1.

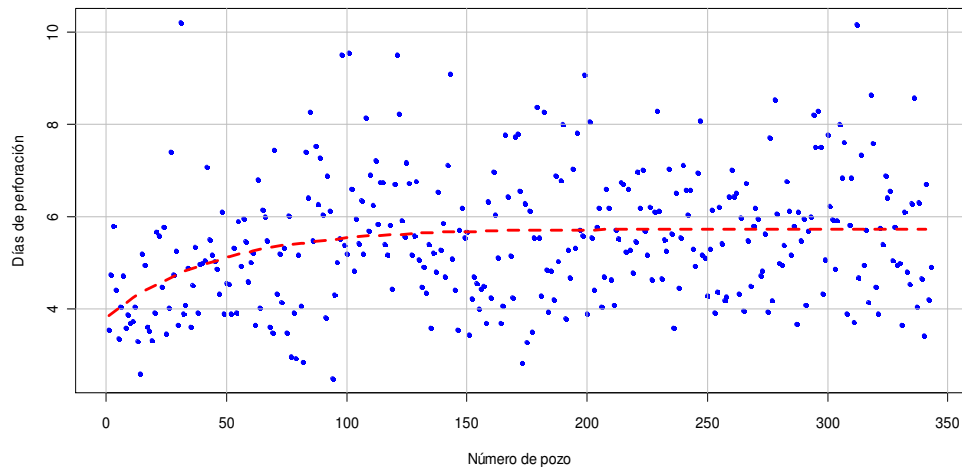


Fig. 3.1-1 Curva de aprendizaje - Todos los pozos

La curva de aprendizaje aplicada a los tiempos de construcción de los pozos en estudio. Se observa una gran dispersión y una aparente estabilidad en el proceso de perforación.

En este modelo se observa que la curva se hace asintótica al valor del parámetro α que es la esperanza del modelo para infinito número de pozos. Al incrementar el número de pozos el término negativo del modelo tiende a anularse. Además, la dispersión de datos es importante

En estas condiciones, la curva de aprendizaje tiene una utilidad muy limitada. Por un lado, la estabilización se logra después del pozo número 100 y además, la gran dispersión hace que las predicciones de pozos nuevos no sean muy precisas.

La dispersión presente remite a las situaciones afrontadas por Noerager y otros (Noerager, y otros, 1987). En el caso citado, la situación fue manifestada, pero no resuelta. En esta tesis se da una solución a este problema mediante agrupamientos que funcionan como criterios de estratificación, lo cual reduce las dispersiones de cada grupo y permite modelos más precisos.

3.1.2 Tasa de construcción

Se analizó inicialmente la variable tasa de construcción total del pozo.

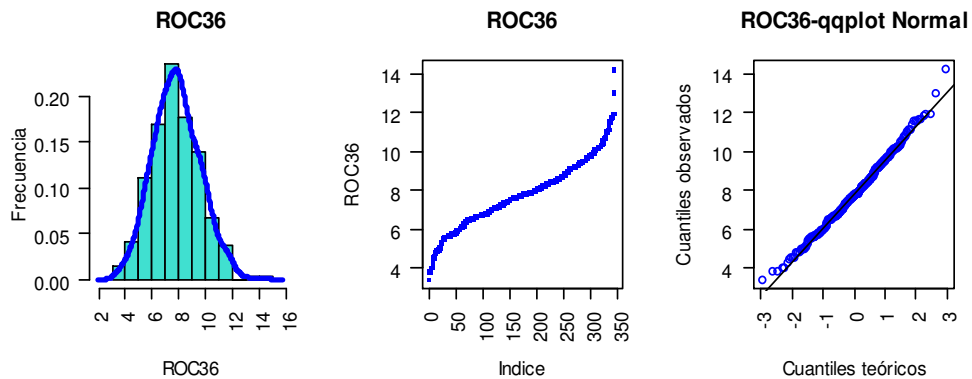


Fig. 3.1-2 ROC36-Tasa de construcción de pozo

Histograma de distribución de ROC36, un gráfico índice de los tiempos ordenados y gráfico de probabilidad normal.

Se representó ROC36 en el histograma de la Fig. 3.1-2, junto con la densidad de la variable y un gráfico de normalidad (qqplot). La prueba de normalidad de Shapiro-Wilks aporta un valor-p de 0.238, de manera que no se rechaza la hipótesis nula de normalidad.

Esto es coherente con el origen de los datos, los cuales son la **adición** de los tiempos de diversas operaciones.

3.1.2.1. Gráficos de caja con la variable ROC36

Gráficos de caja: Las relaciones de la tasa de construcción (ROC36) con las variables Año, Rig, Tipo (de pozo), Diámetro (de trépano) se ven en los gráficos de caja de la Fig. 3.1-3.

Los valores p de las pruebas de normalidad Shapiro-Wilks para la tasa de construcción según los grupos por Año, Rig y Tipo de Pozo resultaron todos no significativos, excepto el grupo del Rig PET-03, cuyo valor p es 0.044. Las variabilidades de la tasa de construcción a través de los agrupamientos por Año, Rig y Tipo de Pozo comparadas con pruebas de Levene presentan valores p en el orden de 0,01, mientras que la variabilidad de ROC frente a los grupos de Diámetro presentó un valor p de 0.400.

Gráficos de enrejados: En el gráfico de la Fig. 3.1-4 se ve el comportamiento de la tasa de construcción según las variables TMD, Año, Rig y Diámetro.

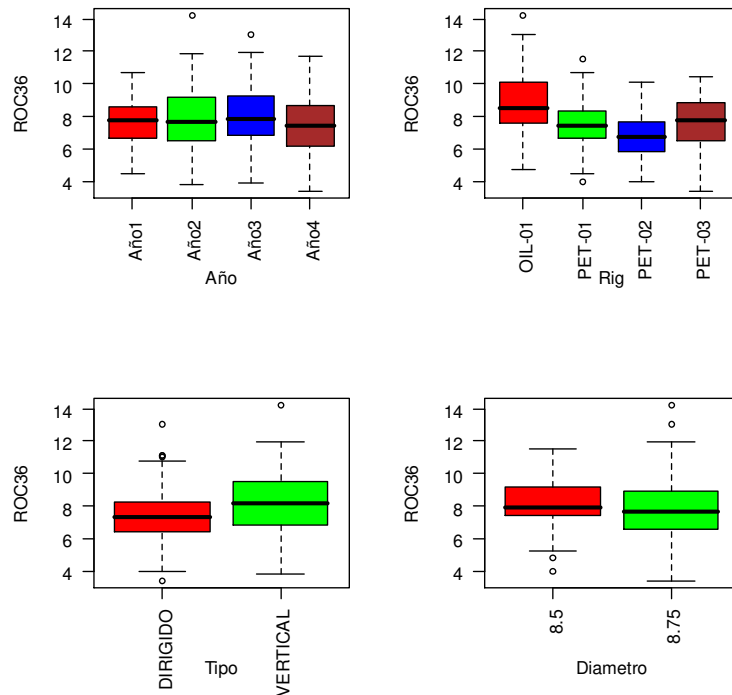


Fig. 3.1-3 ROC36-Tasa de construcción factorizada

Gráficos de caja de ROC36, según las variables categóricas Año, Rig, Tipo, Diámetro.

En este gráfico se ve que los diámetros de trépano de 8.5” corresponden solamente a algunos pozos perforados por un par de equipos en los dos primeros años. Se ha verificado que este diámetro se dejó de utilizar y a la fecha su análisis no reviste interés. Esto invalida la presencia de Diámetro como factor.

3.1.2.2. ANOVA de la tasa de construcción

En el gráfico de diseño de la Fig. 3.1-5 se pueden apreciar las medias de la variable respuesta para los distintos niveles de cada factor. En el gráfico se ve que los rangos más

amplios de medias corresponden a Rig y luego a Tipo. Año y Diametro tienen rangos menores, de manera que es posible que no sean tan influyentes sobre la tasa de construcción. Además, se observa que la media de Diametro 8.75” está muy cerca de la media general, confirmando que Diametro 8.5” no es tan influyente.

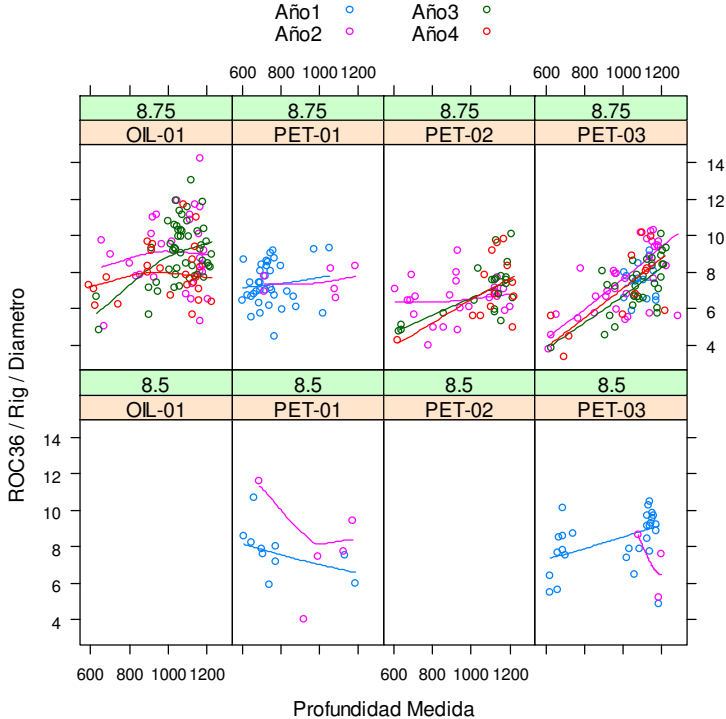


Fig. 3.1-4ROC36-Dispersiones cruzadas

ROC36 clasificada por las variables categóricas Rig, Diametro y Año contra profundidad de pozo.

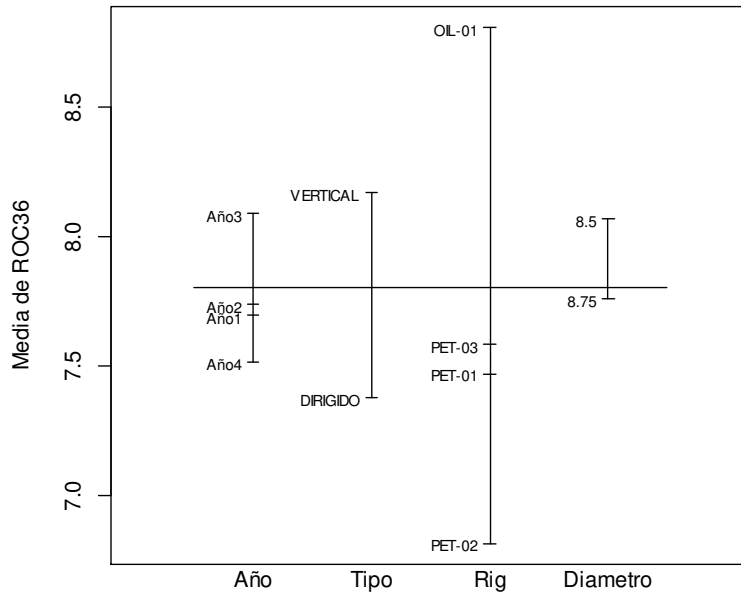


Fig. 3.1-5ROC36-Gráfico de diseño

La media general de ROC36 frente a las medias de los factores Año, Tipo, Rig, Diámetro.

A continuación se ve el análisis de varianza para este sistema.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Año	3	14.00	4.66	2.03	0.1103
Tipo	1	60.50	60.52	26.28	5.13E-07
Rig	3	184.40	61.46	26.69	1.99E-15
Año:Tipo	3	15.90	5.29	2.30	0.0775
Año:Rig	5	21.70	4.33	1.88	0.0973
Tipo:Rig	3	8.20	2.73	1.19	0.3145
Año:Tipo:Rig	5	12.20	2.44	1.06	0.3829
Residuals	319	734.50	2.30		

Las interacciones y el efecto de Año no son significativos, mientras que los factores Tipo y Rig sí son significativos.

En la Fig. 3.1-6 se ve el diagnóstico de supuestos donde no se observan anomalías en los residuales. El test de Shapiro-Wilks presenta un valor p de 0.088, entonces no se rechaza la hipótesis nula de que los residuales distribuyen normalmente.

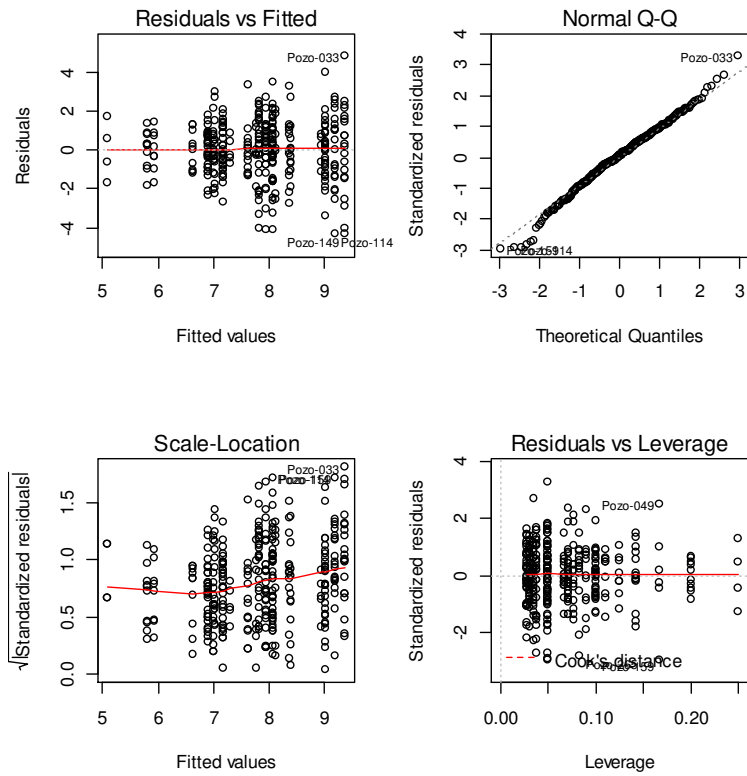


Fig. 3.1-6 Diagnósticos de supuestos-Anova de ROC36

Arriba izquierda: Residuales contra valores ajustados. No se notan anomalías graves.
 Arriba derecha: Grafico cuantil-cuantil de normalidad. Confirma el test Shapiro-Wilks.
 Abajo izquierda: Raiz de valores absolutos de errores y valores ajustados. Simil al primer gráfico.
 Abajo derecha: Residuales estandarizados contra apalancamientos y distancias de Cook.

3.1.2.3. Comparación según variables categóricas

En el análisis de la varianza resultaron factores significativos las variables Rig y Tipo. Las relaciones de estos factores con ROC36 se analizan a continuación.

ROC36 y Rig: Primero se observan los intervalos de confianza simultáneos de las diferencias de ROC36 entre los cuatro equipos.

Se detectaron diferencias significativas entre todos los pares de Rigs, excepto entre PET-01 y PET-03. En la Fig. 3.1-7 se ven los intervalos de confianza múltiples.

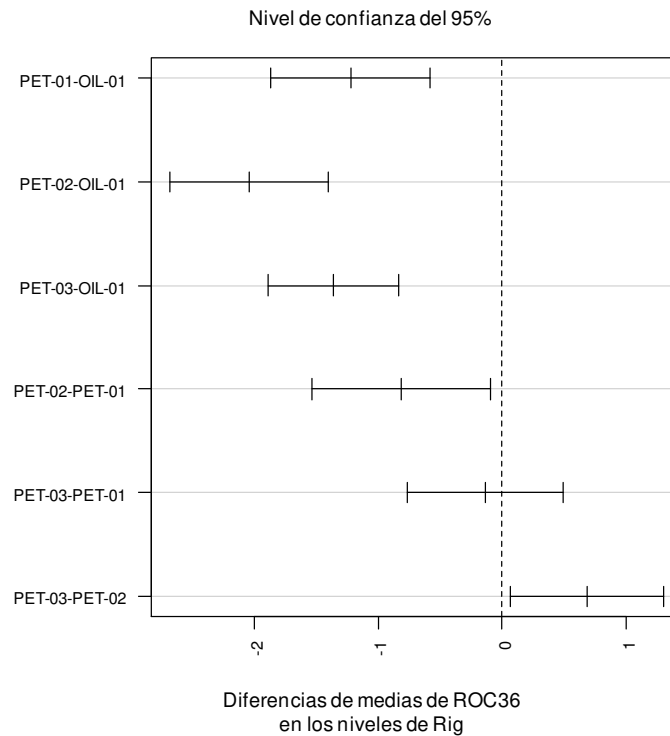


Fig. 3.1-7ROC36-Comparación de diferencias de medias por Rig

Intervalos de confianza de 95% de las diferencias de ROC36 entre los distintos equipos.

ROC36 vs Tipo: Se probó el comportamiento de ROC36 con Tipo. En la Fig. 3.1-8 se ven los qqplot. Las pruebas normalidad de Shapiro-Wilks de ROC36 para los pozos Verticales presentó un valor p de 0.112 y para los pozos Horizontales, de 0.572. Los correspondientes histogramas se ven en la Fig. 3.1-9.

3.1.2.4. Comportamiento de ROC36

En el análisis de ROC36 se ha podido apreciar:

- ✓ **Año:** No hay diferencias apreciables a lo largo del tiempo, es decir que los tiempos de perforación se pueden considerar estancados. Esto amerita revisar los procedimientos operativos de perforación.
- ✓ **Rig:** Se aprecian diferencias de ROC36 según el equipo de perforación. Se puede establecer un orden de efectividad.
- ✓ **Tipo de pozo:** Se aprecian diferencias de ROC36 según el tipo de pozo. Esto sería razonable, ya que los procedimientos de perforación direccional son más complejos y lentos que los de perforación vertical.

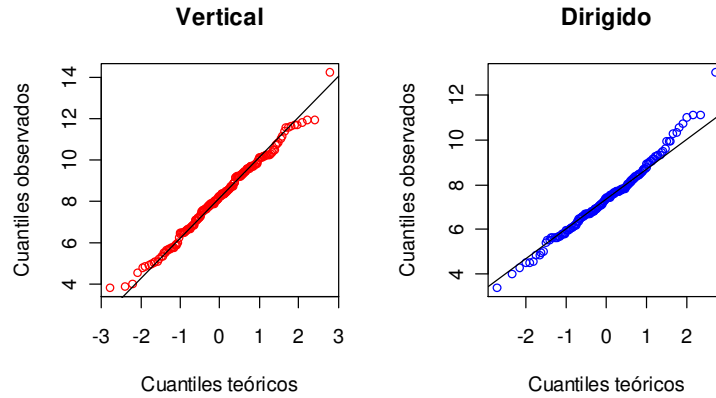


Fig. 3.1-8ROC36-Normalidad por Tipo de pozo

Gráficos cuantil-cuantil de normalidad de ROC36 de pozos Verticales y Dirigidos. Confirma el test Shapiro-Wilks.

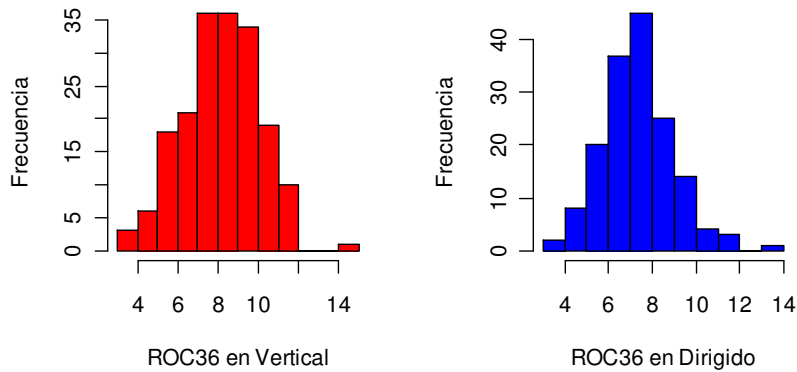


Fig. 3.1-9ROC36-Histogramas de ROC36 por Tipo de pozo

3.1.3 Tasa de penetración

A continuación se desarrolla el análisis de la actividad Perforación Pura en la Fase 600, seleccionada para este estudio.

En la Fig. 3.1-10a y b se presenta el histograma y la densidad de la variable ROP600. Se observa en ambos gráficos que la distribución de la variable es sesgada. En la Fig. 3.1-10 c y d se pueden comparar los qq-plots de las distribuciones normal y gamma de ROP.

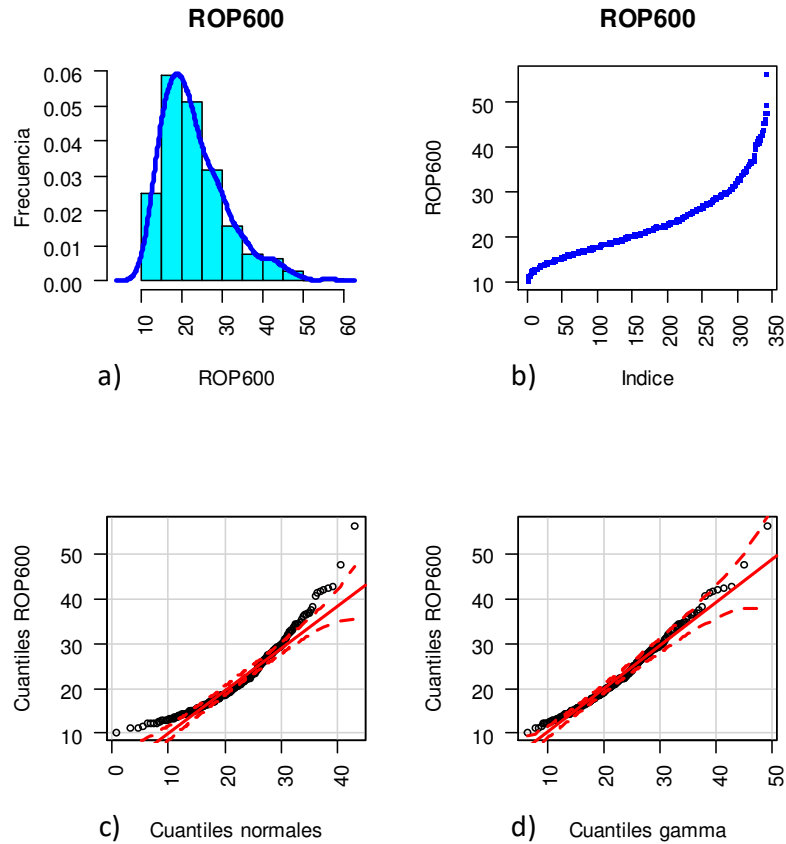


Fig. 3.1-10 Selección del tipo de distribución de ROP600

a) y b) Histograma de ROP600 y gráfico índice de los tiempos ordenados de ROP600.
Se observa que la distribución de la variable es sesgada.
c) y d) qq-plot como distribución normal y como distribución gamma de ROP.

Asumiendo una distribución gamma, se realizó una prueba de Kolmogorov-Smirnov inicial con los parámetros teóricos, tomando

$$\text{Parámetro de forma} = 9.268$$

$$\text{Parámetro de tasa} = 0.4229$$

el valor p resultó ser 0.1487, de modo que no se puede rechazar la hipótesis nula de igualdad en las distribuciones. Se observa mejor ajuste de la distribución gamma.

A fin de mejorar el ajuste, se corrió una aproximación iterativa que proporcionó como mejores parámetros:

$$\text{Parámetro de forma} = 10.79$$

$$\text{Parámetro de tasa} = 0.503$$

Estos nuevos parámetros suministran un valor p de 0.7234

Con el paquete `riskDistributions` de R se obtuvieron los siguientes parámetros:

Parámetro de forma = 10.43

Parámetro de tasa = 0.48

En la Fig. 3.1-11 se puede ver la comparación de los datos originales y de la distribución Gamma con los parámetros estimados.

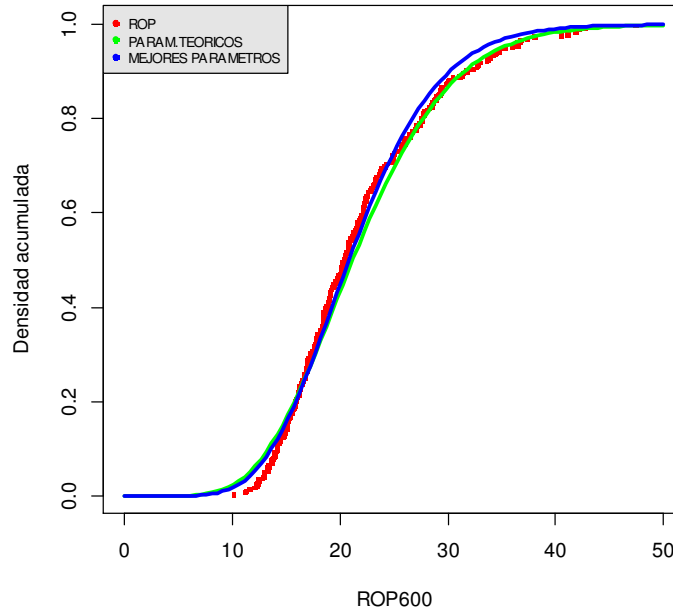


Fig. 3.1-11 Comparación ROP600-Gamma

Comparación de ROP600 original y como distribución Gamma con los parámetros estimados. Confirma el test de Kolmogorov.

3.1.3.1. ROP600 con Factores

Este gráfico se realizó sólo con los pozos perforados en 8.75”, que son los más numerosos y mejor repartidos entre los factores Tipo y Rig. Con la exclusión de los pozos perforados en 8.5”, los 343 pozos iniciales se reducen a 297 pozos.

En la Fig. 3.1-12 se ve la relación de ROP600 según la profundidad de pozo, discriminada por Tipo y Rig. Se observa que hay grupos de datos relacionados y que se dispone de buena cantidad de observaciones en cada grupo. Esto alienta a mantener como factores a las variables Tipo y Rig.

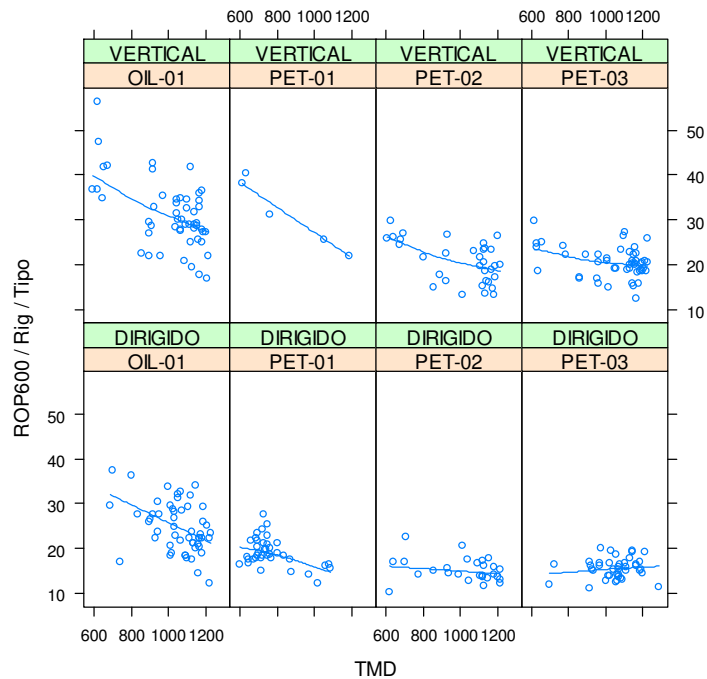


Fig. 3.1-12 Gráfico enrejado de ROP600 por Rig y Tipo

ROP600 clasificada por las variables categóricas Rig y Tipo contra profundidad de pozo.

3.1.3.2. Dispersiones cruzadas

Se realizaron los gráficos cruzados de acuerdo a dos dimensiones importantes: Lodo y Perforación.

Propiedades de Lodo

En la Fig. 3.1-13 se presenta el gráfico de dispersiones cruzadas con coeficientes de correlación donde es posible examinar las relaciones entre las variables del Lodo y la variable en estudio ROP600.

Las tres propiedades de viscosidad presentan coeficientes de correlación positiva relativamente altos entre sí (0.49, 0.51, 0.60). Esto llevó a considerar la posibilidad de tomar solamente alguna de esas variables como representativa de las propiedades de viscosidad del fluido. Se podía optar por Viscosidad, que es un indicador puramente empírico, o bien por las otras dos propiedades, Viscosidad Plástica y Punto Cedente, las cuales corresponden a un modelo de flujo que es el de plásticos de Bingham. A fin de preservar la integridad y coherencia técnica, se conservaron como variables representativas de las condiciones de flujo a la **Viscosidad Plástica** y al **Punto Cedente**.

Las propiedades químicas presentes son el pH y el contenido de Potasio. Entre ambas propiedades el coeficiente de correlación es de 0.53, que para las presentes condiciones es

relativamente alto. La base del lodo es un polímero que se activa con potasio para controlar las arcillas (evaluadas por el **MBT**), entonces se consideró que el contenido de **Potasios** es una variable relevante.

Otras variables que definen las características del lodo son **Densidad** y **Filtrado**.

Propiedades de Perforación

En la Fig. 3.1-14 Dispersiones cruzadas: ROP600 -características de perforación

se presenta el gráfico de dispersiones cruzadas con los correspondientes coeficientes de correlación, donde se pueden ver posibles relaciones entre las variables de Perforación y la variable en estudio ROP600.

Las principales variables relacionadas al flujo que influyen en la tasa de penetración se encuentran el caudal, la velocidad del chorro de fluido en orificios del trépano, el impacto del chorro sobre el fondo y la potencia hidráulica disipada sobre el fondo del pozo, debajo del trépano. Las tres últimas características se calculan a partir de la primera, el caudal, de manera que habrá alguna clase de correlación o dependencia entre ellas, generando una colinealidad. Por este motivo, se descartaron las últimas variables y se conservó como variable representante de las condiciones de flujo al **Caudal**.

Otras variables que gobiernan la perforación son **Presión** (en la bomba), **Peso** (aplicado sobre el trépano), **RPM** (revoluciones por minuto en el trépano).

3.1.4 Análisis de componentes principales

Se realizó un análisis de componentes principales sobre las variables continuas de perforación (ROP600, Peso, RPM, Caudal, Filtrado, Densidad, ViscPlast, PtoCed, Presión, Potasio y MBT). El mayor interés es la relación de ROP600 con el resto de las variables. En la Fig. 3.1-15 se ve el gráfico de las primeras dos componentes principales de las variables. Para mayor claridad se indica el alineamiento a la variable respuesta ROP600 y su correspondiente perpendicular. Se observan comportamientos y relaciones totalmente coherentes con la técnica de perforación.

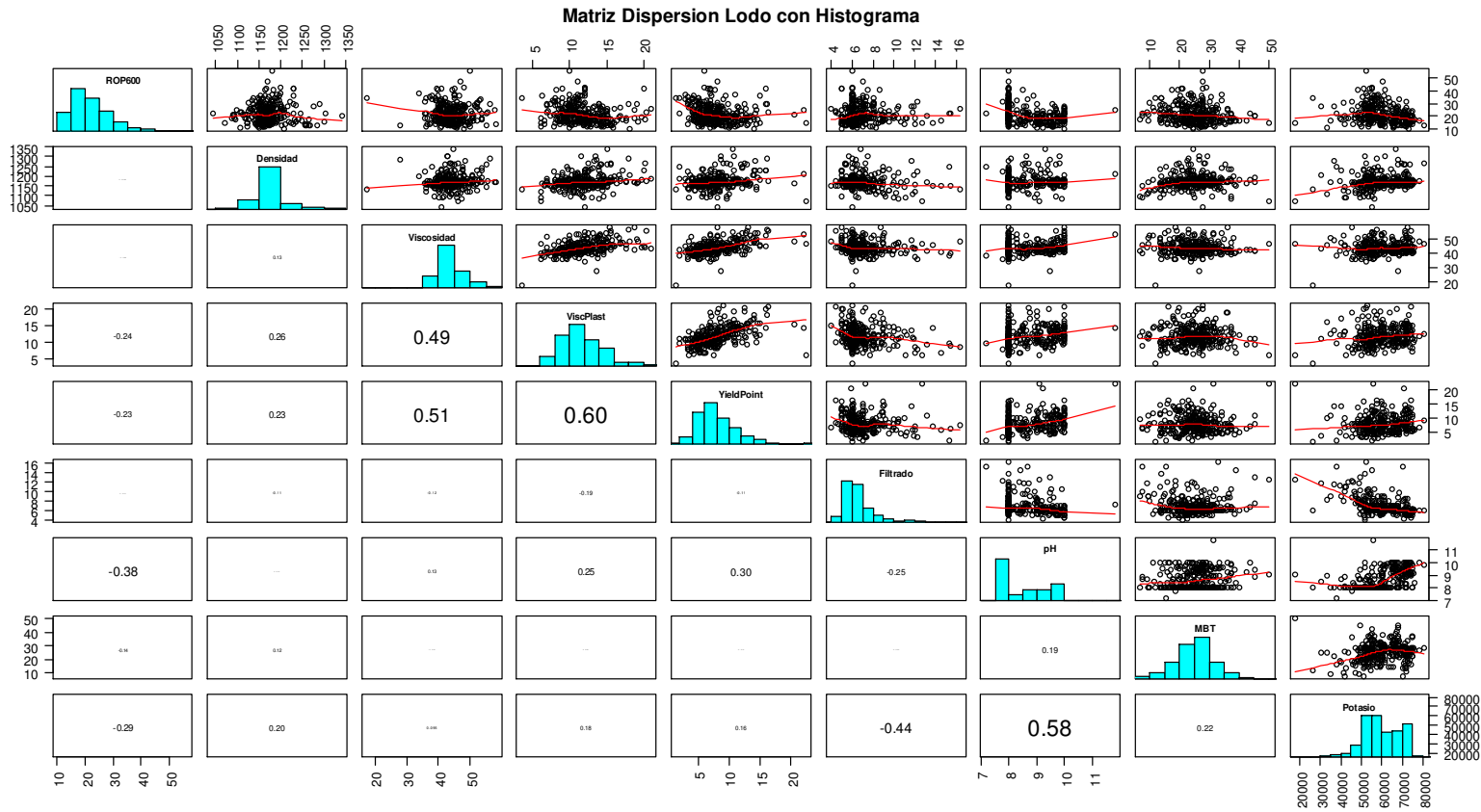


Fig. 3.1-13 Dispersiones cruzadas: ROP600 -propiedades del lodo
 Incluye histograma de cada variable y coeficientes de correlación (fuente proporcional a su valor)

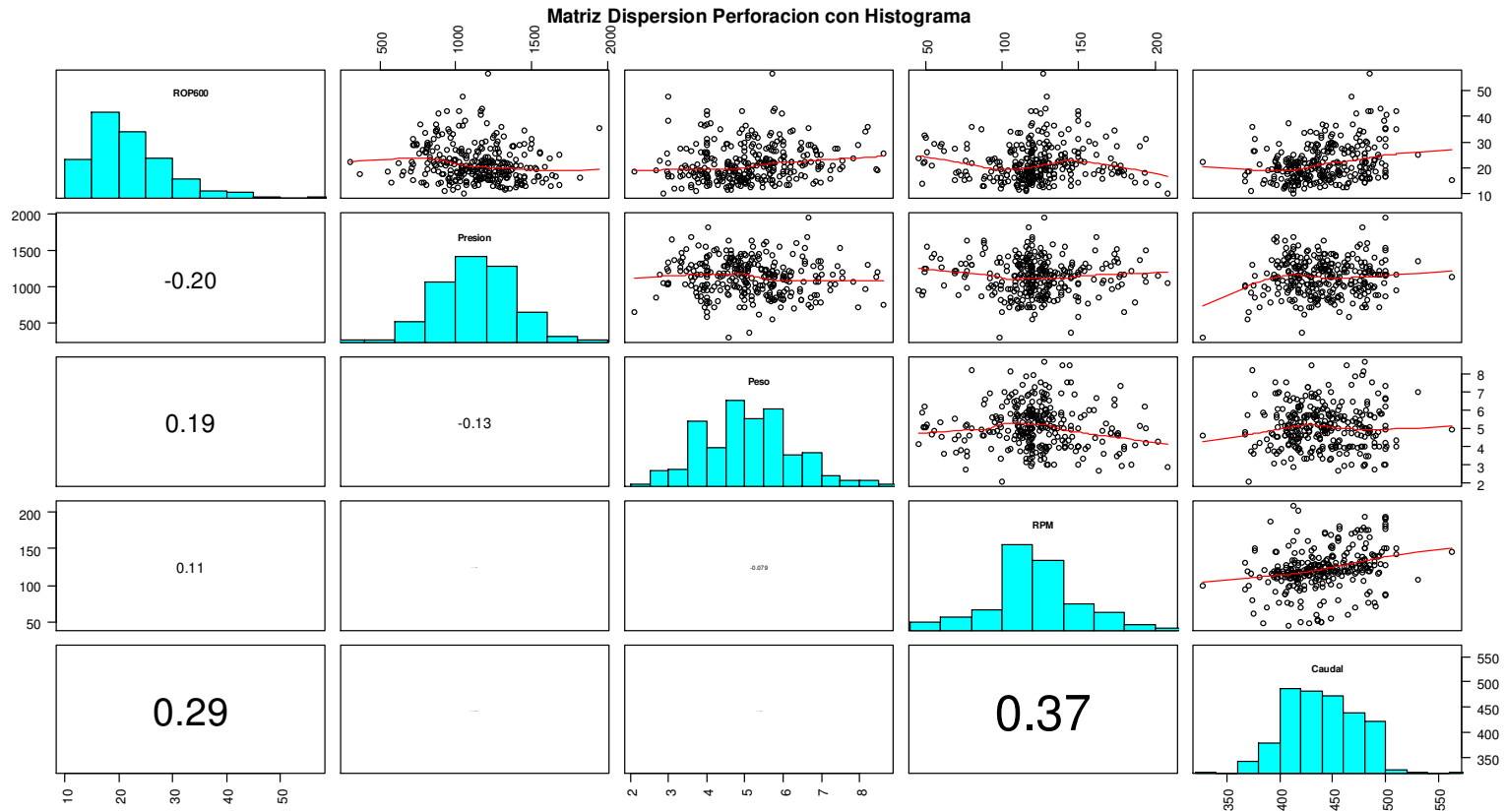


Fig. 3.1-14 Dispersiones cruzadas: ROP600 -características de perforación

Incluye histograma de cada variable y coeficientes de correlación (fuente proporcional a su valor)

En primer lugar se analizaron las variables agrupadas en el cuarto cuadrante: Densidad, PtoCed, ViscPlast y Presion. Esta relación es técnicamente razonable dado que a mayor densidad del fluido de perforación (mayor contenido de sólidos densos) se debe corregir las propiedades reológicas del lodo para tener mayor capacidad de transporte (en condiciones dinámicas) y sustento (en condiciones estáticas). Esto implica aumentar el punto cedente y la viscosidad plástica, lo que a su vez se traduce en una mayor presión en la bomba. La relación de la Presion no es tan evidente porque su valor puede cambiar por otras razones. Todas estas variables afectan negativamente a ROP600.

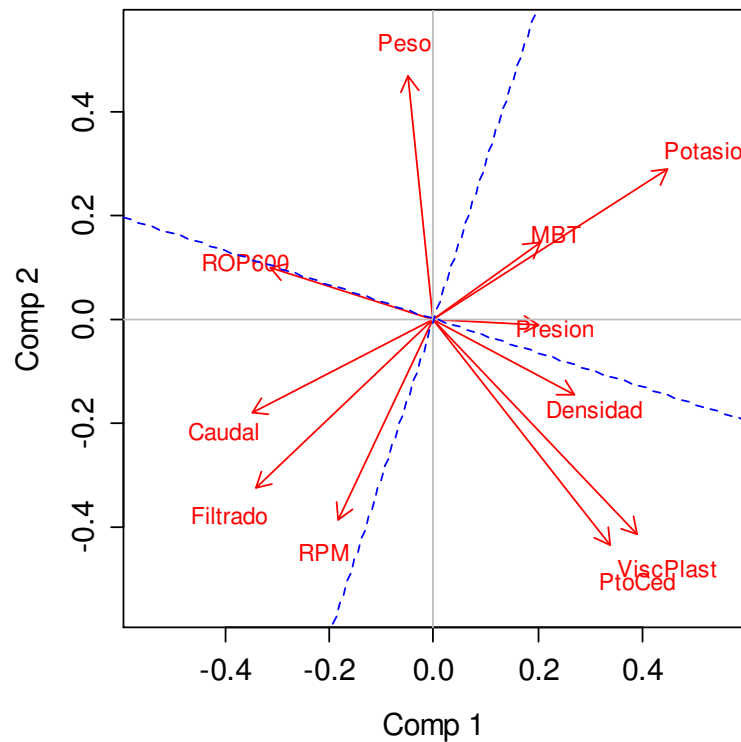


Fig. 3.1-15 Componentes principales de las variables

Se incluye eje alineado con la variable respuesta ROP600 y su perpendicular. Se aprecian las relaciones de ROP600 con las otras variables.

Por otro lado, hay un par de variables de lodo coalineadas en el primer cuadrante: MBT y Potasio. Esta alineación también es absolutamente coherente dado que cuando aumentan las arcillas reactivas, aumenta el MBT y la reacción del ingeniero de lodos es controlar la contingencia con un aumento de Potasio.

Como variable de lodo queda el Filtrado, cuya correlación con ROP600 es tan reducida que no debería afectarla apreciablemente.

La relación de estas variables con ROP600 concuerda con la teoría de perforación (Parr. 1.2.3).

Por último, quedan las variables de las condiciones de perforación: Peso, RPM y Caudal. Éstas deberían tener una relación directa con ROP600, sin embargo, el Peso tiene sólo una moderada influencia sobre ROP600 y RPM, casi ninguna, mientras que Caudal sí pareciera afectar ROP600 en alguna medida. Una explicación plausible sería la siguiente: Cuando se encuentran presiones de formación más elevadas que lo normal, se debe aumentar la Densidad del fluido. Esto implica aumento de la reología y aumento de la presión de bomba. Es posible que haya limitaciones mecánicas u operativas por las cuales no sea posible mantener el Caudal (a esas mayores presiones). Entonces puede ser que se disminuya el Caudal. Al disminuir el Caudal desmejora la limpieza en el fondo y ROP decae aunque se incremente el Peso (por remolienda de recortes sobre el fondo).

Algunas de estas relaciones se pueden resumir en el siguiente esquema:

$$> P_f \rightarrow > D \rightarrow > PC \text{ y } > VP > P_b \rightarrow < Q \rightarrow < ROP$$

donde:

P_f : Presión de formación

P_b : Presión en la bomba

D : Densidad de lodo

Q : Caudal de lodo

PC : Punto cedente de lodo

ROP : Tasa de penetración

VP : Viscosidad plástica de lodo

Se aprecia que la Densidad de lodo (por sí misma e influyendo sobre otras propiedades) tiene un efecto sobre la ROP. Esta premisa es muy indicativa para el ingeniero de perforación porque orienta la optimización de la perforación hacia un aspecto crucial. En este caso resulta preponderante el diseño y cuidado de las propiedades de lodo y la selección del equipo por su capacidad operativa (bomba y circuitos hidráulicos).

Capítulo 3.2 Agrupamientos

3.2.1 Distancias

Como en la tabla de casos y variables existen 297 casos, resulta una matriz de distancias de 297x297, lo cual es imposible de presentar completa aquí. Dado que la matriz es simétrica con diagonal de ceros, se presenta la misma en forma de matriz triangular inferior sin diagonal. La Fig. 3.2-1 es un extracto de extremo inferior derecho de la matriz de disimilaridades.

Pozo-338												
Pozo-339	0.164											
Pozo-340	0.185	0.231										
Pozo-341	0.203	0.258	0.318									
Pozo-343	0.306	0.355	0.349	0.275								
Pozo-344	0.262	0.271	0.211	0.342	0.261							
Pozo-345	0.040	0.178	0.209	0.189	0.326	0.267						
Pozo-347	0.098	0.231	0.279	0.258	0.339	0.265	0.105					
Pozo-348	0.218	0.293	0.098	0.355	0.375	0.242	0.220	0.253				
Pozo-351	0.251	0.279	0.384	0.356	0.347	0.325	0.259	0.248	0.400			
Pozo-352	0.109	0.151	0.231	0.189	0.365	0.313	0.101	0.147	0.280	0.236		
Pozo-354	0.241	0.231	0.230	0.388	0.199	0.234	0.261	0.281	0.267	0.267	0.257	
Pozo-338												
Pozo-339												
Pozo-340												
Pozo-341												
Pozo-343												
Pozo-344												
Pozo-345												
Pozo-347												
Pozo-348												
Pozo-351												
Pozo-352												
Pozo-354												

Fig. 3.2-1 Tabla de disimilaridades

Cabe recordar que las variables numéricas fueron estandarizadas por la diferencia absoluta media y para las distancias se usó la generalización de las distancias de Gower propuesta por Kaufman y Rousseeuw.

3.2.2 Métodos particionales

3.2.2.1. Agrupamiento alrededor de medoides

Número de grupos óptimo

En la Fig. 3.2-2 se ve que el ancho medio de silueta máximo se obtiene con 7 grupos.

En la Fig. 3.2-3 se ve la representación de la partición en 7 grupos realizada con el programa pam.

La primera observación es la presencia de dos grupos principales, representados por dos nubes de puntos a la izquierda y a la derecha del gráfico.

Además, se ve que hay 6 grupos bastante netos y casi excluyentes. Si bien hay pozos en algunas fronteras que podrían generar dudas, se puede decir que los 6 grupos están bien definidos. Sin embargo, en el Grupo 2 se ve que hay 4 pozos aparentemente alejados del

núcleo principal del grupo. Posiblemente el Grupo 2 esté alejado de los restantes grupos en la tercer dimensión. Una alternativa que se consideró fue analizar si es posible alojar estos 4 pozos en un octavo grupo.

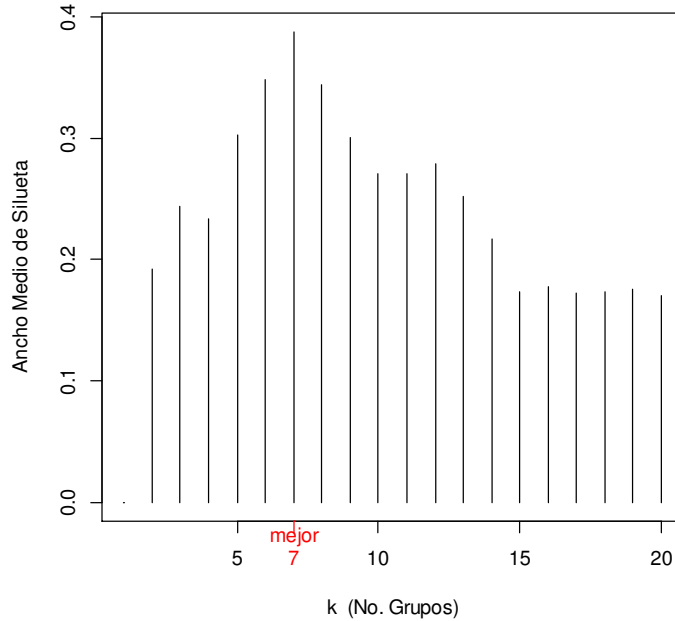
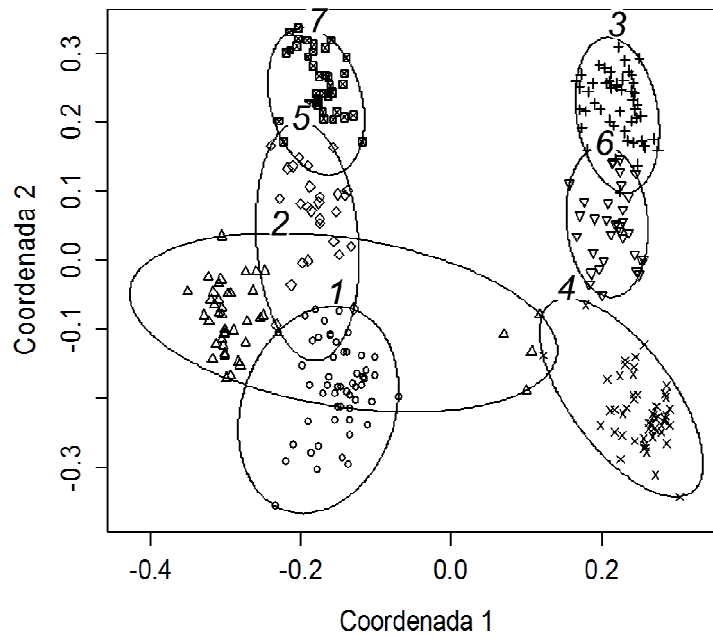


Fig. 3.2-2 Selección del número de grupos

Ancho medio de silueta de 2 a 20 grupos. Se obtiene mayor ancho con 7 grupos.



Estos dos componentes explican el 15.65% de la variabilidad de los puntos

Fig. 3.2-3 Agrupamiento con 7 medoides

Representación de los 7 grupos sobre escalado multidimensional de la matriz de disimilaridades.

En la Fig. 3.2-4 se identificaron esos pozos, que son: Pozo-071, Pozo-014, Pozo-341 y Pozo-113. Se observa que Pozo-172 (grupo 4) parece cercano al grupo 2, lo que sugiere que podría conservar algunas similitudes con los objetos de ese conjunto.

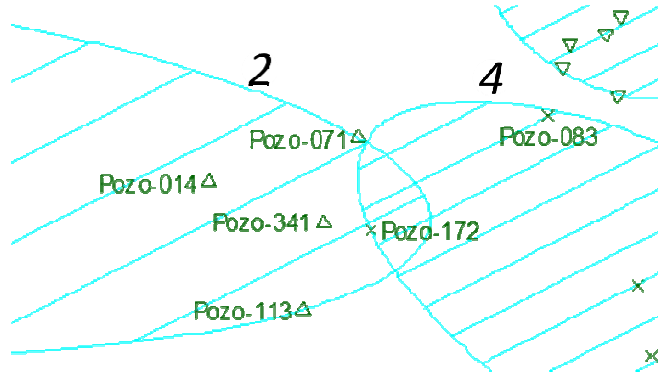


Fig. 3.2-4 Cuatro pozos del Grupo 2

Identificación de los 4 pozos más aislados del grupo 2.

3.2.2.2. Visualización en 3 dimensiones

Estos agrupamientos se graficaron en 3 dimensiones, incluyendo la tercer coordenada. En Fig. 3.2-5 se confirma que seis grupos se ven “compactos”, mientras que el Grupo 2 (en rojo) parece dispersarse en dos pequeñas nubes.

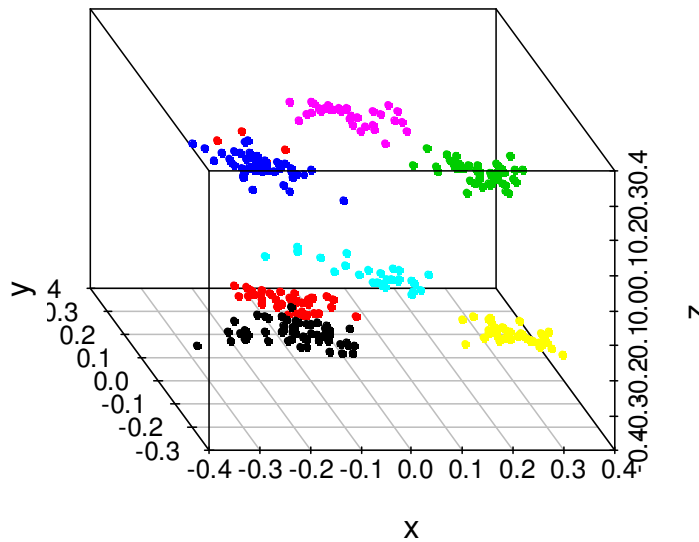


Fig. 3.2-5 Siete grupos a 3 Coordenadas

Presentación 3D de los 7 grupos. El grupo 2 (en rojo) aparece como dos nubes.

En la Fig. 3.2-6 se ve una proyección en 3D de los pozos del Grupo 2 a la derecha y el resto de los pozos en el gráfico de la izquierda. En esta vista, el grupo 2 está formado por dos nubes.

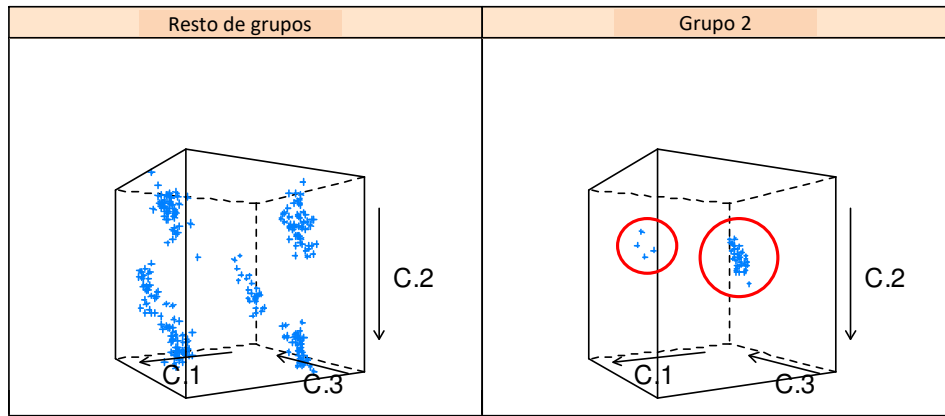


Fig. 3.2-6 Comparación del grupo 2 y el resto de grupos

Presentación 3D del grupo 2 comparada con el resto de los grupos.

En la Fig. 3.2-7 se ha rotado la representación hasta que el Grupo 2 aparece como una sola nube, y se puede ver que tiene real consistencia.

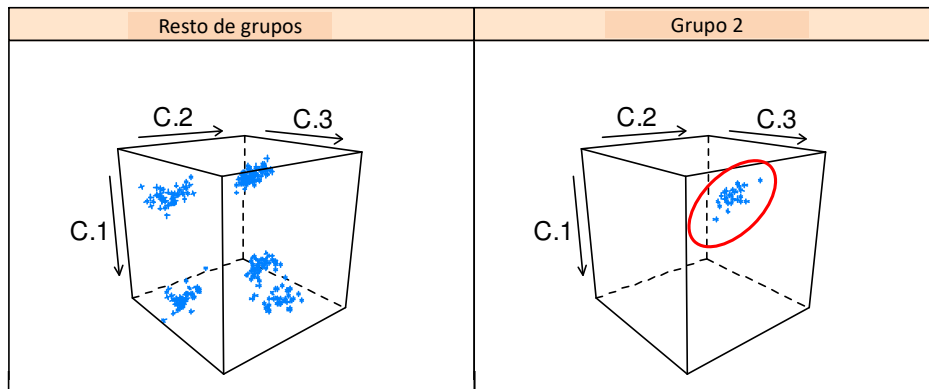


Fig. 3.2-7 Comparación del grupo 2 y el resto de grupos

Girando la presentación 3D el grupo 2 se visualiza como una sola nube.

3.2.2.3. Visualización por matriz ordenada

En la Fig. 3.2-8 se presenta la imagen obtenida para la matriz de disimilaridades de los datos de perforación.

Se nota que los grupos están bastante bien definidos. Se pueden percibir, además de los 7 grupos principales, dos zonas principales que abarcan los grupos (4,6,3) por un lado y (7,5,2,1) por otro lado, resultado que concuerda con la presentación vía coordenadas. También se manifiesta, con un sombreado más oscuro, las vecindades (4,6), (4,3), (6,3), (3,7), (6,5), (7,5), (4,1).

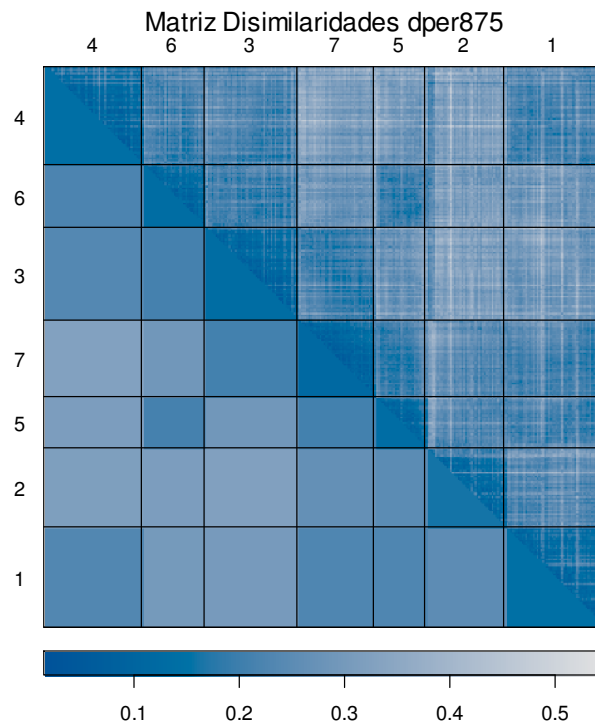


Fig. 3.2-8 Disimilaridades ordenadas

El ordenamiento de la matriz de disimilaridades permite visualizar 7 grupos.

3.2.2.4. Ancho medio de silueta

En la Fig. 3.2-9 se ven los anchos de siluetas para los 7 grupos. Los Grupos 1, 3 y 5 a 7 están bien definidos gracias a que casi todos los pozos de cada grupo tienen anchos de silueta bastante altos y hay una cierta uniformidad de distancias dentro cada grupo y con respecto a los otros grupos. En el grupo 2 se observan cuatro pozos con anchos de silueta relativamente bajos, comparados con los restantes pozos del mismo grupo. Lo mismo ocurre con 1 pozo en el grupo 4. Es decir que la asignación de estos cinco pozos a los grupos 2 y 4 es algo débil. Los anchos de silueta promedio de cada grupo valen entre 0.35 y 0.46 y el ancho de silueta promedio es 0.39.

De acuerdo al criterio subjetivo de Kaufman y Rousseeuw, la calidad del agrupamiento es algo débil y se debería investigar más con otros métodos de agrupamiento.

Estas dudas se trataron y despejaron mediante el análisis de agrupamientos jerárquicos.

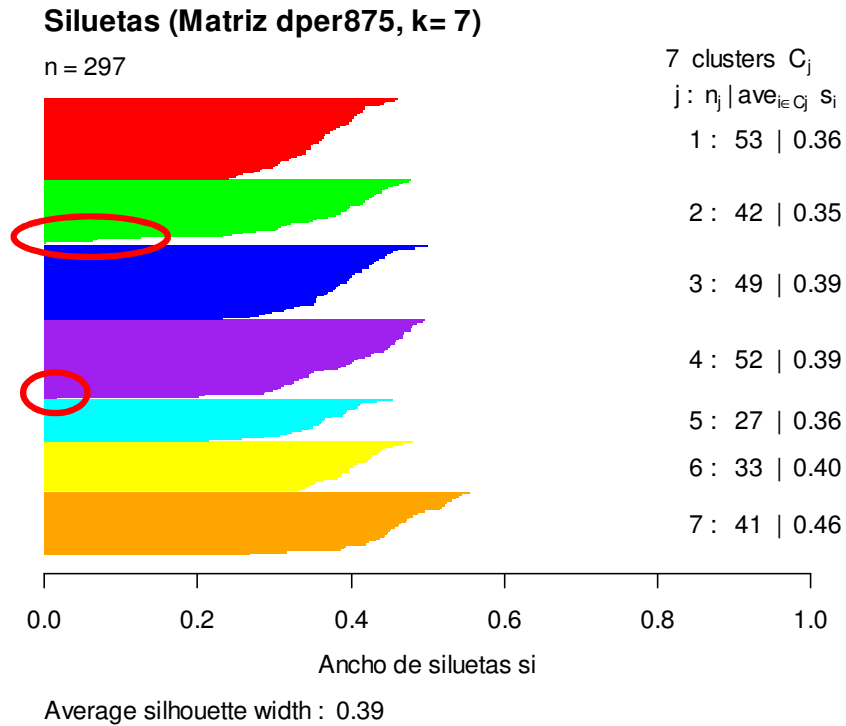


Fig. 3.2-9 Anchos de silueta para 7 grupos

Los anchos de siluetas de los Grupos 1, 3 y 5 a 7 están bien definidos. En el grupo 2 hay cuatro pozos con anchos de silueta relativamente bajos. Lo mismo ocurre con 1 pozo en el grupo 4.

3.2.2.5. Información de los grupos

En la Tabla 3.2-1 hay información adicional acerca de la integración de los grupos. Como se puede ver, la disimilaridad media tiene valores parecidos a los de la separación, lo que sugiere que los grupos no están muy alejados entre sí en relación a su tamaño. Esto confirma lo analizado con las siluetas.

Tabla 3.2-1 Información de los grupos

Grupo	Tamaño	Disimilaridad Máxima	Disimilaridad Media	Diámetro	Separación
1	53	0.200	0.102	0.293	0.114
2	42	0.255	0.107	0.385	0.080
3	49	0.190	0.088	0.265	0.108
4	52	0.186	0.089	0.318	0.080
5	27	0.162	0.090	0.273	0.114
6	33	0.151	0.088	0.219	0.108
7	41	0.164	0.076	0.251	0.110

Otra presentación indicativa de la cohesión de los grupos y de su separación (fortaleza grupal) es la Aislación (isolation) que es un indicativo de la separación o aislación de un

grupo. Ningún grupo presenta categoría L o L*, lo que confirma que el agrupamiento obtenido es algo débil.

3.2.2.6. Identificación de objetos en los grupos

En el Anexo Ase han identificado los pozos que integran cada grupo y los medoides correspondientes. Éstos representarían el caso típico de cada grupo, un prototipo representativo. En la industria se lo llama “pozo tipo” y se utiliza a los fines presupuestarios o comparativos. Además, se han identificado los pozos que integran cada grupo. Esto sirve para analizar cómo están integrados los grupos y las características de los pozos.

Tabla 3.2-2 Relación de Pozos Medoides con Rig y Tipo

Rig	Tipo	Pozo
OIL-01	DIRIGIDO	Pozo-300
OIL-01	VERTICAL	Pozo-061
PET-01	DIRIGIDO	Pozo-146
PET-02	DIRIGIDO	Pozo-291
PET-02	VERTICAL	Pozo-084
PET-03	DIRIGIDO	Pozo-255
PET-03	VERTICAL	Pozo-251

Los medoides de cada grupo son Pozo-300, Pozo-146, Pozo-251, Pozo-061, Pozo-291, Pozo-084 y Pozo-255. En la Tabla 3.2-2 se observa que cada pozo medoide corresponde a una pareja Rig-Tipo.

3.2.3 Agrupamientos jerárquicos

3.2.3.1. Métodos de clasificación

Se realizaron los árboles y luego se hicieron comparaciones.

3.2.3.2. Clasificación aglomerativa

Método “complete”

En la Fig. 3.2-10 se ve el árbol obtenido con el método complete.

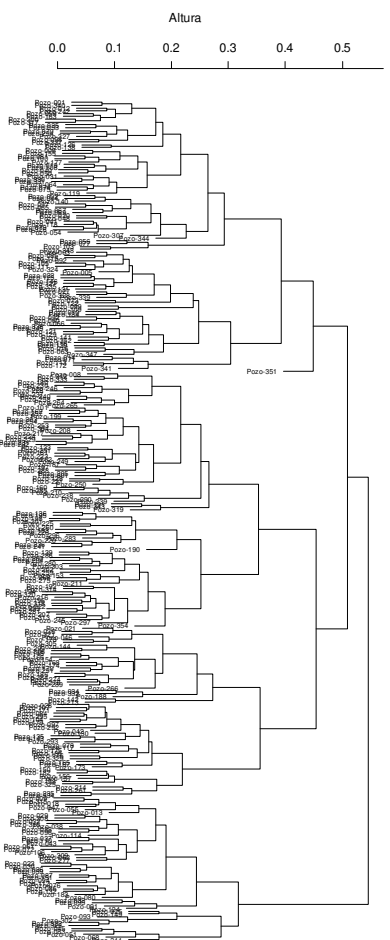


Fig. 3.2-10 Método Complete

Método "average"

En la Fig. 3.2-11 se ve el árbol obtenido con el método average.

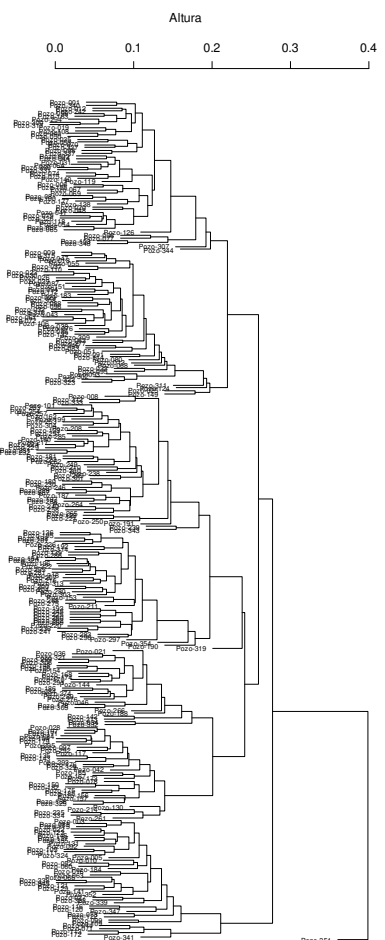


Fig. 3.2-11 Método Average

Método “single”

En la Fig. 3.2-12 se ve el árbol obtenido con el método single.

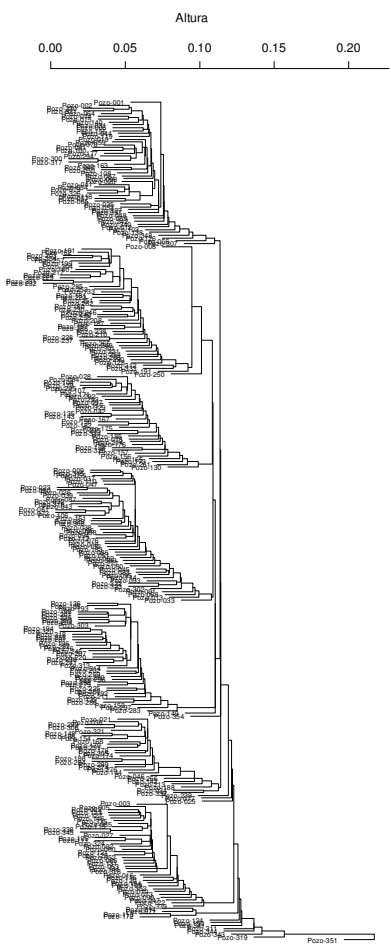


Fig. 3.2-12 Método Single

Método “weighted”

En la Fig. 3.2-13 se ve el árbol obtenido con el método weighted.

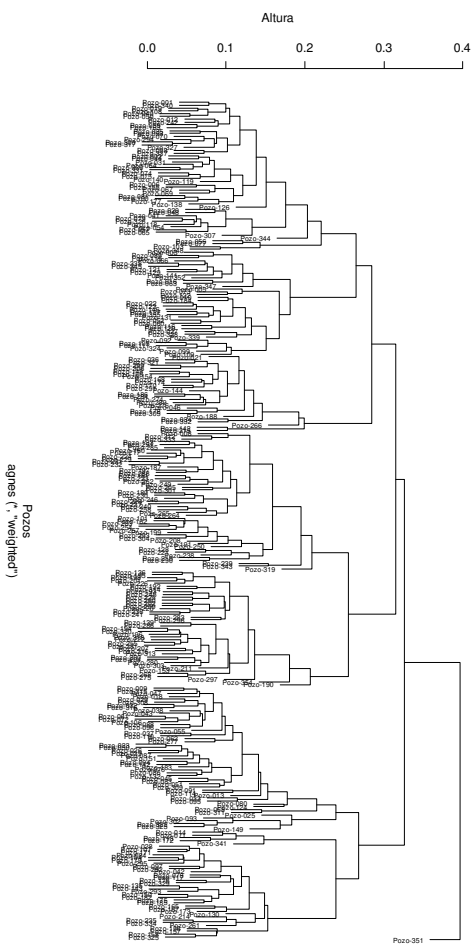


Fig. 3.2-13 Método Weighted

Método “flexible”

En la Fig. 3.2-14 se ve el árbol obtenido con el método flexible.

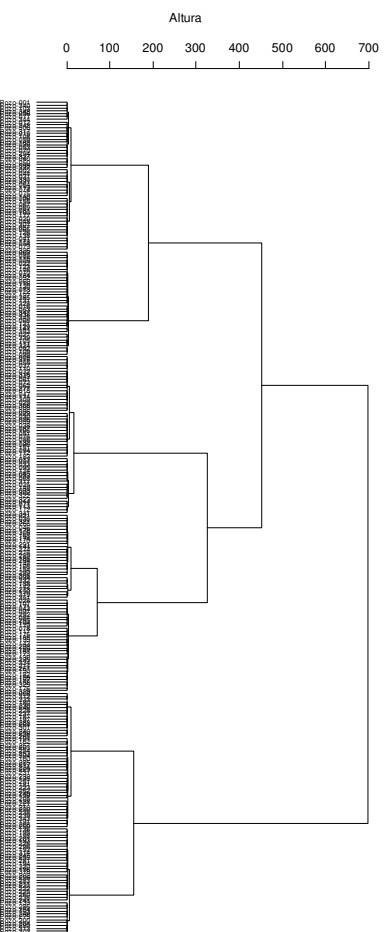


Fig. 3.2-14 Método Flexible

3.2.3.3. Clasificación divisiva

Método “diana”

En la Fig. 3.2-15 se ve el árbol obtenido con el método diana.

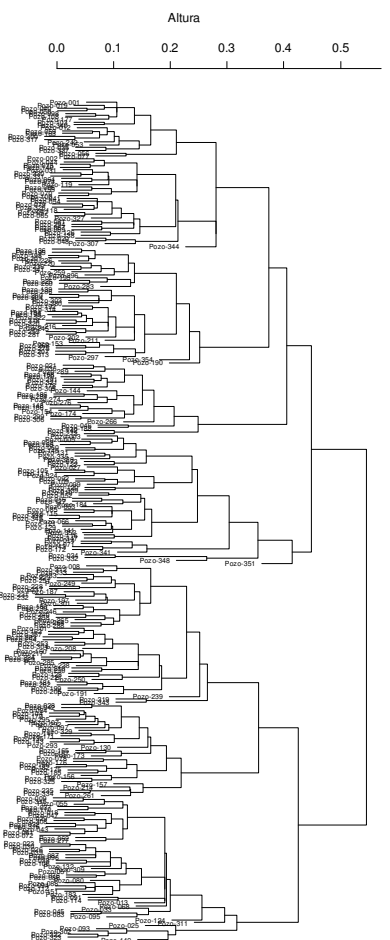


Fig. 3.2-15 Método diana

3.2.3.4. Tabla de correlaciones

La Tabla 3.2-3 presenta los coeficientes de correlación entre la matriz de distancias original y la matriz de distancias cofenéticas obtenida de los agrupamientos y los coeficientes de agregación o división según sea el caso. Se observa que los

métodos diana y average presentan las mayores correlaciones, aunque no se destaca uno de ellos como el más apropiado.

Tabla 3.2-3 Coeficientes de correlación cofenética

MÉTODO	Correlación Cofenética
Divisivo	0.717
Average	0.715
Weighted	0.692
Single	0.663
Complete	0.651
Flexible	0.525

3.2.3.5. Identificación de los grupos

Se identificaron los grupos o ramas de acuerdo a los factores Rig y Tipo sobre los dendrogramas obtenidos con los métodos diana y average (Fig. 3.2-16 y Fig. 3.2-17). En la rama de más alta jerarquía el método divisivo clasifica por Tipo y en un nivel inferior, por Rig. En cambio, el método average termina el nivel superior clasificando por Rig y el nivel inmediato inferior, por Tipo. En ambos casos las jerarquías superiores son determinadas por los factores Tipo y Rig. Según el criterio experto del perforador, hay acuerdo en que la clasificación es enteramente razonable. Atento a esta visión, los agrupamientos deberían ser según Tipo y Rig, lo que determina un total de 8 grupos.

En el árbol divisivo se da una situación particular con el Rig PET-01: Esta rama alberga tanto los pozos verticales como los dirigidos a pesar de que la primera jerarquía Tipo ya clasificó en un nivel más alto. En el árbol average esta situación no se destaca debido a que la primera jerarquía (la más alta) es Rig.

También se observa en ambos árboles que el pozo Pozo-351 se integra en un nivel excepcionalmente alto. En el árbol average este pozo no se pudo albergar en ninguna rama y debió ser aislado.

En la partición realizada con pam se obtuvieron 7 grupos, uno de los cuales incluía 4 pozos extraños (ver párrafo 3.2.2.1). Estos pozos corresponderían a la octava ramificación jerárquica, puesto que están entre los VERTICALES perforados por el Rig PET-01.

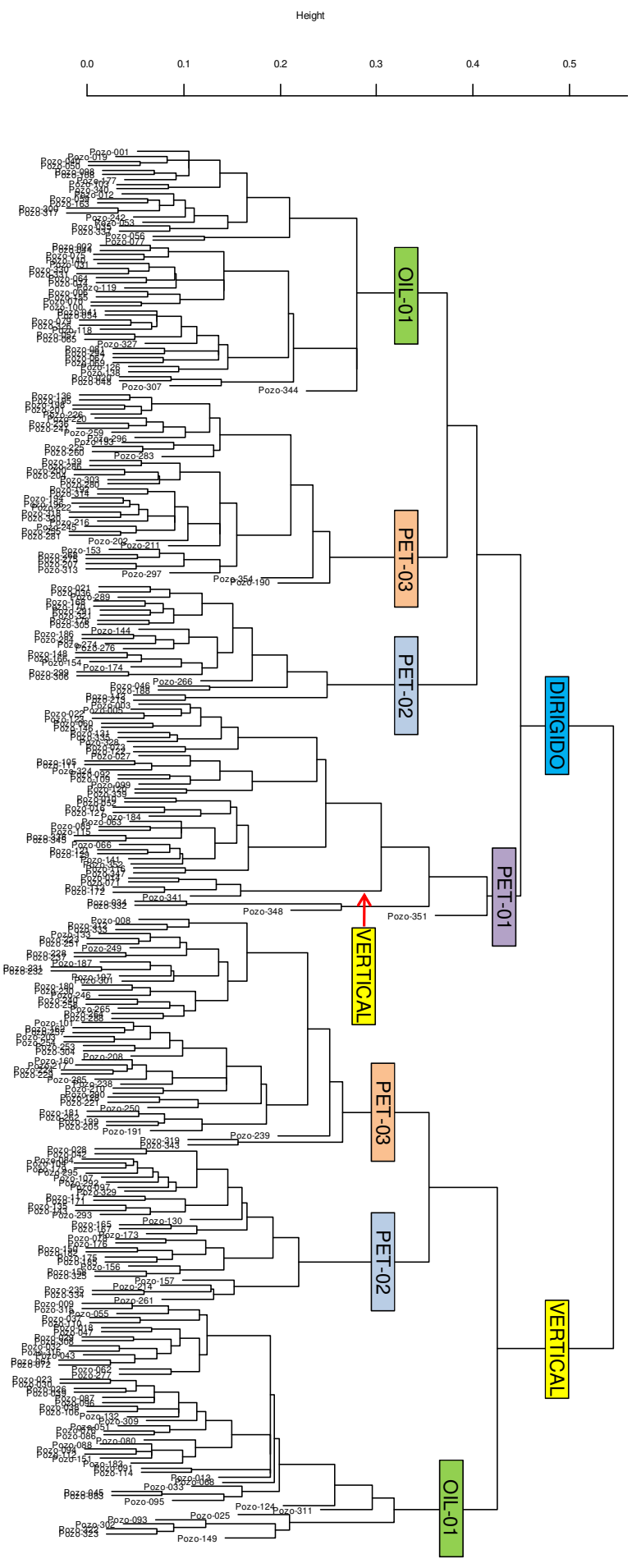


Fig. 3.2-16 Agrupamiento jerárquico - Método divisivo

Se identifican primero 2 grandes grupos según Tipo y luego, 7 grupos según Fig. Un pequeño grupo de pozos verticales incluso en el grupo de pozos dirigidos.

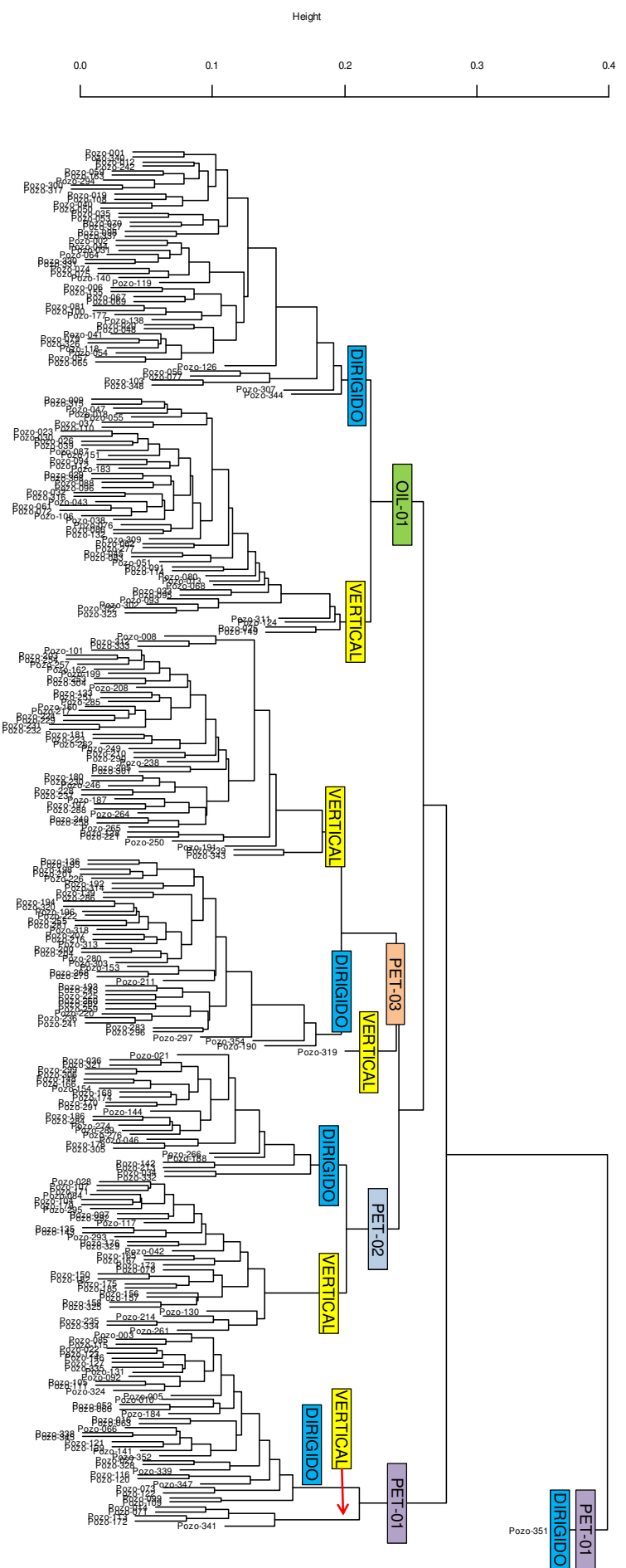


Fig. 3.2-17 Agrupamiento jerárquico - Método Average

Se identifican primero 4 grupos según Fig y luego, 8 grupos según Tipo, con algunos pozos aislados.

3.2.3.6. Selección del agrupamiento

Dado que se obtuvieron dos tipos de agrupamientos, en 7 y en 8 grupos, se debió decidir cuál conservar. En la dilucidación de esta cuestión, y siguiendo nuevamente el pensamiento de Kaufman y Rousseeuw, intervino el criterio del experto perforador. En razón de conservar la “naturalidad” del agrupamiento, se preservó la estructura de 8 grupos, tal como sugirió la clasificación jerárquica. Los grupos adoptados finalmente se ven en la Tabla 3.2-4.

Tabla 3.2-4 Grupos finales

Grupo	Rig	Tipo	Pozos
1	OIL-1	DIRIGIDO	53
2	PET-01	DIRIGIDO	38
3	PET-03	VERTICAL	49
4	OIL-1	VERTICAL	51
5	PET-02	DIRIGIDO	27
6	PET-02	VERTICAL	33
7	PET-03	DIRIGIDO	41
8	PET-01	VERTICAL	5

3.2.4 Diagnósticos

En la Fig. 3.2-18 se presenta un gráfico de cajas del logaritmo de ROP600 contra los 8 grupos. Se aplicó la transformación ‘log’ a la variable ROP600 para aproximar la normalidad.

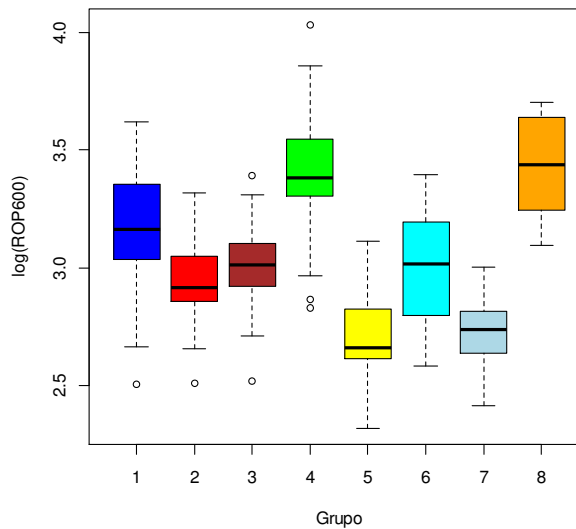


Fig. 3.2-18 Boxplot de log(ROP600) según 8 grupos

Se ve que hay disparidades notables de ROP600 entre los grupos. En la siguiente tabla se ve el análisis de la varianza del modelo.

	GL	Suma de Cuadrados	Cuadrados Medios	Valor F	Pr(>F)
Grupo	7	16.140	2.306	56.130	<2e-16
Error	289	11.870	0.041		

Resultaron diferencias significativas en las ROP600 y se rechazó la hipótesis de igualdad de tasas de penetración. En la Fig. 3.2-19 se muestra la distribución de los residuales del modelo lineal y un q-q plot para verificar la normalidad de los mismos. La prueba de normalidad de Shapiro-Wilks ofrece un valor p de 0.390. Este valor confirma la efectividad de la transformación.

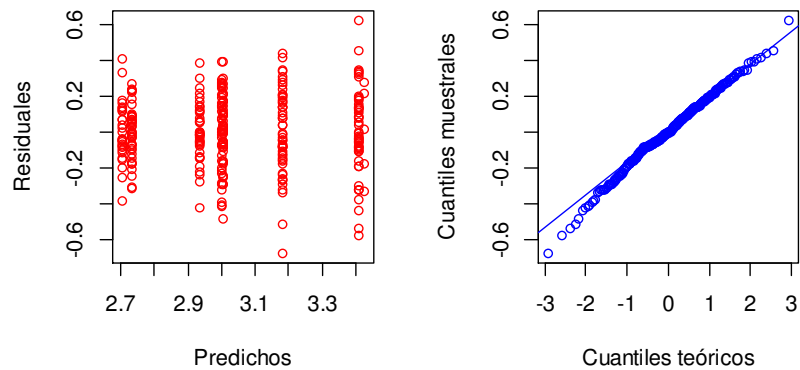


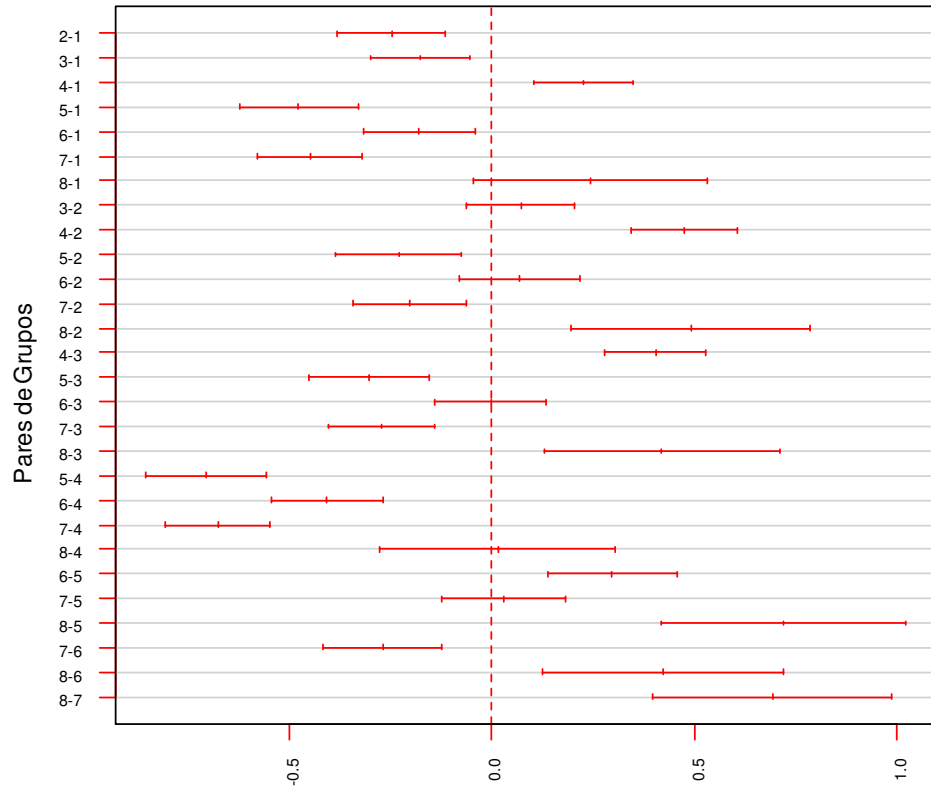
Fig. 3.2-19 Distribución de residuales

La distribución de los residuales del modelo lineal indica cierta homogeneidad de la varianza y el gráfico cuantil-cuantil de los mismos sugiere normalidad en los residuos.

3.2.4.1. Comparación múltiple

Se hicieron las comparaciones múltiples aplicando el test HSD de Tukey, las cuales se presentan gráficamente en la Fig. 3.2-20.

En la mayoría de los pares de grupos se ven diferencias, pero hay algunos grupos donde no hay diferencias apreciables, de manera que podemos decir que los agrupamientos no se debieron exclusivamente a los niveles de las medias de ROP600 de los grupos sino que además intervienen otras características importantes, como pueden ser Tipo de pozo y Rig. Esta última presunción se patentiza entre los grupos 6-4 y 5-1, donde los grupos 6 y 4 están integrados por pozos Verticales y los grupos 5 y 1, por pozos Dirigidos.



Diferencias de medias de log(ROP600) en los niveles de Grupo

Fig. 3.2-20 Intervalos de confianza de las diferencias de medias

El test HSD de Tukey presenta diferencias en la mayoría de los pares de grupos.

Se puede observar que, en general, cada grupo presenta más diferencias con respecto a los demás grupos, que similitudes. Por ejemplo, el Grupo 1 presenta diferencias con los Grupos 2 a 7, o sea con 6 grupos, y no es diferente al Grupo 8, o sea un solo grupo. Esto es indicativo de que pareciera haber alguna relación

3.2.5 Conclusión

Un aspecto fundamental en el análisis de agrupamientos se refiere a la necesidad de comprender el significado de los datos originales y recordar que debemos ver o entender lo que los datos nos están diciendo (Kaufman, y otros, 2005).

En el presente caso se observó que los niveles de Rig y Tipo de pozo de agrupamiento o clasificación fueron los que determinaron los agrupamientos. Esto es totalmente coherente con la situación real en estudio. Intuitivamente o en base a la experiencia del ingeniero de perforación, es esperable que esos dos

factores tengan una influencia o un peso considerable en el resultado de las perforaciones, sobretodo en cuanto a las tasas de penetración. Según las condiciones de perforación es total y razonablemente esperable que haya alguna diferencia entre los pozos Dirigidos y los Verticales, porque las herramientas, la técnica y la operativa son distintas. También es típico encontrar diferencias en la performance de los equipos. Cada uno de ellos tiene características, desenvolvimientos propios. Esto se puede deber al diseño del equipo, antigüedad, tamaño, potencia, cuadrillas que lo operan, etc. Así, se puede atribuir las formas de agrupamiento a una combinación del efecto de las variables Tipo de pozo, Rig y ROP600.

De esta manera, se considera que la clasificación presentada en la Tabla 3.2-4 es satisfactoria y que sucesivos estudios pueden referirse a este ordenamiento.

La detección de estos 8 grupos, gobernados por los factores Rig y Tipo puede ser de gran utilidad a la hora de analizar el desempeño de las perforaciones. Se pueden comparar diversos equipos y obrar en consecuencia; por ejemplo, arbitrar los medios para mejorar la performance de los más lentos, eventualmente, separar de la operación a los más ineficientes e irremediables. Otra utilidad importantísima es la aplicación de estos grupos a la generación presupuestaria, ya que está claro que hay significativas diferencias según el equipo en operación y según el tipo de pozo que se perfore.

En base al conocimiento brindado por las exploraciones y agrupamientos, quedó el campo abierto al estudio de modelos, tema que se presenta en el próximo capítulo, con la aplicación de Modelos Lineales Generalizados, Redes Neuronales Artificiales y Regresión por Árboles Recursivos.

Capítulo 3.3 Modelos de tasa de penetración

3.3.1 Modelos Lineales Generalizados

3.3.1.1. Modelo completo

Antes de iniciar el modelado, se comprobaron las posibles interacciones entre las variables categóricas. El resultado se presenta en la Fig. 3.3-1, donde se aprecia que la interacción entre las variables Tipo y Rig con respecto a la variable respuesta HrsF600 no es importante.

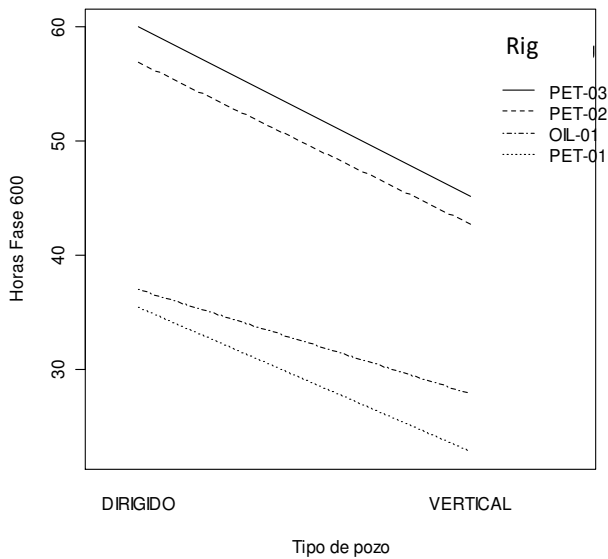


Fig. 3.3-1 Interacción de Tipo y Rig con HrsF600

Se aprecia un cierto paralelismo entre los efectos de las variables Tipo y Rig con respecto a la variable respuesta HrsF600. La interacción no sería importante.

También se exploraron las posibles interacciones entre las variables continuas mediante el mismo método gráfico, pero la cantidad de datos impidió la obtención de gráficos expresivos. Ante la duda de existencia de interacciones, se corrió un modelo con las interacciones dobles completas y luego se depuró por el método de selección “stepwise”. Se obtuvo un modelo complejo que consumía numerosos grados de libertad y alejado de lo que se pretende sea un modelo parsimonioso. Ante esta situación, se corrió un modelo sin interacciones y se comparó con el anterior. El modelo sin interacciones resultó ser un modelo más simple, que consumía pocos grados de libertad, se percibía parsimonioso y que además presentaba un AIC razonablemente bajo.

El modelo resultante fue el siguiente:

	Coefficiente	Error Estandar	Valor t	Pr(> t)
(Intercept)	-3.2100	0.4445	-7.2220	0.000
TipoVERTICAL	-0.2800	0.0306	-9.1550	< 2e-16
RigPET-01	0.3073	0.0516	5.9610	0.000
RigPET-02	0.4805	0.0383	12.5420	< 2e-16
RigPET-03	0.4198	0.0384	10.9360	< 2e-16
Densidad	0.0004	0.0004	1.0300	0.304
VisPlas	-0.0054	0.0058	-0.9290	0.354
PtoCed	0.0089	0.0046	1.9380	0.054
Filtrado	0.0032	0.0089	0.3630	0.717
MBT	0.0026	0.0019	1.3410	0.181
Potasio	0.0000	0.0000	0.0440	0.965
Presion	-0.0001	0.0001	-1.0470	0.296
Peso	0.0336	0.0123	2.7280	0.007
RPM	0.0002	0.0005	0.3020	0.763
Caudal	-0.0014	0.0005	-3.0370	0.003

Parámetro de dispersión de la familia Gamma se tomó como: **0.0362**

Devianza nula: **22.0276** con **237** grados de libertad

Devianza residual: **8.0365** con **223** grados de libertad

AIC: **1644.7**

En estos resultados, el parámetro de dispersión se estima con $X^2/(n - p)$, donde X^2 es el residual estandarizado de Pearson.

En este tipo de modelos la variable offset no se muestra, pero está presente como el logaritmo de la variable, con un coeficiente unitario.

En este modelo se puede ver que la devianza del modelo se reduce con respecto a la devianza del modelo nulo y en relación al número de grados de libertad disponibles, lo que asegura que el modelo ajusta mejor que el modelo nulo. En efecto, una prueba χ^2 sobre la devianza y los grados de libertad presenta un p-valor prácticamente unitario (debido al número de grados de libertad), lo cual revela un buen ajuste.

En la Fig. 3.3-2 se presentan gráficos de diagnóstico en los cuales se puede ver que los residuales de devianza no presentan mayores problemas, aunque se observan algunos casos con cierto nivel de apalancamiento.

El primer gráfico (a) presenta los residuales contra los valores ajustados, mientras que el tercer gráfico (c) duplica la resolución aplicando la raíz de los valores absolutos de los residuos. En este último no se aprecian anomalías.

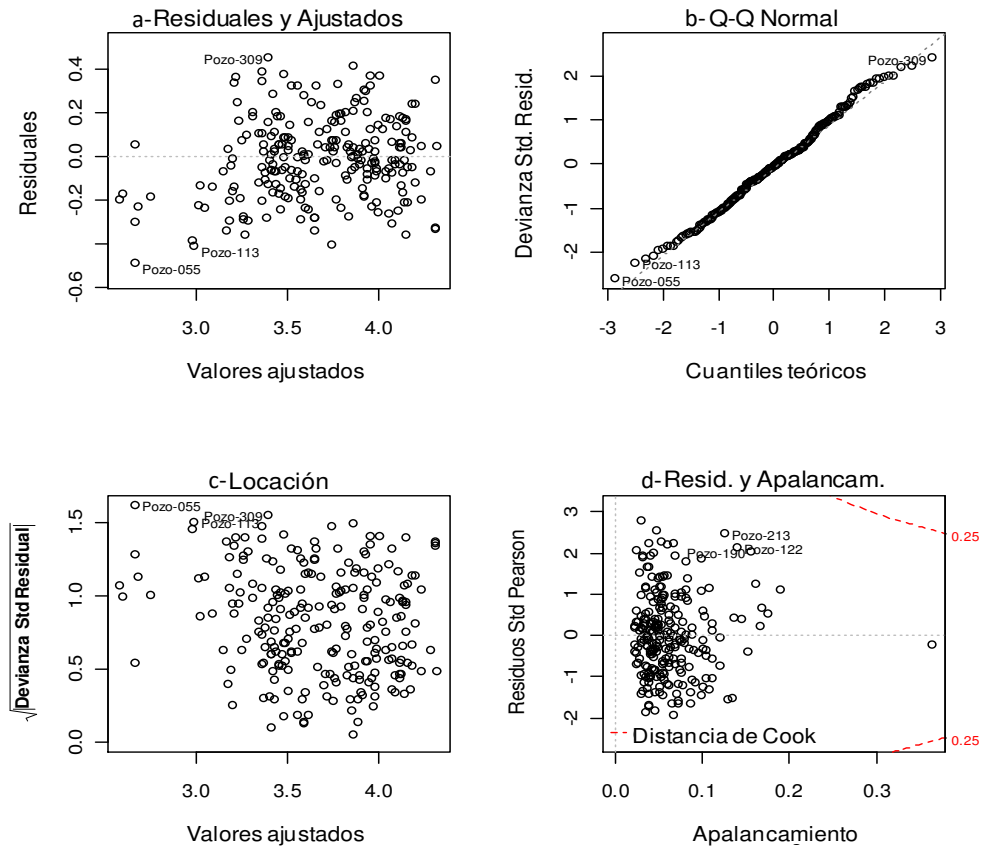


Fig. 3.3-2 Grafico de diagnóstico - Modelo completo

Arriba izquierda: Residuales contra valores ajustados. No se notan anomalías graves.
 Arriba derecha: Grafico cuantil-cuantil de normalidad.
 Abajo izquierda: Raiz de residuos de desviación estandarizados y valores ajustados.
 Abajo derecha: Residuales estandarizados contra apalancamientos y distancias de Cook. Se observan algunos casos con cierto nivel de apalancamiento.

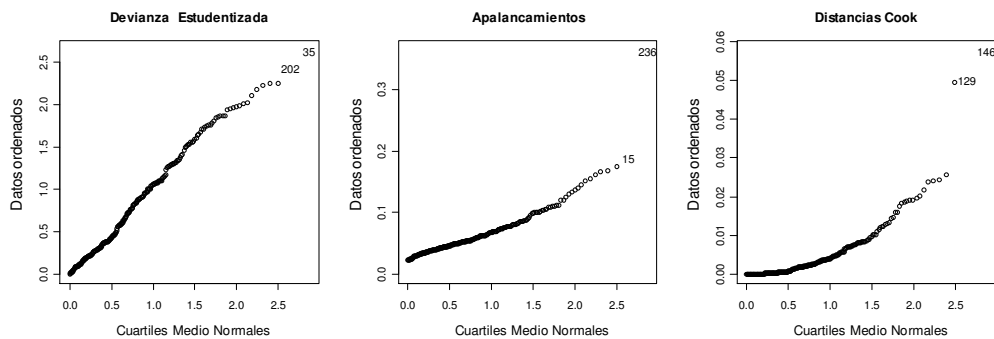


Fig. 3.3-3 Q-Q plot medio normal - Modelo completo

Para verificar la existencia de casos con valores extremos, en la Fig. 3.3-3 se presentan los Q-Q plots medio normales. Se puede ver que hay algunos casos

que salen de la tendencia y se apartan del conjunto de datos sugiriendo que pueden ser casos extremos, como el caso 236 cuyo apalancamiento sale de la tendencia.

3.3.1.2. Modelo reducido

Se procedió a identificar y descartar los casos detectados en la sección anterior y se repitió el modelado. El índice 236 en el gráfico central de la Fig. 3.3-3 corresponde al Pozo-351, que en la clasificación jerárquica debió ser aislado por tener algunos valores anormales. Se revisó la información de este pozo, hallándose que durante la perforación hubo problemas de pérdidas de circulación. Esto motivó una operación dificultosa y con muchas demoras. Estos eventos son los que ocasionan los desvíos en la ROP y colocan este pozo en posición totalmente anormal. Así, resulta coherente proceder a la eliminación de este caso del sub conjunto extraído para estudio.

	Coefficiente	Error Estandar	Valor t	Pr(> t)
(Intercept)	-3.1930	0.4540	-7.0320	0.000
TipoVERTICAL	-0.2803	0.0307	-9.1370	< 2e-16
RigPET-01	0.3080	0.0517	5.9530	0.000
RigPET-02	0.4815	0.0387	12.4560	< 2e-16
RigPET-03	0.4206	0.0386	10.9010	< 2e-16
Densidad	0.0004	0.0004	0.9930	0.322
VisPlas	-0.0058	0.0061	-0.9480	0.344
PtoCed	0.0094	0.0051	1.8310	0.068
Filtrado	0.0032	0.0090	0.3580	0.721
MBT	0.0027	0.0020	1.3390	0.182
Potasio	0.0000	0.0000	-0.0180	0.986
Presion	-0.0001	0.0001	-1.0430	0.298
Peso	0.0336	0.0123	2.7260	0.007
RPM	0.0002	0.0005	0.3030	0.763
Caudal	-0.0014	0.0005	-3.0350	0.003

Parámetro de dispersión de la familia Gamma se tomó como: **0.0363**
 Devianza nula: **21.9268** con **236** grados de libertad
 Devianza residual: **8.0347** con **222** grados de libertad
 AIC: **1638.3**

Se observa que el parámetro de dispersión se mantiene, las devianzas y el AIC se reducen, lo cual es alentador.

Con la repetición de los gráficos de diagnóstico de la Fig. 3.3-4 se aprecia que se ha obtenido un resultado aceptable.

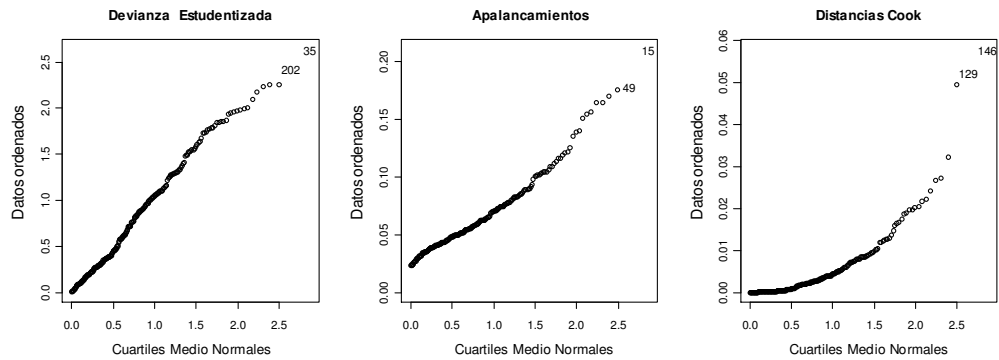


Fig. 3.3-4Influencia, apalancamiento y dist. Cook-Modelo completo

Considerando que el presente modelo representa un modelo “completo”, en el sentido que incluye todas las variables disponibles, se aplicó una selección paso a paso con reducción del AIC para quitar variables del modelo, llegando al siguiente modelo reducido:

	Coefficiente	Error Estandar	Valor t	Pr(> t)
(Intercept)	-2.7721	0.2016	-13.7490	< 2e-16
TipoVERTICAL	-0.2681	0.0271	-9.9100	< 2e-16
RigPET-01	0.2997	0.0450	6.6560	0.000
RigPET-02	0.4685	0.0358	13.1000	< 2e-16
RigPET-03	0.4156	0.0322	12.8980	< 2e-16
PtoCed	0.0074	0.0040	1.8580	0.064
Peso	0.0336	0.0118	2.8540	0.005
Caudal	-0.0014	0.0004	-3.3180	0.001

Parámetro de dispersión de la familia Gamma se tomó como: **0.0359**
 Devianza nula: **21.9268** con **236** grados de libertad
 Devianza residual: **8.2126** con **229** grados de libertad
 AIC: **1629.5**

En este caso, se observa que se redujeron el parámetro de dispersión y el AIC, lo cual indica que el modelo reducido se optimiza. Por otro lado, aunque la devianza del modelo aumentó, también aumentaron los grados de libertad, de manera que su relación tuvo una disminución, lo cual representa una mejora. Todo esto habla de la buena calidad del modelo reducido.

En la Fig. 3.3-5 se presentan gráficos de diagnóstico en los cuales se puede ver que los residuales de devianza no presentan mayores problemas, aunque se observan algunos casos con cierto nivel de apalancamiento. Asimismo, se observa que las distancias de Cook se han atenuado.

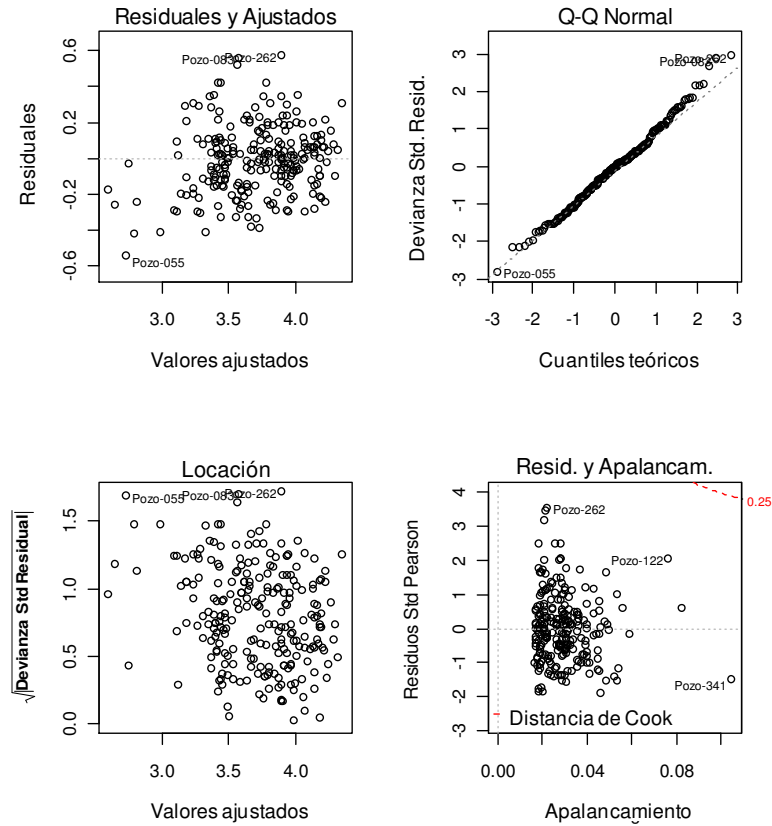


Fig. 3.3-5Gráfico de diagnóstico - Modelo reducido

Abajo derecha: Ya no se observan casos con apalancamiento importante.

Con la repetición de los gráficos de diagnóstico de la Fig. 3.3-6 se aprecia que se ha obtenido un resultado aceptable.

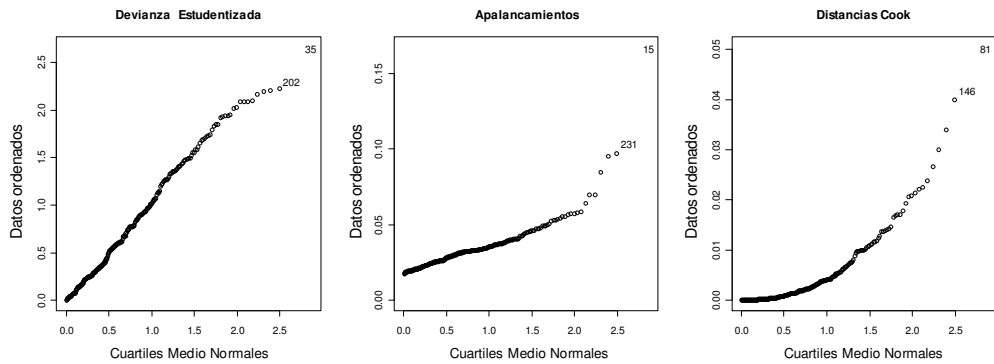


Fig. 3.3-6Influencia, apalancamiento y dist. Cook-Modelo Reducido

Una prueba χ^2 sobre la devianza y los grados de libertad presenta un valor prácticamente unitario, confirmando que el ajuste es apropiado.

Se pueden construir intervalos de confianza aproximados para los parámetros de la regresión usando aproximaciones normales sobre las estimaciones de los parámetros.

	2.5%	97.5%
(Intercept)	-3.1646	-2.3794
TipoVERTICAL	-0.3216	-0.2146
RigPET-01	0.2107	0.3892
RigPET-02	0.3984	0.5389
RigPET-03	0.3524	0.4787
PtoCed	-0.0004	0.0152
Peso	0.0103	0.0570
Caudal	-0.0023	-0.0006

El intervalo de confianza $100(1-\alpha)\%$ sería:

$$\hat{\beta}_i \pm z_{\alpha/2}se(\beta_i)$$

donde $z_{\alpha/2}$ es un cuantil de la distribución normal. Para un 95% de confianza,

para β_1 (el factor Tipo VERTICAL) se tiene:

$$\text{Límite Inferior: } -0.2681 - 1.96 * -0.0271 = -0.3212$$

$$\text{Límite Superior: } -0.2681 + 1.96 * -0.0271 = -0.2150$$

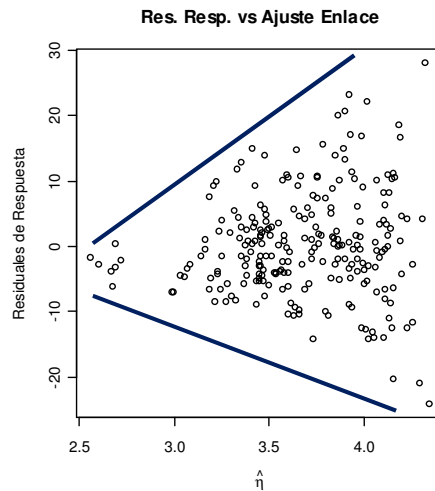


Fig. 3.3-7 Diagnóstico de residuales de respuesta

Los residuales de respuesta contra los predictores lineales η_i . Se observa un patrón incremental.

La Fig. 3.3-7 muestra los residuales de respuesta contra los predictores lineales η_i , donde se observa un patrón incremental. Este comportamiento de embudo abierto hacia la derecha es indicativo de que la varianza es una función creciente de η_i (Montgomery, y otros, 2004).

En la Fig. 3.3-8 se ven los residuales de devianza contra las respuestas

predichas μ y contra los predictores lineales η donde no se observan anomalías. Esto muestra que el enlace elegido para el modelo es adecuado, ya que corrige el comportamiento observado en la Fig. 3.3-7.

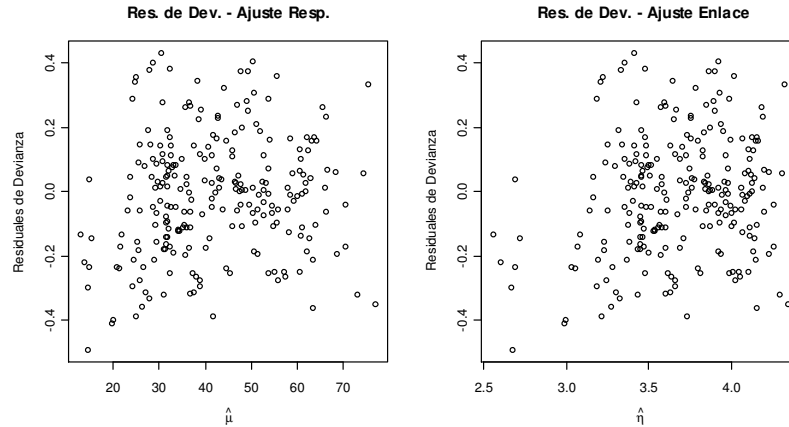


Fig. 3.3-8 Diagnóstico de residuales de devianza

Residuales de devianza contra las respuestas predichas μ y contra los predictores lineales η . No se observan anomalías, porque el modelo corrige ese comportamiento.

3.3.1.3. Ajustes sobre los sub conjuntos

Ajustes sobre el sub conjunto de modelado

En la Fig. 3.3-9 se presentan los ajustes del sub conjunto utilizado para modelar. Para mayor claridad, los casos se han factorizado y presentado por separado. Se presentaron conjuntamente los valores observados y los ajustados, para comparación. Se observan, en general, buenos ajustes, aunque algunos casos presentan ciertas anomalías. Es difícil asegurar a qué se deben, pero por su aspecto se presume que interviene algún factor imprevisto no siempre presente u ocasional, propio de la actividad de perforación de pozos. Es posible que una revisión específica sobre estos casos particulares podría aclarar estas diferencias. Esta sería una utilidad adicional de esta investigación, la de hallar factores disturbantes “accidentales” no deseados. Esta detección permitiría tomar acciones para evitarlas.

PARTE 3 – RESULTADOS Y DISCUSIÓN

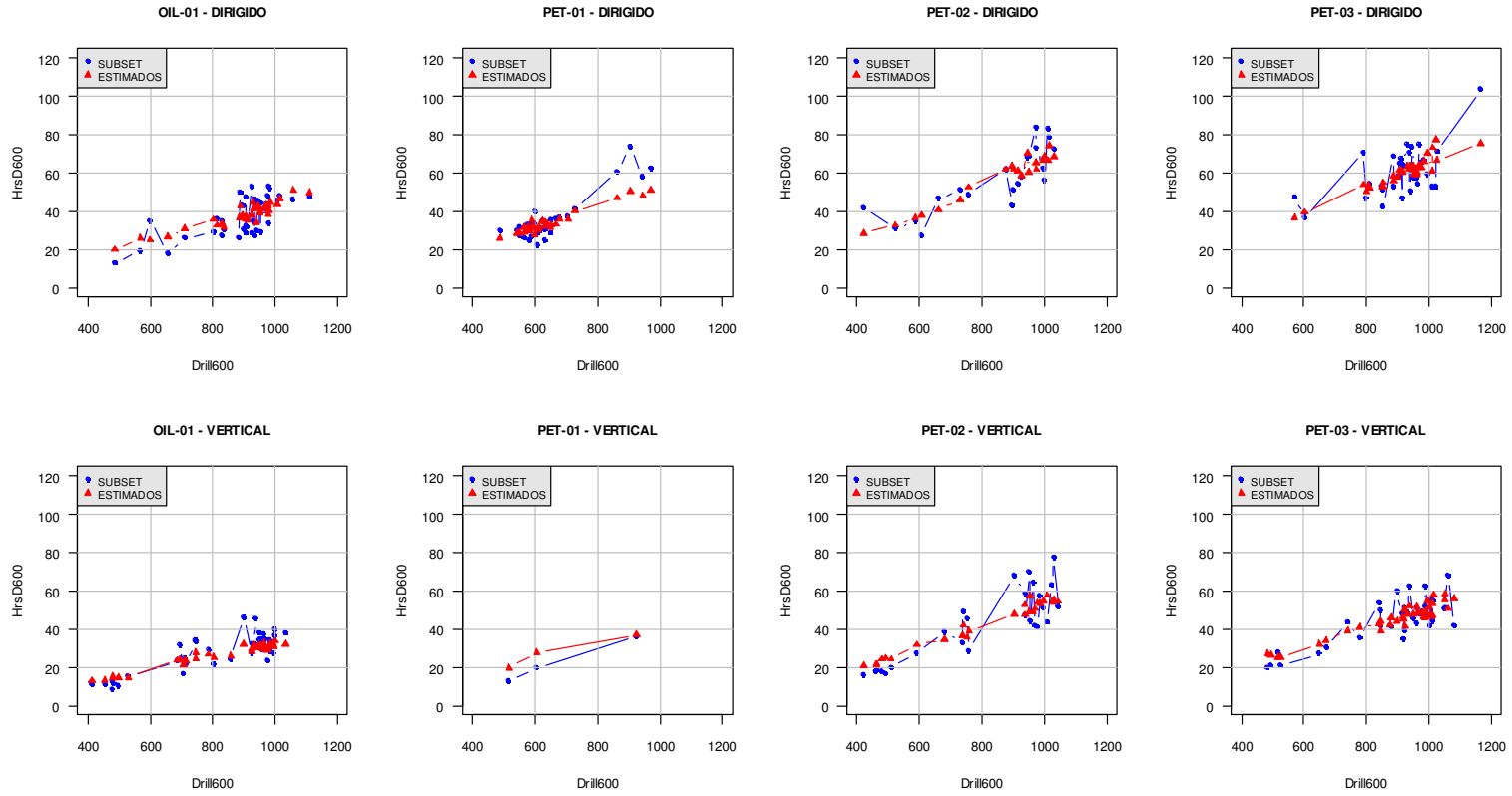


Fig. 3.3-9 Ajustes del subset - Modelo reducido

Comparación de HrsF600 real con estimados clasificados por Rig y por Tipo. Se observa una apreciable concordancia.

PARTE 3 – RESULTADOS Y DISCUSIÓN

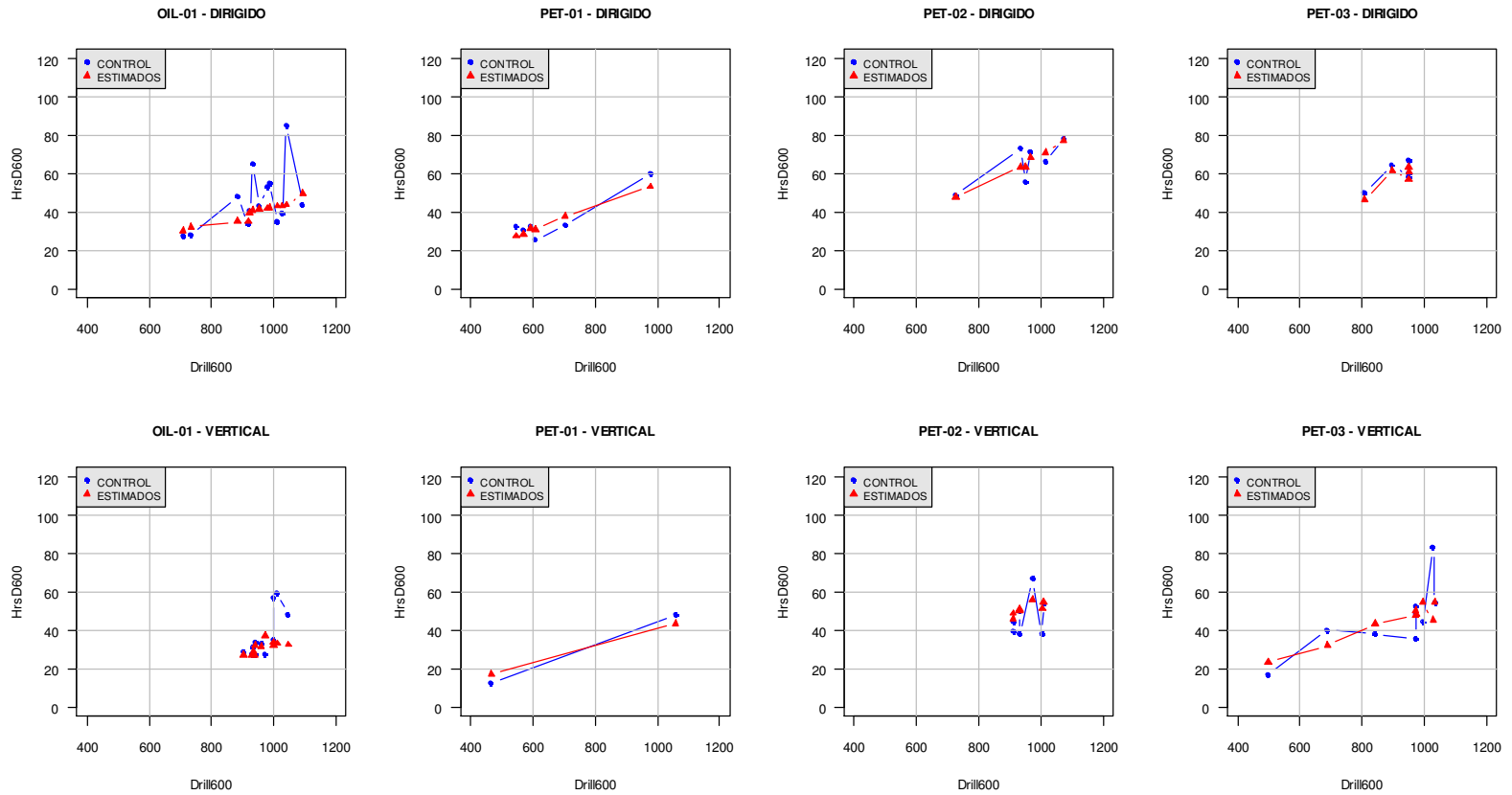


Fig. 3.3-10 Ajustes de control-Modelo reducido

Comparación de HrsF600 real con estimados clasificados por Rig y por Tipo. Se observa una apreciable concordancia.

Ajustes sobre sub conjunto de control

En la Fig. 3.3-10 se presentan los ajustes realizados con el sub conjunto de control. Para mayor claridad, los casos se factorizaron y presentaron por separado. Se puede decir que, en general, se observa que los ajustes son satisfactorios y valen los mismos comentarios de los ajustes sobre el sub conjunto de modelado. En este caso, los ajustes sobre el conjunto de control han servido para corroborar el buen funcionamiento del modelo propuesto con el conjunto de estimación.

3.3.1.4. Interpretación del MLG

En definitiva, el modelo propuesto quedó de la siguiente manera:

		Coficiente	Error Estandar	Valor t	Pr(> t)
α	(Intercept)	-2.7721	0.2016	-13.7490	< 2e-16
β	TipoVERTICAL	-0.2681	0.0271	-9.9100	< 2e-16
γ_1	RigPET-01	0.2997	0.0450	6.6560	0.000
γ_2	RigPET-02	0.4685	0.0358	13.1000	< 2e-16
γ_3	RigPET-03	0.4156	0.0322	12.8980	< 2e-16
δ_{pc}	PtoCed	0.0074	0.0040	1.8580	0.064
δ_w	Peso	0.0336	0.0118	2.8540	0.005
δ_q	Caudal	-0.0014	0.0004	-3.3180	0.001

$$HrsF600_i = \exp \left(-2.7721 - 0.2681 * T_i + 0.2997 * P1_i + 0.4685 * P2_i + 0.4156 * P3_i + 0.0074 * PC_i + 0.0336 * W_i - 0.0014 Q_i + \log(D_i) \right) \quad \text{Ec. 3.3-1}$$

3.3-1

con:

- T_i Tipo de Pozo del caso i. Si es: Dirigido= 0; Vertical= 1
- $P1_i$ Rig del caso i. Si es PET-01: Ausente= 0; Presente= 1
- $P2_i$ Rig del caso i. Si es PET-02: Ausente= 0; Presente= 1
- $P3_i$ Rig del caso i. Si es PET-03: Ausente= 0; Presente= 1
- PC_i Valor de la variable de lodo PtoCed del caso i.
- W_i Valor de la variable de perforación Peso del caso i.
- Q_i Valor de la variable de perforación Caudal del caso i.
- D_i Tramo perforado en la Fase 600 (Drill600) del caso i.
- $HrsF600$ Tiempo de perforación en la Fase 600 del caso i

Variables categóricas

Tal como era de prever, después de lo observado en las exploraciones y en los agrupamientos, el **Tipo** de pozo y el **Rig** o equipo de perforación, que entraron como variables indicadoras, resultaron de gran importancia y son todas significativas.

Tipo de pozo VERTICAL tiene un comportamiento que equivale a $\frac{3}{4}$ partes de los tiempos de Tipo de pozo DIRIGIDO ($\exp(-0.2681) = 0.76$)

Los equipos PET-01, PET-02 y PET-03 tienen comportamientos que equivalen a 1.35, 1.6 y 1.5 veces los tiempos del equipo OIL-01.

Variables continuas

En este modelo, las variables numéricas presentesson unavariable de lodo y dos variables de perforación. Esto sugiere que ambos aspectos de la actividad afectan los resultados en cuanto a tiempos de perforación.

Es posible trazar un paralelismo entre estos resultados y los presentados en el análisis de componentes principales (variables continuas) de la Fig. 3.1-15. Allí se ve que las variables de lodo PtoCed y ViscPlast tienen la mayor correlación negativa con ROP600, mientras que Peso y Caudal son las que tienen mayor correlación positiva con ROP600.

Recordando el Capítulo 2 dedicado a Tecnología de Perforación, se puede relacionar esos conceptos con el análisis de las variables del modelo y sus coeficientes.

PtoCed: Esta variable representa el *Punto Cedente*. Esta es una propiedad del lodo que junto a ViscPlast (o sea, la *Viscosidad Plástica*) tiene injerencia en el comportamiento del fluido, en las condiciones de flujo y en las pérdidas de carga. Estas dos propiedades determinan la Viscosidad Aparente del fluido al régimen de flujo, la cual interviene inversamente en el valor del número de Reynolds (Re), que es una propiedad indicativa del nivel de hidráulica en el trépano. El nivel de hidráulica influye sobre la tasa de penetración (Eckel, 1968). A mayores Punto Cedente y Viscosidad Plástica, mayor Viscosidad Aparente, lo cual reduce la tasa de penetración. Es posible que este efecto se deba a una menor eficiencia en la limpieza del pozo. Este efecto sería aditivo al efecto del peso aplicado sobre el trépano. El coeficiente es positivo, indicando que a mayor Punto Cedente, aumenta el tiempo de perforación (se reduce la tasa de penetración). De todas

maneras su efecto es relativo, ya que su valor p no llega a ser significativo al 5% (su coeficiente tiene un valor p de 0.064).

Peso: En este modelo, el coeficiente de la variable Peso es positivo, lo cual significa que cuando aumenta el peso sobre el trépano se incrementa el tiempo de perforación (se reduce la tasa de penetración).

Por el contrario, en la sección 1.2.3.4 del capítulo de perforación (Bourgoyne, y otros, 1986) se vio que, en condiciones normales y hasta cierto punto, cuando aumenta el peso aplicado al trépano, la velocidad de perforación incrementa (los tiempos disminuyen). Al margen del desgaste de los dientes, este comportamiento opuesto del peso se puede explicar de la siguiente manera: (1) El peso aplicado clava los dientes completamente en el fondo del pozo, (2) El fluido de perforación no alcanza a evacuar todos los recortes producidos por la perforación y el trépano se entretiene en remoler los recortes. Como los pesos aplicados en la mayoría de los casos no son demasiado elevados, se descarta la primera explicación. La segunda explicación resulta más apropiada porque es coherente con el efecto de la Viscosidad Aparente. Este efecto (llamado “floundering” o estancamiento) también se explica en el párrafo 1.2.3.4.

Caudal: El comportamiento del Caudal dentro del modelo es el esperado por la técnica y la experiencia, ya que, como el coeficiente es negativo, mayores caudales ocasionan una reducción en los tiempos de perforación. La disminución de los tiempos con el incremento del caudal puede deberse propiamente al efecto de la hidráulica (potencia disipada y fuerza de impacto sobre el fondo del pozo), caso en el cual tendría preponderancia la limpieza de recortes generados por el trépano en el fondo del pozo.

Resumen: La composición de los coeficientes de las variables categóricas y continuas resulta coherente. En el ACP se observó que ROP600 podría ser afectada mayormente por el concatenamiento de variables de lodo [presión de formación–densidad–reología–presión de bomba] y por el Peso y el Caudal. En este modelo, se han presentado representantes de esas variables.

Podría ser que el peso aplicado al trépano es suficiente para generar una cantidad de recortes que el flujo de lodo es inhábil de retirar, ya sea por elevada reología o escaso Caudal o una combinación de ambas propiedades. La evacuación de recortes del fondo es incompleta y el trépano continúa remoliendo recortes, reduciendo su avance.

En estas circunstancias, cabría recomendar al Ingeniero de Perforación la modificación de las condiciones de flujo, reduciendo la reología y/o aumentando el Caudal de trabajo, observando el comportamiento de la tasa de penetración. Eventualmente, si los resultados de esta acción son propicios, será posible hacer nuevos ajustes. Sería razonable pensar que incrementando la limpieza del fondo se logren mayores tasas de penetración, lo que daría lugar al estudio del efecto de aplicar luego mayores pesos al trépano.

3.3.2 Regresión por Redes Neuronales Artificiales

Se analizó ROP600. El número de neuronas recomendado es $\frac{297}{2*16} < n < \frac{2*297}{16}$ y resulta $10 < n < 37$. Inicialmente, se adoptaron números de neuronas ocultas en ese rango, pero luego se redujo el número a 5 neuronas, debido a los mejores resultados obtenidos.

Se halló que con 12 o más neuronas en la capa oculta se lograron coeficientes de determinación muy cercanos a la unidad en el grupo de entrenamiento, pero coeficiente muy pobres en el grupo de validación, cercanos a 0. Esto se puede deber a un problema de sobre-entrenamiento. Se mejoró el coeficiente del grupo de validación reduciendo el número de neuronas en la capa oculta hasta 5, a costa de reducir el coeficiente de determinación en el grupo de entrenamiento. Los coeficientes resultantes se pueden ver en el Anexo B. En estas condiciones los mejores resultados obtenidos son:

	Coeficiente de determinación
Grupo de Entrenamiento	0.885
Grupo de Validación	0.393

Se destaca que en el grupo de validación el coeficiente es relativamente bajo, y esto indicaría que el modelo no es muy eficiente. Posiblemente esto se deba a la complejidad de las relaciones entre las variables y a que la cantidad de casos disponibles para este tipo de análisis resulte escasa.

Actualmente estos métodos se están aplicando en la industria en monitoreos en línea recibiendo y analizando la información colectada por sensores en los puntos estratégicos de la operación.

3.3.3 Regresión por Particiones Recursivas

El método de particiones recursivas (Recursive partitioning, paquete rpart de R) aplicado a la matriz de datos ofreció los resultados que se pueden ver en el Anexo C. Se analiza ROP600 y el árbol correspondiente se presenta en la Fig. 3.3-11. En cada nodo, la aceptación de la condición ofrecida por el nodo conduce por la rama de la izquierda. Ejemplo: Un pozo perforado por el rig “a” (OIL-01), a un caudal de 480 gpm, con una presión de bomba de 1100 psi y un peso de 5 tons, conduce por las ramas de la derecha en todos los nodos y el valor esperado de ROP600 sería de 38.9 m/hr.

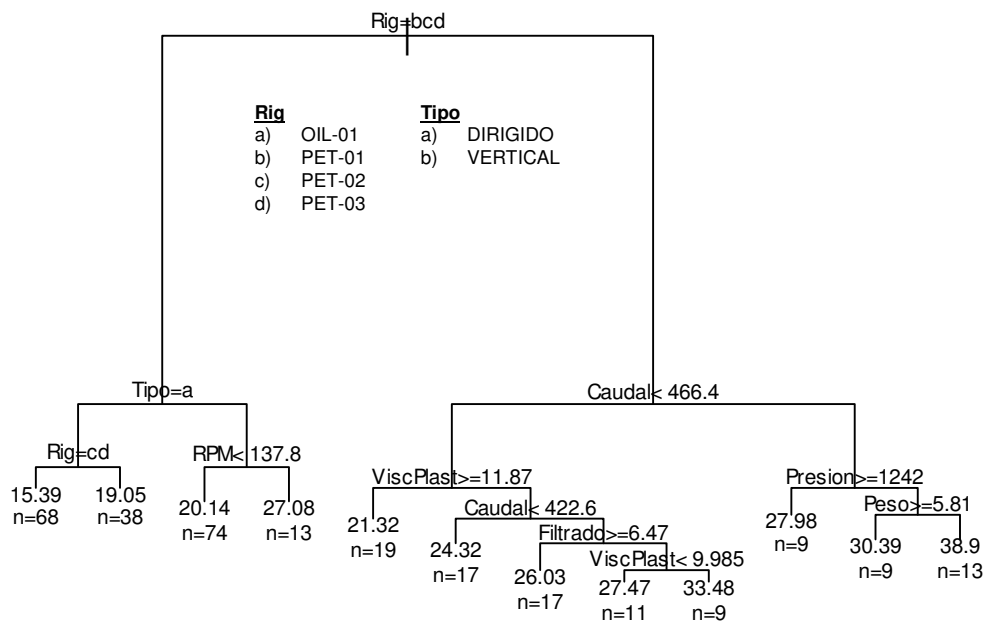


Fig. 3.3-11Árbol de regresión

En cada nodo, la condición del mismo conduce por la rama de la izquierda.

Se observa que el primer nodo se ejecuta sobre Rig, quedando los Rig PET-01, PET-02 y PET-03 en la rama izquierda y el Rig OIL-01 en la rama derecha. Esta partición absorbe una suma de cuadrados de residuos importante, representada por la altura de las ramas. En el nodo 1, es SCR=15336, mientras que la rama 2 preserva SCR=4117 y la rama 3 preserva 5594, sumando 9711. La partición absorbió 5625, o sea un tercio de la SCR del nodo 1. En las siguientes particiones se observan menores disminuciones de SCR.

En la Tabla 3.3-1 se presenta el parámetro de complejidad y en la Fig. 3.3-12 el correspondiente gráfico. Se probó una partición con sólo 4 divisiones, es decir con $cp > 0.029$

Tabla 3.3-1 Ceficiente de complejidad

CP	No.Part	Error Relat	Error cruzado	xstd
0.367	0	1.000	1.012	0.115
0.085	1	0.633	0.650	0.083
0.062	2	0.549	0.621	0.080
0.035	3	0.486	0.537	0.063
0.030	4	0.452	0.547	0.064
0.024	5	0.422	0.532	0.062
0.021	7	0.374	0.520	0.060
0.012	8	0.352	0.512	0.069
0.011	9	0.341	0.534	0.070
0.010	11	0.319	0.546	0.071

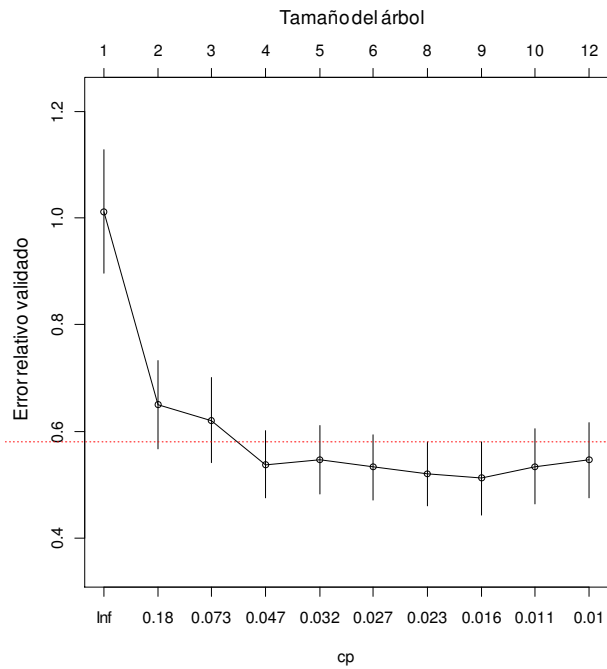


Fig. 3.3-12 Parámetro de complejidad de la partición

Por encima del nivel 4, no se obtienen mejoras importantes en el error.

El nuevo árbol se ve en la Fig. 3.3-13. La primera división es Rig=bcd (PET-01, PET-02, PET-03). Si es verdadero, la segunda división es por Tipo=a (Dirigido). Si es verdadero, la ROP-600 es de 16.7 m/hr. Es decir que estos tres equipos, perforando pozos dirigidos tienen la performance más baja. Cuando estos equipos perforan pozos verticales, tienen rendimientos relacionados a

las revoluciones de la mesa (RPM). Es de notar que el número de casos con altas RPM y mejor performance (27.8 m/hr) son solamente 13 pozos.

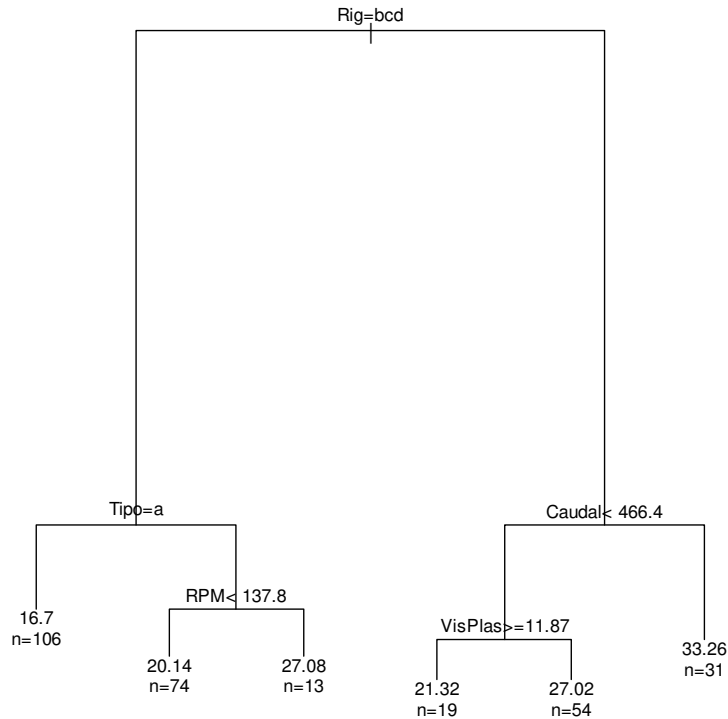


Fig. 3.3-13 Árbol podado

Por el otro lado, en la primera división, el equipo OIL-01 conduce por la rama derecha, donde un segundo nodo discrimina por Caudal. Caudales bajos están relacionados a ViscPlast con las tasas de perforación más bajas del equipo OIL-01. Caudales altos producen las ROP más altas.

3.3.1.5. Conclusiones de Particiones Recursivas

1. Hay diferencia de rendimientos entre los equipos OIL-01 y los equipos PET-01/02/ 03.
2. Los equipos PET-01/02/ 03 son sensiblemente más lentos perforando pozos dirigidos.
3. En los pozos verticales, los equipos PET-01/02/ 03 perforan a bajas RPM en la mayoría de los casos, lo que conduce a bajos rendimientos.
4. El rendimiento del equipo OIL-01 está condicionado a caudales bajos y éstos, a su vez, están relacionados a condiciones de reología (ViscPlast). Esto se condice con el análisis previo de componentes principales (Parr. 3.1.4).

5. A caudales altos, el equipo OIL-01 obtiene las mejores tasas de penetración.

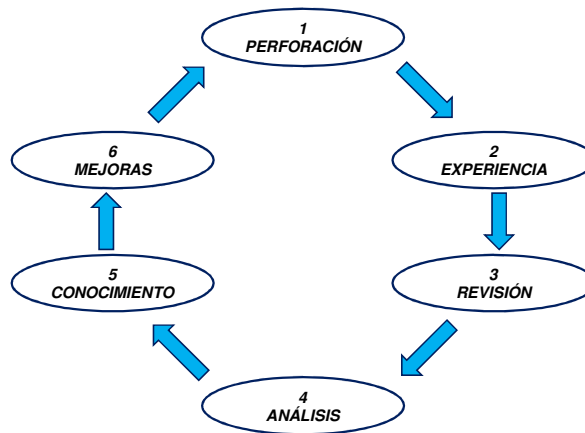
Estos resultados respaldan los obtenidos con el modelo lineal generalizado propuesto en cuanto que hay diferencias entre los equipos de perforación y en cuanto que las ROP600 están influenciadas por las propiedades de lodo y sobre todo por el caudal.

PARTE 4 - CONCLUSIONES

Capítulo 4.1 Resumen y Conclusiones

4.1.1 Desarrollo y perforación

La perforación de pozos es un proceso de **mejora continua**: Por un lado, intervienen la experiencia (cantidad de pozos perforados, cantidad de problemas encontrados y solucionados, cantidad y tipo de equipamientos y técnicas nuevas que se han introducido) y por otro lado, el conocimiento aportado por la revisión y el análisis de las experiencias. El siguiente esquema representa el ciclo:



En este estudio se delinearón procedimientos idóneos para la revisión y el análisis de los múltiples factores que intervienen en la perforación con el fin de detectar cuáles son los más importantes, de modo de poder optimizarlos. Otro objetivo es el desarrollo de modelos que permitan hacer predicciones de comportamiento futuro.

4.1.2 Exploración de los Datos

La base de datos utilizada comprende la actividad de perforación de 3 años y medio. Dado que los datos eran pre-existentes, se desarrolló un análisis **observacional** de los tiempos de perforación.

Se realizaron exploraciones sobre las variables ROC36 y ROP600. Para un estudio más detallado fue seleccionada la actividad Perforación Pura en la Fase 600 por las siguientes razones: (1) La perforación de la Fase 600 ocupa el 53.1% del tiempo, es decir, más de la mitad del tiempo, (2) Esta actividad contaba con suficientes datos adecuados. Análisis

adicionales sobre las otras actividades podrán seguir los mismos procedimientos, en caso de disponer de suficiente información.

4.1.3 Agrupamientos

En el presente caso se observó que los niveles de Rig y Tipo de pozo fueron los que determinaron los agrupamientos. Esto es totalmente coherente con la situación real en estudio. Intuitivamente o en base a la experiencia del ingeniero de perforación, es esperable que esos dos factores tengan una influencia o un peso considerable en el resultado de las perforaciones, sobretodo en cuanto a las tasas de penetración. Según las condiciones de perforación es total y razonablemente esperable que haya alguna diferencia entre los pozos Dirigidos y los Verticales, porque las herramientas, la técnica y la operativa son distintas. También es típico encontrar diferencias en la performance de los equipos, puesto que cada uno de ellos tiene características y desenvolvimientos propios.

La detección de estos 8 grupos, mayormente gobernados por los factores Rig y Tipo fue de gran utilidad a la hora de analizar el desempeño de las perforaciones. Se pueden comparar diversos equipos y obrar en consecuencia; por ejemplo, arbitrar los medios para mejorar la performance de los más lentos, eventualmente, separar de la operación a los más ineficientes e irremediables. Otra utilidad importantísima es la aplicación de estos grupos a la generación presupuestaria, ya que está claro que hay significativas diferencias según el equipo en operación y según el tipo de pozo que se perfore.

Los agrupamientos actuaron como un factor de aglomeración, lo que permitió seccionar los datos en grupos más homogéneos y facilitó el modelado y la interpretación. Estas dificultades las enfrentaron Noerager y otros sin solución, tal como se aprecia en la Fig. 1.2-8(Noerager, y otros, 1987). Los agrupamientos permitieron identificar los equipos más eficientes, de manera que el ingeniero de perforación puede tomar acciones para mejorar las performances o seleccionar los equipos (Rig) más aventajados. Además, los agrupamientos se pueden aplicar en la tarea de presupuestación, ya que discriminan diferencias de tiempos de perforación (performance) e indirectamente, los costos de perforación. En lugar de valerse de un solo pozo tipo como base presupuestaria, en este caso sería recomendable considerar 8 pozos tipo. El comportamiento de las variables también se analizó mediante Análisis de Componentes Principales.

4.1.4 Modelado

Se analizó el modelado de la variable tasa de penetración definida como variable (ROP600) o fraccionada en sus componentes, el tiempo de perforación (HrsF600) y la longitud del

tramo perforado (Drill600). Se trata de una variable que aparentemente responde a una distribución gamma, la cual tiene varianza no constante. Además, como se trata de una tasa, pone algunas condiciones al modelado, por ejemplo, no tolera respuestas negativas.

Esta situación se resolvió con un **modelo lineal generalizado sobre la variable tiempo de perforación (HrsF600), con distribución gamma, enlace log y con una variable offset del log de la variable tramo perforado (Drill600) con coeficiente unitario.**

Este modelo es un aporte completamente original cuyo mérito principal ha sido atribuir la distribución gamma a la variable de interés, la tasa de penetración. Esta característica valida el procedimiento, le da soporte y lo hace sólido. Este modelado se podría extender al resto de las operaciones de perforación.

También se verificaron modelos vía Redes Neuronales Artificiales y Árboles de Particiones Recursivas. Éstos últimos respaldaron las conclusiones del MLG.

En esta tesis se utilizó ampliamente el sistema R (R Development Core Team, 2009), lo que muestra la versatilidad y aptitud del sistema para tratar con este tipo de circunstancias.

4.1.5 Conclusiones

Se comenzó con técnicas tradicionales de exploración, se continuó estudiando los datos mediante técnicas de agrupamiento y finalmente se modeló una de las variables más importantes de la perforación. Este modelado permitió detectar cuáles son los factores determinantes en el rendimiento. Aquí también el ingeniero de perforación puede tomar iniciativas para mejorar la performance.

Los objetivos propuestos en el párrafo 1.1.6 se han logrado satisfactoriamente.

1. Se aplicaron técnicas de análisis de datos multivariados a la revisión de campañas de perforación. A saber:
 - 1.1. Exploración de datos por medios gráficos
 - 1.1.1. Histogramas de datos
 - 1.1.2. Densidades
 - 1.1.3. Boxplots
 - 1.1.4. Gráficos enrejadas y factorizados
 - 1.1.5. Matrices gráficas
 - 1.2. Análisis de la varianza, gráficos de diseño
 - 1.3. Comparaciones múltiples
 - 1.4. Análisis y comparación de distribuciones

- 1.5. Agrupamientos particionales
 - 1.5.1. Selección de número de grupos, agrupamientos
 - 1.5.2. Representaciones gráficas de agrupamientos (componentes principales bi y tridimensionales, densidad de matrices ordenadas)
 - 1.5.3. Evaluación de grupos (Siluetas, índices de disimilaridades, identificación de objetos y medoides de grupos)
- 1.6. Agrupamientos jerárquicos
 - 1.6.1. Aplicación de diversos métodos de partición
 - 1.6.2. Comparación de resultados, selección de un óptimo
- 1.7. Diagnóstico de agrupamientos
 - 1.7.1. Anova
 - 1.7.2. Comparaciones múltiples
- 1.8. Componentes principales: Análisis de variables
- 1.9. Modelos lineales Generalizados
 - 1.9.1. Análisis de interacciones
 - 1.9.2. Modelos gamma con variables dummy y variable offset
 - 1.9.3. Modelo completo y reducido
 - 1.9.4. Validación sobre conjunto de datos de control
 - 1.9.5. Interpretación
- 1.10. Redes Neuronales Artificiales
 - 1.10.1. Selección de neuronas ocultas
 - 1.10.2. Entrenamiento y validación
 - 1.10.3. Interpretación
- 1.11. Regresión por Particiones Recursivas
 - 1.11.1. Árbol completo, coeficiente de complejidad
 - 1.11.2. Poda de árboles
 - 1.11.3. Interpretación
- 2. Del punto de vista de la práctica de la perforación
 - 2.1. Se caracterizaron y clasificaron las perforaciones según su eficiencia y variables relacionadas al proceso de perforación de pozos.
 - 2.2. Se detectaron y evaluaron las variables determinantes en los tiempos de perforación.
 - 2.3. Se evaluaron modelos estadísticos para la predicción de tiempos de perforación.

3. Se desarrollaron e implementaron rutinas en lenguaje R para cada etapa de la revisión de campañas de perforación. Las rutinas (scripts) se pueden ver en el siguiente enlace:

<https://skydrive.live.com/redirect?page=view&resid=44F1F16F7978AFE7!927&authkey=!AHL5xzdNhrTheH8>

Anexo A - Agrupamientos

PARTICIONES: Pozos por grupo y pozos medoides

Grupo1	Grupo2	Grupo3	Grupo4	Grupo5	Grupo6	Grupo7
Pozo-001	Pozo-003	Pozo-008	Pozo-009	Pozo-021	Pozo-028	Pozo-136
Pozo-002	Pozo-005	Pozo-101	Pozo-013	Pozo-034	Pozo-042	Pozo-139
Pozo-006	Pozo-010	Pozo-128	Pozo-018	Pozo-036	Pozo-078	Pozo-153
Pozo-012	Pozo-014	Pozo-133	Pozo-023	Pozo-046	Pozo-084	Pozo-190
Pozo-019	Pozo-016	Pozo-160	Pozo-025	Pozo-142	Pozo-097	Pozo-192
Pozo-020	Pozo-022	Pozo-162	Pozo-026	Pozo-144	Pozo-104	Pozo-193
Pozo-031	Pozo-027	Pozo-180	Pozo-029	Pozo-148	Pozo-107	Pozo-194
Pozo-035	Pozo-052	Pozo-181	Pozo-030	Pozo-154	Pozo-117	Pozo-195
Pozo-040	Pozo-060	Pozo-187	Pozo-032	Pozo-166	Pozo-130	Pozo-196
Pozo-041	Pozo-063	Pozo-191	Pozo-033	Pozo-168	Pozo-135	Pozo-198
Pozo-044	Pozo-066	Pozo-197	Pozo-037	Pozo-170	Pozo-143	Pozo-200
Pozo-048	Pozo-071	Pozo-199	Pozo-038	Pozo-174	Pozo-150	Pozo-201
Pozo-050	Pozo-073	Pozo-203	Pozo-039	Pozo-178	Pozo-156	Pozo-202
Pozo-053	Pozo-085	Pozo-205	Pozo-043	Pozo-186	Pozo-157	Pozo-204
Pozo-054	Pozo-092	Pozo-208	Pozo-045	Pozo-188	Pozo-158	Pozo-207
Pozo-056	Pozo-099	Pozo-210	Pozo-047	Pozo-213	Pozo-165	Pozo-211
Pozo-057	Pozo-105	Pozo-217	Pozo-051	Pozo-266	Pozo-167	Pozo-216
Pozo-059	Pozo-109	Pozo-221	Pozo-055	Pozo-274	Pozo-171	Pozo-220
Pozo-064	Pozo-111	Pozo-223	Pozo-061	Pozo-276	Pozo-173	Pozo-222
Pozo-065	Pozo-113	Pozo-224	Pozo-062	Pozo-284	Pozo-175	Pozo-225
Pozo-067	Pozo-115	Pozo-228	Pozo-068	Pozo-289	Pozo-176	Pozo-226
Pozo-069	Pozo-116	Pozo-229	Pozo-072	Pozo-291	Pozo-179	Pozo-236
Pozo-070	Pozo-120	Pozo-230	Pozo-076	Pozo-299	Pozo-182	Pozo-241
Pozo-074	Pozo-121	Pozo-231	Pozo-080	Pozo-305	Pozo-185	Pozo-245
Pozo-075	Pozo-122	Pozo-232	Pozo-083	Pozo-306	Pozo-214	Pozo-255
Pozo-077	Pozo-123	Pozo-237	Pozo-086	Pozo-321	Pozo-235	Pozo-259
Pozo-079	Pozo-127	Pozo-238	Pozo-087	Pozo-332	Pozo-261	Pozo-260
Pozo-081	Pozo-129	Pozo-239	Pozo-088		Pozo-292	Pozo-268
Pozo-098	Pozo-131	Pozo-240	Pozo-091		Pozo-293	Pozo-275
Pozo-100	Pozo-141	Pozo-246	Pozo-093		Pozo-295	Pozo-280
Pozo-103	Pozo-146	Pozo-249	Pozo-094		Pozo-325	Pozo-281
Pozo-108	Pozo-184	Pozo-250	Pozo-095		Pozo-329	Pozo-283
Pozo-118	Pozo-324	Pozo-251	Pozo-096		Pozo-334	Pozo-286
Pozo-119	Pozo-328	Pozo-253	Pozo-106			Pozo-296
Pozo-126	Pozo-335	Pozo-254	Pozo-110			Pozo-297
Pozo-138	Pozo-338	Pozo-257	Pozo-112			Pozo-303
Pozo-140	Pozo-339	Pozo-258	Pozo-114			Pozo-313
Pozo-155	Pozo-341	Pozo-262	Pozo-124			Pozo-314
Pozo-163	Pozo-345	Pozo-264	Pozo-132			Pozo-318
Pozo-177	Pozo-347	Pozo-265	Pozo-149			Pozo-320
Pozo-242	Pozo-351	Pozo-285	Pozo-151			Pozo-354
Pozo-294	Pozo-352	Pozo-288	Pozo-172			
Pozo-300		Pozo-290	Pozo-183			
Pozo-307		Pozo-301	Pozo-277			
Pozo-317		Pozo-304	Pozo-302			
Pozo-326		Pozo-312	Pozo-308			
Pozo-327		Pozo-319	Pozo-309			
Pozo-330		Pozo-333	Pozo-311			
Pozo-331		Pozo-343	Pozo-315			
Pozo-337			Pozo-316			
Pozo-340			Pozo-322			
Pozo-344			Pozo-323			
Pozo-348						
53	42	49	52	27	33	41

Anexo B - Red Neuronal Artificial

Red con una capa de entrada de 16 neuronas, 1 capa oculta de 5 neuronas y una capa de salida de 1 neurona.

Coefficientes de la red

b → h1	i1 → h1	i2 → h1	i3 → h1	i4 → h1	i5 → h1	i6 → h1	i7 → h1	i8 → h1	i9 → h1	i10 → h1	i11 → h1	i12 → h1	i13 → h1	i14 → h1	i15 → h1	i16 → h1
-9.98	1.65	2.26	29.2	56.73	-54.03	43.44	47.64	10.48	7.8	4.42	55.09	-11.85	-11.49	-38.98	-18.99	18.02
b → h2	i1 → h2	i2 → h2	i3 → h2	i4 → h2	i5 → h2	i6 → h2	i7 → h2	i8 → h2	i9 → h2	i10 → h2	i11 → h2	i12 → h2	i13 → h2	i14 → h2	i15 → h2	i16 → h2
-20.57	-9.14	-10.33	3.37	39.71	30.53	10.46	27.56	-31.93	41.36	21.25	-13.19	7.71	-6.69	14.17	26.27	-27.24
b → h3	i1 → h3	i2 → h3	i3 → h3	i4 → h3	i5 → h3	i6 → h3	i7 → h3	i8 → h3	i9 → h3	i10 → h3	i11 → h3	i12 → h3	i13 → h3	i14 → h3	i15 → h3	i16 → h3
-22.4	18.13	-11.32	-40.18	3.78	20.16	-8.85	-30.09	-1.24	1.51	-0.86	4.95	22.26	-27.72	1.28	9.33	-9.14
b → h4	i1 → h4	i2 → h4	i3 → h4	i4 → h4	i5 → h4	i6 → h4	i7 → h4	i8 → h4	i9 → h4	i10 → h4	i11 → h4	i12 → h4	i13 → h4	i14 → h4	i15 → h4	i16 → h4
86.5	8.16	-1	3.72	19.26	0.48	24.73	-17.64	24.42	-29.56	-3.59	-13.05	9.07	12.74	-2.78	20.7	-20.6
b → h5	i1 → h5	i2 → h5	i3 → h5	i4 → h5	i5 → h5	i6 → h5	i7 → h5	i8 → h5	i9 → h5	i10 → h5	i11 → h5	i12 → h5	i13 → h5	i14 → h5	i15 → h5	i16 → h5
15.83	37.63	-32.72	43.7	17.74	-35.65	-7.16	13.02	29.17	-4.43	-62.02	25.19	10.29	2.22	-35.71	5.16	-5.4
b → o	h1 → o	h2 → o	h3 → o	h4 → o	h5 → o											
2.04	1.08	-0.83	0.73	-2.3	-0.57											

Nomenclatura:

b: Bias (sesgo o intercepto)

i: Input (Entrada)

h: Hiden (oculta)

o: Output (salida)

Fórmula de operación

`nnet.formula(formula = ROP600 ~ ., data = spr, size = 5, linout = TRUE, trace = FALSE)`

Significado:

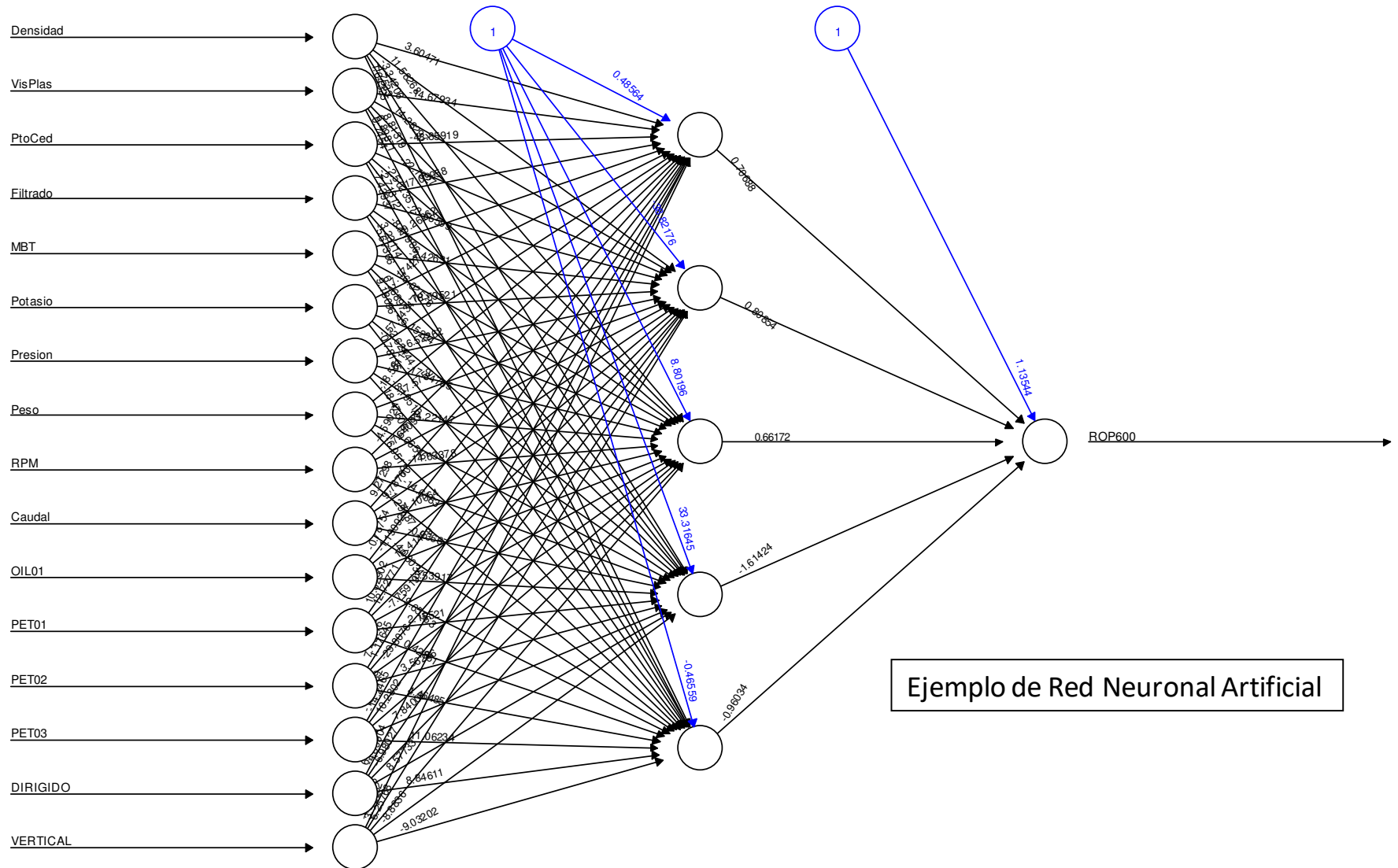
b → h1 representa el coeficiente de bias a la primera neurona escondida

i3 → h5 representa el coeficiente de la tercera neurona de entrada a la quinta neurona escondida

b → o representa el coeficiente de bias de la capa escondida a la capa de salida

h2 → o representa el coeficiente de la segunda neurona escondida a la neurona de salida

ANEXO B – RED NEURONAL ARTIFICIAL



Anexo C - Partición Recursiva

n= 297

Nodo	Partición	n	Devianza	Valor y	
1)	Raiz	297	15335.7	21.9	
2)	Rig=PET-01,PET-02,PET-03	193	4117.0	18.7	
4)	Tipo=DIRIGIDO	106	1089.0	16.7	
8)	Rig=PET-02,PET-03	68	392.3	15.4	*
9)	Rig=PET-01	38	370.2	19.0	*
5)	Tipo=VERTICAL	87	2071.0	21.2	
10)	RPM<137.845	74	1064.5	20.1	*
11)	RPM>=137.845	13	473.9	27.1	*
3)	Rig=OIL-01	104	5594.2	27.8	
6)	Caudal<466.43	73	2525.4	25.5	
12)	VisPlas>=11.865	19	330.7	21.3	*
13)	VisPlas<11.865	54	1738.1	27.0	
26)	Caudal<422.57	17	537.4	24.3	*
27)	Caudal>=422.57	37	1018.8	28.3	
54)	Filtrado>=6.47	17	438.7	26.0	*
55)	Filtrado<6.47	20	422.4	30.2	
110)	VisPlas<9.985	11	176.9	27.5	*
111)	VisPlas>=9.985	9	66.9	33.5	*
7)	Caudal>=466.43	31	1772.7	33.3	
14)	Presion>=1241.61	9	147.1	28.0	*
15)	Presion<1241.61	22	1272.4	35.4	
30)	Peso>=5.81	9	197.9	30.4	*
31)	Peso<5.81	13	689.4	38.9	*

Bibliografía

A New Approach for the Prediction of Rate of Penetration (ROP) Values.

Bilgesu, H. I., Altmis, U. y Mohaghegh, S. 1997. 1997, S.P.E.

Adeleye, A.R., y otros. 2004.*Improving Drilling Performance Through Systematic Analysis of Historical Data: Case Study of a Canadian Field.* s.l. : IADC/SPE, 2004. IADC/SPE 87177.

Agresti, A. 2002.*Categorical Data Analysis.* New York : Wiley, 2002.

Algoritmos de Partición en el Análisis de Conglomerados: Un Estudio Comparativo. **Leiva, Susana y Torres, Francisco. 2008.** 2008, VIII CLATSE.

An Interactive Drilling Simulator for Teaching and Research. **Cooper, G. A., Cooper, Alexander G. y Bihn, Gregory. 1995.** 1995, S.P.E.

Arnold, M. S., y otros. 1991.*How a Computerized Drilling System has Improved Drilling Performance.* s.l. : SPE/IADC, 1991. SPE/IADC 21925.

Bernardis, A. 2005.*Elaboración De Un Proyecto De Tesis.* 2005.

Bond, D. F., y otros. 1996.*Step Change Improvement and High Rate Learning are Delivered by Targeting Technical Limits on Sub-Sea Wells.* s.l. : IADC/SPE, 1996. IADC/SPE 35077.

Bourgoyne, A. T. y Young, F. S. 1974.*A Multiple Regression Approach to Optimal Drilling and Abnormal Pressure Detection.* s.l. : S.P.E., 1974.

Bourgoyne, A., y otros. 1986.*Applied Drilling Engineering.* Richardson, TX : Society of Petroleum Engineers, 1986.

Box, G. P., Hunter, W. G. y Hunter, J. S. 1978.*Statistics for Experimenters.* New York : Wiley, 1978.

Bramardi, S. 2004.*Apuntes de Análisis Multivariado.* Neuquén : s.n., 2004.

Brett, J. F. y Millheim, K. K. 1986.*The Drilling Performance Curve: A Yardstick for Judging Drilling Performance.* s.l. : SPE, 1986. SPE 15362.

Brett, J. y Yoder, D. L. 1985.*Use of Drilling Simulator for Drilling and Designing Drilling Rigs.* Richardson, TX : S.P.E, 1985.

Canavos, G.C. 1998.*Probabilidad y Estadística.* s.l. : McGraw-Hill, 1998.

Comparison of the Penetration Rate Models Using Field Data for One of the Gas Fields in Persian Gulf Area. **Rahimzadeh, H., y otros. 2010.** 2010, S.P.E.

Cook, R. D. y Cross-Dabrera, R. 1998.*Partial Residual Plots in Generalized Linear Models.* 1998. págs. 442, 730-739.

Cook, R. Dennis., 1977.*Detection of Influential Observation in Linear Regression.* 1977.

Cook, R. Dennis. 1979.*Influential Observations in Linear Regression.* 1979. págs. 365, 169-174.

Crawley, Michael J. 2007.*The R book.* Chichester, England : John Wiley & Sons Ltd, 2007.

Curry, D., y otros. 2005.*Technical Limit Specific Energy: An Index to Facilitate Drilling Performance Evaluation.* s.l. : SPE/IADC, 2005. SPE/IADC 92318.

Davidian, M. 2007.*Applied Longitudinal Data Analysis.* s.l. : NCSU, 2007.

Dobson, A. 2002.*An Introduction to Generalized Linear Models.* London : Chapman and Hall, 2002.

Drilling Optimization Based on the ROP Model in One of the Iranian Oil Fields. **Mostofi, M., y otros. 2010.** 2010, S.P.E.

Drilling Optimization Software Verified in the North Sea. **Bratli, R. K., y otros. 1997.** 1997, S.P.E.

Drilling Optimization Technique - Using Real Time Parameters. **Iqbal, Fauzan. 2008.** 2008, S.P.E.

Drilling Optimization Using Data and Available Technology. **Rampersad, P. R., Hareland, G. y Boonyapaluk, P. 1994.** 1994, S.P.E.

Eckel, John R. 1968.*Microbit Studies of the Effect of Fluid Properties and Hydraulics on Drilling Rate.* Dallas, Texas, USA : SPE, 1968.

Everitt, B. S. y Dunn, G. 2001.*Applied Multivariate Data Analysis.* London : Arnold, 2001.

Everitt, Brian S. y Hothorn, Torsten. 2006.*A Handbook of Statistical Analysis Using R.* Boca Raton : Chapman & Hall/CRC, 2006.

- Everitt, Brian S., y otros. 2011.***Cluster Analysis*. West Sussex, United Kingdom : John Wiley & Sons Ltd, 2011.
- Faraway, J. J. 2006.***Extending the Linear Model with R*. Boca Raton : Chapman & Hall/CRC, 2006.
- Faraway, J. 2004.***Linear Models with R*. Boca Raton : Chapman and Hall/CRC, 2004.
- Fayyad, U. M., y otros. 1996.***Advances in Knowledge Discovery and Data Mining*. Menlo Park : AAAI Press / The MIT Press, 1996.
- Fernández Loureiro de Pérez, Emma. 2001.***Teoría de los conjuntos borrosos*. Buenos Aires : Ediciones Cooperativas, 2001.
- Freund, J.E. y Walpole, R.E. 1990.***Estadística Matemática con Aplicaciones*. s.l. : Prentice-Hall, 1990.
- Gabriel, K. R. 1971.***The Biplot Graphical Display of Matrices with Application to Principal Components Analysis*. 1971. págs. 58, 453-467.
- Gelman, A. y Hill, J. 2007.***Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge : Cambridge University Press, 2007.
- Gower, J.C. 1971.***A General Coefficient of Similarity and Some of Its Properties*. s.l. : Biometrics, 27(4), 1971. págs. 857-871.
- Hahsler, Michael y Hornik, Kurt. 2009.***Dissimilarity PLOTS: A Visual Exploration Tool for Partitional Clustering*. Wien : Department of Statistics and Mathematics-Wirtschafts Universität, 2009.
- Hair, Joseph F., y otros. 1999.***Análisis Multivariante*. Madrid : Prentice Hall Iberia, 1999. ISBN 84-8322-035-0.
- Hampel, F.R., y otros. 1986.***Robust Statistics: The Approach Based on Influence Functions*. New York : Wiley, 1986.
- Han, J. y Kamber, M. 2006.***Data Mining: Concepts and Techniques*. San Francisco, CA, USA : Morgan Kaufmann Publishers, 2006.
- Hartigan, J. A. 1975.***Clustering Algorithms*. New York : John Wiley & Sons, 1975.
- How to Improve Rate of Penetration in Field Operations.* **Fear, M. J. 1999.** 1999, S.P.E.

Identifying Inefficient Drilling Conditions Using Drilling Specific Energy. **Armenta, Miguel.** 2008. 2008, S.P.E.

Ikoku, Chi U. 1978. *Application of Learning Curve Models to Oil and Gas Well Drilling.* s.l. : SPE-AIME, 1978. SPE 7119.

Implementation of ROP Management Process in Qatar North Field. **Remmert, Stephen M., Witt, Joseph W. y Dupriest, Fred E.** 2007. 2007, S.P.E.

Integrated Petroleum Engineering Simulation and Decision Making Program. **Cooper, George A., Mota, Jose F. y Cooper, Alexander G.** 1996. 1996, S.P.E.

Irrgang, R., y otros. 1999. *A Case-Based System to Cut Drilling Costs.* s.l. : SPE, 1999. SPE 56504.

Iyoho, A. W., y otros. 2005. *Lessons From Integrated Analysis of GOM Drilling Performance.* s.l. : SPE, 2005. SPE 97464.

Iyoho, A. W., y otros. 2004. *Methodology and Benefits of a Drilling Analysis Paradigm.* s.l. : IADC/SPE, 2004. IADC/SPE 87121.

Iyoho, W., y otros. 2005. *Managing Drilling Performance: Technical Analysis Well Challenges and Field Engagement.* s.l. : SPE, 2005. SPE 95314.

Johnson, D. 2000. *Métodos multivariados aplicados al análisis de datos.* Mexico : International Thomson Editores, 2000.

Kadaster, A.G., Townsend, C.W. y Albaugh, E.K. 1992. *Drilling Time Analysis: A Total Quality Management Tool for Drilling in the 1990's.* s.l. : SPE, 1992. SPE 24559.

Kaufman, L. y Rousseeuw, P.J. 2005. *Finding Groups in Data - An Introduction to Cluster Analysis.* Hoboken, New Jersey : John Wiley & Sons, Inc., 2005. ISBN 0-471-73578-7.

Koederitz, William L. y Weis, Jeff. 2005. *A Real Time Implementation of MSE.* Houston, TX : AADE, 2005.

Kravis, S., y otros. 2004. *The Trouble Probability Plot: A New Measure of Drilling Maturity.* s.l. : SPE, 2004. SPE 89959.

Lagreca, A. J., y otros. 2004. *The Role of Knowledge, Tools and Systems for Drilling Analysis.* Kuala Lumpur : SPE, 2004. IADC/SPE 87996.

Landmark. 2000. *DIMS-Drilling Information Management System.* 2000.

- Lindsey, J. K. 1997.***Applying Generalized Linear Models*. New York : Springer, 1997.
- Lummus, J. L. 1971.***Acquisition and Analysis of Data for Optimized Drilling*. s.l. : SPE-AIME, 1971. SPE 3716.
- Lummus, James L. 1970.***Drilling Optimization*. s.l. : SPE-AIME, 1970. SPE 2744.
- Maechler, M., y otros. 2005.***Cluster Analysis Basics and Extensions*. s.l. : unpublished, 2005.
- Maindonald, J. y Braun, J. 2003.***Data Analysis and Graphics Using R*. Cambridge, UK : Cambridge University Press, 2003.
- Martín Martín, Quintín y De Paz Santana, Yanira del Rosario. 2007.***Aplicación de las Redes Neuronales Artificiales a la Regresión*. Madrid : La Muralla, 2007.
- McCullagh, W. y Nelder, J. 1989.***Generalized Linear Models*. London : Chapman and Hall, 1989.
- McCulloch, W. y Searle, S. 2002.***Generalized, Linear, and Mixed Models*. New York : Wiley, 2002.
- Mendenhall, W. y Sincich, T. 1997.***Probabilidad y Estadística para Ingeniería y Ciencias*. s.l. : Prentice-Hall Hispanoamericana, S.A., 1997.
- Millheim, K. K. y Huggins, R. L. 1983.***An Engineering Simulator for Drilling: Part I*. Dallas, TX : S.P.E., 1983.
- Millheim, K., Maidla, E. y Kravis, S. 1998.***An Example of the Drilling Analysis Process for Extended Reach Wells*. s.l. : SPE, 1998. SPE 49111.
- Millheim, Keith K. y Huggins, R. L. 1983-2.***An Engineering Simulator for Drilling: Part II*. Dallas, TX : S.P.E., 1983-2.
- Montgomery, D.C. y Runger, G.C. 1996.***Probabilidad y Estadística aplicadas a la Ingeniería*. s.l. : McGraw-Hill, 1996.
- Montgomery, Douglas C., Peck, Elizabeth A. y Vining, G. Geoffrey. 2004.***Introducción al Análisis de Regresión Lineal*. México : Compañía Editorial Continental S.A., 2004.
- Moscoloni, N. 2005.***Apuntes de Metodología de la Investigación*. 2005.

Multiple Regression Approach To Optimize Drilling Operations in the Arabian Gulf Area. **Al-Betairi, Emad A., Moussa, Mohamed M. y Al-Otaibi, Saud. 1988.** 1988, S.P.E.

Murrell, P. 2005.*R Graphics.* Florida : Chapman and Hall/CRC, 2005.

Myatt, Glenn J. 2007.*Making Sense of Data - A Practical Guide to Exploratory Data Analysis and Data Mining.* Hoboken, New Jersey : John Wiley & Sons, Inc., 2007. ISBN-13: 978-0-470-07471-8.

Nakagawa, E. y Miura, K. 2005.*What is the Source of Drilling and Completion Data.* s.l. : SPE, 2005. SPE 93822.

Nelder, J. y Wedderburn, R. 1972.*Generalized Linear Models.* 1972. págs. 370-384.

Noerager, J.A., y otros. 1987.*Drilling Time Predictions From Statistical Analysis.* s.l. : SPE/IADC, 1987. SPE/IADC 16164.

Optimization Techniques for Minimizing Drilling Costs. **Wilson, David C. y Bentsen, Ramon G. 1972.**1972, S.P.E.

Peña, Daniel. 2002.*Análisis de datos multivariantes.* Madrid : McGraw-Hill, 2002.

Pérez, C. 2004.*Técnicas de Análisis Multivariante de Datos.* Madrid : PEARSON EDUCACIÓN, 2004.

Peterson, S. K., Murtha, J. A. y Roberts, R. W. 1995.*Drilling Performance Predictions: Case Studies Illustrating The Use of Risk Analysis.* s.l. : SPE/IADC, 1995. SPE/IADC 29364.

Pinhero, J. C. y Bates, D. M. 2000.*Mixed-Effects Models in S and S-PLUS.* New York : Springer, 2000.

Pison, Greet, Struyf, Anja y Rousseeuw, Peter J. 1999.*Displaying a Clustering With CLUSPLOT.* 1999. págs. 381-392.

R Development Core Team. 2009.*R: A Language and Environment for Statistical Computing.* Vienna, Austria : R Foundation for Statistical Computing, 2009. ISBN 3-900051-07-0.

Rabe-Hesketh, S. y Skrondal, A. 2008.*Miltilevel and Longitudinal Modeling Using Stata.* College Station : Stata Press, 2008.

- Rodríguez, G. 2007.** Lecture Notes on Generalized Linear Models. [En línea] 2007. <http://data.princeton.edu/wws509/notes/> .
- Ross, S.M. 2002.** *Probabilidad y Estadísticas para Ingeniería y Ciencias*. s.l. : McGraw-Hill, 2002.
- Samaja, J. 2004.** *Epistemología y metodología: elementos para una teoría de la investigación científica*. Buenos Aires : EUDEBA, 2004.
- Schabenberger, Oliver y Pierce, Francis J. 2002.** *Contemporary statistical models for the plant and soil sciences*. s.l. : CRC Press LLC, 2002.
- Scheffé, H. 1959.** *The Analysis of Variance*. New York : Wiley, 1959.
- Sneath, P. H. A y Sokal, R. R. 1973.** *Numerical Taxonomy*. San Francisco : W. H. Freeman and Company, 1973.
- Sokal, R. R. y Rohlf, J. 1962.** *The Comparison of Dendrograms by Objective Methods*. s.l. : Taxon, 1962.
- Sokal, R.R. y Michener, C.D. 1958.** *A Statistical Method for Evaluating Systematic Relationships*. 1958. págs. 38, 1409-1438.
- Taboga, Marco. 2010.** Lectures on probability and statistics. [En línea] 2010. <http://www.statlect.com..>
- Tan, P. N., Steinbach, M. y Kumar, V. 2005.** *Introduction to Data Mining*. s.l. : Addison-Wesley, 2005.
- Team, R Development Core. 2005.** An Introduction to R. <http://www.R-project.org>. [En línea] 2005. <http://www.R-project.org>.
- The "Perfect-Cleaning" Theory of Rotary Drilling.* **Maurer, W. C. 1962.** s.l. : S.P.E., 1962, Journal of Petroleum Technology.
- Tukey, J. 1977.** *Exploratory Data Analysis*. New York : Addison Wesley, 1977.
- Tukey, J. W. 1953.** *The problem of multiple comparisons*. Princeton University : Unpublished manuscript, 1953.
- Uriel Jiménez, Ezequiel y Aldás Manzano, Joaquín. 2005.** *Análisis multivariante aplicado*. Madrid : Thomson, 2005.

BIBLIOGRAFÍA

Valdez, H. y Sager, J. 2005.*Benchmarking Drilling Performance: Achieving Excellence in MODU's Operating Practices for Deep Water Drilling.* s.l. : SPE/IADC, 2005. SPE/IADC 92235.

Venables, W. y Ripley, B. 2002.*Modern Applied Statistics with S.* New York : Springer, 2002.

Vercellis, Carlo. 2009.*Business Intelligence: Data Mining and Optimization for Decision Making.* s.l. : John Wiley & Sons, Ltd., 2009.

Verzani, J. 2004.*Using R for Introductory Statistics.* Boca Raton : Chapman and Hall/CRC, 2004.

Vose Software. 2007. Help File for ModelRisk Version 4. [En línea] 2007. <http://www.vosesoftware.com>.

Wood, S. 2006.*An Introduction to Generalized Additive Models.* Boca Raton : CRC Press, 2006.