

**DETERMINACIÓN DE LA RELACIÓN ENTRE LOS PARÁMETROS  
PETROFÍSICOS Y DE PRODUCCIÓN MEDIANTE UN ANÁLISIS DE  
SENSIBILIDAD A CUATRO CLUSTERS DEL CAMPO IMPERIAL**

**CARLOS ALBERTO ARCINIEGAS RIVAS**

**FUNDACIÓN UNIVERSIDAD DE AMÉRICA  
FACULTAD DE INGENIERÍAS  
PROGRAMA DE INGENIERÍA DE PETRÓLEOS  
BOGOTÁ D.C  
2020**

**DETERMINACIÓN DE LA RELACIÓN ENTRE LOS PARÁMETROS  
PETROFÍSICOS Y DE PRODUCCIÓN MEDIANTE UN ANÁLISIS DE  
SENSIBILIDAD A CUATRO CLUSTERS DEL CAMPO IMPERIAL**

**CARLOS ALBERTO ARCINIEGAS RIVAS**

**Proyecto Integral de Grado para optar al título de:  
INGENIERO DE PETRÓLEOS**

**Director:  
Gilberto Alejandro Méndez Bohórquez  
Ingeniero de Petróleos**

**FUNDACIÓN UNIVERSIDAD DE AMÉRICA  
FACULTAD DE INGENIERÍAS  
PROGRAMA DE INGENIERÍA DE PETRÓLEOS  
BOGOTÁ D.C  
2020**

Nota de aceptación

---

---

---

---

---

---

Firma del presidente del jurado

---

Firma de jurado 1

---

Firma de jurado 2

Bogotá D.C., Mayo de 2020

## **DIRECTIVAS DE LA UNIVERSIDAD**

Presidente de la Universidad y Rector del Claustro

Dr. Mario Posada García-Peña

Consejero Institucional

Dr. Luis Jaime Posada García-Peña

Vicerrectora Académica y de Investigaciones

Dra. María Claudia Aponte González

Vicerrector Administrativo y Financiero

Dr. Ricardo Alfonso Peñaranda Castro

Secretario General

Dra. Alexandra Mejía Guzmán

Decano de la Facultad de Ingenierías

Ing. Julio Cesar Fuentes Arismendi

Director Programa Ingeniería de Petróleos.

Ing. Juan Carlos Rodríguez Esparza

Las directivas de la Universidad de América, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento. Estos corresponden únicamente a los autores.

## **DEDICATORIA**

Dedico este trabajo de grado y el logro de ser profesional, a mis padres Gloria y Alberto por ser mi motivación y ejemplos por seguir, a mis hermanos Trilce, Jaime Leonardo y Sebastián por brindarme una mano siempre que lo necesite.

A mi director de tesis Gilberto Alejandro Méndez Bohórquez y al equipo de petrofísica, Edith Coromoto y Alexandra Mesa por haberme brindado la oportunidad de trabajar con ellos y de continuar siendo parte de sus proyectos con este trabajo.

A todas esas personas especiales que conocí en la Fundación Universidad América que día a día me hicieron una mejor persona, a Sonia Amaya y Diego Rodríguez por su incondicionalidad y apoyo. A mis amigos que fueron una gran ayuda por sus grandes conocimientos, David Martin, Nicolas Cabrera y Felipe Hernández.

CARLOS ALBERTO ARCINIEGAS

## **AGRADECIMIENTOS**

El agradecimiento póstumo a mi abuela Rosalba es el más especial para mí, debido a que fue una persona maravillosa con un amor incondicional. Agradezco a mi familia, que han sido siempre mi motivación para seguir adelante, a mis padres Gloria y Alberto por haberme brindado tanto amor, a mis hermanos Trilce, Jaime Leonardo y Sebastián por ser esas personas tan ejemplares y a mi tía Margarita por su apoyo incondicional.

Quiero agradecerle de forma especial a mi director de tesis Gilberto Alejandro Méndez Bohórquez por haberme dado la oportunidad de formar parte de su equipo y sus proyectos que me hicieron crecer como persona y profesional.

A esos amigos con los que pase buenos momentos, Sonia, Diego R., Diego M., Martín, Iván, Juan, Felipe, Justin, Daniel, Alejandra, Fabio, Gabriela y a todas las personas que me acompañaron durante esta etapa, muchas gracias.

CARLOS ALBERTO ARCINIEGAS

## CONTENIDO

	pág.
INTRODUCCIÓN	22
1. MARCO TEÓRICO	25
1.1 PARÁMETROS PETROFÍSICOS Y OTROS	25
1.1.1 Porosidad.	25
1.1.2 Permeabilidad.	26
1.1.3 Volumen de arcilla y modelo de gamma ray lineal.	27
1.1.4 Net pay.	27
1.1.5 Modelo dual water.	28
1.1.6 Coeficiente de heterogeneidad.	28
1.1.7 Ubicación del pozo.	29
1.2 PARÁMETROS DE PRODUCCIÓN	30
1.2.1 Producción acumulada y pico máximo de producción.	30
1.2.2 Índice de productividad.	30
1.2.3 Presión en cabeza de pozo (THP).	31
1.2.4 Agua y sedimentos (BSW).	31
1.2.5 Estimado de recobro final (EUR).	31
1.3 REGRESIÓN LINEAL MÚLTIPLE	32
1.3.1 Coeficiente de determinación (R cuadrado).	32
1.3.2 Potencia de la prueba (p-valor).	33
1.3.3 Coeficiente de correlación lineal de Pearson.	33
1.3.4 Selección de los mejores predictores.	34
1.3.5 Colinealidad.	34
1.3.6 Normalidad.	34
1.3.7 Homocedasticidad.	34
1.3.8 Autocorrelación.	35
1.3.9 Observaciones influyentes.	35
1.3.10 Coeficientes estandarizados.	36
1.4 COLUMNA ESTRATIGRÁFICA	36



2. METODOLOGÍA Y DATOS	39
2.1 PARÁMETROS DE SELECCIÓN DE LA MUESTRA	41
2.2 SELECCIÓN DE LAS MUESTRAS	41
2.2.1 Muestra de estudio.	41
2.2.2 Muestra de validación.	41
2.3 COMPILACIÓN DE LOS DATOS	42
2.3.1 Heterogeneidad.	43
2.3.2 Distancia con el acuífero.	43
2.3.3 Días de producción.	43
2.3.4 Pico máximo de producción.	43
2.4 CONCATENACIÓN DE LOS DATOS	46
2.4.1 Modelo de tabla completa.	46
2.4.2 Modelos de promedio y mediana.	46
2.5 PROGRAMACIÓN EN EL LENGUAJE “RSTUDIO”	47
2.6 CÁLCULO DE LOS PARÁMETROS DE PRODUCCIÓN	52
2.7 COMPARACIÓN DE LOS RESULTADOS	53
2.8 ESTABLECER LAS CONCLUSIONES	53
3. RESULTADOS	54
3.1 DESARROLLO DE LAS REGRESIONES LINEALES MÚLTIPLES	54
3.2 ANÁLISIS DE LA RELACIÓN ENTRE PARÁMETROS PETROFÍSICOS PARA EL MODELO PROMEDIO	56
3.3 SELECCIÓN DE LOS MEJORES PREDICTORES PARA EL MODELO PROMEDIO	58
3.4 VALIDACIÓN DE LA NORMALIDAD PARA EL MODELO PROMEDIO CON MEJORES PREDICTORES	71
3.5 VALIDACIÓN DE LA HOMOCEDASTICIDAD PARA EL MODELO PROMEDIO CON MEJORES PREDICTORES	80
3.6 VALIDACIÓN DE LA AUTOCORRELACIÓN PARA EL MODELO PROMEDIO CON MEJORES PREDICTORES	83
3.7 IDENTIFICACIÓN DE VALORES ATÍPICOS PARA EL MODELO PROMEDIO CON MEJORES PREDICTORES	84
3.8 IDENTIFICACIÓN DE OBSERVACIONES INFLUYENTES PARA EL	

MODELO PROMEDIO CON MEJORES PREDICTORES	86
3.9 BETAS ESTANDARIZADOS PARA EL MODELO PROMEDIO CON MEJORES PREDICTORES	88
3.10 RESULTADOS DEL MODELO PROMEDIO CON MEJORES PREDICTORES	95
3.11 VALIDACIÓN DEL MODELO PROMEDIO CON MEJORES PREDICTORES	97
3.12 COMPORTAMIENTO TEMPORAL DEL MODELO	101
4. CONCLUSIONES	102
5. RECOMENDACIONES	103
BIBLIOGRAFÍA	104
ANEXOS	106

## LSTA DE CUADROS

	pág.
<b>Cuadro 1.</b> Comandos del Código	50
<b>Cuadro 2.</b> Parámetros petrofísicos en modelo con mejores predictores	59

## LISTA DE ECUACIONES

	pág.
<b>Ecuación 1.</b> Porosidad	25
<b>Ecuación 2.</b> Ley de Darcy	26
<b>Ecuación 3.</b> Índice de Gamma Ray	27
<b>Ecuación 4.</b> Saturación de agua total	28
<b>Ecuación 5.</b> Heterogeneidad	29
<b>Ecuación 6.</b> Índice de productividad	30
<b>Ecuación 7.</b> BSW	31
<b>Ecuación 8.</b> Regresión lineal múltiple	32
<b>Ecuación 9.</b> Coeficiente de determinación	33
<b>Ecuación 10.</b> Valor crítico	33
<b>Ecuación 11.</b> Coeficiente de Pearson	33
<b>Ecuación 12.</b> Factor de inflación de varianza (VIF)	34
<b>Ecuación 13.</b> Estadístico W	34
<b>Ecuación 14.</b> Regresión de residuos	35
<b>Ecuación 15.</b> Estadístico D-W	35
<b>Ecuación 16.</b> Leverages (hat)	35
<b>Ecuación 17.</b> Distancia Cook	36
<b>Ecuación 18.</b> Coeficientes estandarizados	36
<b>Ecuación 19.</b> Ecuación modelo de la regresión lineal múltiple.	52

## LISTA DE FIGURAS

	pág.
<b>Figura 1.</b> Porosidad de la roca	26
<b>Figura 2.</b> Net pay	28
<b>Figura 3.</b> Curva de Lorenz	29
<b>Figura 4.</b> BSW	31
<b>Figura 5.</b> Formaciones en el campo Imperial	37
<b>Figura 6.</b> Columna estratigráfica	38
<b>Figura 7.</b> Diagrama de flujo del trabajo de grado	39
<b>Figura 8.</b> Datos de los parámetros petrofísicos de la muestra de estudio	44
<b>Figura 9.</b> Datos de los parámetros de producción de la muestra de estudio	45
<b>Figura 10.</b> Diagrama de flujo para la programación en el lenguaje “Rstudio”	48
<b>Figura 11.</b> Gráficos cruzados de parámetros petrofísicos	57
<b>Figura 12.</b> Correlación entre parámetros petrofísicos	57
<b>Figura 13.</b> Diagrama de dispersión entre los parámetros petrofísicos y los residuos del EUR	71
<b>Figura 14.</b> Diagrama de dispersión entre los parámetros petrofísicos y los residuos del BOPD	72
<b>Figura 15.</b> Diagrama de dispersión entre los parámetros petrofísicos y los residuos del BSW	72
<b>Figura 16.</b> Diagrama de dispersión entre los parámetros petrofísicos y los residuos del BWPD	73
<b>Figura 17.</b> Diagrama de dispersión entre los parámetros petrofísicos y los residuos del Oil	74
<b>Figura 18.</b> Diagrama de dispersión entre los parámetros petrofísicos y los residuos del IP	74
<b>Figura 19.</b> Diagrama de dispersión entre los parámetros petrofísicos y los residuos del Water	75
<b>Figura 20.</b> Diagrama de dispersión entre los parámetros petrofísicos y los residuos del THP	76
<b>Figura 21.</b> Diagrama de dispersión entre los parámetros petrofísicos y los residuos del Pico	77
<b>Figura 22.</b> Cuantiles teóricos contra cuantiles normales para los parámetros de producción	78
<b>Figura 23.</b> Residuos frente a los valores ajustados del modelo para los parámetros de producción	81
<b>Figura 24.</b> Distribución de residuos frente a los valores predichos por el modelo para los parámetros de producción	85
<b>Figura 25.</b> Residuos frente a los valores “Leverages (hat)” del modelo para los parámetros de producción	87
<b>Figura 26.</b> Histogramas de modelo promedio con mejores predictores y datos de campo	96

**Figura 27.** Datos de producción de campo frente a los valores calculados por las ecuaciones del modelo

98

## LISTA DE GRÁFICAS

	pág.
<b>Gráfica 1.</b> $R^2$ para EUR	60
<b>Gráfica 2.</b> $R^2$ para BOPD	61
<b>Gráfica 3.</b> $R^2$ para BWPD	61
<b>Gráfica 4.</b> $R^2$ para BSW	62
<b>Gráfica 5.</b> $R^2$ para Oil	62
<b>Gráfica 6.</b> $R^2$ para Water	63
<b>Gráfica 7.</b> $R^2$ para IP	63
<b>Gráfica 8.</b> $R^2$ para THP	64
<b>Gráfica 9.</b> $R^2$ para Pico	64
<b>Gráfica 10.</b> VIF de los parámetros petrofísicos del EUR	66
<b>Gráfica 11.</b> VIF de los parámetros petrofísicos del BOPD	66
<b>Gráfica 12.</b> VIF de los parámetros petrofísicos del BWPD	67
<b>Gráfica 13.</b> VIF de los parámetros petrofísicos del BSW	67
<b>Gráfica 14.</b> VIF de los parámetros petrofísicos del Oil	68
<b>Gráfica 15.</b> VIF de los parámetros petrofísicos del Water	68
<b>Gráfica 16.</b> VIF de los parámetros petrofísicos del IP	69
<b>Gráfica 17.</b> VIF de los parámetros petrofísicos del THP	69
<b>Gráfica 18.</b> VIF de los parámetros petrofísicos del Pico	70
<b>Gráfica 19.</b> Shapiro-Wilk test para los parámetros de producción	79
<b>Gráfica 20.</b> Breusch-Pagan test para los parámetros de producción	82
<b>Gráfica 21.</b> Durbin-Watson test para los parámetros de producción	84
<b>Gráfica 22.</b> Betas estandarizados para EUR	88
<b>Gráfica 23.</b> Betas estandarizados para BOPD	89
<b>Gráfica 24.</b> Betas estandarizados para BWPD	89
<b>Gráfica 25.</b> Betas estandarizados para BSW	90
<b>Gráfica 26.</b> Betas estandarizados para Oil	90
<b>Gráfica 27.</b> Betas estandarizados para Water	91
<b>Gráfica 28.</b> Betas estandarizados para IP	91
<b>Gráfica 29.</b> Betas estandarizados para THP	92
<b>Gráfica 30.</b> Betas estandarizados para Pico	92
<b>Gráfica 31.</b> BOPD para el pozo Impero 3	101

## LSTA DE TABLAS

pág.

<b>Tabla 1.</b> Resultados de los modelos tabla completa, mediana y promedio	55
------------------------------------------------------------------------------	----



## LISTA DE ANEXOS

pág.

<b>Anexo A.</b> Capacidad acumulada y normalizada de la permeabilidad contra la capacidad acumulada y normalizada de la porosidad	107
<b>Anexo B.</b> Ecuaciones del modelo tabla completa	108
<b>Anexo C.</b> Ecuaciones del modelo tabla mediana	109
<b>Anexo D.</b> Ecuaciones del modelo promedio	110
<b>Anexo E.</b> Ecuaciones del modelo tabla promedio con los mejores predictores	111
<b>Anexo F.</b> Gráficos cruzados	112

## LISTA DE ABREVIATURAS

**PHIE:** Porosidad efectiva.  
**VCL:** Volumen de arcilla.  
**K:** Permeabilidad.  
**Netpay:** Espesor de explotación.  
**Heteroge:** Heterogeneidad.  
**Acu:** Distancia con el acuífero.  
**No:** Número de pozos por cluster.  
**Dias:** Días de producción.  
**EUR:** Estimado de recobro final.  
**BOPD:** Producción de petróleo diaria.  
**BWPD:** Producción de agua diaria.  
**BSW:** Corte de agua.  
**Oil:** Producción de petróleo acumulada.  
**Water:** Producción de agua acumulada.  
**IP:** Índice de productividad.  
**THP:** Presión en cabeza de pozo.  
**Pico:** Pico máximo de producción.  
**WOC:** Contacto del agua.  
**bbbl:** Barril.  
**Mbbbl:** Millones de barriles.  
**md:** Mili Darcies.  
**ft:** Pies.  
**%:** Porcentaje.  
**psi:** Libras por pulgada cuadrada.  
**/:** Fracción.  
**#:** Numero adimensional.  
**R:** Leguaje de computación estadística.

**R cuadrado:** Coeficiente de determinación.

**p-valor:** Potencia de la prueba.

**W:** Estadístico Wilk.

**BP:** Estadístico Breusch-Pagan.

**VIF:** Factor de inflación de varianza.

**D-W:** Estadístico Durbin-Watson.

**Pwf:** Presión dinámica en fondo de pozo.

**Swb:** Saturación de agua inmersa en la arcilla.

**MD:** Profundidad total del pozo.

**TVD:** Profundidad vertical del pozo.

## RESUMEN

El campo Imperial es un área de explotación de hidrocarburos en Colombia con un gran potencial productivo, en donde los pozos son perforados con un patrón estrella compuesto por un pozo vertical (cluster) y un determinado número de pozos horizontales que depende del volumen de petróleo extraíble del área cercana.

Debido a la alta heterogeneidad litológica del yacimiento, predecir el comportamiento productivo de los pozos perforados en esta área es un reto para los ingenieros y geólogos que trabajan en el campo. Además, se suma el hecho de que el yacimiento es empujado por un acuífero activo, el cual condiciona al pozo en sus distintas etapas, en la perforación cualquier irrupción o cercanía al WOC o un completamiento desfasado, puede repercutir en un alto corte de agua en la producción y durante la producción el BSW es un factor conocido que es directamente proporcional al tiempo.

Con el fin de mejorar la capacidad predictiva de producción de los pozos se desarrolló el presente trabajo de grado, que busca identificar relaciones entre los parámetros petrofísicos y los de producción a partir de una muestra de pozos del campo. Finalmente, se planteó un modelo de ecuaciones en las cuales a partir de los parámetros petrofísicos seleccionados es posible predecir los parámetros de producción evaluados con un grado de certidumbre.

**PALABRAS CLAVE:** PETROFÍSICA, PRODUCCIÓN, MODELO, PREDICCIÓN.

## **ABSTRACT**

The Imperial field is a hydrocarbon exploitation area in Colombia with great productive potential, where the wells are drilled with a star pattern consisting of a vertical well (cluster) and a certain number of horizontal wells that depends on the recoverable volume of oil from the nearby area.

Due to the high lithological heterogeneity of the reservoir, predicting the productive behavior of wells drilled in this area is a challenge for engineers and geologists working in the field. In addition, the fact that the reservoir is pushed by an active aquifer, which conditions the well in its different stages, in drilling any irruption or proximity to the WOC or an out of phase completion can result in a high water cut in production and during production phase the BSW is a known factor that is directly proportional to time.

In order to improve the predictive capability of production of the wells, this thesis aims to identify relationships between the petrophysical and the production parameters, from a sample of wells in the field. Finally, a model of equations was proposed in which from the selected petrophysical parameters it is possible to predict the production parameters evaluated with a degree of certainty.

**KEY WORDS:** PETROPHYSICS, PRODUCTION, MODEL, PREDICTION.

## INTRODUCCIÓN

Uno de los problemas más recurrentes durante la etapa de desarrollo de los Campos de hidrocarburos se relaciona a la incertidumbre asociada a los modelos predictivos de producción implementados por las Compañías Operadoras, lo anterior afecta la sostenibilidad técnica y financiera de los proyectos alrededor de las metas de producción. El problema ha sido estudiado por el equipo de desarrollo y los orígenes de este, guardan una relación con las áreas de petrofísica y yacimientos las cuales buscan mejorar la capacidad predictiva de producción de los pozos a partir de información como los registros de pozo, historial de producción de pozos cercanos, entre otros.

En este estudio se evaluará la capacidad predictiva de producción del Campo Imperial, ubicado en la Cuenca de los Llanos Orientales de Colombia a partir de parámetros petrofísicos. Las condiciones de alta heterogeneidad litológica y variación en la saturación de fluidos en el yacimiento han llevado a tener una incertidumbre en las zonas que se encuentran fuera del rango de investigación de los registros y a su vez esto provoca una complejidad geométrica en los pozos por cambios en la inclinación y azimut en la geo navegación de la sección horizontal. Estas condiciones traen como resultado la posibilidad de que se obtengan pozos con buenas propiedades petrofísicas y una producción por debajo de la estimada y de forma análoga, pozos con malas propiedades petrofísicas y una producción que supere la expectativa.

Debido a esto, el presente trabajo de grado tiene como tema de estudio, la relación entre parámetros petrofísicos y de producción, mediante un análisis de sensibilidad a cuatro clusters productores, los cuales comprenden un total de cuarenta y seis (46) pozos horizontales y cuatro (4) pozos verticales que abarcan múltiples escenarios de desarrollo y explotación con el fin de tener una muestra de estudio heterogénea.

Con el fin de establecer una metodología para la investigación, se consultó el artículo "SPE 161343" de la autoría de Charles Cunningham, Lisa Cooley, Gregory Wozniak y Jim Pancake<sup>1</sup> donde la investigación se basa en determinar cuáles parámetros de diseño de pozo (longitud del pozo, espaciamento entre pozos del mismo cluster, entre otros) y características de la formación (porcentaje de arcilla, espesor de la formación, entre otros), contribuyen a la producción de los pozos que se producen de la formación arcillosa Marcellus ubicada en la cuenca Appalachian. La conclusión más relevante de los investigadores es que de los parámetros estudiados, los de mayor incidencia sobre el EUR de los pozos son el espaciamento entre los clusters, espaciamento lateral de los pozos y la varianza en el azimut. Es

---

<sup>1</sup> CUNNINGHAM, Charlie. COOLEY, Lisa. WOZNIAK, Gregory. PANCAKE, Jim. *Using Multiple Linear Regression to Model EURs of Horizontal Marcellus Shale Wells. Pennsylvania: SPE International, 2012. 161343.*

importante resaltar que los pozos estudiados en este trabajo pertenecen a un yacimiento no convencional, por lo cual hay que evaluar las diferencias.

También se tuvo en cuenta la investigación realizada por O. Buj y J. Gisbert<sup>2</sup> donde se emplea un estudio estadístico de análisis de varianza (ANOVA) para valorar el comportamiento de las areniscas Uncastillo, Ayerbe y Alcañiz. Una de las conclusiones es que los materiales estudiados se pueden definir como homogéneos desde el punto de vista de su comportamiento mecánico y físico, destacando la arenisca de Alcañiz, la cual posee la menor variabilidad evidenciado por el método estadístico.

Finalmente se tuvo en cuenta el artículo “SEG-2001-1560” de la autoría de Reinaldo J. Michelena, Juan R. Jiménez <sup>3</sup>, el cual presenta un método que realiza la estimación de propiedades petrofísicas a partir de atributos de la sísmica en dos dimensiones. Dentro de las conclusiones se encuentra que el número de atributos o propiedades del pozo a utilizar y el número de pozos en el estudio debe guardar una proporción.

La investigación se realizó con la intención de identificar el impacto que tienen parámetros petrofísicos sobre parámetros de producción y cuantificar la magnitud de estas relaciones. Entender este fenómeno permitirá a los ingenieros maximizar el recobro de los pozos por medio de la identificación de las zonas con mayor potencial productivo y así aumentar la rentabilidad de las campañas. Al identificar zonas no favorables desde el punto de vista de la petrofísica, pero con buen rendimiento productivo, se puede ampliar las reservas y la producción acumulada del campo, dándole una mayor longevidad al proyecto y aumentar el margen de utilidad, para así guiar el futuro desarrollo del campo a las mejores zonas y entonces, poder extender esta metodología a campos con condiciones litológicas y fisicoquímicas similares.

Para el desarrollo de la investigación se evaluó la relación de los siguientes parámetros petrofísicos, porosidad, permeabilidad, el volumen de arcilla, el coeficiente de heterogeneidad, net pay, distancia con el acuífero y el número de pozos completados en el cluster con los parámetros de producción que se presentan a continuación, producción acumulada, pico máximo de producción, índice de productividad, EUR, presión en la cabeza del pozo y corte de agua. El estudio se realizó sobre cuatro clusters productores del campo, cluster A, cluster B, cluster C, y cluster D, los cuales suman un total de cincuenta pozos, los cuales han sido perforados entre el año 2009 y 2018 bajo un patrón de estrella y la formación a

---

<sup>2</sup> BUJ, O. GISBERT, J. *Petrophysical characterization of three commercial varieties of miocene sandstones from the Ebro valley. Zaragoza: Universidad de Zaragoza, 2007. ISSN: 0465-2746.*

<sup>3</sup> MICHELENA, Reinaldo. JIMÉNEZ, Juan. *Estimation of Petrophysical Properties Using Multiple Attributes: Generalizing Linear Regressions. Texas: SEG Exposition an annual meeting. 2001.*

evaluar es la zona Objetivo. Los criterios de selección tanto de los clusters como de los pozos fueron, el cubrimiento extensional del campo, amplio rango en el espesor de la columna petrolífera, diferentes coeficientes de heterogeneidad, que se incluyeran todos los años de desarrollo del campo, diferencias entre las intensidades de desarrollo local y periférica (número de pozos por cluster) y que tanto los parámetros petrofísicos y de producción se encontraran disponibles.

Debido a la gran cantidad de información que se manejó, se programó un código en el lenguaje "R" que permitió el manejo de la base de datos, con el fin de aplicar el método estadístico de regresión lineal múltiple a cada uno de los parámetros de producción a evaluar. Posterior ello, se tomarán diez pozos del mismo campo para así obtener de estos los parámetros petrofísicos y de esta forma de calcular los parámetros de producción con las ecuaciones obtenidas con el método estadístico y compararlo con el valor de campo.

Previo a realizar esta investigación se presume que para los pozos del Campo Imperial existen correlaciones entre la petrofísica y la producción, las cuales se pueden determinar a partir de regresiones lineales múltiples involucrando parámetros de ambas áreas, lo cual permita mejorar la capacidad predictiva de producción de los pozos del campo.

El objetivo general de la investigación es determinar la relación entre los parámetros petrofísicos y de producción de los pozos del Campo Imperial por medio de un análisis de sensibilidad a cuatro clusters de producción. Los objetivos específicos son:

- Consolidar una base de datos con los parámetros petrofísicos y de producción de los pozos en estudio.
- Correlacionar los parámetros petrofísicos y de producción de los pozos por medio de regresiones lineales múltiples, mediante un código de programación.
- Realizar los cálculos de los parámetros de producción con las ecuaciones obtenidas en el método estadístico a pozos del campo.
- Comparar los valores calculados con datos de producción actual de los pozos.

Este trabajo de grado se desarrolló con la intención de determinar la relación entre parámetros petrofísicos y de producción, por medio de la selección de una muestra de pozos de estudio y la aplicación de regresiones lineales múltiples para cada una de las variables de producción. Cada una de las ecuaciones de variables de producción en función de parámetros de petrofísicos tienen una capacidad explicativa igual al valor del R cuadrado de cada uno de los modelos.



## 1. MARCO TEÓRICO

En esta sección se presentan los conceptos teóricos más relevantes en la investigación, de las áreas de petrofísica, producción y estadística. Se realiza una descripción de los parámetros que se desean correlacionar y el método estadístico a emplear.

### 1.1 PARÁMETROS PETROFÍSICOS Y OTROS

Son aquellas variables por medio de las cuales se logra definir y evaluar las propiedades físicas y químicas de la roca y su interacción con los fluidos que la saturan, con el fin dar una explicación acerca de la acumulación y migración de los líquidos y/o gases presentes en la roca.

**1.1.1 Porosidad.** Es una medida de espacios vacíos que toma valores entre cero (0) y cien (100) por ciento, la cual se representa en la industria de hidrocarburos como el porcentaje de volumen de poros o de volumen de roca que puede contener fluidos, ya sean líquidos o gaseosos. Dentro de la clasificación de la porosidad encontramos dos tipos, según su origen y comunicación de sus poros.

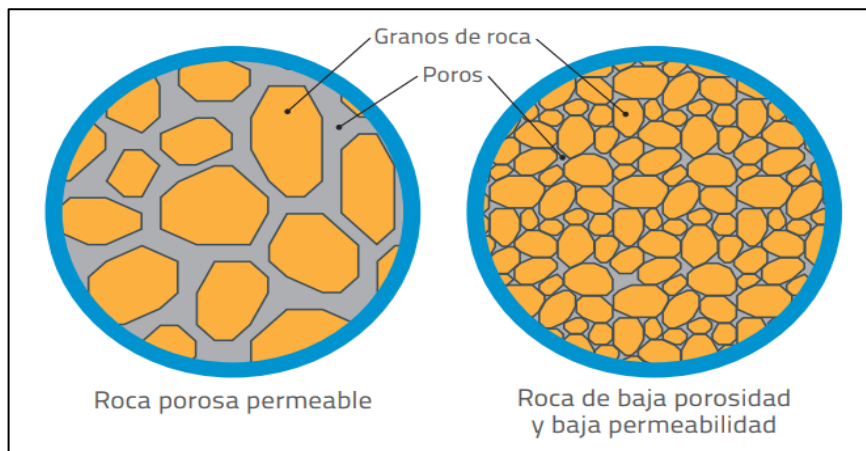
Según su origen encontramos la porosidad primaria, la cual hace referencia a aquellos espacios vacíos que resultan entre los granos de la roca después de que los sedimentos se acumulen y la porosidad secundaria resulta debido a la interacción de un agente externo el cual provoca fracturas que inicialmente no se encontraban en la roca. Al clasificar la porosidad según su comunicación entre poros encontramos la porosidad efectiva, la cual evidencia el volumen de poros que tiene la capacidad de almacenar fluidos y se encuentran interconectados entre sí y la porosidad total es el espacio poroso total presente en la roca el cual no está ocupado por matriz. Existen distintos métodos para calcular la porosidad, dentro de estos encontramos las distintas pruebas de laboratorio que se le realizan a los tapones y múltiples registros eléctricos, los comúnmente usados son el registro sísmico, registro lito-densidad, registro neutrón, registro de propagación electromagnética, entre otros (estos registros se pueden correlacionar con pozos cercanos).

#### Ecuación 1. Porosidad

$$\phi_{total} = \frac{\text{Volumen poroso}}{\text{Volumen de roca}}$$

**Fuente:** SERRA, Oberto. Fundamentals of well-log interpretation. Francia: Elsevier, 1984. 423p. ISBN 0444421327

**Figura 1.** Porosidad de la roca



**Fuente:** ASOCIACIÓN COLOMBIA DE PETRÓLEO, Todo sobre fracking.[Sitio WEB]. Colombia [28, octubre, 2019] Disponible en: <https://acp.com.co/web2017/en/todo-sobre-el-fracking/955-que-son-los-hidrocarburos-en-reservorios-no-convencionales>

**1.1.2 Permeabilidad.** Es la capacidad de un material, que permite que un fluido lo atraviese sin alterar su estructura interna, esto fue demostrado por Henry Darcy (1856). En la industria del petróleo se define como la propiedad que permite a la roca el paso de fluidos a través de su medio poroso. Dentro de la clasificación de la permeabilidad encontramos, permeabilidad absoluta, la permeabilidad efectiva, y la permeabilidad relativa.

En la industria existen diversos métodos para el cálculo de esta propiedad, dentro de las rocas consideradas como permeables encontramos las areniscas (común objetivo de explotación), las cuales tienen la capacidad de almacenar y permitir el paso de fluidos, de forma análoga las lutitas o limolitas se clasifican como rocas no permeables, lo cual favorece a la formación de un sello, la ley de Darcy viene dada por.

### **Ecuación 2.** Ley de Darcy

$$V_{fase\ en\ dir.\ x} = - \frac{K_{fase\ en\ dir.\ x}}{\mu_{fase}} * \frac{\partial \phi}{\partial x}$$

**Fuente:** ECUACIÓN DE DIFUSIVIDAD. (10, octubre, 2019: Bogotá D.C.) RAMOS, Jorge Luis. Bogotá D.C. Universidad América, 2019. 10 p.

**1.1.3 Volumen de arcilla y modelo de gamma ray lineal.** El volumen de arcilla se define como el porcentaje neto de arcilla presente en un volumen de roca específico y es un parámetro importante para la evaluación de formaciones, debido a que afecta de forma negativa otras propiedades de la roca como la porosidad, permeabilidad y saturación de hidrocarburos. Para estimar su valor en la roca, existen diferentes métodos, donde se puede destacar las pruebas a nivel de laboratorio en taponos, registros de pozo de indicadores simples o dobles y ecuaciones matemáticas para su interpretación.

El modelo de gamma ray lineal implementa el registro de Gamma Ray, el cual indica el nivel de radioactividad de la roca emitido por el potasio, torio y uranio, para realizar un cálculo de volumen de arcilla, debido a que estas dos variables poseen una tendencia similar. La presencia de arcilla tiene un gran impacto en la estimación de reservas y producción del pozo, siendo un parámetro indispensable para la evaluación petrofísica.

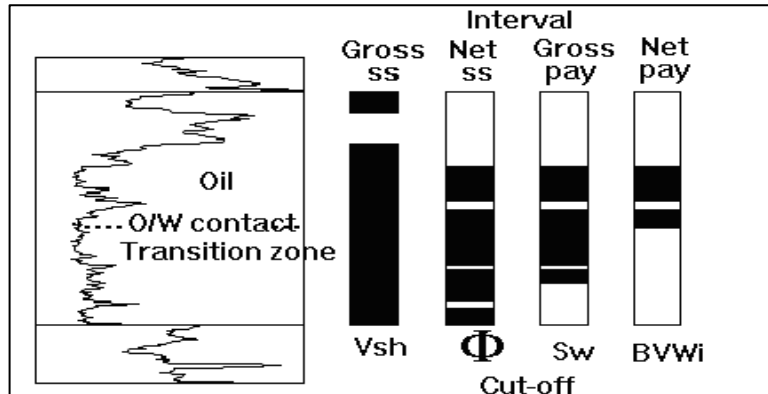
**Ecuación 3.** Índice de Gamma Ray

$$I_{GR} = \frac{GR_{log} - GR_{min}}{GR_{max} - GR_{min}}$$

**Fuente:** SCHLUMBERGER, Interactive Petrophysics. Manual. 2018.

**1.1.4 Net pay.** El net pay es la parte del reservorio el cual contiene hidrocarburos económicamente recuperables bajo una tecnología actual y que generan un margen de utilidad, para determinar estas zonas, se realiza una interpretación y evaluación petrofísica a la formación con ayuda de las pruebas de campo realizadas (registros de pozo, recolección de recortes, corazonamientos y pruebas de laboratorio, entre otros) sobre un modelo ya establecido para el campo o zona en cuestión. Para clasificar una zona como “*Net pay*” sus propiedades deben encontrarse dentro de un rango de aceptación, la porosidad debe ser mayor al veinte (20) por ciento, el volumen de arcilla menor al diez (10) por ciento y la saturación de agua no mayor al cuarenta (40) por ciento. Estos rangos son específicos para los pozos horizontales del campo Imperial, donde las medidas en profundidad se cuantifican en MD.

**Figura 2. Net pay**



**Fuente:** GEOLOGICAL SURVEY, Pay determination.[Sitio WEB]. Kansas [25, octubre, 2019] Disponible en: <http://www.kgs.ku.edu/software/PfEFFER-java/HELP/PfEFFER/Pfeffer-theory5.html>

**1.1.5 Modelo dual water.** El modelo “*dual water*” es un método matemática que tiene como fin determinar valores para la saturación de agua de la formación a partir de una información base, este se basa en los datos obtenidos por Waxman, Smits y Thomas (WST), quienes dicen que la saturación de agua se calcula en cada punto de datos del reservorio mediante una solución iterativa de ecuaciones complejas multiparamétricas, a partir de mediciones de resistividad, porosidad y saturación en el laboratorio de rocas reales. La ecuación empleada para la determinación de la saturación de agua por este método se describe a continuación.

**Ecuación 4.** Saturación de agua total

$$\frac{1}{R_t} = \frac{\phi_T^m * SwT^n}{a} * \left( \frac{1}{R_w} + \frac{Swb}{SwT} \left( \frac{1}{Rwb} - \frac{1}{R_w} \right) \right)$$

**Fuente:** SCHLUMBERGER, Interactive Petrophysics. Manual. 2018.

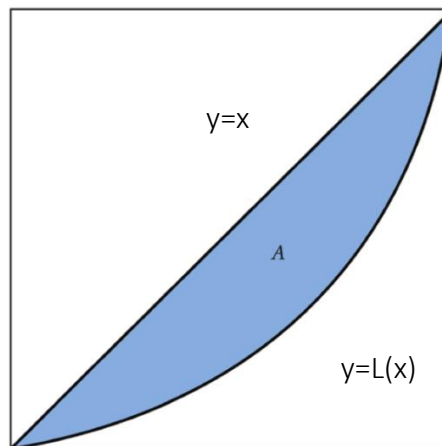
**1.1.6 Coeficiente de heterogeneidad.** Es un valor adimensional el cual evidencia la variación en las propiedades de las rocas de acuerdo con la ubicación espacial, dentro de las propiedades más relevantes que pueden cambiar son la porosidad y el volumen de arcilla. Esta heterogeneidad supone un reto en el modelamiento de sistemas petroleros, la evaluación de la formación y la simulación del yacimiento.

Uno de los procedimientos empleados para el cálculo de la heterogeneidad de los pozos y yacimientos de hidrocarburos, es el coeficiente de Lorenz, el cual toma valores entre cero (0) y uno (1) donde el mínimo valor hace referencia a una arena

completamente homogénea y el máximo valor a un cuerpo heterogéneo en su totalidad. Para llevar a cabo este modelo se requieren datos de porosidad y permeabilidad asociados a un espesor y realizar los siguientes pasos.

- 1-Ordenar los valores de permeabilidad de mayor a menor.
- 2-Multiplicar la permeabilidad y la porosidad por su espesor asociado.
- 3-Calcular la capacidad acumulada de la porosidad y la permeabilidad.
- 4-Normalizar ambas capacidades acumuladas entre los valores de cero y uno.
- 5-Graficar la capacidad acumulada y normalizada de la porosidad contra la capacidad acumulada y normalizada de la permeabilidad.
- 6-Obtener una ecuación que represente a la gráfica.
- 7-Aplicar la ecuación del índice de Gini para obtener un valor de heterogeneidad.

**Figura 3.** Curva de Lorenz



**Fuente:** elaboración propia

### **Ecuación 5.** Heterogeneidad

$$IG = 2 \int (x - L(x)) dx$$

**Fuente:** VADULLI, Rolly Buccioni. Estimación del coeficiente de concentración de Gini a partir de la curva estimada de Lorenz. 2012. Chile. Revista Chilena de Economía y Sociedad, 5(1-2).

**1.1.7 Ubicación del pozo.** La ubicación de los pozos en el campo se determina a partir de los rendimientos productivos de una zona, la intensidad de desarrollo y la delimitación geográfica del campo. Al perforarse bajo un patrón estrella, existe una gran cercanía en superficie para los pozos pertenecientes al mismo cluster.

## 1.2 PARÁMETROS DE PRODUCCIÓN

Son las variables que cuantifican el desempeño global productivo de pozos productores de hidrocarburos, por medio de mediciones, muestras y cálculos matemáticos. Estos son determinados a partir de mediciones directas en campo, estimaciones por modelos matemáticos, pruebas de laboratorio, correlación con pozos cercanos, entre otros.

**1.2.1 Producción acumulada y pico máximo de producción.** La producción acumulada ( $N_p$ ) hace referencia a la cantidad total de petróleo y gas recuperados del subsuelo en un tiempo determinado, siendo medido a condiciones de superficie y fiscalizado por la autoridad competente bajo unos estándares de calidad. El reporte generado desde campo se realiza de forma diaria, con el fin de llevar un control sobre el volumen de producción de los pozos, campo o bloque, la unidad estándar de medida son los barriles. Esta producción de hidrocarburos pasa por múltiples fases de tratamiento con el fin de sustraer sustancias no deseadas y posteriormente es transportada a refinerías (comúnmente por oleoductos).

Una vez puesto en producción un pozo, el comportamiento típico de esta es el ascenso por efecto de los trabajos de optimización y acondicionamiento y posteriormente una caída progresiva con el tiempo la cual guarda gran similitud con la caída de presión. El pico máximo de la producción es un valor volumétrico que refleja el potencial del pozo en sus condiciones óptimas de operación y comúnmente condiciona la vida productiva del pozo, esta variable es medida a lo largo del tiempo y también puede ser calculada de forma previa a explotación por medio de modelos matemáticos.

**1.2.2 Índice de productividad.** El índice de productividad es un indicador de la capacidad productiva de un pozo bajo ciertas condiciones operacionales. Este se expresa como la relación entre el caudal producido y una caída de presión. *“Cuando la presión dinámica de fondo de pozo ( $P_{wf}$ ) es más grande que la presión de punto de burbuja ( $P_b$ ), el flujo de fluido es similar a un flujo monofásico.”*<sup>4</sup>

### Ecuación 6. Índice de productividad

$$IP = \frac{Q_o}{P_r - P_{wf}}$$

**Fuente:** BROWN, Kermit. The technology of artificial lift methods. Denver: PennWell Books, 2007. 1087 p. ISBN 0878141375

---

<sup>4</sup> Índice de productividad, marzo 19 de 2010, La comunidad petrolera.

**1.2.3 Presión en cabeza de pozo (THP).** Es la presión que se experimenta en la cabeza del pozo, donde se tienen presiones estáticas y dinámicas las cuales son medidas por diferentes accesorios. Esta presión depende de la condición operativa del pozo, la presión del yacimiento, la profundidad del pozo, el levantamiento artificial, tasa de producción y de la densidad de los fluidos en la columna.

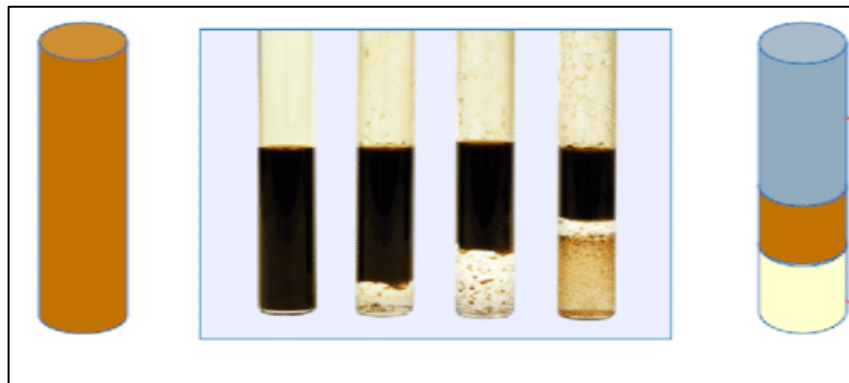
**1.2.4 Agua y sedimentos (BSW).** Es una fracción del líquido de producción el cual representa el contenido de agua y sedimentos, esta se cuantifica a partir de una muestra y pruebas de laboratorio y es una especificación técnica del crudo referente a impurezas de este.

**Ecuación 7. BSW**

$$BSW = \frac{V_{sedimentos} + V_{agua}}{V_{muestra}}$$

**Fuente:** BROWN, Kermit. The technology of artificial lift methods. Denver: PennWell Books, 2007. 1087 p. ISBN 0878141375

**Figura 4. BSW**



**Fuente:** RESEARCH GATE, Liquid-Liquid Separation [Sitio WEB]. Palermo [22, octubre, 2019] Disponible en: [https://www.researchgate.net/figure/MBT-results-for-a-water-in-oil-emulsion-at-30-BSW-Dots-experimental-results\\_fig5\\_254534820](https://www.researchgate.net/figure/MBT-results-for-a-water-in-oil-emulsion-at-30-BSW-Dots-experimental-results_fig5_254534820)

**1.2.5 Estimado de recobro final (EUR).** Es una cantidad aproximada de petróleo o gas que es potencialmente recuperable o ya ha sido producido, este puede ser calculado a través de múltiples métodos dependiendo de los objetivos de este, es un concepto muy importante en la industria ya que los proyectos deben cumplir un umbral de producción, con el fin de generar una rentabilidad. La clasificación de estas reservas es dependiente del grado de probabilidad de que el petróleo pueda recuperarse utilizando la tecnología y recursos actuales. En las reservas probadas

se tiene más de noventa (90) por ciento, en las reservas probables se tiene entre de noventa (90) y cincuenta (50) por ciento y posibles reservas es cuando se tiene menos de cincuenta (50) por ciento de oportunidad de extraer los hidrocarburos.

### 1.3 REGRESIÓN LINEAL MÚLTIPLE

La regresión lineal múltiple tiene como fin ajustar modelos bajo una ecuación, a partir de una variable dependiente y múltiples variables independientes. Por medio de este método se pueden identificar que variable le dan explicación a otra, además de comparar y comprobar modelos causales. La finalidad del método es predecir valores de una variable a partir de otras con un nivel de confiabilidad. Es necesario descartar en este modelo los problemas de heterocedasticidad, multicolinealidad y la especificación.

Dentro de los valores que se obtiene aplicando un modelo de regresión lineal múltiple, encontramos un R cuadrado quien nos dice la capacidad explicativa del modelo, unos coeficientes por variable independiente, los cuales reflejan el impacto sobre la variable dependiente, unos p-valores por variable independiente que reflejan la significancia de cada una sobre la variable dependiente y un estadístico F quien representa la significancia conjunta de las variables independientes.

El modelo de ecuación que se obtiene tiene la siguiente forma.

#### Ecuación 8. Regresión lineal múltiple

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

**Fuente:** WALPOLE, Myers. Probabilidad y estadística. México: Pearson, 2012. 816 p. ISBN 9786073214179

**1.3.1 Coeficiente de determinación (R cuadrado).** El valor del coeficiente de determinación oscila entre cero (0) y uno (1), entre más cercano a uno (1), mayor es el ajuste del modelo, es decir una mayor capacidad explicativa de las variables independientes sobre la dependiente. De forma análoga, cuanto más cercano es su valor del cero (0), menos ajustado serán los resultados del modelo, es decir un mayor grado de incertidumbre. La ecuación para calcular el R cuadrado es la siguiente donde el numerador es la suma de cuadrados de la regresión y el denominador la suma de cuadrados total.



### Ecuación 9. Coeficiente de determinación

$$R^2 = \frac{\sum_{i=1}^n (y^*i - y)^2}{\sum_{i=1}^n (yi - y)^2}$$

**Fuente:** WALPOLE, Myers. Probabilidad y estadística. México: Pearson, 2012. 816 p. ISBN 9786073214179

**1.3.2 Potencia de la prueba (p-valor).** Se denomina potencia de la prueba, la cual es usada para determinar la validez de dos hipótesis, la nula y alternativa planteadas por el autor del problema analizado, y se calcula por medio de un valor crítico o estadístico de prueba. El valor de estadístico ayuda a diferenciar los resultados que son productos del azar de los resultados estadísticamente significativos y se calcula a partir de la siguiente fórmula.

### Ecuación 10. Valor crítico

$$t = \frac{bj - \beta_j}{s \sqrt{c_{jj}}}$$

**Fuente:** WALPOLE, Myers. Probabilidad y estadística. México: Pearson, 2012. 816 p. ISBN 9786073214179

**1.3.3 Coeficiente de correlación lineal de Pearson.** Es un parámetro estadístico el cual permite cuantificar el grado de correlación o variación conjunta entre dos variables. Su rango de valores oscila entre menos uno (-1) y uno (1), donde los valores negativos hacen referencia a una relación inversamente proporcional y los positivos a una relación directamente proporcional. La cercanía a los valores de los extremos del rango indican una mayor correlación lineal entre variables.

### Ecuación 11. Coeficiente de Pearson

$$P_{xy} = \frac{cov_{xy}}{\sigma_x \sigma_y}$$

**Fuente:** WALPOLE, Myers. Probabilidad y estadística. México: Pearson, 2012. 816 p. ISBN 9786073214179

**1.3.4 Selección de los mejores predictores.** Esta es una operación que se realiza sobre las regresiones lineales múltiples para determinar cuáles de los predictores tienen una mayor capacidad explicativa sobre la variable de salida. De los métodos para calcularlos encontramos el jerárquico que se basa en criterios de análisis, de entrada forzada el cual introduce todos los predictores de forma simultánea y el paso a paso que emplea criterios matemáticos para decidir cuál de los predictores tiene una mayor contribución.

**1.3.5 Colinealidad.** Este es uno de los problemas que puede presentar la base de datos previo a ser ejecutado un modelo de regresión lineal múltiple y se evidencia cuando una variable independiente es una combinación lineal de otras. Para determinar la existencia de este problema existen diferentes métodos, uno de ellos es el factor de inflación de varianza que se calcula de la siguiente forma.

**Ecuación 12.** Factor de inflación de varianza (VIF)

$$VIF = \frac{1}{1 - R^2}$$

**Fuente:** WALPOLE, Myers. Probabilidad y estadística. México: Pearson, 2012. 816 p. ISBN 9786073214179

**1.3.6 Normalidad.** Es una propiedad de las variables continuas la cual indica que los datos de esta se distribuyen de forma normal, es decir que al graficar la función de densidad esta tiene una forma acampanada y simétrica lo cual permite modelar múltiples fenómenos. Para verificar la normalidad de las variables se emplea el estadístico de Shapiro-Wilk el cual se calcula mediante la siguiente ecuación.

**Ecuación 13.** Estadístico W

$$W = \frac{(\sum_{i=1}^n ai * xi)^2}{\sum_{i=1}^n (xi - x^*)^2}$$

**Fuente:** WALPOLE, Myers. Probabilidad y estadística. México: Pearson, 2012. 816 p. ISBN 9786073214179

**1.3.7 Homocedasticidad.** Es una propiedad del modelo donde la varianza de los residuos debe de ser constante en todo el rango de observaciones, distribuyéndose de forma aleatoria pero mantenido la misma dispersión. Uno de los métodos para verificar esta propiedad es mediante el test de Breusch-Pagan el cual examina la regresión de los residuos mediante la siguiente ecuación.

#### **Ecuación 14.** Regresión de residuos

$$u^2 = y_0 + y_{1x} + v$$

**Fuente:** WALPOLE, Myers. Probabilidad y estadística.  
México: Pearson, 2012. 816 p. ISBN 9786073214179

**1.3.8 Autocorrelación.** Es una propiedad del modelo donde los valores de cada observación son independientes de los otros, es decir que no ven afectados unos por otros, afectando de manera especial a las mediciones temporales. Dentro de los métodos existentes para la validación de la autocorrelación se destaca el test de hipótesis de Durbin-Watson el cual calcula el estadístico D-W mediante la siguiente ecuación.

#### **Ecuación 15.** Estadístico D-W

$$D-W = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$$

**Fuente:** WALPOLE, Myers. Probabilidad y estadística.  
México: Pearson, 2012. 816 p. ISBN 9786073214179

**1.3.9 Observaciones influyentes.** Al realizar un modelo de regresión lineal múltiple, las observaciones ingresadas a este tienen un peso sobre los resultados. Es importante identificar cuáles observaciones tienen una influencia superior a las demás para evaluar su extracción, esta influencia se puede cuantificar a partir de los valores de “Leverages” y “Distancia Cook” los cuales pueden ser calculados mediante las siguientes ecuaciones.

#### **Ecuación 16.** Leverages (hat)

$$H_{ii} = \frac{\delta y_i^*}{\delta y_i}$$

**Fuente:** WALPOLE, Myers. Probabilidad y estadística.  
México: Pearson, 2012. 816 p. ISBN 9786073214179

### Ecuación 17. Distancia Cook

$$D_i = \frac{(b - b_i)'X'X(b - b_i)}{ps^2}$$

**Fuente:** WALPOLE, Myers. Probabilidad y estadística. México: Pearson, 2012. 816 p. ISBN 9786073214179

**1.3.10 Coeficientes estandarizados.** Son los coeficientes que definen la ecuación de regresión una vez esta ha sido obtenida tras estandarizar las variables predictoras, es decir cuando se convierten las puntuaciones directas en típicas y la forma de calcularlos es mediante la siguiente ecuación.

### Ecuación 18. Coeficientes estandarizados

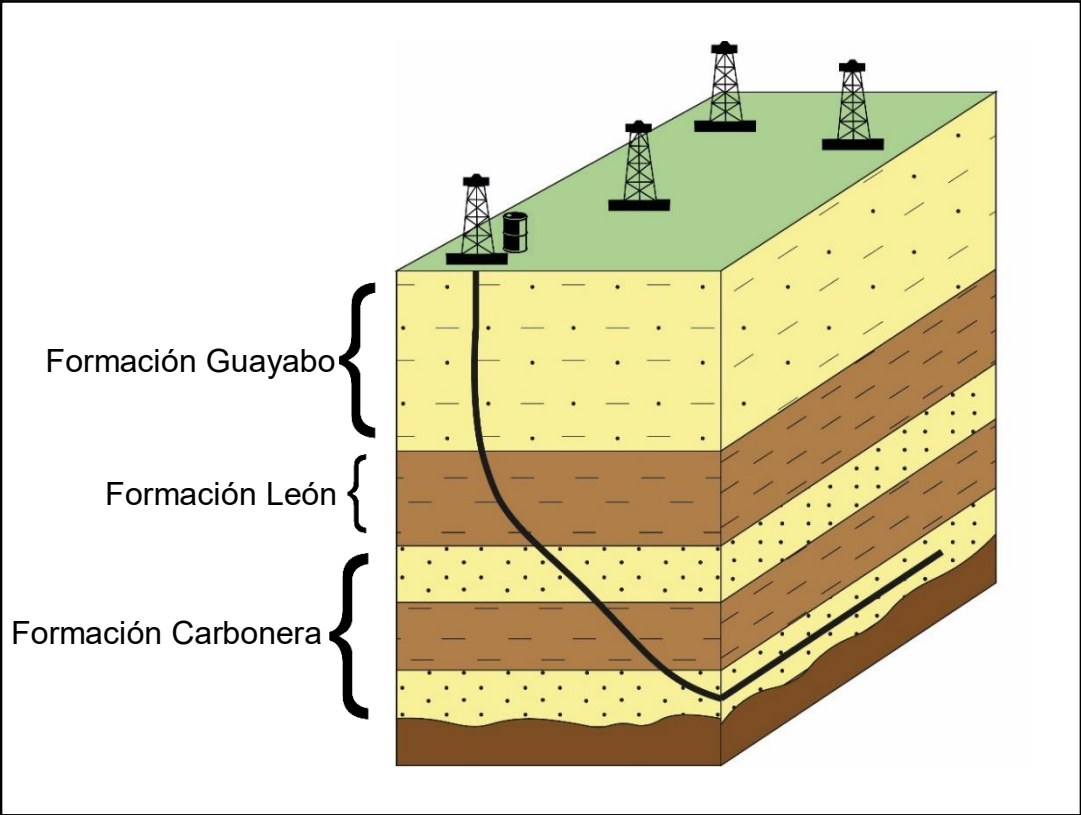
$$\beta_{st-1} = \beta_1 \frac{S_x}{S_y}$$

**Fuente:** WALPOLE, Myers. Probabilidad y estadística. México: Pearson, 2012. 816 p. ISBN 9786073214179

## 1.4 COLUMNA ESTRATIGRÁFICA

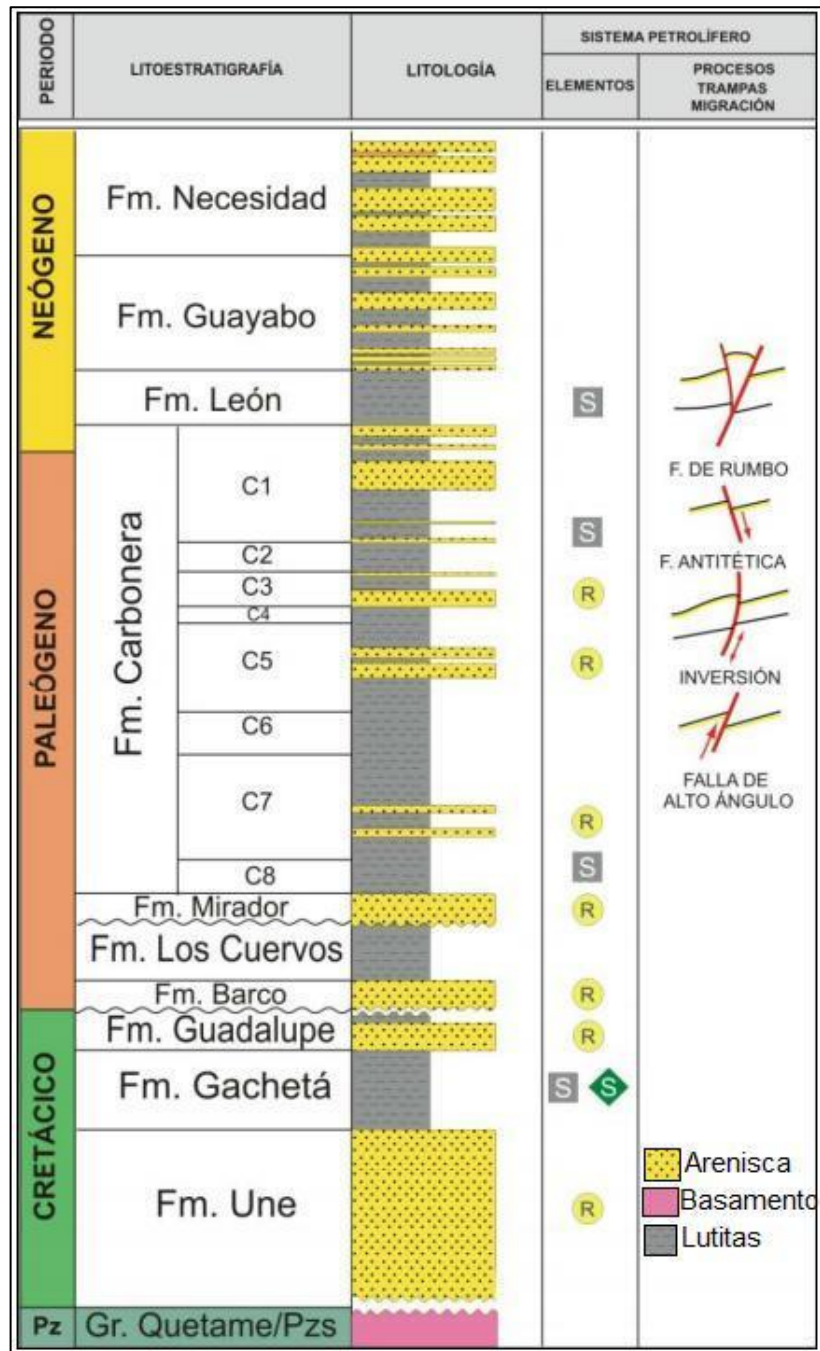
Dentro de las formaciones que atraviesan los pozos perforados en el campo Imperial encontramos la formación Guayabo, la cual está compuesta por arcillas, areniscas, y arcillas arenosas donde se encuentran acuíferos de superficie, su espesor varío alrededor de los quinientos (500) metros. La formación León, la cual está compuesta principalmente por lodolitas grises y gris verde, con escasas intercalaciones de areniscas de cuarzo, su granulometría va de grano fino a medio y su espesor varia alrededor de los doscientos cincuenta (250) metros. La formación Carbonera es considerada como el basamento económico, está compuesta por arcillolitas grises a gris verdoso intercaladas con areniscas (posible presencia de hidrocarburos) y con algunos mantos de carbón en su parte inferior y superior, siendo esta la zona objetivo de los pozos del campo Imperial. Por debajo de Carbonera se encuentran rocas del Paleozoico donde no existe interés petrolífero en el sector de la cuenca. Las formaciones descritas anteriormente se pueden apreciar junto a la trayectoria de un pozo en la siguiente figura.

**Figura 5.** Formaciones en el campo Imperial



**Fuente:** elaboración propia

**Figura 6.** Columna estratigráfica

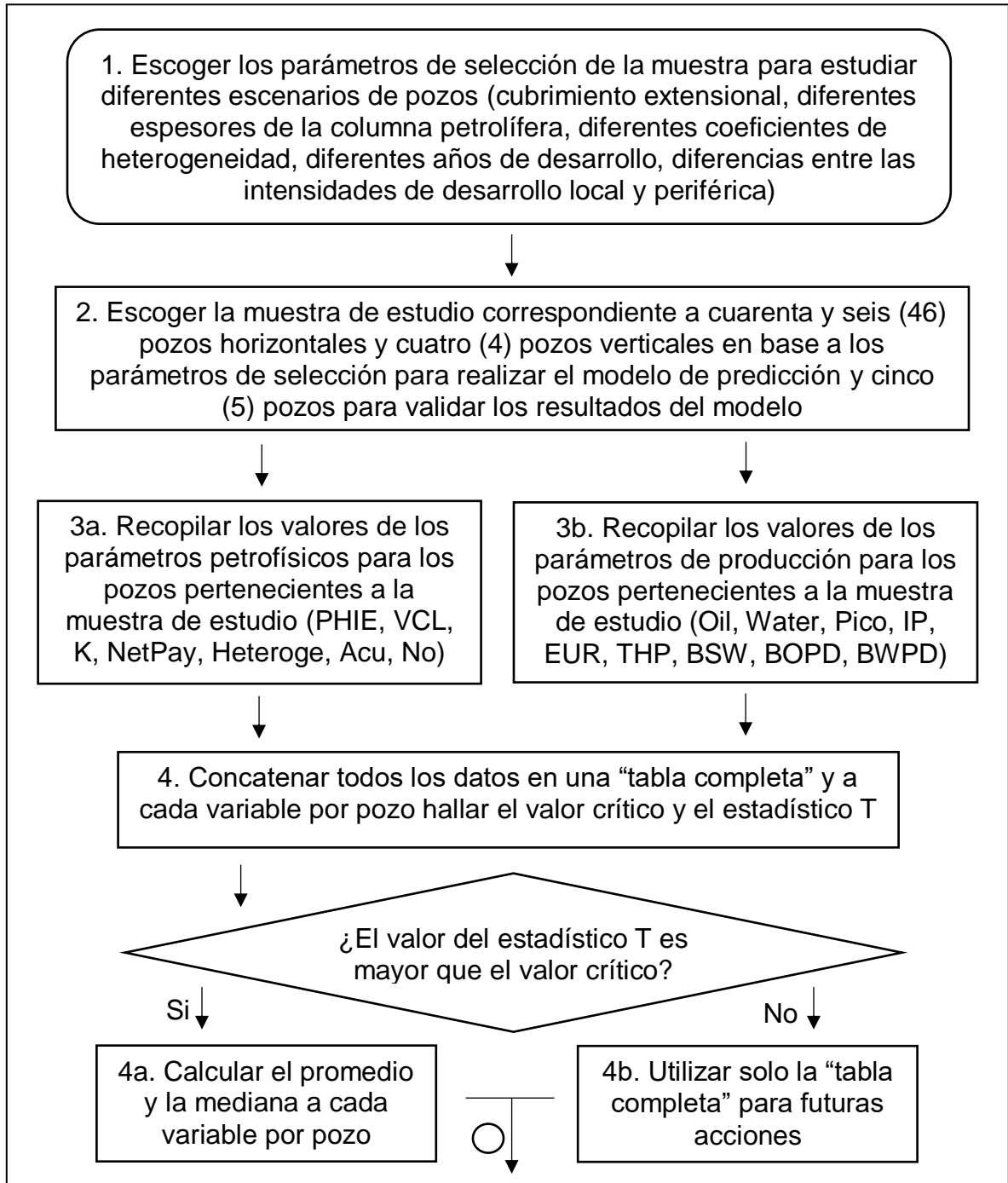


**Fuente:** ACEROS GARCÍA, Savieth. Diseño del módulo de selección de conectividad para el manual general de completamiento de pozos de Ecopetrol s.a. [en lineal] Tesis. Fundación Universidad De América, Bogotá D.C.: 2017. Disponible en: <https://repository.uamerica.edu.co/bitstream/20.500.11839/896/1/5121333-2017-1-IP.pdf>

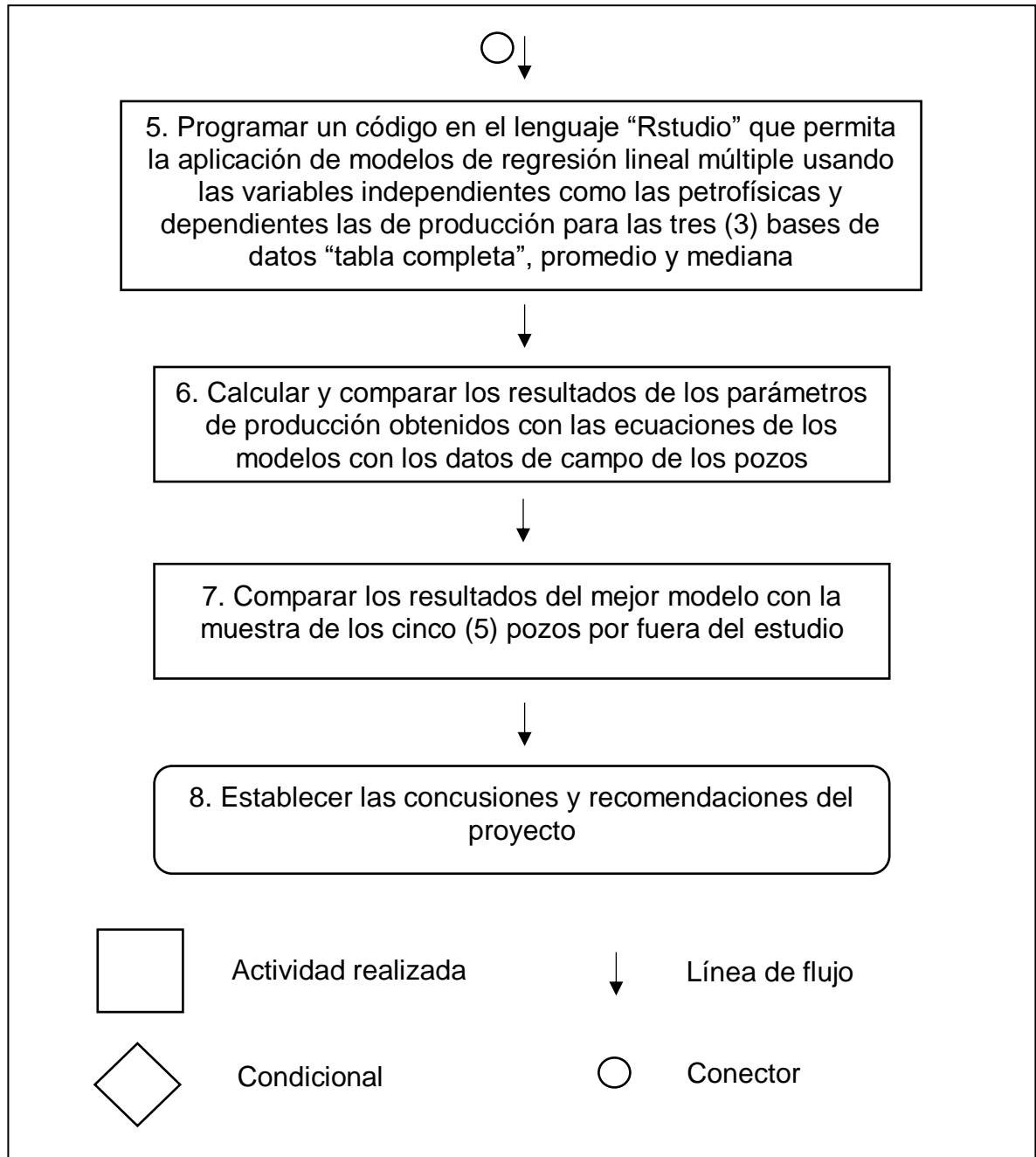
## 2. METODOLOGÍA Y DATOS

La figura 7 muestra la metodología empleada para el desarrollo del trabajo de grado en forma de diagrama de flujo.

**Figura 7.** Diagrama de flujo del trabajo de grado



**Figura 7.** (Continuación)



**Fuente:** elaboración propia



## 2.1 PARÁMETROS DE SELECCIÓN DE LA MUESTRA

Una de las principales causas de la problemática en estudio es la alta heterogeneidad litológica y variación en la saturación de fluidos, por esta razón la muestra de estudio (conjunto de pozos), debía abarcar varios escenarios tanto productivos como petrofísicos. Dentro de los escenarios se encuentra, el cubrimiento extensional del campo lo cual hace referencia a que el área delimitada por pozos tuviese una proporción mayor al sesenta (60) por ciento con respecto al área total del campo, amplio rango en el espesor de la columna petrolífera, lo que nos indica que las zonas de explotación de los pozos se encuentran en un rango entre los cien (100) y mil trecientos (1300) pies, diferentes coeficientes de heterogeneidad adquiriendo valores en un margen entre quince (15) y sesenta y cinco (65) por ciento, que se incluyeran todos los años de desarrollo del campo desde el año dos mil nueve (2009) hasta el dos mil diecinueve (2019), diferencias entre las intensidades de desarrollo local y periférica, lo cual indica el número de pozos por cluster que puede variar entre dos (2) a nueve (9) pozos y que tanto los parámetros petrofísicos y de producción se encontraran disponibles.

Estos parámetros fueron seleccionados para obtener una muestra de pozos heterogénea y crear un modelo que sea una herramienta predictora sobre el comportamiento productivo de estos y finalmente evaluar esta capacidad predictiva sobre pozos por fuera de la muestra de estudio, pero pertenecientes al campo.

## 2.2 SELECCIÓN DE LAS MUESTRAS

Con el fin de que el conjunto de datos por variable presentase un comportamiento normal desde el punto de vista estadístico, es necesario tener un número de observaciones mayor a treinta (30), y con el objetivo de abarcar todos los parámetros de selección de la muestra, se estableció un número de cincuenta (50) pozos para la muestra de estudio y por disponibilidad de la información para la muestra de validación se emplearon los mismos cincuenta (50) pozos de la muestra de estudio junto con cinco (5) pozos por fuera de la muestra de estudio, pero pertenecientes al campo.

**2.2.1 Muestra de estudio.** De la totalidad de pozos presentes en el campo, se seleccionaron cuatro (4) clusters de producción, cluster A, cluster B, cluster C, y cluster D los cuales suman un total de cincuenta (50) pozos completados que avanzan secuencial y unitariamente desde el pozo “Imperio 1” hasta el pozo “Imperio 50”, estos han sido perforados entre el año 2009 y 2018 bajo un patrón de estrella (consta de un pozo vertical en el medio y múltiples pozos horizontales desplegados de forma radial, con una gran cercanía en superficie) y la formación a evaluar es la zona Objetivo. De la totalidad de los pozos mencionados, cuarenta y seis (46) son horizontales y cuatro (4) son verticales.

**2.2.2 Muestra de validación.** Con el fin de comparar los resultados obtenidos con los valores ya calculados en campo y laboratorio, se seleccionó una muestra de

cincuenta y cinco (55) pozos para conformar la muestra de validación, estos se componen por los mismo cincuenta (50) pozos pertenecientes a la muestra de estudio y cinco (5) pozos por fuera de ella (pertenecientes al Campo Imperial), estos últimos avanzan secuencial y unitariamente desde el pozo “Imperio 51” hasta el pozo “Imperio 55”.

Esta muestra tiene como objetivo obtener los parámetros petrofísicos por pozo y así calcular los parámetros de producción con las ecuaciones obtenidas con el método estadístico de regresión lineal múltiple. Los criterios de selección de los cinco (5) pozos de la muestra de validación se basan en los mismos que para la muestra de estudio.

### **2.3 COMPILACIÓN DE LOS DATOS**

Bajo la premisa de cumplir el objetivo general del proyecto, determinar la relación entre los parámetros petrofísicos y de producción, se establecieron cuáles serían las variables a involucrar en el estudio con base a su disponibilidad y peso en la industria, por lo tanto, los parámetros predictores seleccionados fueron la porosidad efectiva “PHIE (/)”, volumen de arcilla “VCL (/)”, permeabilidad “K (md)”, espesor de explotación “Net pay (ft)”, heterogeneidad “Heteroge (/)”, distancia con el acuífero “Acu (ft)”, número de pozos por cluster “No (#)” y días de producción “Días (#)”, estos dos últimos (No, Días) no son parámetros petrofísicos, sin embargo se consideran aspectos importantes para la predicción de los parámetros productivos de los pozos. Los valores de estos parámetros empleados en el desarrollo del trabajo se pueden visualizar en la figura 8.

Así mismo se seleccionaron múltiples parámetros productivos que cuantifican de diversas formas el desempeño productivo del pozo, estos son el estimado de recobro final “EUR (Mbbl)”, producción de petróleo acumulada “Oil (Mbbl)”, producción de agua acumulada “Water (Mbbl)”, corte de agua “BSW (%)”, producción de petróleo diaria “BOPD (bbl)”, producción de agua diaria “BWPD (bbl)”, presión en cabeza de pozo “THP (psi)”, índice de productividad “IP (bbl/psi)” y pico máximo de producción “Pico (bbl)”. Los valores de estos parámetros empleados en el desarrollo del trabajo se pueden visualizar en la figura 9.

Los parámetros tanto petrofísicos como de producción mencionados previamente en su mayoría fueron suministrados por la empresa soporte del trabajo de grado, sin embargo, la heterogeneidad de los pozos, distancia con el acuífero, días de producción y pico máximo de producción fueron calculados bajo los procedimientos descritos a continuación.

**2.3.1 Heterogeneidad.** La heterogeneidad de un pozo representa la variación en las propiedades de las rocas de acuerdo con la ubicación espacial, las propiedades que esta evalúa son la porosidad y la permeabilidad las cuales pueden ser calculadas a partir de registros LWD de los pozos, pruebas de laboratorio, entre otros.

Una vez compilada la información necesaria, se ordenan los valores de permeabilidad de mayor a menor, luego se multiplica la permeabilidad y la porosidad por su espesor asociado, se calcula la capacidad acumulada de la porosidad y la permeabilidad, se normaliza ambas capacidades acumuladas entre los valores de cero (0) y uno (1), se grafica la capacidad acumulada y normalizada de la porosidad contra la capacidad acumulada y normalizada de la permeabilidad, se obtiene la ecuación que represente a la gráfica y finalmente se aplica la ecuación del índice de Gini para obtener un valor de heterogeneidad.

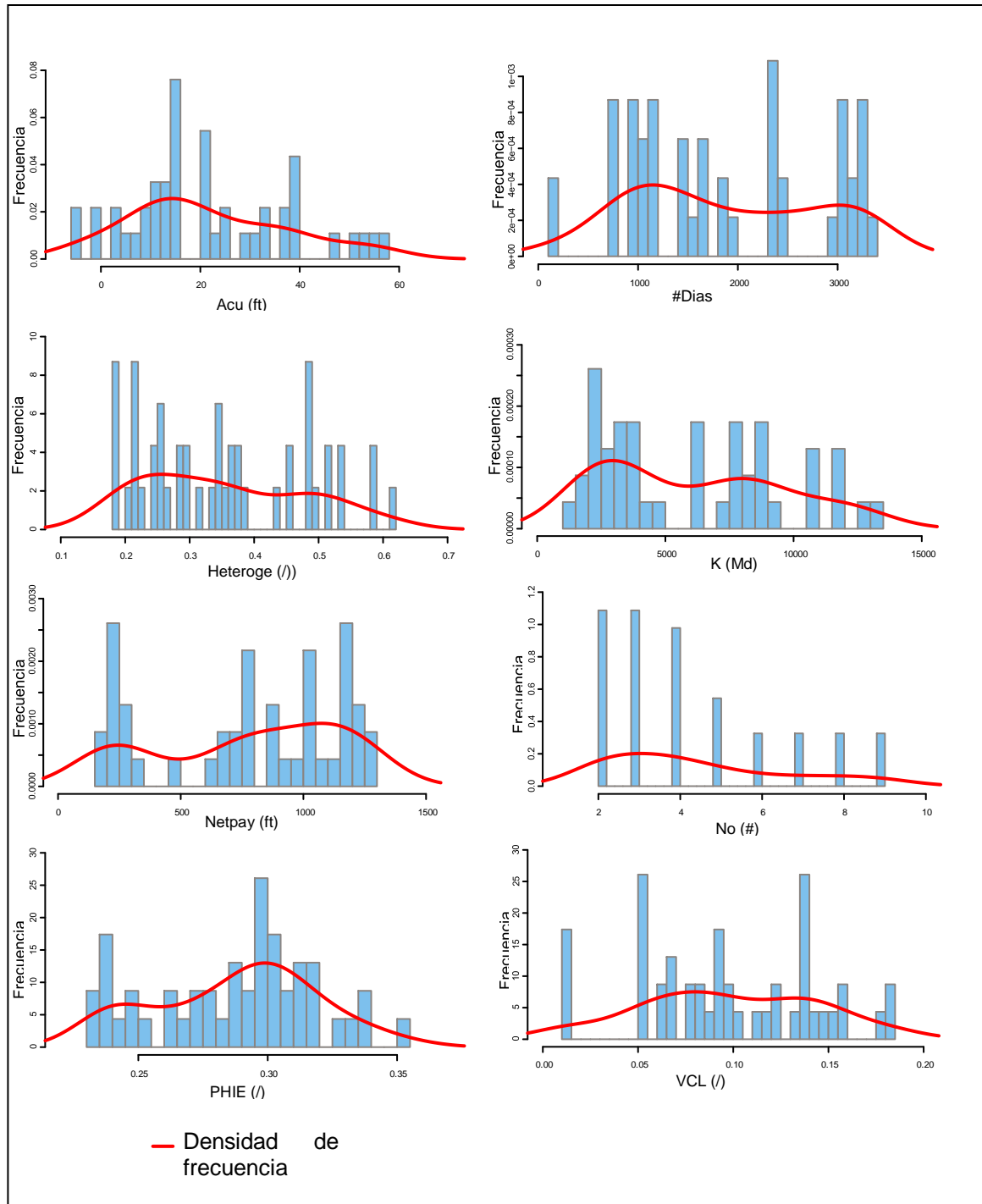
La grafica que resulta de plasmar los valores de la capacidad acumulada y normalizada de la porosidad contra la capacidad acumulada y normalizada de la permeabilidad se debe realizar para todos los pozos de la muestra de estudio, dentro de los anexos del proyecto en la Anexo A se puede observar este comportamiento para los pozos “Imperio 1”, “Imperio 2” e “Imperio 4”.

**2.3.2 Distancia con el acuífero.** Por medio de la interpretación de los registros de un pozo vertical, los petrofísicos pueden estimar un contacto de agua dentro de la arena que se desea navegar, este se puede determinar como el WOC y es reportado a una profundidad vertical en pies. Posterior a ello la información de registro de pozo LWD se integra con el survey proporcionado por el MWD, y se obtiene los registros medidos tanto en profundidad como en profundidad vertical (TVD). Finalmente, el parámetro de distancia con el acuífero se calcula a partir de la resta de la profundidad vertical del WOC y la navegación del pozo en profundidad vertical (TVD), asumiendo que el contacto del agua no tiene variación significativa en las zonas cercanas al pozo.

**2.3.3 Días de producción.** Diariamente en los pozos del campo Imperial se toman los registros de los parámetros de producción como los barriles de petróleo, barriles de agua, corte de agua, presión en cabeza, entre otros. Al conocer la fecha del primer día de producción del pozo, se puede determinar cuántos días lleva produciendo el pozo a una fecha determinada. Para calcular los días de producción a las diferentes fechas que se realizaron los registros, se procede a restar la fecha del registro menos la fecha del primer día de producción del pozo.

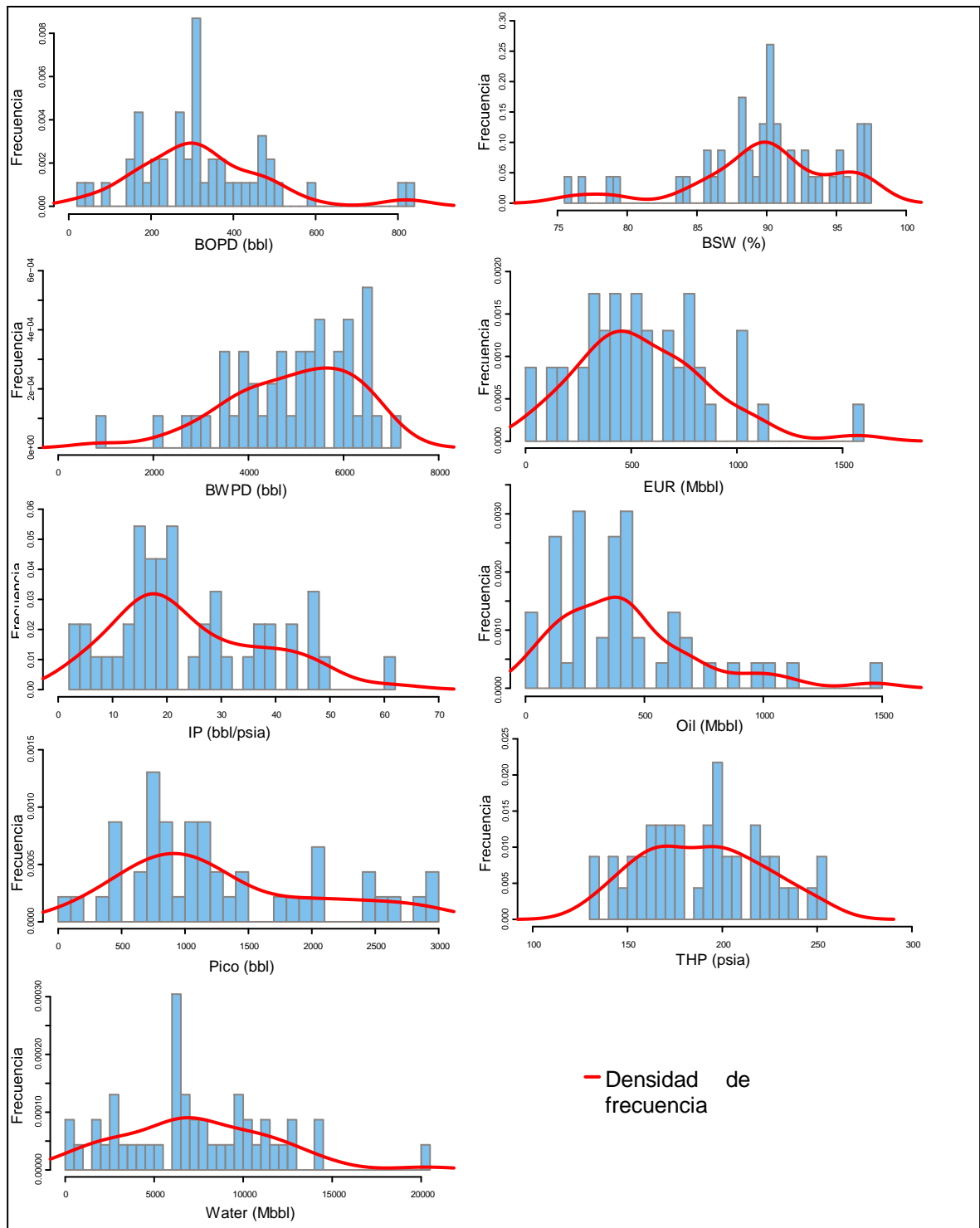
**2.3.4 Pico máximo de producción.** De los datos de producción histórica del pozo, se tiene la variable de barriles de petróleos diarios producidos (BOPD), estos son medidos en campo para su posterior fiscalización. Una vez se posea un historial de producción significativo (aproximadamente un año), al graficar los barriles de producción diaria contra el tiempo se puede estimar el valor máximo de estos el cual se denomina pico máximo de producción.

**Figura 8.** Datos de los parámetros petrofísicos de la muestra de estudio



Fuente: elaboración propia

**Figura 9.** Datos de los parámetros de producción de la muestra de estudio



**Fuente:** elaboración propia

## 2.4 CONCATENACIÓN DE LOS DATOS

Uno de los mayores retos del trabajo de grado, fue la correlación de variables espaciales como lo son los parámetros petrofísicos, es decir que su valor está asociado a una unidad de volumen ubicada en un sistema, con variables temporales como lo son parámetros de producción, los cuales están referenciados a una fecha en específico. Debido a la diferencia de referenciación entre los parámetros, el número de observaciones por variable también es diferente, razón por la cual varios de los métodos estadísticos multivariantes no pueden ser ejecutados para estos datos, debido a que su desarrollo se realiza por medio de matrices cuadradas, a este problema se le dio solución a través de la generación de los siguientes modelos.

**2.4.1 Modelo de tabla completa.** Se empleo una técnica estadística denominada “*Left join*”, la cual tiene como objetivo combinar los valores de dos (2) tablas a partir de una variable de referencia, que en este caso fue el nombre del pozo. La compilación de los datos de petrofísica dio como resultado una tabla compuesta por diez (10) variables y treinta y cuatro mil doscientos noventa y seis (34,296) observaciones por cada uno de los parámetros y la tabla de producción se conformó por doce (12) variables y siete mil ochocientos sesenta y cuatro (7,864) observaciones por parámetro.

Finalmente, tras realizar la combinación de ambas tablas por medio de la técnica estadística y tomando la variable de referencia como el nombre del pozo, se obtuvo una tabla completa que cuenta con veintiún (21) variables y dieciséis millones cuatrocientos ochenta y siete mil doscientos treinta y cinco (16'487,235) observaciones, con la cual el método estadístico multivariable de regresión lineal múltiple puede llevarse a cabo y generar el modelo de tabla completa.

**2.4.2 Modelos de promedio y mediana.** Bajo la intención de comparar distintos modelos de regresión lineal múltiple, se procedió a calcular el promedio y la mediana por cada uno de los parámetros tanto petrofísicos como de producción. Esto se realizó con el fin de tener tres (3) modelos (promedio, mediana y tabla completa) y compararlos entre sí para determinar cuál de estos tiene los mejores resultados.

Sin embargo, para poder emplear tanto los valores de los promedios como las medianas de los parámetros, se debe corroborar su nivel de significancia por medio de la diferencia entre el estadístico T y el valor crítico de la variable, el procedimiento a emplear se describe a continuación.

Cabe resaltar que todos los cálculos descritos se deben realizar para cada una de las variables y a su vez por cada pozo. Lo primero es calcular la desviación estándar (cuantifica la dispersión de un conjunto de datos) y número de observaciones. A partir de los datos previos se estima el valor del estadístico T, el cual es usado para verificar la distribución normal de una muestra pequeña, siguiendo una distribución t de Student.

Posterior a ello a partir del número de grados de libertad de la muestra y un nivel de significancia del uno (1) por ciento se determina el valor crítico y finalmente, se compara la magnitud de ambos resultados:

Estadístico  $T >$  valor crítico: Se acepta la hipótesis de que los valores de promedio y mediana son significativos y se generan los dos (2) nuevos modelos.

Estadístico  $T <$  valor crítico: No se pueden tomar los valores de promedio y mediana y se procede a trabajar solo con el modelo de tabla completa.

Tras haber realizado estas operaciones, se verifico que a para todos los parámetros por pozo el estadístico T es mayor que su valor crítico, razón por la cual los modelos de promedio y mediana son estadísticamente significativos y se pueden emplear en el método de regresión lineal múltiple.

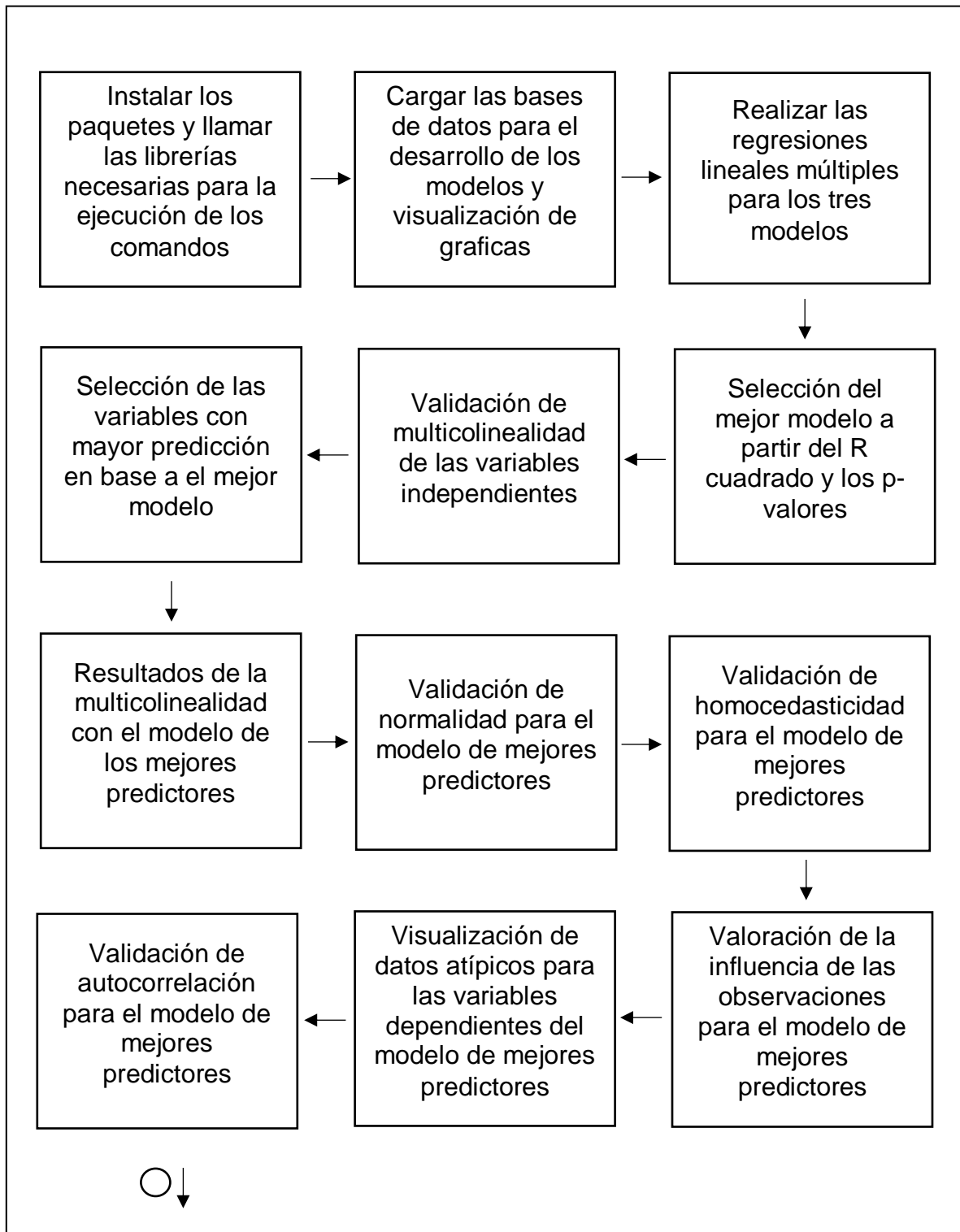
## **2.5 PROGRAMACIÓN EN EL LENGUAJE “RSTUDIO”**

Uno de los factores que conlleva a emplear esta herramienta computacional (código R) es el interés por utilizar distintas ramas del conocimiento para el desarrollo del trabajo de grado, ya que un ingeniero debe ser capaz de apoyarse en diferentes herramientas para poder resolver los problemas que se le presenten, por esta razón se incorporan diferentes técnicas y conocimientos de la ingeniería de petróleos, la estadística y la computación.

El proyecto maneja una gran cantidad de información, es por esto que la herramienta estadística computacional como lo es R y su extensión R Studio resulta muy útil por ser un lenguaje de programación de código libre con un enfoque estadístico y grandes virtudes a la hora de desarrollar un modelos y visualización de los datos y resultados. En mil novecientos noventa y tres (1993) se dio a conocer el programa por Ross Ihaka y Robert Gentleman, tiene sus bases en el lenguaje S, con una gran ampliación en el aspecto de aprendizaje automático.

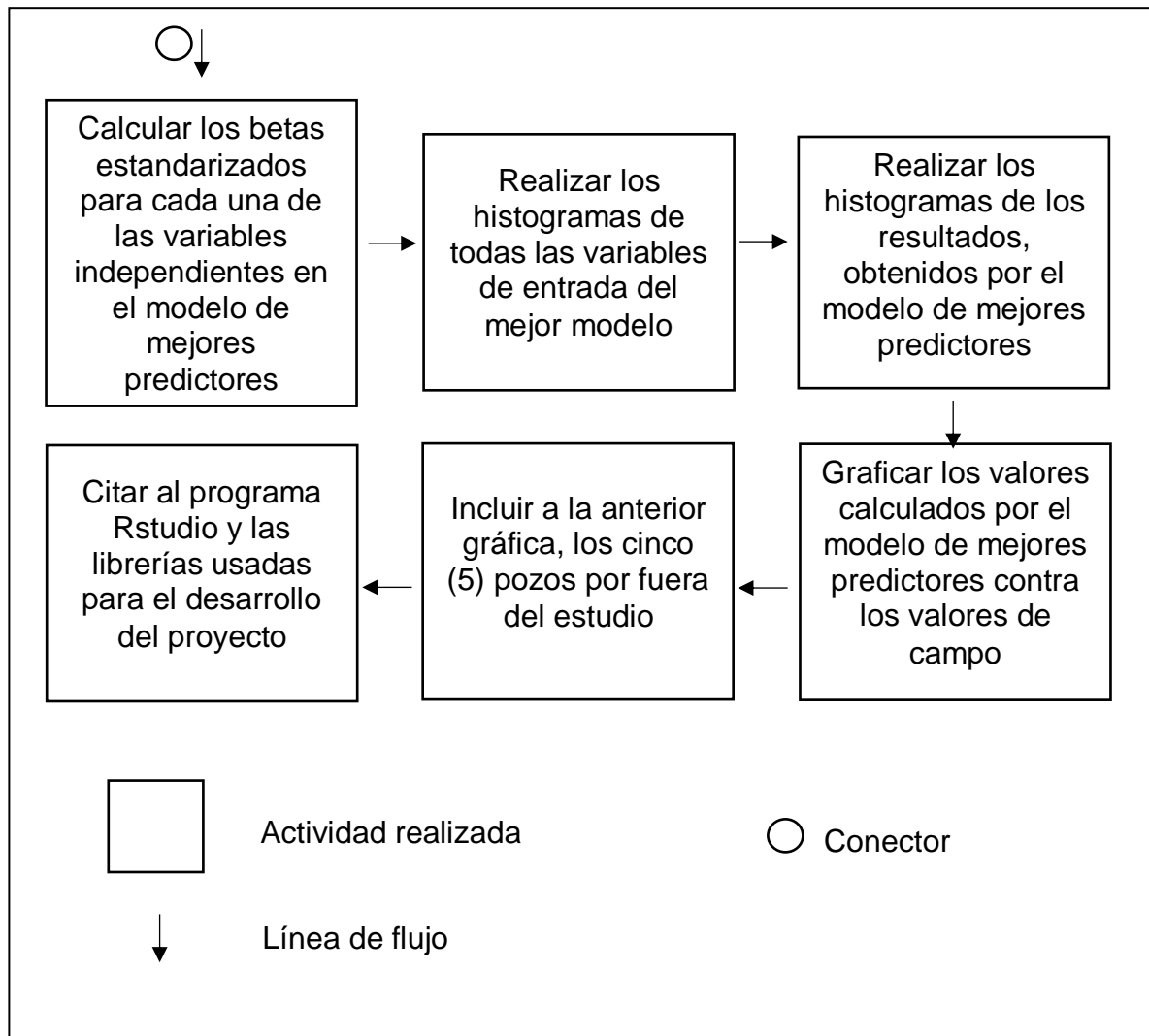
Esta es la actividad principal del trabajo de grado debido a que es donde se ingresan los datos, estos son modificados, se generan las regresiones lineales múltiples por cada modelo, se comparan los modelos, se descartan múltiples problemas y finalmente se visualizan los resultados. El desarrollo de esta actividad se describe paso a paso por medio de un diagrama de flujo que se puede visualizar en la figura 10.

**Figura 10.** Diagrama de flujo para la programación en el lenguaje “Rstudio”





**Figura 10.** (Continuación)



**Fuente:** elaboración propia

Esta serie de pasos son la descripción de un código completo el cual consta de setecientos cincuenta y siete (757) líneas de programación y un total de treinta y un (31) comandos diferentes, los cuales pueden ser visualizados junto con su descripción en el cuadro 1.

**Cuadro 1.** Comandos del Código

Comando	Descripción
<code>install.packages("paquete")</code>	Instala en el programa Rstudio los paquetes requeridos para desarrollar los objetivos del autor.
<code>library(libreria)</code>	Trae de los paquetes las librerías que contiene los códigos que se emplearan.
<code>setwd("ruta")</code>	Establece una ruta en el computador para insertar archivos al código.
<code>excel &lt;- read_excel("nombre")</code>	El programa reconoce archivos Excel como los valores de variables.
<code>colnames(excel)</code>	Despliega los nombres de las variables de un archivo Excel.
<code>Excelf = left_join(Excel 1,Excel 2,by="variable")</code>	Unifica dos tablas Excel usando una variable de referencia por medio de la técnica "left join".
<code>modelo &lt;- lm(`variable` ~ predictores, excel)</code>	Ejecuta un modelo de regresión lineal multivariable a partir de la definición de la variable dependiente y las independientes.
<code>summary(modelo)</code>	Muestra los resultados obtenidos por el modelo de regresión lineal multivariable.
<code>ggpairs(excel, descripción), diag = list(descripción), ejes)</code>	Realiza un gráfico cruzado entre un grupo de variables mostrando su valor de correlación.
<code>corrplot(cor(select(excel,predictores)), descripción)</code>	Clasifica la correlación entre un grupo de variables
<code>step(object = modelo, descripción)</code>	Evalúa un modelo de regresión lineal multivariada para establecer los mejores predictores.
<code>vif(modelo)</code>	Asigna valores a cada variable para establecer su correlación lineal e inflación de varianza.
<code>plot&lt;-ggplot(excel, aes(variables))+geom_point()+ geom_smooth(descripción)+geom_hline(intercepto)+ theme_bw()</code>	Grafica la dispersión por cada variable dependiente del modelo en torno al cero para evaluar la normalidad.
<code>grid.arrange(plots)</code>	Unifica diferentes plots en solo una imagen.

**Cuadro 1.** (Continuación)

Comando	Descripción
qqnorm(modelo)	Grafica los residuos frente a los valores teóricos de los cuantiles para evaluar la normalidad.
qqline(modelo)	Grafica la línea de normalidad.
shapiro.test(modelo)	Realiza un test de hipótesis para verificar la normalidad de las variables.
bptest(modelo)	Realiza el test estandarizado de Breusch-Pagan para verificar la homocedasticidad del modelo.
dwt(modelo, alternative = "two.sided")	Calcula el estadístico D-W para evaluar la autocorrelación de las variables.
modelo\$residuals <- rstudent(modelo)	Asigna los resultados de los residuos del modelo a una variable.
ggplot(excel,aes(x = predict(modelo), y=abs(modelo\$residuals))) +geom_hline (descripción)+geom_point(aes(descripción)) +scale_color_identity()+labs(ejes)+theme_bw() + theme(titulo))	Grafica los valores atípicos por cada una de las variables con el fin de identificar observaciones anormales.
which(abs(modelo\$residuals) > 3)	Identifica cuales y cuantos datos atípicos hay en los datos por cada variable.
summary(influence.measures(modelo))	Reconoce las observaciones que son significativamente influyentes para los predictores por medio de los valores de Leverages y distancias de Cook.
influencePlot(modelo)	Muestra de forma gráfica las observaciones que son significativamente influyentes.
lm.beta (modelo)	Calcula los betas estandarizados de las variables independientes del modelo.
hist(excel, freq =false, ejes, escalas,descripción) density(excel), lines(density(excel),descripción) legend(posición, descripción)	Realiza un histograma de frecuencia por cada una de las variables ingresadas al modelo con su línea de densidad.
predict(modelo, interval = "prediction")	Calcula el intervalo de confianza del 95% por observación

**Cuadro 1.** (Continuación)

Comando	Descripción
<code>cbind(Excel, x)</code>	Concatena dos tablas de datos
<code>ggplot(Excel, ejes)+geom_point() + stat_smooth(method = lm)+ geom_line()+theme_bw()</code>	Grafica los valores calculados contra los actuales mostrando los intervalos de confianza y predicción al 95%.
<code>citation()</code>	Muestra la forma de citar al programa Rstudio.
<code>citation(paquete)</code>	Muestra la forma de citar los paquetes usados en el código.

**Fuente:** elaboración propia

## 2.6 CÁLCULO DE LOS PARÁMETROS DE PRODUCCIÓN

A partir de la metodología de la programación en el lenguaje “Rstudio”, se obtienen para cada modelo y cada uno de los parámetros de producción una ecuación que se encuentra en función de los parámetros petrofísicos, la cual tiene la siguiente forma.

**Ecuación 19.** Ecuación modelo de la regresión lineal múltiple.

$\text{Parámetro de producción} = \beta_0 + \beta_1 \text{PHIE} + \beta_2 \text{VCL} + \beta_3 \text{K} + \beta_4 \text{Acu} + \beta_5 \text{Netpay} + \beta_6 \text{Heteroge} + \beta_7 \text{No} + \beta_8 \text{Dias}$
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Fuente:** elaboración propia

Una vez plateado los tres (3) modelos (promedio, mediana y tabla completa) se determina cuál de estos cuenta con los valores de R cuadrado más altos para tener una mejor capacidad explicativa del parámetro de producción y los p-valores más bajos para que los parámetros petrofísicos tengan un alto nivel de significancia, para así, denominar el mejor modelo.

El modelo de promedio se compone de los valores que son característicos de todas las observaciones por parámetro evaluado y se obtiene a partir de la sumatoria de todas las observaciones sobre el número de sumando. El modelo de mediana corresponde a los valores por parámetro que se encuentra en la posición central del conjunto de observaciones ordenadas por magnitud.

A partir del mejor modelo escogido y la estimación de los mejores predictores sobre este, se obtiene las ecuaciones finales para el cálculo de los parámetros de producción para la muestra de validación, estos resultados pueden ser visualizados

en la figura 26. Todas las ecuaciones obtenidas para los modelos pueden ser visualizadas en los Anexos B, C y D.

## **2.7 COMPARACIÓN DE LOS RESULTADOS**

Como método de validación de los resultados, en el presente trabajo de grado se realizó la comparación de los valores obtenidos por las ecuaciones del mejor modelo de mejores predictores con los datos de campo para los pozos pertenecientes a la muestra de validación. Esto se realizó a partir de graficar los valores actuales de los pozos contra los valores que fueron predichos por el modelo, además de la creación de un intervalo de predicción al noventa y cinco (95) por ciento el cual indica la zona donde individualmente se pueden encontrar los resultados y un intervalo de confianza al noventa y cinco (95) por ciento el cual delimita el área dentro de la gráfica donde se encuentra el valor promedio de la muestra, estas graficas se pueden visualizar en la figura 27.

## **2.8 ESTABLECER LAS CONCLUSIONES**

Finalmente, a partir de las gráficas de valor actual contra valor predicho, betas estandarizados, valores de R cuadrado y p-valores de los modelos se establecen las conclusiones de la investigación con el fin de satisfacer el cumplimiento de los objetivos general y específicos del trabajo de grado y dar unas recomendaciones a otros investigadores para darle una continuidad al proyecto.

### 3. RESULTADOS

Con el fin de cumplir los objetivos propuestos en el trabajo de grado, con la información de los parámetros petrofísicos y de producción disponibles y los modelos de tabla completa, mediana y promedio, se procedió a realizar las regresiones lineales múltiples para saber cuál de estos generaba los mejores resultados.

El desarrollo estadístico se realizó por medio del código programado en el lenguaje R, obteniendo los valores de un R cuadrado y un p-valor por cada modelo y por cada variable, a partir de estos resultados se determinó que de los tres (3) modelos propuestos, el del promedio fue el que mayores valores de R cuadrado y menores de p-valor obtuvo (esto resultados se pueden visualizar de la gráfica 1 a la gráfica 9), por esta razón todas las posteriores operaciones y validación de problemas se hicieron para este modelo.

#### 3.1 DESARROLLO DE LAS REGRESIONES LINEALES MÚLTIPLES

Una vez las medidas estadísticas poblacionales son validadas con el estadístico T, se establecen los tres (3) modelos (tabla completa, mediana y promedio) a emplear en las regresiones lineales múltiples. El procedimiento consta de ingresar las bases de datos al lenguaje R mediante el código y por medio de las instalación de una serie de paquetes, el método estadístico de regresión lineal múltiple es corrido para los tres (3) modelos mediante unos comandos.

Las regresión lineal múltiple nos permiten generar un modelo lineal, en el cual los parámetros de producción son determinados a partir del conjunto de parámetros petrofísicos, brindando la opción de determinar las respuestas de forma cuantitativa y/o para evaluar la influencia de los predictores, donde la ecuación modelo tiene la forma de la ecuación 9. Los resultados del modelo son el R cuadrado quien evidencia la capacidad explicativa del conjunto de parámetros petrofísicos sobre los de producción en modo de fracción, p-valores quienes evidencian la significancia y los betas que se asocian a cada parámetro petrofísico guardando las unidades de la variable dependiente en el numerador y las unidades de su variable en el denominador, razón por la cual no son comparables entre sí.

El procedimiento general que realiza el programa R internamente para la generación de los modelos de regresión lineal múltiple se evidencia en una serie de pasos.

Multiplicar cada observación de los parámetros petrofísicos por su valor correspondiente en el parámetro de producción evaluado y realizar la sumatoria.

Multiplicar por observación, todos los parámetros petrofísicos y realizar la sumatoria.

Elevar al cuadrado por observación, todos los parámetros petrofísicos y realizar la sumatoria.

Elevar al cuadrado por observación, el parámetro de producción evaluado y realizar la sumatoria.

Realizar la sumatoria de las observaciones de todos los parámetros petrofísicos.

Realizar la sumatoria de las observaciones del parámetro de producción evaluado.

Calcular el número de observaciones.

Genera un sistema de ecuaciones cuadrado, donde el tamaño de este es igual al número de parámetros petrofísicos más uno (1).

Resolver el sistema de ecuaciones por métodos numéricos para obtener los valores de los betas y el intercepto.

Calcular la suma de cuadrados total y suma de cuadrados de regresión.

Dividir la suma de cuadrados de regresión entre la suma de cuadrados total para obtener el valor del R cuadrado con la ecuación 9.

Calcular el valor del estadístico T con la ecuación 10.

Hallar la potencia de prueba para el estadístico T, que se tomara como el valor de p-valor.

Ya habiendo realizado las regresiones lineales múltiples para todos los modelos y todos los parámetros petrofísicos, los resultados pueden ser visualizados en R mediante un comando y estos se encuentra sintetizados en la tabla 1.

**Tabla 1.** Resultados de los modelos tabla completa, mediana y promedio

EUR			BOPD			BWPD		
MODELO	R <sup>2</sup>	p-valor	MODELO	R <sup>2</sup>	p-valor	MODELO	R <sup>2</sup>	p-valor
Tabla completa	0.5338	< 2.2E-16	Tabla completa	0.2919	< 2.2E-16	Tabla completa	0.214	< 2.2E-16
Mediana	0.4663	0.003624	Mediana	0.5223	0.003796	Mediana	0.6634	0.001286
Promedio	0.6824	8.24E-08	Promedio	0.6315	3.76E-06	Promedio	0.7507	4.42E-09
BSW			Oil			Water		
MODELO	R <sup>2</sup>	p-valor	MODELO	R <sup>2</sup>	p-valor	MODELO	R <sup>2</sup>	p-valor
Tabla completa	0.2243	< 2.2E-16	Tabla completa	0.6312	< 2.2E-16	Tabla completa	0.8074	< 2.2E-16
Mediana	0.5238	0.007308	Mediana	0.672	0.00177	Mediana	0.7496	0.004447
Promedio	0.6725	5.05E-07	Promedio	0.7844	3.40E-10	Promedio	0.8654	7.30E-14
IP			THP			Pico		
MODELO	R <sup>2</sup>	p-valor	MODELO	R <sup>2</sup>	p-valor	MODELO	R <sup>2</sup>	p-valor
Tabla completa	0.2325	< 2.2E-16	Tabla completa	0.0802	< 2.2E-16	Tabla completa	0.6659	< 2.2E-16
Mediana	0.7907	0.002241	Mediana	0.3688	0.007173	Mediana	0.5628	0.001843
Promedio	0.8042	6.14E-11	Promedio	0.6749	4.46E-07	Promedio	0.7318	3.90E-09

**Fuente:** elaboración propia

Los datos que nos muestra la tabla 1 son los resultados tras realizar las regresiones lineales múltiples para los modelos iniciales, donde se buscan valores altos del R cuadrado y bajos para el p-valor. Una observación importante es que los valores del p-valor para el modelo tabla completa siempre es el mismo ( $<2.2E-16$ ), siendo a priori un valor positivo por su pequeña magnitud, sin embargo esto se debe a un problema de autocorrelación de los datos temporales, razón por la cual este valor no tiene significancia y en el caso de haber escogido este modelo, hubiese sido necesario corregir este problema.

Al realizar la comparación de los valores de R cuadrado y los p-valores de los modelos, se observa que en todos los parámetros de producción el mayor valor para el R cuadrado es el modelo de promedio, demostrando que la capacidad explicativa de las variables petrofísicas en este modelo es el mejor siempre y cuando los p-valores sean menores a cero punto cero uno (0.01) lo cual representa un uno (1) por ciento. Al compara los valores del p-valor para los modelos (menos el de tabla completa por problemas de autocorrelación) se puede apreciar que los menores valores se encuentran en el modelo promedio, y debido a la orden de los resultados se le da un gran peso al valor del R cuadrado de este modelo. Finalmente, observando los resultados de los modelos, se establece que el modelo con mejores resultados es el promedio y será el modelo para trabajar en las siguientes etapas.

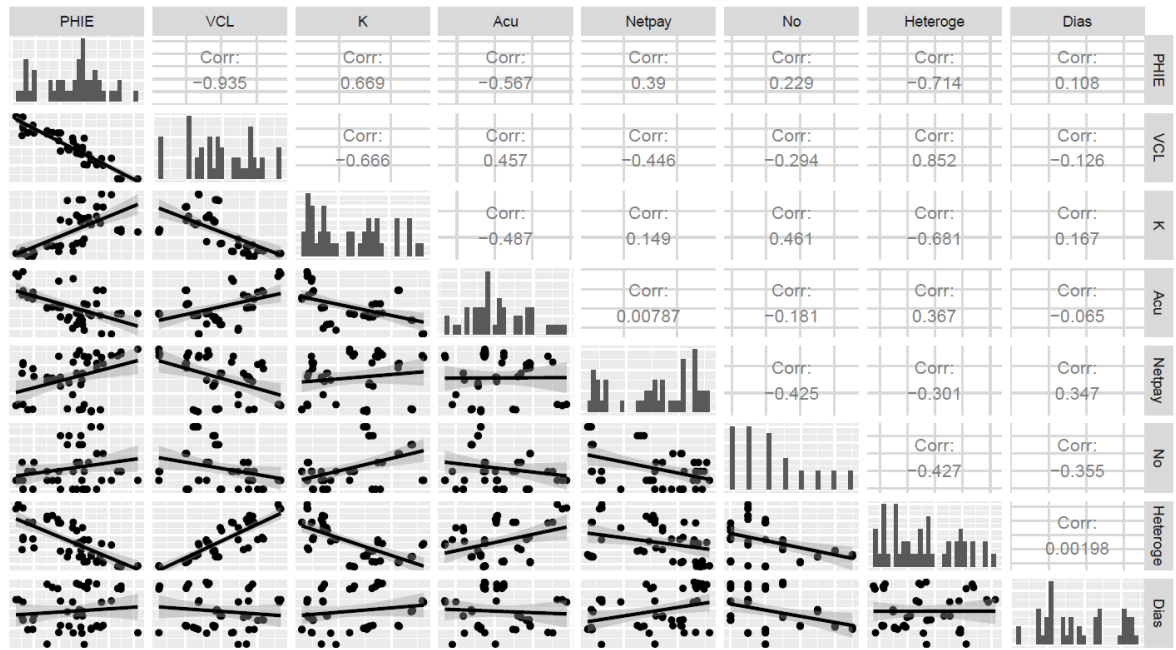
### **3.2 ANÁLISIS DE LA RELACIÓN ENTRE PARÁMETROS PETROFÍSICOS PARA EL MODELO PROMEDIO**

Ya definido el modelo de promedio como el mejor modelo, el primer tratamiento al que se sometió, fue la evaluación de las correlaciones existentes entre los parámetros petrofísicos, ya que estos son las variables independientes. Esta operación es fundamental pues busca conocer cuáles de estos parámetros guardan una relación estrictamente lineal entre sí, debido a que ingresar dos (2) o más variables independientes que presenten un comportamiento similar, restaría peso explicativo a estas variables.

El fundamente para realizar esta operación que los predictores (parámetros petrofísicos) deben ser independientes, es decir que no debe existir colinealidad entre ellos, debido a que no se puede identificar de forma precisa el impacto individual de las variables que guarden una relación lineal entre ellas sobre el parámetro petrofísico evaluado y finalmente, conllevaría a un aumento de la varianza de los coeficientes (betas) y dependiendo la magnitud de la divergencia, podría ser prácticamente imposible determinar su significancia estadística. Se realizaron todos los gráficos cruzados de los parámetros petrofísicos y se calculó su nivel de correlación para poder observar si guardaban una relación lineal significativa que pudiese repercutir en un problema a momento de evaluar el modelo. Los resultados los podemos visualizar en las figuras 11 y 12.

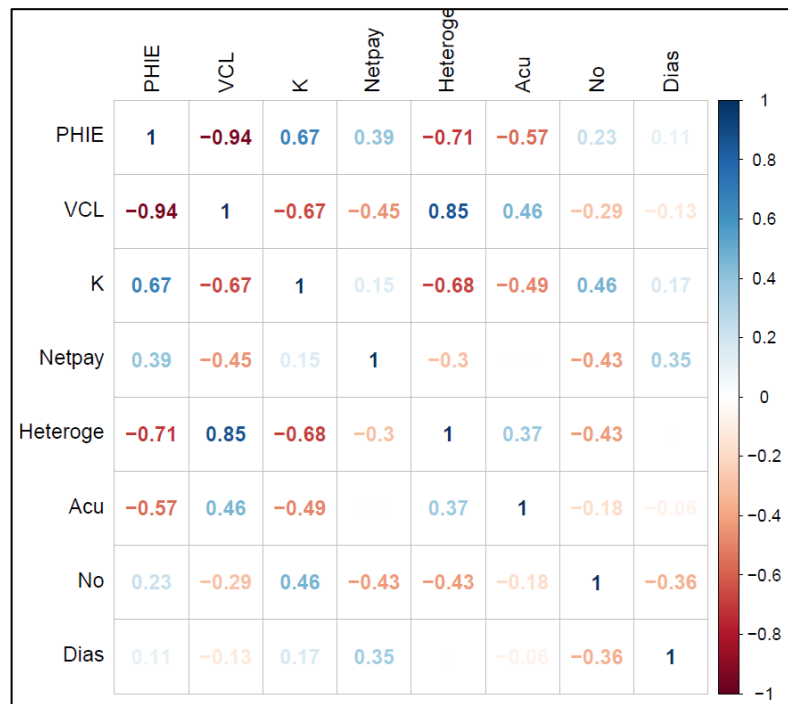


Figura 11. Gráficos cruzados de parámetros petrofísicos



Fuente: elaboración propia

Figura 12. Correlación entre parámetros petrofísicos



Fuente: elaboración propia

Observando los resultados de las gráficas 11 y 12 podemos determinar que existe una colinealidad en algunas variables, lo cual se evidencia por medio de la tendencia lineal entre los gráficos cruzados y la cercanía a los valores de correlación a el uno (1) y menos uno (-1). Dentro de los casos más críticos encontramos la relación entre la porosidad efectiva y el volumen de arcilla, el volumen de arcilla y la heterogeneidad, la permeabilidad y el volumen de arcilla y por último la permeabilidad y la porosidad. Estas correlaciones lineales no son un fenómeno nuevo, es conocido el hecho de que las propiedades petrofísicas de la roca son dependientes unas de otras como es el ejemplo de que a mayor volumen de arcilla menor porosidad. Los gráficos cruzados de la figura 11 pueden ser visualizados en el Anexo F.

Se pudo identificar que de las relaciones de colinealidad el parámetro más común es el volumen de arcilla, razón por la cual se puede llegar a considerar el hecho de omitir este parámetro en el estudio, sin embargo se optó por evaluar cuales son los mejores predictores por cada modelo y observar si en los modelo finales estas correlaciones se encuentran implícitas.

### **3.3 SELECCIÓN DE LOS MEJORES PREDICTORES PARA EL MODELO PROMEDIO**

En la búsqueda de encontrar la existencia de una correlación entre variables, inicialmente se plantearon ocho (8) parámetros petrofísicos y nueve (9) parámetros de producción, sin embargo, el modelo de regresión lineal múltiple requiere verificar que todas las variables predictoras tienen un nivel mínimo de significancia sobre la variable por predecir. Por esta razón se hace necesario aplicar un método que permita conocer cuáles son los mejores parámetros petrofísicos para predecir cada uno de los parámetros de producción, además de que ayuda a solventar el problema de la colinealidad entre variables independientes.

Dentro de los posibles métodos que se pueden emplear encontramos el jerárquico, de entrada forzada, paso a paso, entre otros. El método que se empleo fue el paso a paso “*stepwise*”, donde se emplean criterios matemáticos para discernir qué parámetros petrofísicos contribuyen significativamente al modelo y en qué orden. Para el desarrollo de este método se pueden emplear tres (3) estrategias diferentes, dirección hacia adelante, dirección hacia atrás y dirección mixta.

Tras seleccionar el método paso a paso y usando la estrategia de dirección mixta, el procedimiento para seleccionar los mejores predictores para el modelo promedio fue, empezar con un modelo inicial que solamente constase del intercepto ( $\beta_0$ ), luego se generan todos los posibles modelos de un solo parámetro petrofísico y aquel que mejore en mayor medida el modelo se selección para conformar la primera variable y primer beta ( $\beta_1$ ). Una vez seleccionado el primer parámetro predictor, este es sometido a un test de extracción de predictores no útiles para evaluar si este deja de contribuir al modelo y es eliminado. Esta operación de añadir

parámetros de forma individual y realizar el test se repite con todos los parámetros petrofísicos hasta encontrar el modelo promedio con mejores predictores.

Es importante aclarar que existen distintos criterios matemáticos para comparar modelos, dentro de los más aceptados encontramos  $C_p$ , AIC, BIC y  $R^2$  ajustado. En el desarrollo del trabajo se empleó el AIC donde el mejor modelo es aquel que presente en valor absoluto el mayor valor del criterio matemático (AIC). Por medio de este procedimiento se obtuvo el modelo promedio con mejores predictores donde los parámetros de producción están en función de algunos parámetros petrofísicos como esta descrito en el cuadro 2, donde el color verde indica que la ecuación cuenta con esa variable y el color rojo evidencia que la variable fue eliminada del modelo. Las ecuaciones del modelo promedio con mejores predictores se puede visualizar en el Anexo E .

**Cuadro 2.** Parámetros petrofísicos en modelo con mejores predictores

	PHIE	VCL	K	Acu	Netpay	No	Heteroge	Dias
EUR	Green	Green	Green	Red	Green	Green	Green	Red
BOPD	Red	Red	Green	Green	Green	Green	Green	Green
BWPD	Green	Red	Green	Green	Green	Green	Green	Green
BSW	Red	Green	Green	Green	Green	Red	Green	Green
Oil	Red	Red	Red	Green	Green	Green	Green	Green
Water	Red	Green	Green	Green	Green	Green	Green	Green
IP	Red	Red	Red	Green	Red	Green	Green	Green
THP	Green	Green	Green	Green	Green	Green	Green	Green
Pico	Green	Red	Green	Red	Green	Green	Green	Red

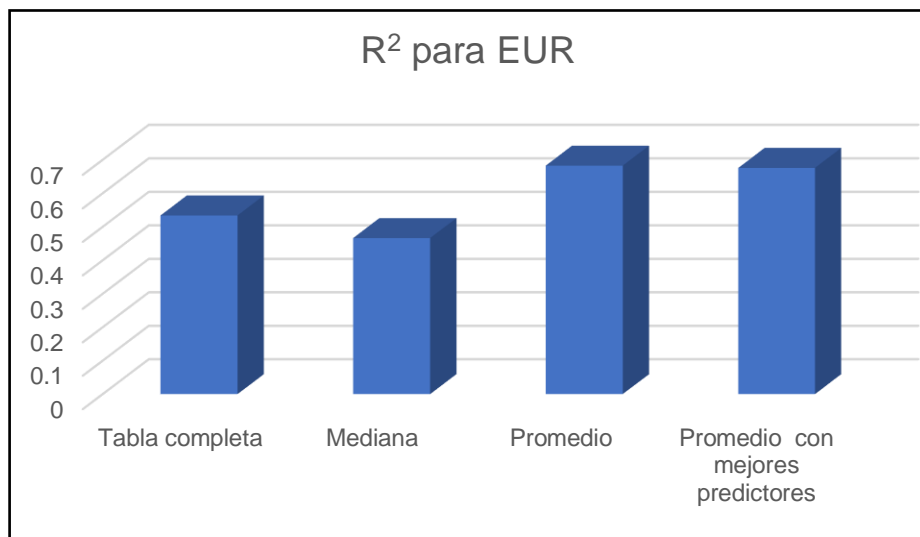
**Fuente:** elaboración propia

A partir de los resultados de la cuadro 2 se puede observar que la variable heterogeneidad no fue omitida en ninguno de los parámetros de producción, dando un indicio de ser una variable clave a la hora de describir la producción de un pozo, así como el número de pozos por cluster y el espesor de explotación que solo fueron omitidas una vez. De forma análoga la porosidad efectiva y el volumen de arcilla fueron omitidos en cinco (5) de los nueve (9) parámetros de producción, evidenciando el problema de colinealidad de estas variables y que no son parámetros tan importantes como otros, estadísticamente hablando, para el modelo. Cabe aclarar que la porosidad y el volumen de arcilla son variables que se encuentran inmersas de forma implícita en otros parámetros, como el espesor de explotación, la heterogeneidad y la permeabilidad, razón por la cual siguen siendo parámetros fundamentales para el proyecto.

Para la presión en cabeza de pozo (THP) ninguno de los parámetros petrofísicos fue omitido, dando a entender que es un parámetro complejo de predecir y que requiere de más variables para poder ser predicha con una alta precisión. De forma análoga el índice de productividad omitió cuatro (4) de los ocho (8) parámetros petrofísicos disponibles dando la evidencia de que solo requiere de pocas variables para ser descrita o que requiere de otro tipo de variables para poder entender su comportamiento.

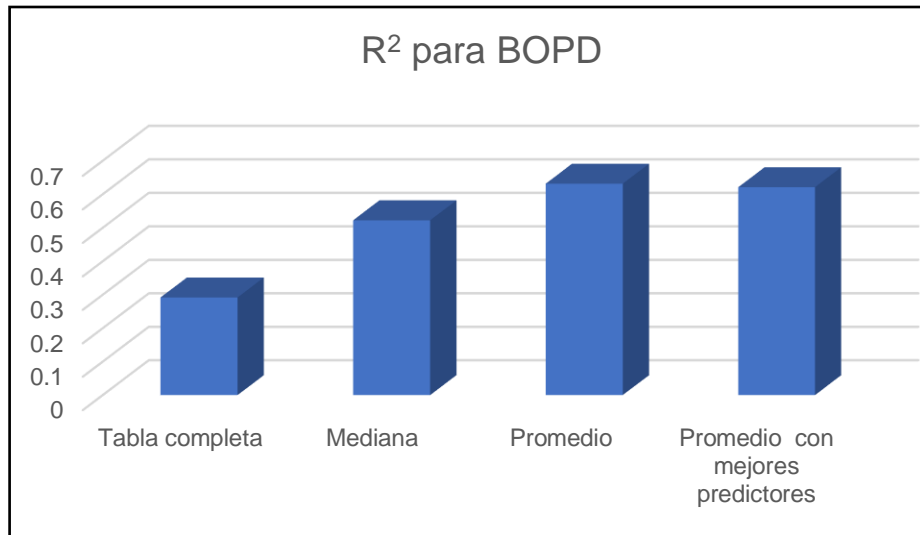
Una vez establecido el modelo promedio con mejores predictores, este es nuevamente corrido en el programa R para poder visualizar sus resultados y poderlos comparar con los tres (3) modelos iniciales (tabla completa, mediana y promedio), esperando así que el valor del R cuadrado sea mayor con el fin de mejorar la capacidad explicativa de los parámetros petrofísicos sobre los parámetros de producción, esta comparación se puede visualizar entre la gráfica 1 y gráfica 9.

**Gráfica 1.** R<sup>2</sup> para EUR



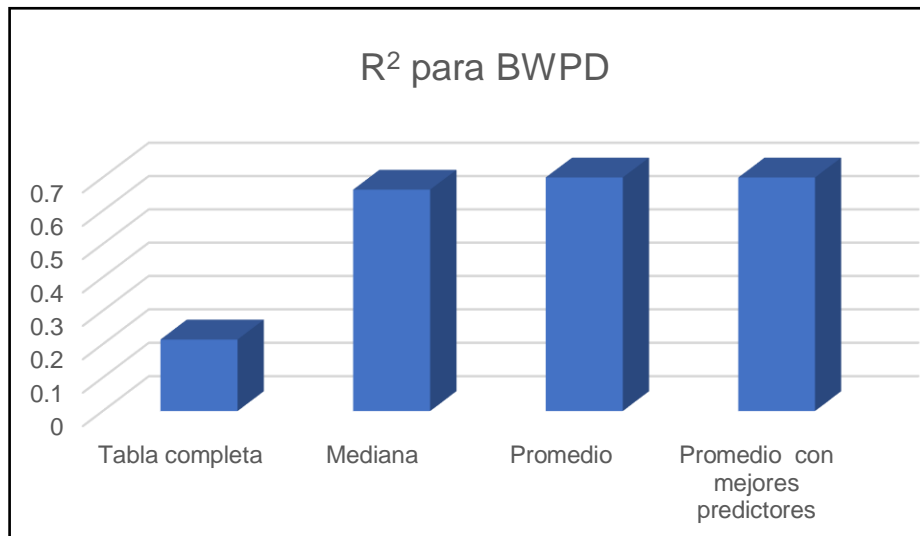
**Fuente:** elaboración propia

**Gráfica 2.**  $R^2$  para BOPD



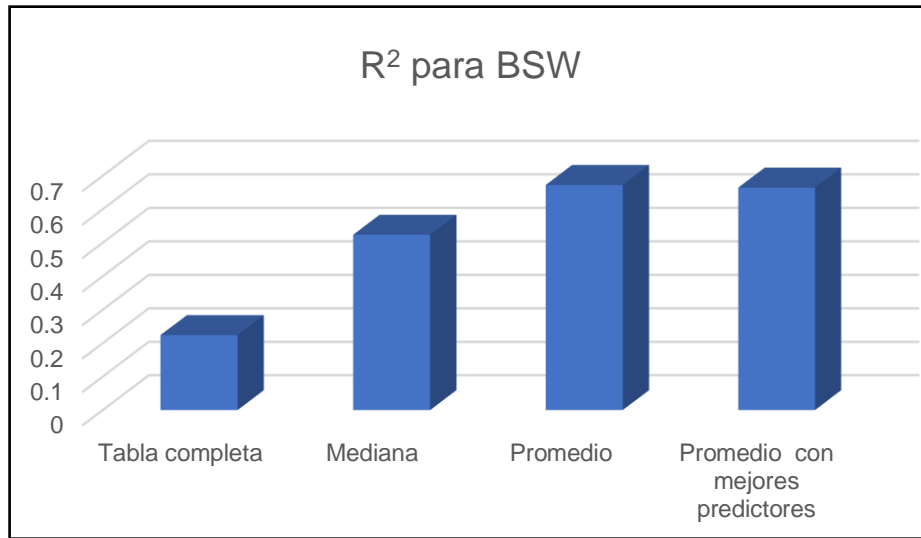
**Fuente:** elaboración propia

**Gráfica 3.**  $R^2$  para BWPD



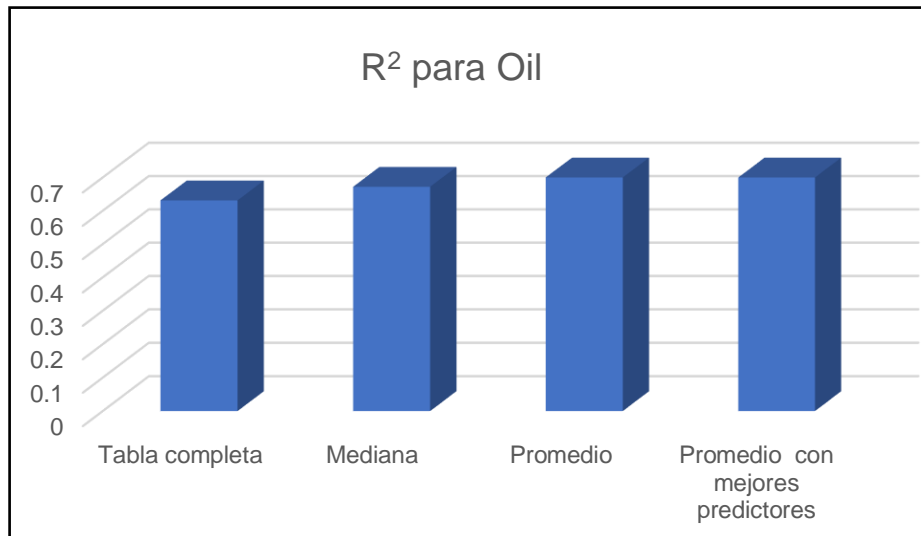
**Fuente:** elaboración propia

**Gráfica 4.**  $R^2$  para BSW



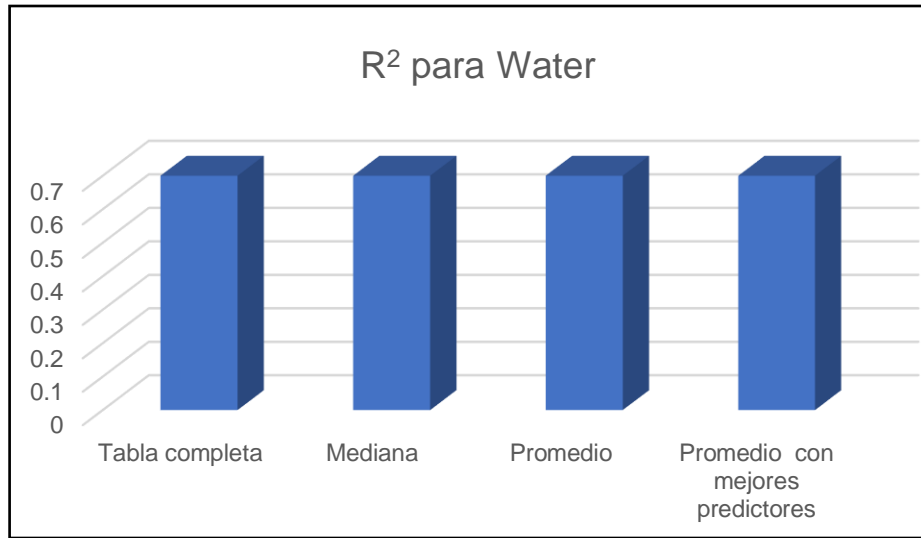
**Fuente:** elaboración propia

**Gráfica 5.**  $R^2$  para Oil



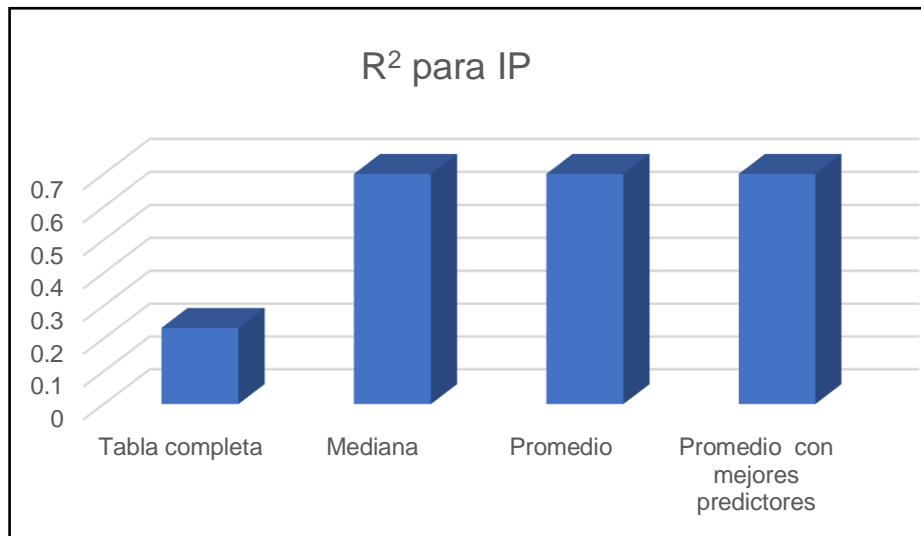
**Fuente:** elaboración propia

**Gráfica 6.**  $R^2$  para Water



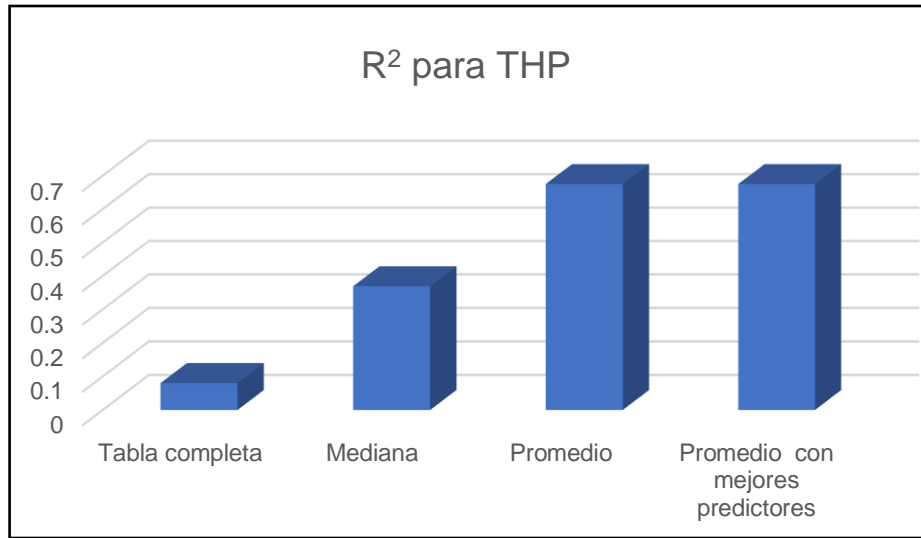
**Fuente:** elaboración propia

**Gráfica 7.**  $R^2$  para IP



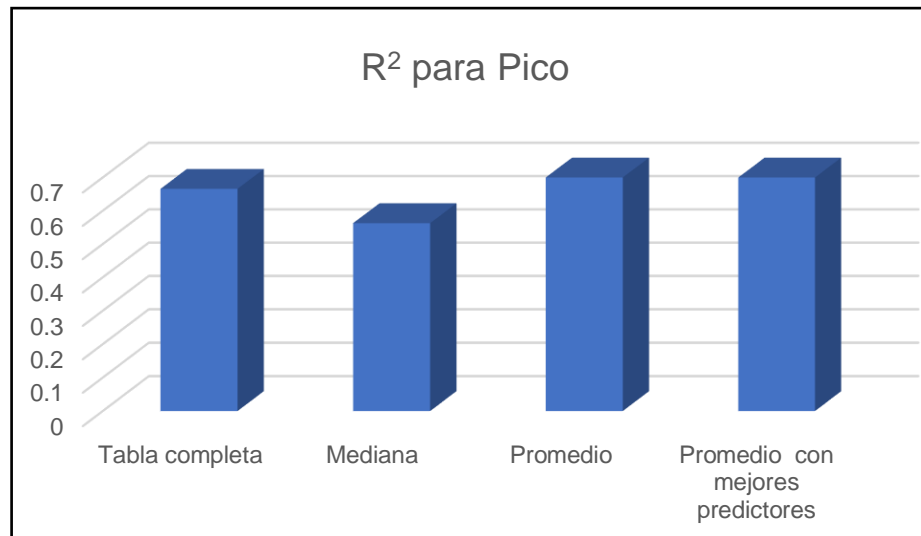
**Fuente:** elaboración propia

**Gráfica 8.**  $R^2$  para THP



**Fuente:** elaboración propia

**Gráfica 9.**  $R^2$  para Pico



**Fuente:** elaboración propia



A partir de las gráficas (gráfica 1 a gráfica 9) se puede evidenciar que los mejores modelos con base al valor de los R cuadrado son el promedio y el promedio con mejores predictores sin una diferencia clara entre ellos, los contrastes más evidentes entre modelos fueron para la THP, IP, BSW, BOPD y BWPD. Sin embargo, es importante aclarar que el modelo promedio con mejores predictores requiere menos parámetros petrofísicos para describir los parámetros de producción que el modelo del promedio, lo cual representa una ventaja. Para las posteriores acciones por ejecutar en el proyecto se realizarán sobre el modelo promedio con mejores predictores.

Retomando la problemática de la colinealidad de los datos ahora sobre el modelo promedio con mejores predictores, se puede observar que las parejas de variables que presentaban colinealidad fueron retiradas de forma parcial. Bajo la intención de conocer la capacidad explicativa de los parámetros petrofísicas sobre los parámetros de producción se continuo el proyecto sin eliminar ninguna de las variables independientes, teniendo el conocimiento de que estas parejas de variables pueden restar capacidad explicativa individual unas a otras, sin embargo se deseó evaluar el problema de forma cuantitativa de la siguiente manera.

Después de haber realizado los gráficos cruzados de los parámetros petrofísicos y el cálculo de la matriz de correlaciones para el modelo promedio bajo la intención de evaluar la colinealidad entre parámetros, se aplica un análisis de varianza (VIF) a el modelo promedio con los mejores predictores con el fin de analizar de forma cuantitativa la colinealidad del nuevo moldeo, la variable VIF es calculada mediante la ecuación 11.

Para la evaluación de la colinealidad de los parámetros petrofísicos del modelo promedio con mejores predictores se emplean los rangos de valores de VIF más usados en la estadística.

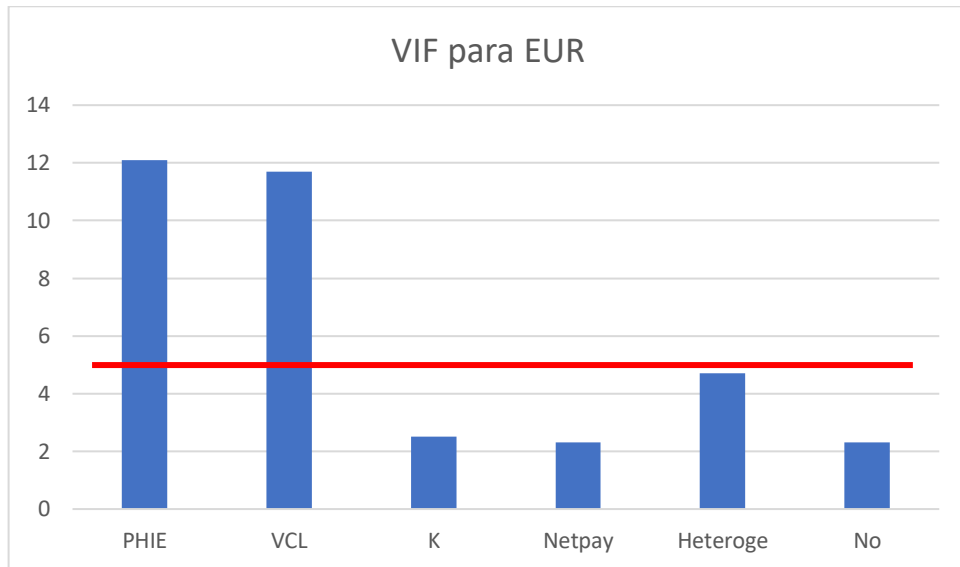
Ausencia total de colinealidad ( $VIF = 1$ ).

La regresión puede verse afectada por cierta colinealidad ( $1 < VIF < 5$ ).

Existencia de la colinealidad ( $VIF > 5$ )

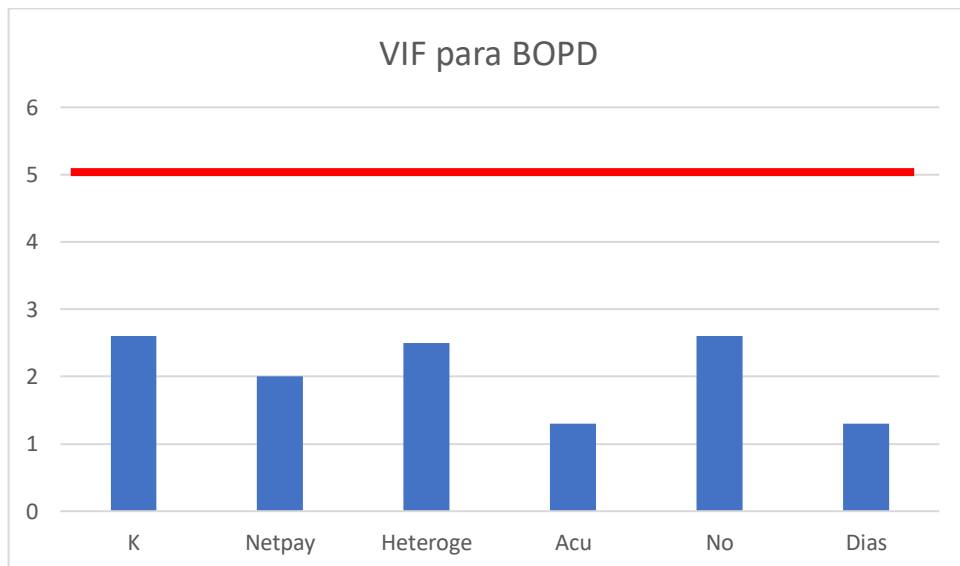
Después de haber realizado el cálculo del VIF para cada uno de los parámetros petrofísicas los resultados pueden ser visualizados de la gráfica 10 a la gráfica 18, donde se estableció el valor de cinco (5) para el factor de inflación de varianza como un indicador.

**Gráfica 10.** VIF de los parámetros petrofísicos del EUR



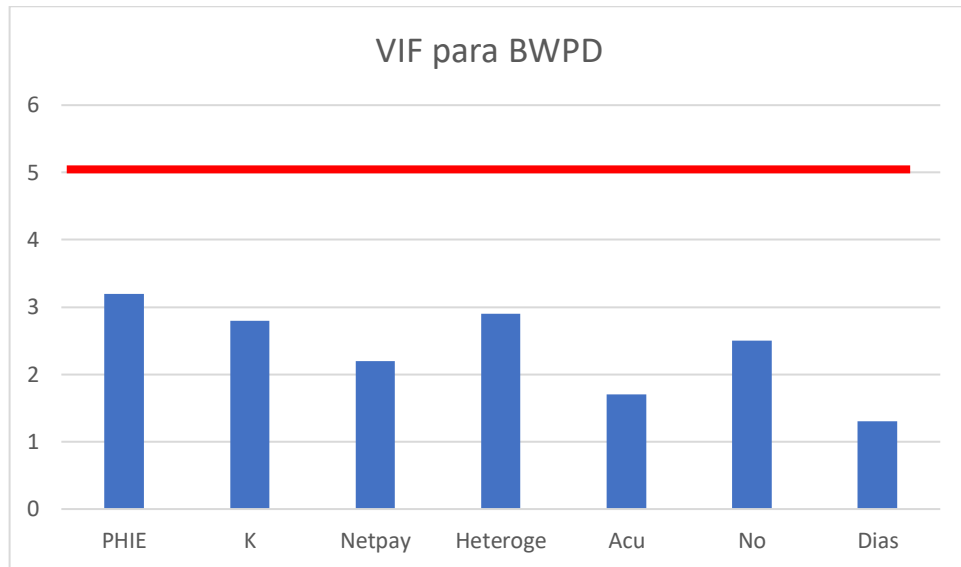
Fuente: elaboración propia

**Gráfica 11.** VIF de los parámetros petrofísicos del BOPD



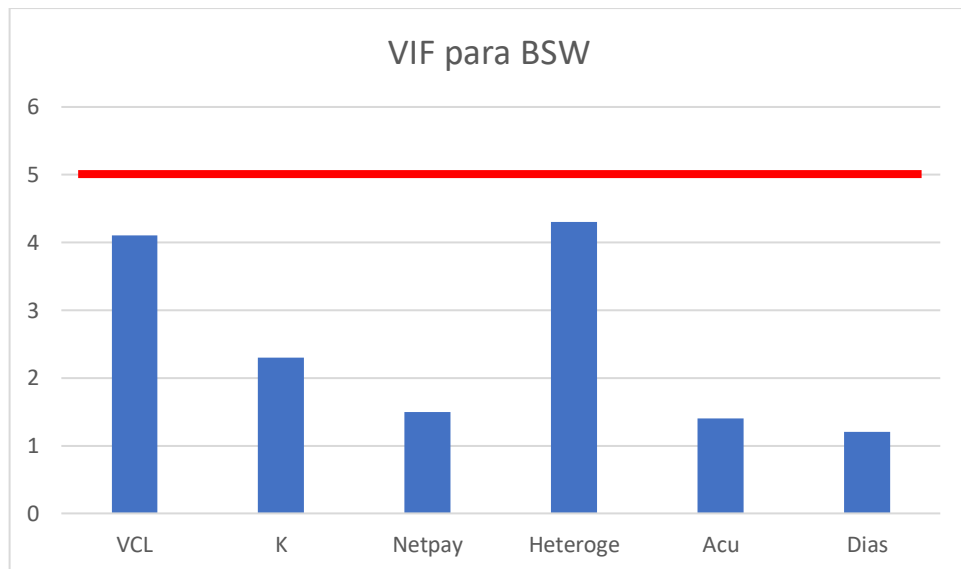
Fuente: elaboración propia

**Gráfica 12.** VIF de los parámetros petrofísicos del BWPD



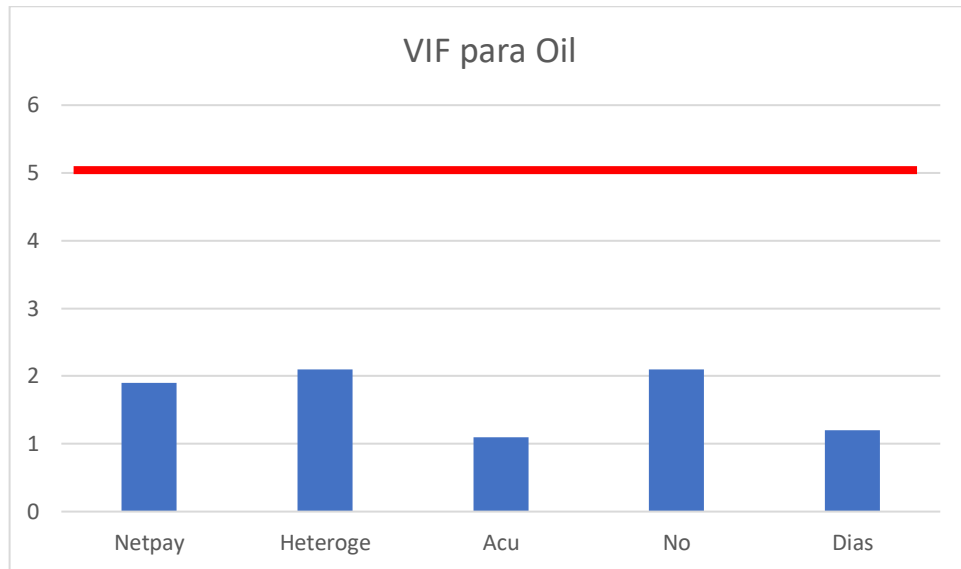
**Fuente:** elaboración propia

**Gráfica 13.** VIF de los parámetros petrofísicos del BSW



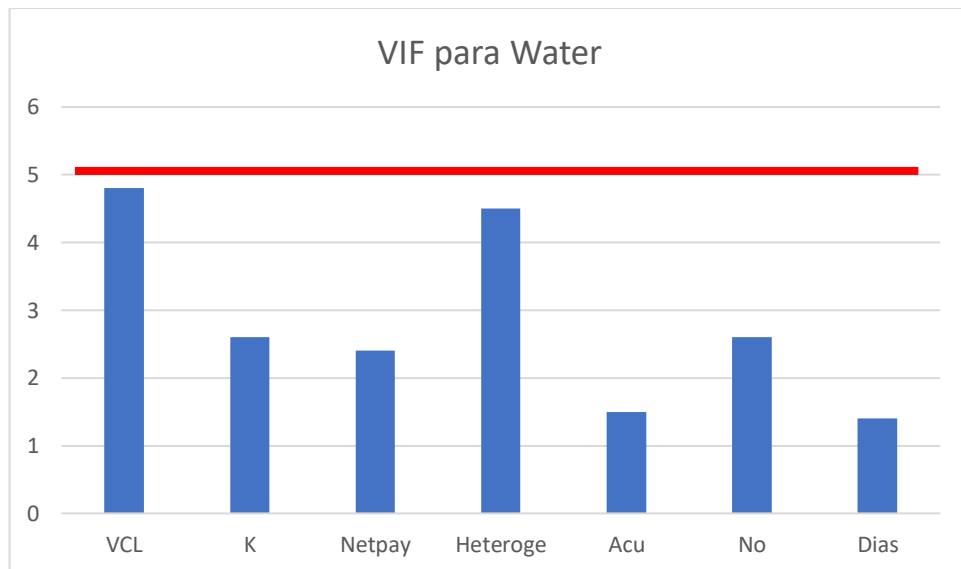
**Fuente:** elaboración propia

**Gráfica 14.** VIF de los parámetros petrofísicos del Oil



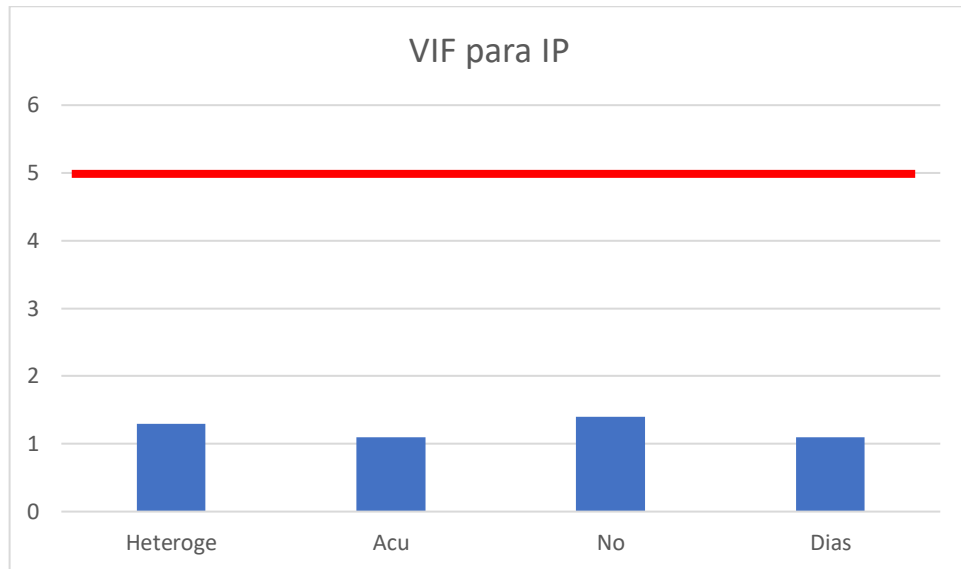
**Fuente:** elaboración propia

**Gráfica 15.** VIF de los parámetros petrofísicos del Water



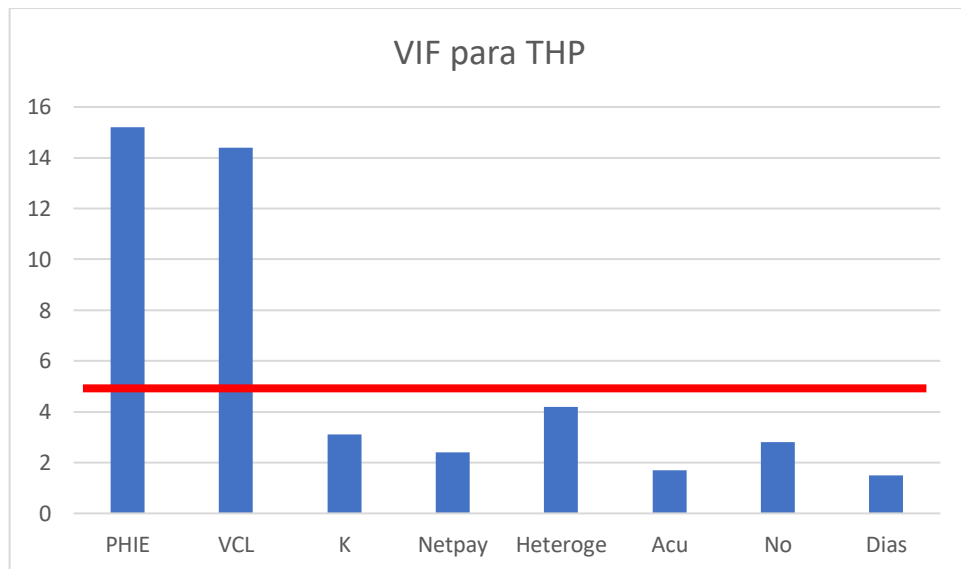
**Fuente:** elaboración propia

**Gráfica 16.** VIF de los parámetros petrofísicos del IP



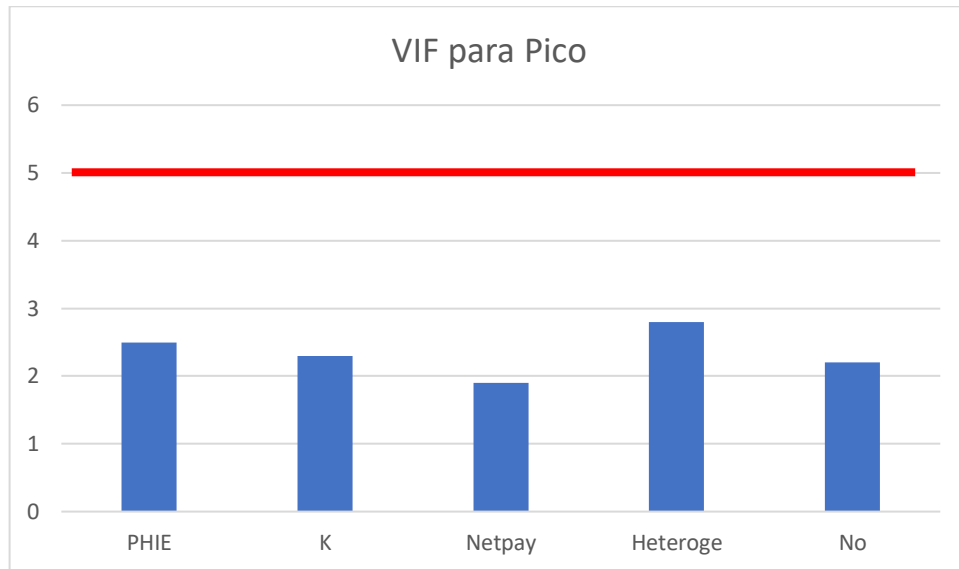
**Fuente:** elaboración propia

**Gráfica 17.** VIF de los parámetros petrofísicos del THP



**Fuente:** elaboración propia

**Gráfica 18.** VIF de los parámetros petrofísicos del Pico



**Fuente:** elaboración propia

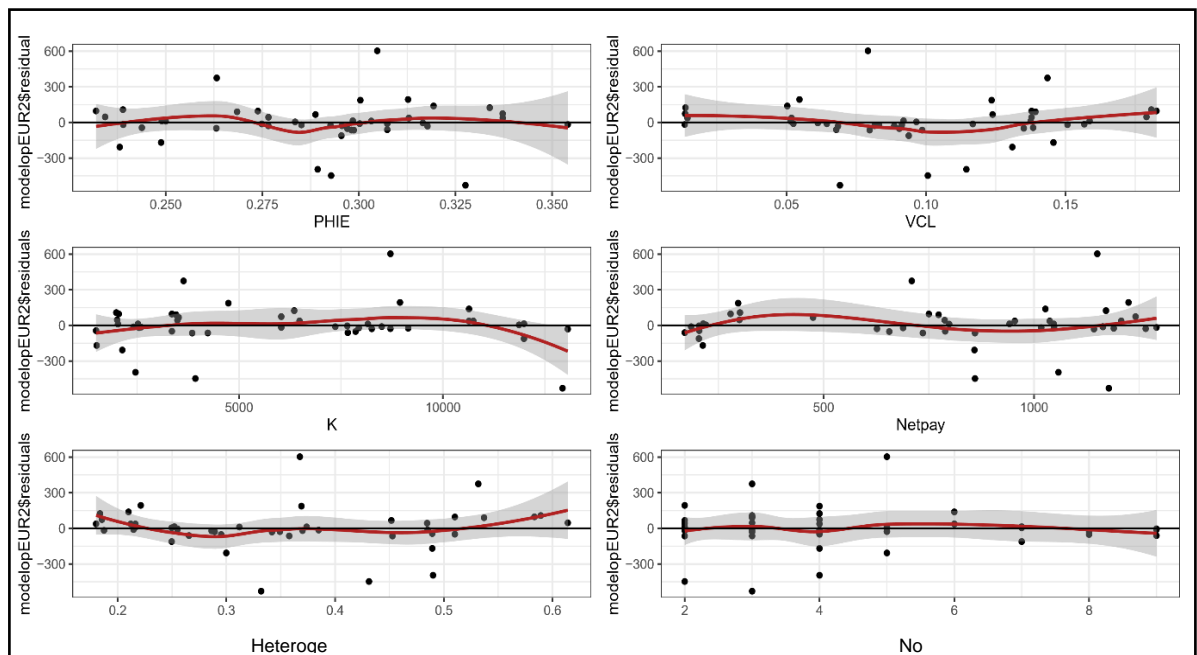
Observando los resultados plasmados entre la gráfica 10 y la gráfica 18, se aprecia que la colinealidad entre variables en el modelo promedio con mejores predictores es menor que en el modelo promedio pero persiste. Dentro de los casos críticos se encuentran los parámetros de producción de EUR y THP, donde la porosidad efectiva y el volumen de arcilla presentan una correlación lineal. Como ya se mencionó, no se procederá a eliminar ninguna de las variables predictoras pero si se tendrá en cuenta para el análisis de los resultados de estos dos (2) parámetros de producción

Analizando los demás parámetros de producción, se destaca una independencia entre parámetros predictores, lo cual favorece a los resultados del modelo. Para los parámetros de producción de petróleo diaria, producción de agua diaria, producción de petróleo acumulada, índice de productividad y pico máximo de producción, sus valores del VIF se encuentran alejados del indicador por lo cual se puede afirmar que los indicios de colinealidad sobre estos modelos son bajos o nulos. Para los parámetros de producción de agua diaria y corte de agua, se observa unos posibles indicios de colinealidad sin ser casos críticos, razón por la cual no es necesario tomar medidas sobre ellos.

### 3.4 VALIDACIÓN DE LA NORMALIDAD PARA EL MODELO PROMEDIO CON MEJORES PREDICTORES

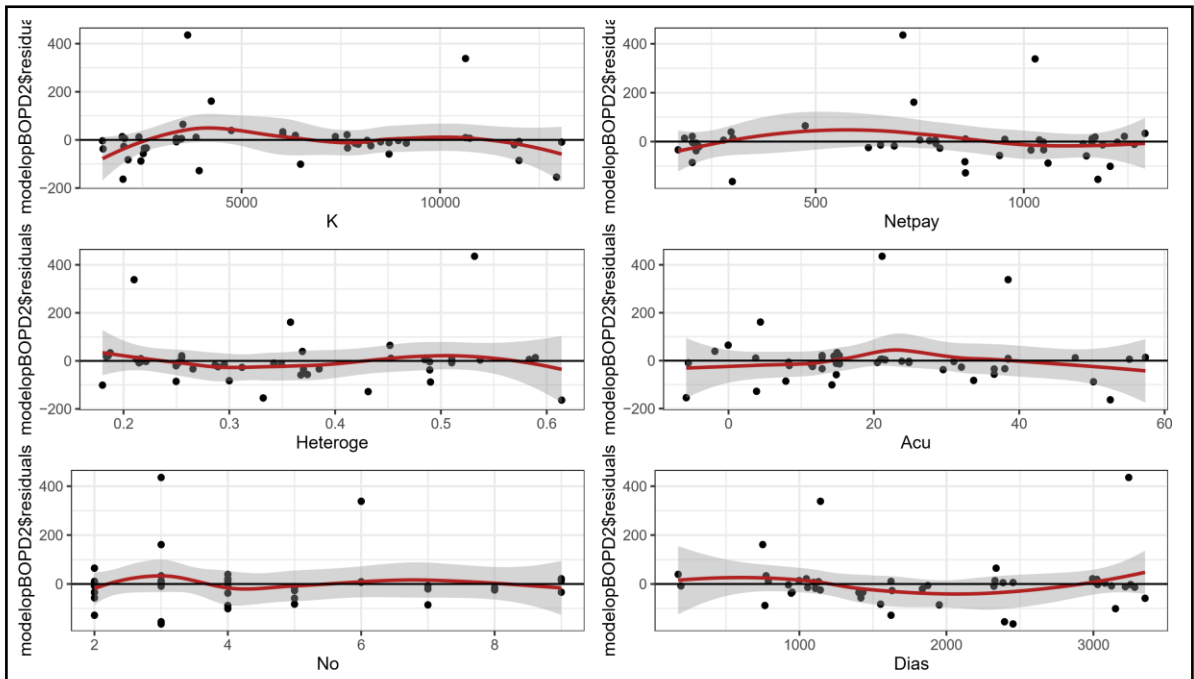
Para obtener un modelo sólido con buenos resultados, se requiere verificar la normalidad de este mediante diagramas de dispersión entre el parámetro de producción evaluado y los parámetros petrofísicos que lo describen o con diagramas de dispersión entre los parámetros petrofísicos y los residuos del modelo (se realizaron el segundo tipo de gráficos) donde los residuos se deben distribuir de forma homogénea en torno al cero, y los datos desfazados pueden ser considerados datos atípicos que se evaluarán posteriormente. Los diagramas de dispersión entre los parámetros petrofísicos y los residuos del modelo promedio con mejores predictores por cada uno de los parámetros de producción, se pueden apreciar de la figura 13 a la figura 21.

**Figura 13.** Diagrama de dispersión entre los parámetros petrofísicos y los residuos del EUR



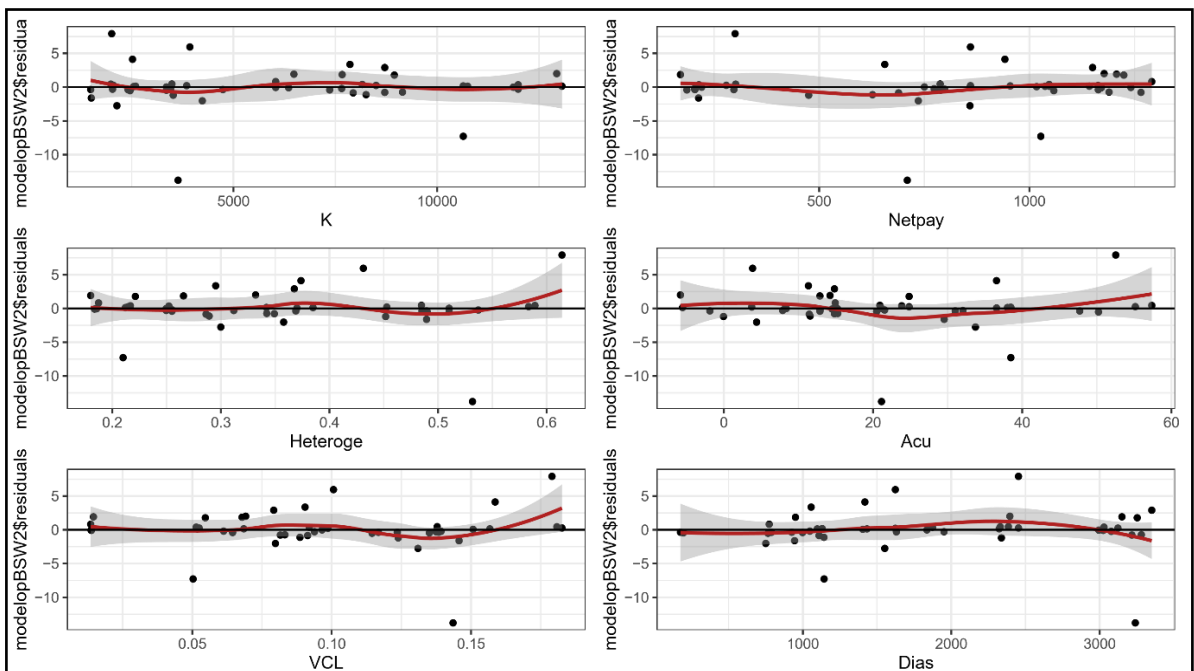
Fuente: elaboración propia

**Figura 14.** Diagrama de dispersión entre los parámetros petrofísicos y los residuos del BOPD



Fuente: elaboración propia

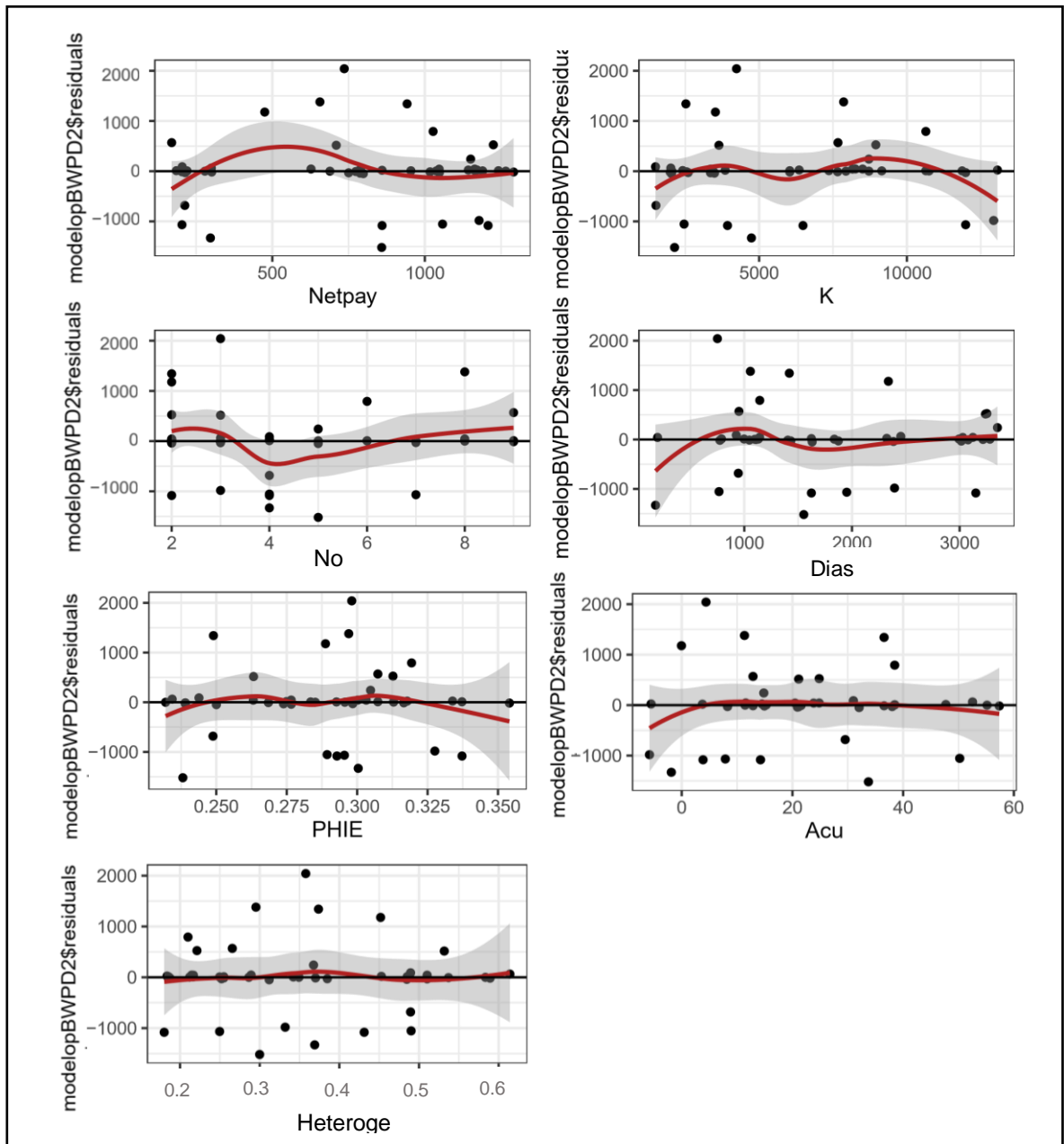
**Figura 15.** Diagrama de dispersión entre los parámetros petrofísicos y los residuos del BSW



Fuente: elaboración propia

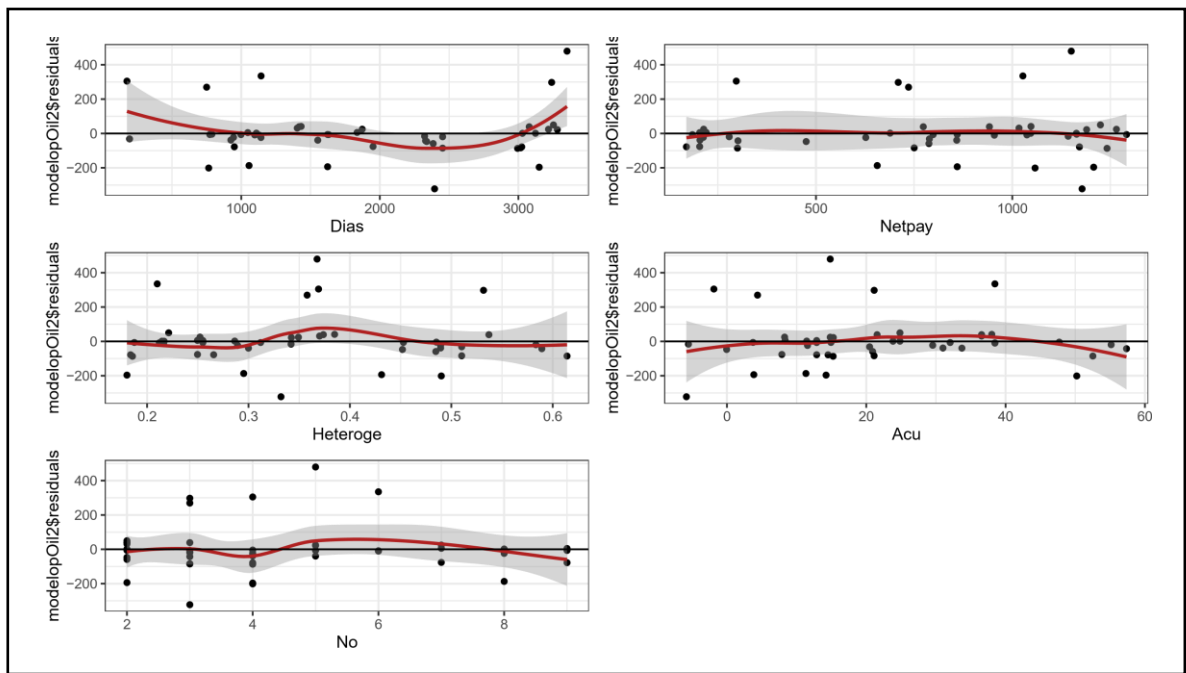


**Figura 16.** Diagrama de dispersión entre los parámetros petrofísicos y los residuos del BWPD



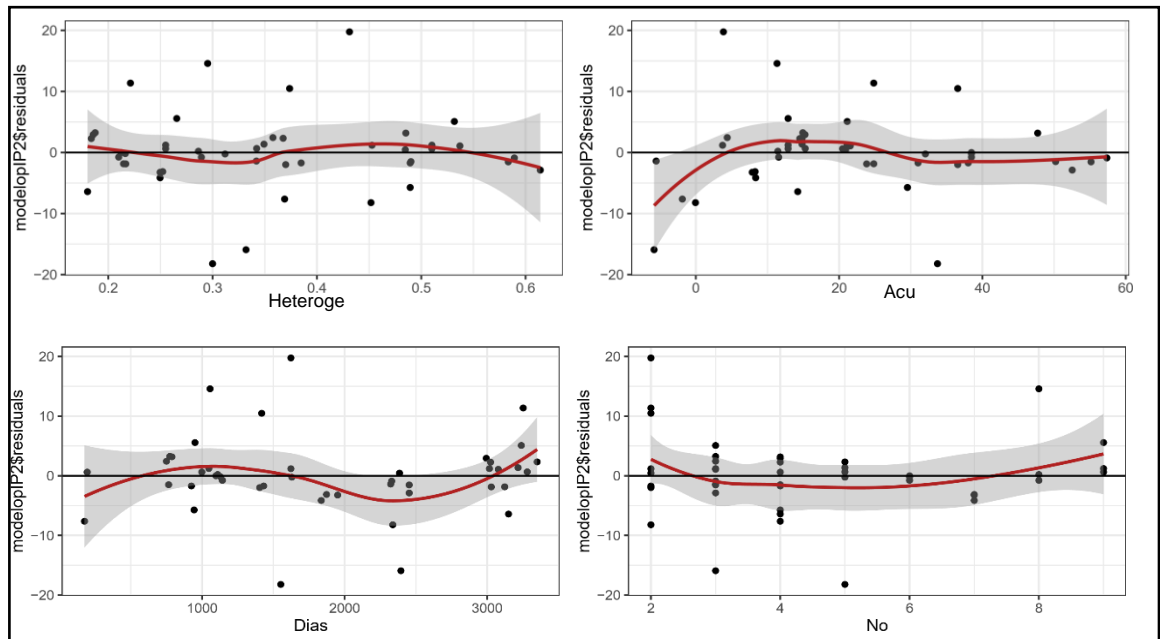
**Fuente:** elaboración propia

**Figura. 17** Diagrama de dispersión entre los parámetros petrofísicos y los residuos del Oil



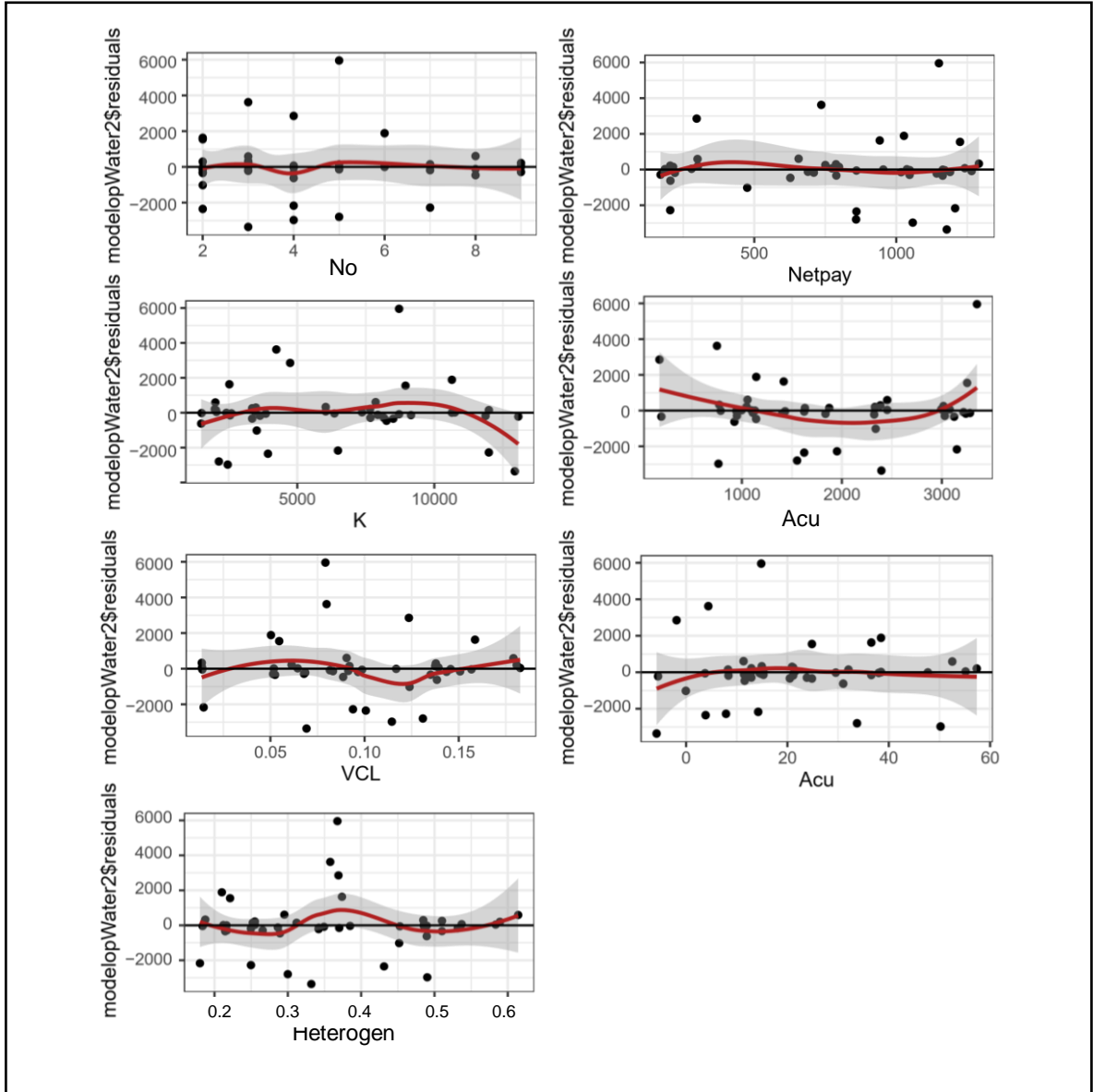
Fuente: elaboración propia

**Figura 18.** Diagrama de dispersión entre los parámetros petrofísicos y los residuos del IP



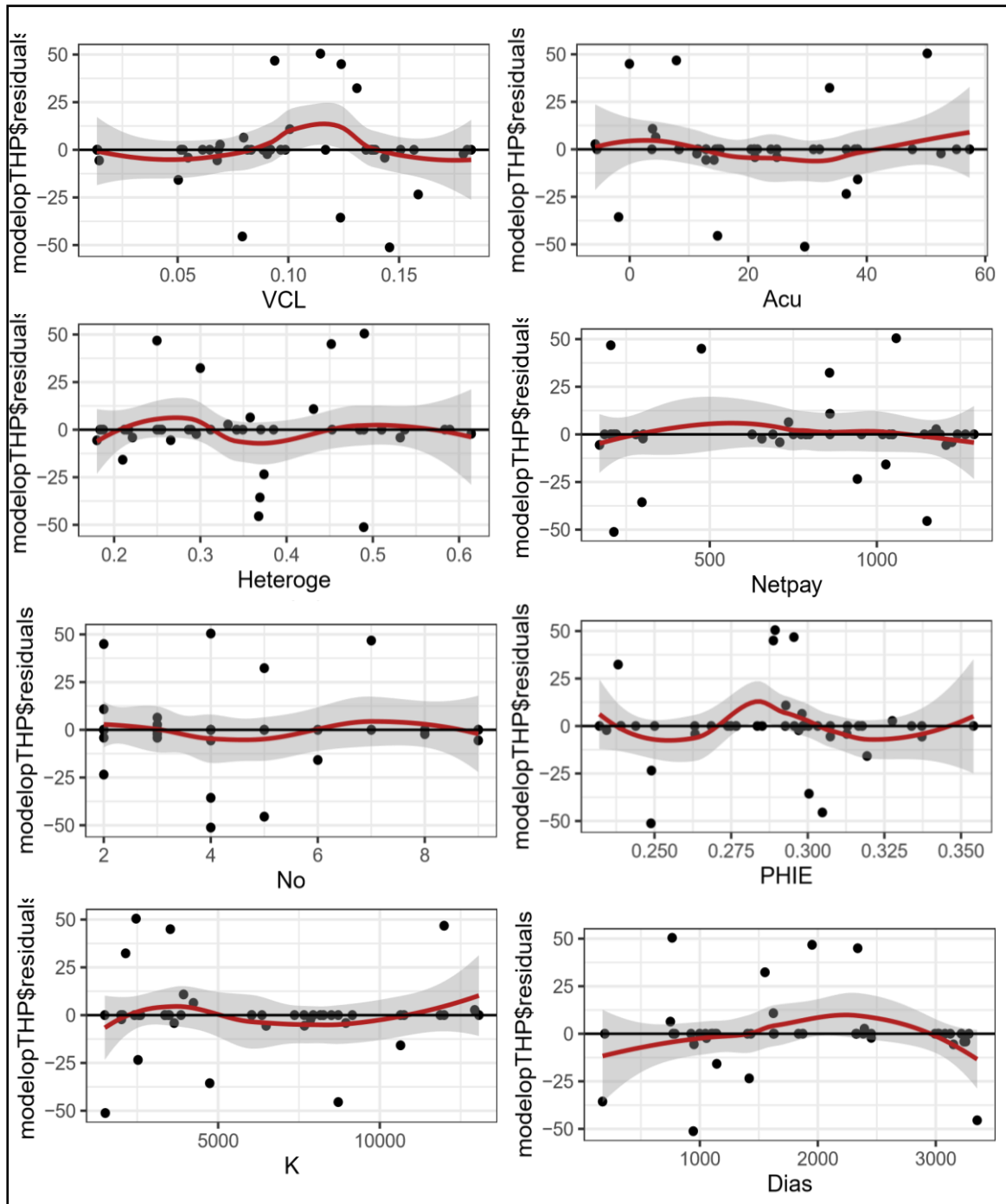
Fuente: elaboración propia

**Figura 19.** Diagrama de dispersión entre los parámetros petrofísicos y los residuos del Water



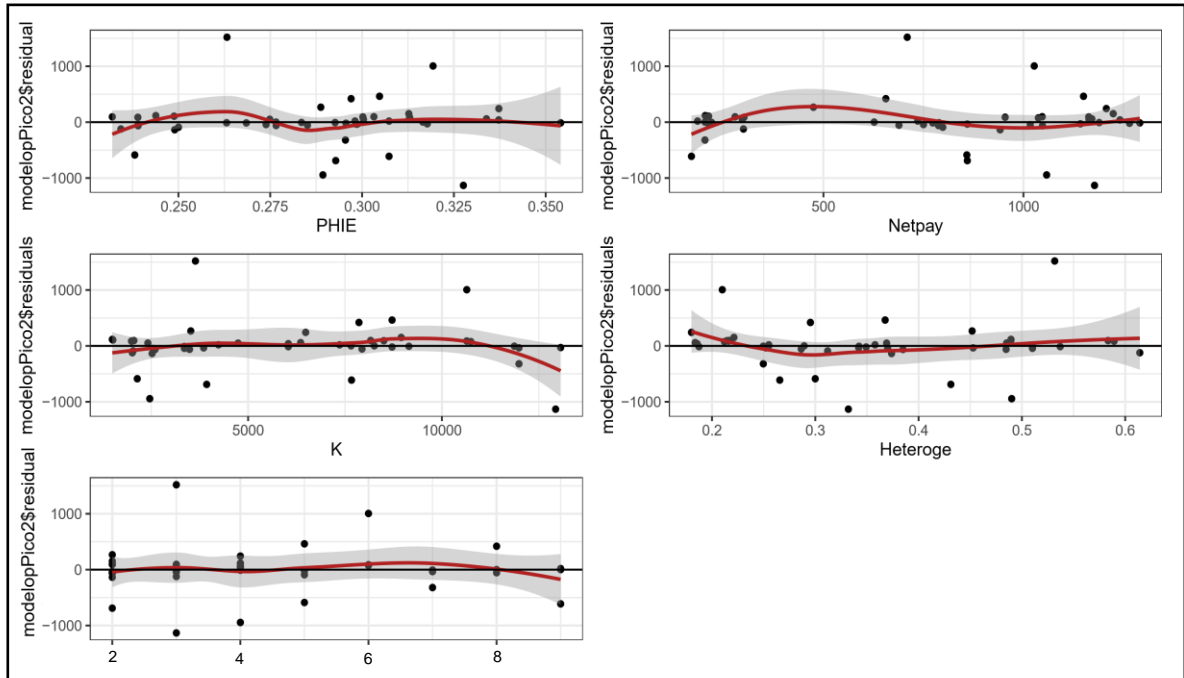
**Fuente:** elaboración propia

**Figura 20.** Diagrama de dispersión entre los parámetros petrofísicos y los residuos del THP



Fuente: elaboración propia

**Figura 21.** Diagrama de dispersión entre los parámetros petrofísicos y los residuos del Pico

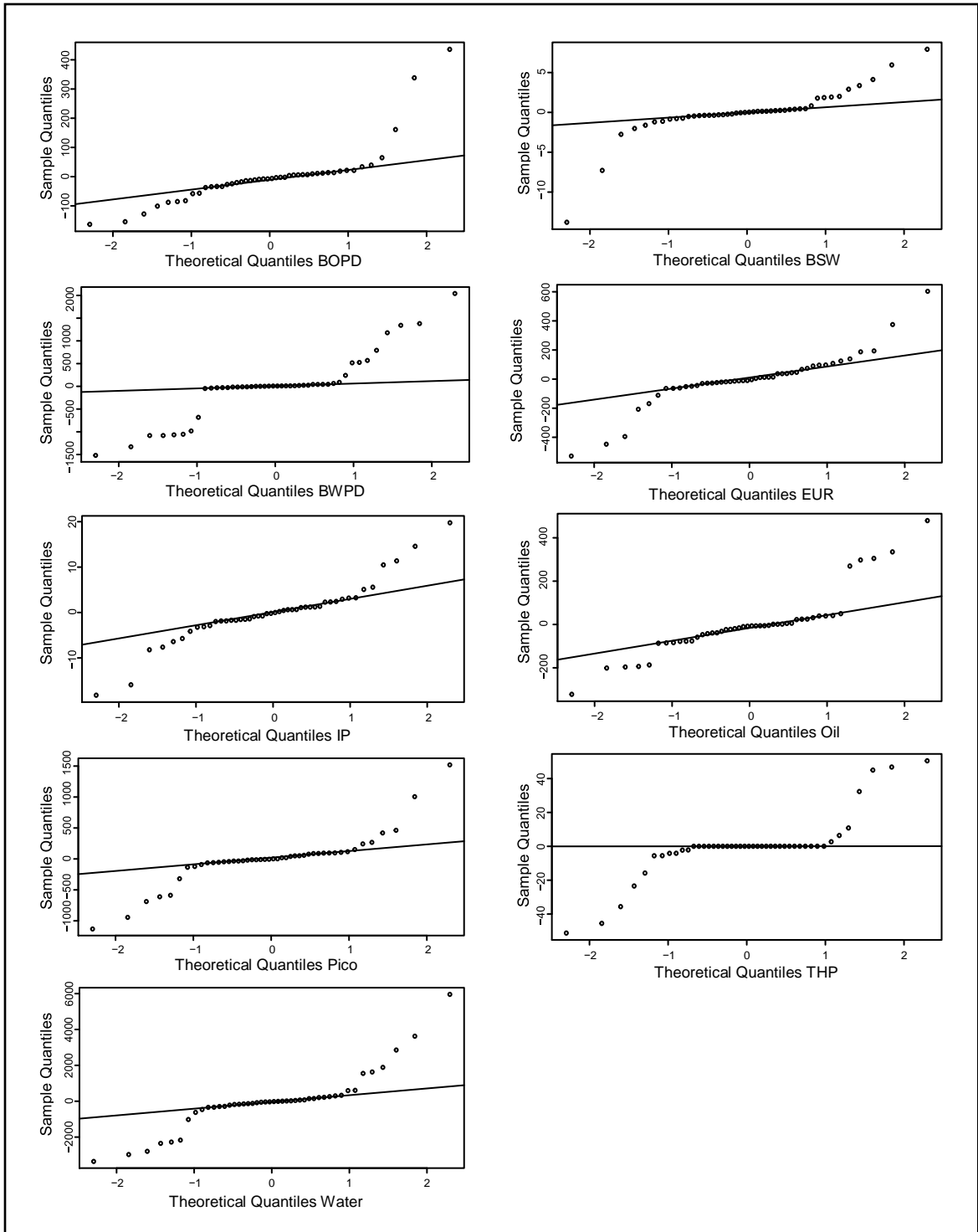


**Fuente:** elaboración propia

Al observar de la figura 13 a la figura 21 se evidencia que los parámetros petrofísicos de las variables EUR, BOPD, BSW, Oil y Pico, presentan un comportamiento normal (la línea roja tiende sobreponerse a la horizontal o línea negra) debió a la distribución homogénea de los residuos. Para las variables BWPD, Water, IP y THP se observa que los residuos de sus variables independientes presentan una dispersión que se acumula en uno de los extremos de forma transitoria, razón por la cual es un indicio de una posible falta de normalidad en sus datos.

Debido a los indicios de falta de normalidad de algunos parámetros se procedió a verificar el supuesto por medio de la gráfica de cuantiles normales contra los cuantiles teóricos, esperando que los puntos se sitúen sobre la recta que tiene como pendiente uno (la cual representa el comportamiento ideal de los cuantiles), estas gráficas las podemos apreciar en la figura 22.

**Figura 22.** Cuantiles teóricos contra cuantiles normales para los parámetros de producción



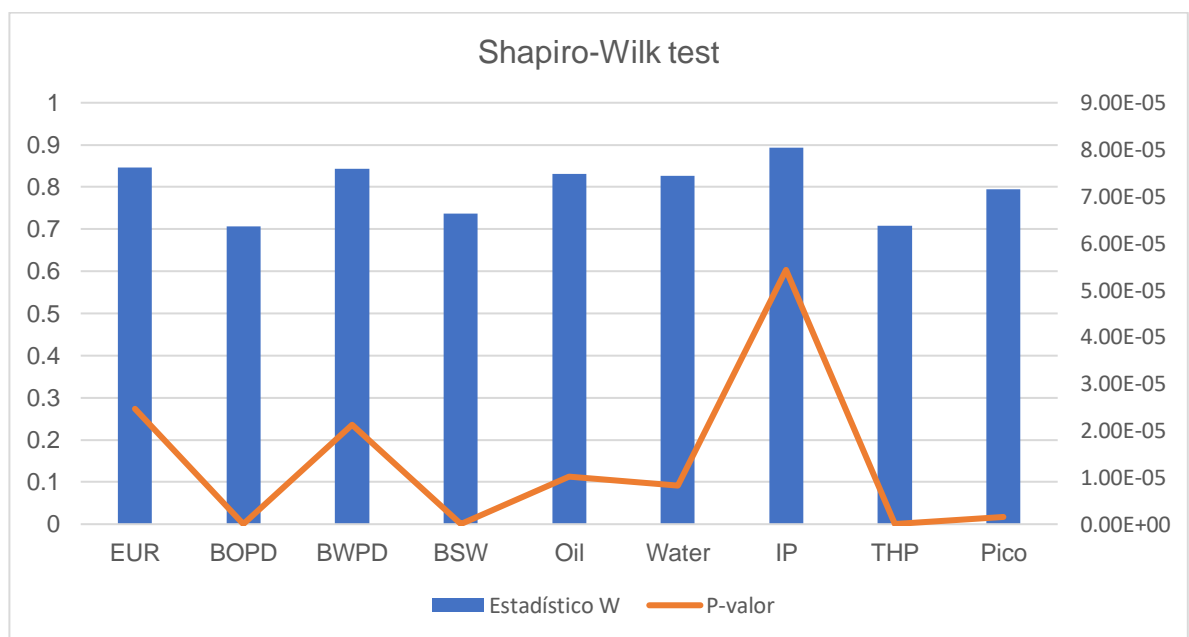
Fuente: elaboración propia

Por medio de los resultados de la figura 22 podemos afirmar que la mayoría de los cuantiles normales presenta un comportamiento muy similar a los cuantiles teóricos lo cual es argumento para poder afirmar que los datos presentan una distribución normal, sin embargo, tanto en la cola inferior como en la cola superior de todos los parámetros se distinguen una cierta cantidad de observaciones que no se adhieren al comportamiento ideal, razón por la cual se decidió realizar la verificación de la normalidad por medio de un test.

Tras haber realizado tanto los diagramas de dispersión de los residuos y los gráficos de cuantiles teóricos contra cuantiles normales para realizar la validación de la normalidad de los datos, se aplica el test de Shapiro-Wilk el cual también tiene la finalidad de contrastar la normalidad de un conjunto de datos por medio del planteamiento de una hipótesis nula (los datos provienen de una población normalmente distribuida) y una hipótesis alternativa (los datos no provienen de una población normalmente distribuida).

El test se basa en el cálculo del estadístico W para corroborar cuál de las hipótesis es aceptada por medio de la comparación con un p-valor (valor crítico) calculado, el valor del estadístico puede oscilar entre cero (0) y uno (1). Con un valor del estadístico W mayor al p-valor, se puede confirmar la hipótesis nula de que los datos provienen de una población normalmente distribuida. Bajo este conocimiento el test de Shapiro-Wilk fue aplicado a cada uno de los parámetros de producción del modelo promedio con los mejores predictores, los resultados los podemos visualizar en la gráfica 19 donde la escala de izquierda es para el estadístico W (calculado a partir de la ecuación 12) y la escala de la derecha es para el p-valor.

**Gráfica 19.** Shapiro-Wilk test para los parámetros de producción



Fuente: elaboración propia

Finalmente, tras haber realizado el test a cada uno de los parámetros de producción del modelo, se determina que por medio de la comparación del estadístico  $W$  y el  $p$ -valor todos los datos del modelo presentan un comportamiento normal, verificando que la distribución de los residuos en torno al cero (figura 13 a figura 21) es homogénea y que los cuantiles normales (gráfica 19) presenta un comportamiento similar a los cuantiles teóricos estadísticamente hablando.

### **3.5 VALIDACIÓN DE LA HOMOCEDASTICIDAD PARA EL MODELO PROMEDIO CON MEJORES PREDICTORES**

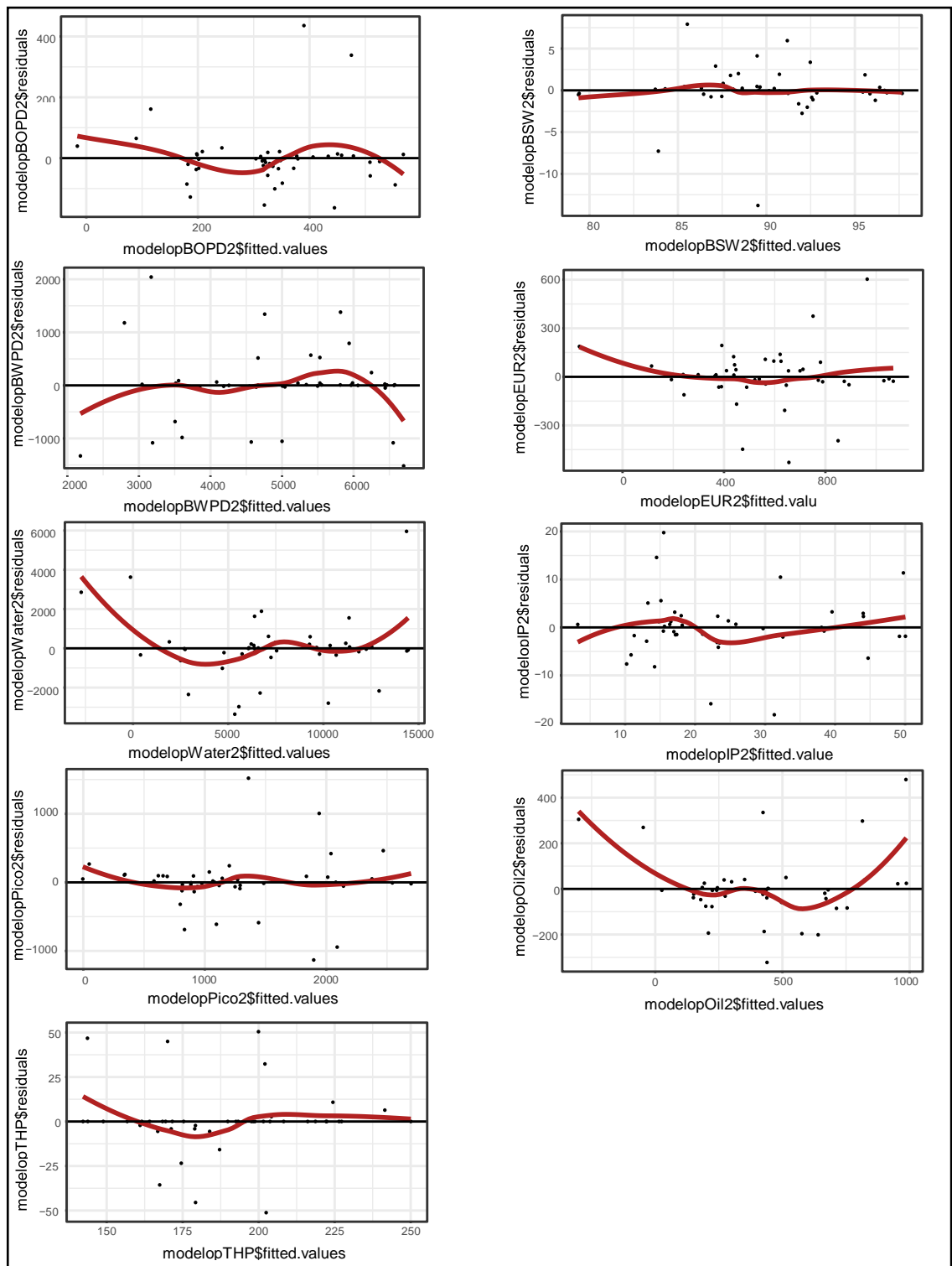
La homocedasticidad es una propiedad de los modelos la cual indica que la varianza del error condicional de las variables independientes tiene poca variación a lo largo de todas las observaciones de la muestra. Uno de los métodos para comprobar si el modelo cuenta con esta propiedad se realiza a través de los residuos del modelo, ya que si estos presentan poca variación o son constantes, se distribuirán de forma aleatoria manteniendo una misma dispersión y sin ningún patrón específico.

Uno de los casos más comunes de la falta de homocedasticidad, es cuando se presenta una distribución cónica en los residuos y esto puede ser producto de que las variables predictoras tiene un gran recorrido con respecto a su propia media, cambio de estructura, uso de variables no relativizadas, entre otras causa. Esta falta de homocedasticidad afecta los resultados del modelo, por lo tanto sería un aspecto para corregir.

Con el fin de verificar la homocedasticidad para el modelo promedio con mejores predictores se procedió a graficar los residuos del modelo frente a los valores ajustados por el modelo, esperando así que los valores se distribuyan de forma aleatoria en torno al cero (0), mantenido la misma variabilidad a lo largo de la horizontal, estos resultados se pueden apreciar en la figura 23.



**Figura 23.** Residuos frente a los valores ajustados del modelo para los parámetros de producción



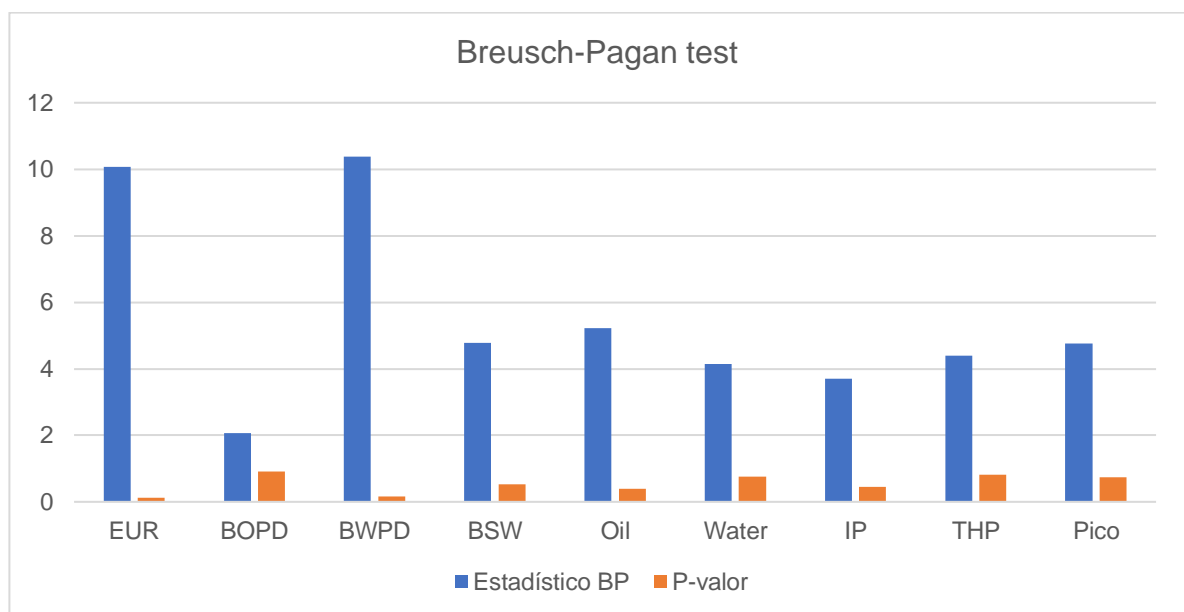
Fuente: elaboración propia

Por medio de los resultados presentes en la figura 23 se evidencia una distribución aleatoria y homogénea entorno al cero (0) para los parámetros petrofísicos de BSW, EUR y Pico debido a la superposición continua de la línea roja (tendencia de los residuos frente a los valores ajustados) sobre la horizontal, confirmando la homocedasticidad en estas variables. Sin embargo para el BOPD, BWPD, Water, IP, Oil y THP la tendencia de los valores tiene constantes variaciones, lo cual es un indicio de la falta de homocedasticidad, razón por la cual se procede a realizar un test para evaluar de forma cualitativa los resultados.

Tras haber realizado los gráficos de los residuos del modelo frente a los valores ajustados por el modelo para realizar la validación de la homocedasticidad, se aplica el test de Breusch-Pagan el cual se utiliza para determinar la homocedasticidad en un modelo de regresión lineal de una sola variable o multivariable por medio de la aplicación de tres (3) pasos. Se empieza por aplicar mínimos cuadrados ordinarios sobre el modelo y calcular los residuos de la regresión, posterior a ello se realiza una regresión auxiliar por medio de la ecuación 13 y finalmente se calcula el estadístico de prueba que es el resultado del coeficiente de determinación de la regresión auxiliar y el tamaño de la muestra.

Tras realizar esta serie de pasos obtenemos el estadístico BP el cual se compara un p-valor (valor crítico), donde las condiciones son, si el estadístico es mayor al p-valor se confirma la homocedasticidad en el modelo, de lo contrario se confirma la ausencia de la homocedasticidad y se procede a solventar este problema, ya que esto influye de forma negativa en los resultados. Los resultados del estadístico BP y los p-valores se puede apreciar en la gráfica 20.

**Gráfica 20.** Breusch-Pagan test para los parámetros de producción



Fuente: elaboración propia

Al comparar los valores del estadístico B-P con sus p-valores correspondientes (gráfica 20) se determina que para todos los parámetros de producción el estadístico es mayor que el valor crítico, razón por la cual se establece que para el modelo promedio con mejores predictores, todas las regresiones lineales múltiples cuentan con la homocedasticidad de sus datos, debido a que la varianza del error condicional tiene poca o nula variación.

### **3.6 VALIDACIÓN DE LA AUTOCORRELACIÓN PARA EL MODELO PROMEDIO CON MEJORES PREDICTORES**

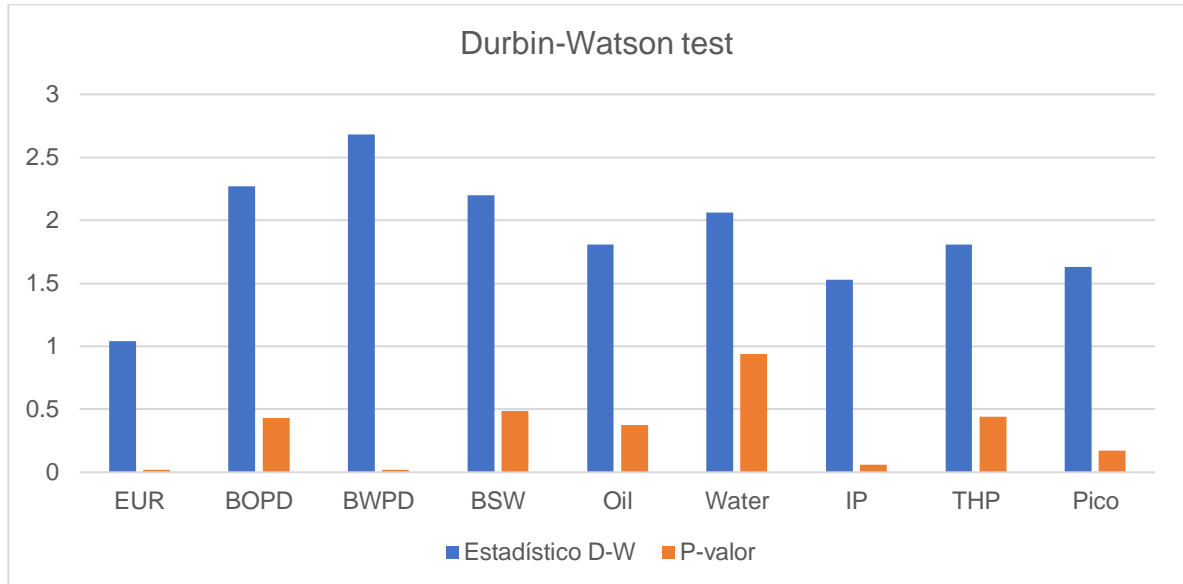
La validación de la autocorrelación es una parte fundamental del proyecto, debido a que tiene especial impacto sobre variables temporales (algunos parámetros de producción), lo que se busca encontrar es que las observaciones son independientes unas de otras es decir, el dato “ $n$ ” no afecta el valor del dato “ $n+1$ ” ni el “ $n-1$ ”.

Para el caso de que alguna de las variables predictoras se encontrase ligada al tiempo, se recomienda representar los residuos ordenados acorde al tiempo de registro de las observaciones para poder evaluar si existe un patrón de indicios de autocorrelación. Debido a que en el proyecto las observaciones son los pozos de estudio, los parámetros petrofísicos no tienen ninguna relación, únicamente son dependiente del pozo.

Otro método empleado para validar la autocorrelación de los datos es por medio del test de Durbin-Watson donde se plantean dos (2) hipótesis, la hipótesis nula donde los errores no están correlacionados temporalmente en una serie y la hipótesis alternativa donde se afirma que siguen un proceso de primer orden autorregresivo.

El procedimiento para emplear el test es el cálculo del estadístico D-W a partir de la ecuación 14 y un p-valor (valor crítico), luego se comparan ambos valores, aceptándose la hipótesis nula en el caso de que el estadístico D-W sea mayor al p-valor y la hipótesis alternativa en el caso contrario. Estos valores fueron calculados por el programa R y se visualizan en la gráfica 21.

**Gráfica 21.** Durbin-Watson test para los parámetros de producción



**Fuente:** elaboración propia

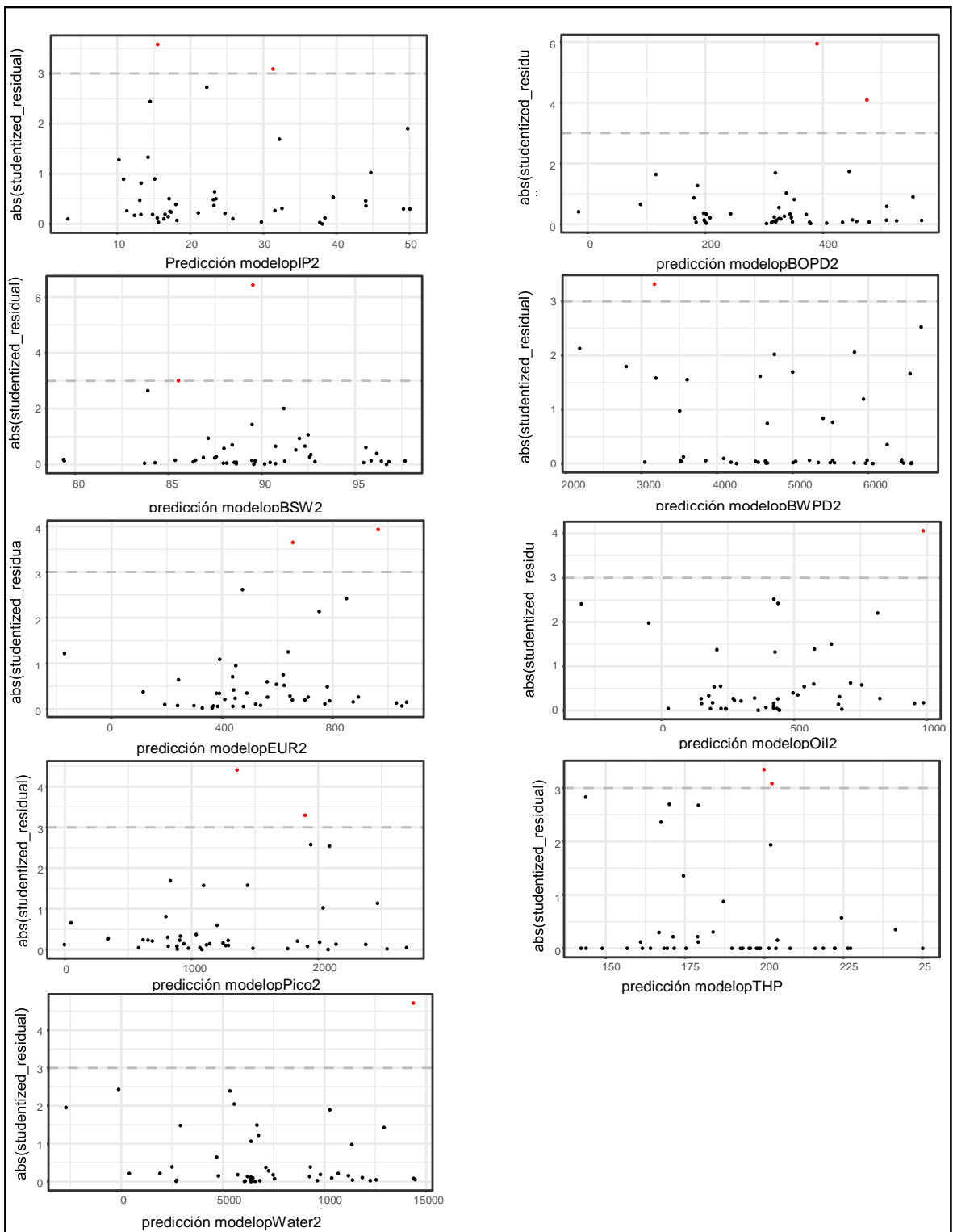
Después de la visualización de los resultados del test (gráfica 21), se puede apreciar que para todos los parámetros de producción del modelo promedio con mejores predictores, los valores del estadístico D-W son mayores al p-valor, razón por la cual se acepta la hipótesis nula de que los errores no están correlacionados temporalmente en una serie y se afirma que no hay evidencia de autocorrelación en el modelo.

### **3.7 IDENTIFICACIÓN DE VALORES ATÍPICOS PARA EL MODELO PROMEDIO CON MEJORES PREDICTORES**

Es importante la identificación de valores atípicos dentro del modelo, debido a que pueden generar una falsa correlación entre variables u ocultar una correlación existente, estos valores pueden ser identificados por varios efectos, uno de ellos es el no ajustarse bien al modelo debido a que su valor real se aleja mucho del predicho, por lo que su residuo es bastante grande.

Teniendo en cuenta lo anterior, se procedió graficar la distribución de los residuos contra los valores predichos por el modelo promedio con mejores predictores, donde se estableció para el eje y un valor de tres (3) como máximo de tolerancia e identificación de valores atípicos. Estos gráficos se visualizan en la figura 24.

**Figura 24.** Distribución de residuos frente a los valores predichos por el modelo para los parámetros de producción



Fuente: elaboración propia

Tras analizar los resultados de los valores atípicos (figura 24) se determina que para cada parámetro de producción se presenta entre uno (1) y dos (2) valores por fuera del rango, teniendo en cuenta que la muestra de estudio es de cincuenta (50) pozos, los valores atípicos no representan más del cuatro (4) por ciento de la población. Es importante evaluar estos valores de forma individual y entender porque la predicción difiere tanto del pronóstico en estos pozos, sin embargo al no ser un numero alto con respecto al total, no es necesario extraerlos del modelo.

Si se evalúa puntualmente los resultados se puede deducir que la mayoría de estos valores atípicos se presentan en escenarios altos, como por ejemplo el parámetro Water, Oil y BOPD, lo que indicaría que el modelo tiene un mejor ajuste en escenarios medios y bajos.

### **3.8 IDENTIFICACIÓN DE OBSERVACIONES INFLUYENTES PARA EL MODELO PROMEDIO CON MEJORES PREDICTORES**

Las observaciones que son influyentes sobre el modelo resultante son aquellas que condicionan en mayor medida los resultados que otros datos, debido a esto se deben identificar y dependiendo la finalidad del proyecto, se evalúa su posible extracción. Por medio de un código en R se identificaron las observaciones significativamente influyentes sobre el modelo promedio con mejores predictores.

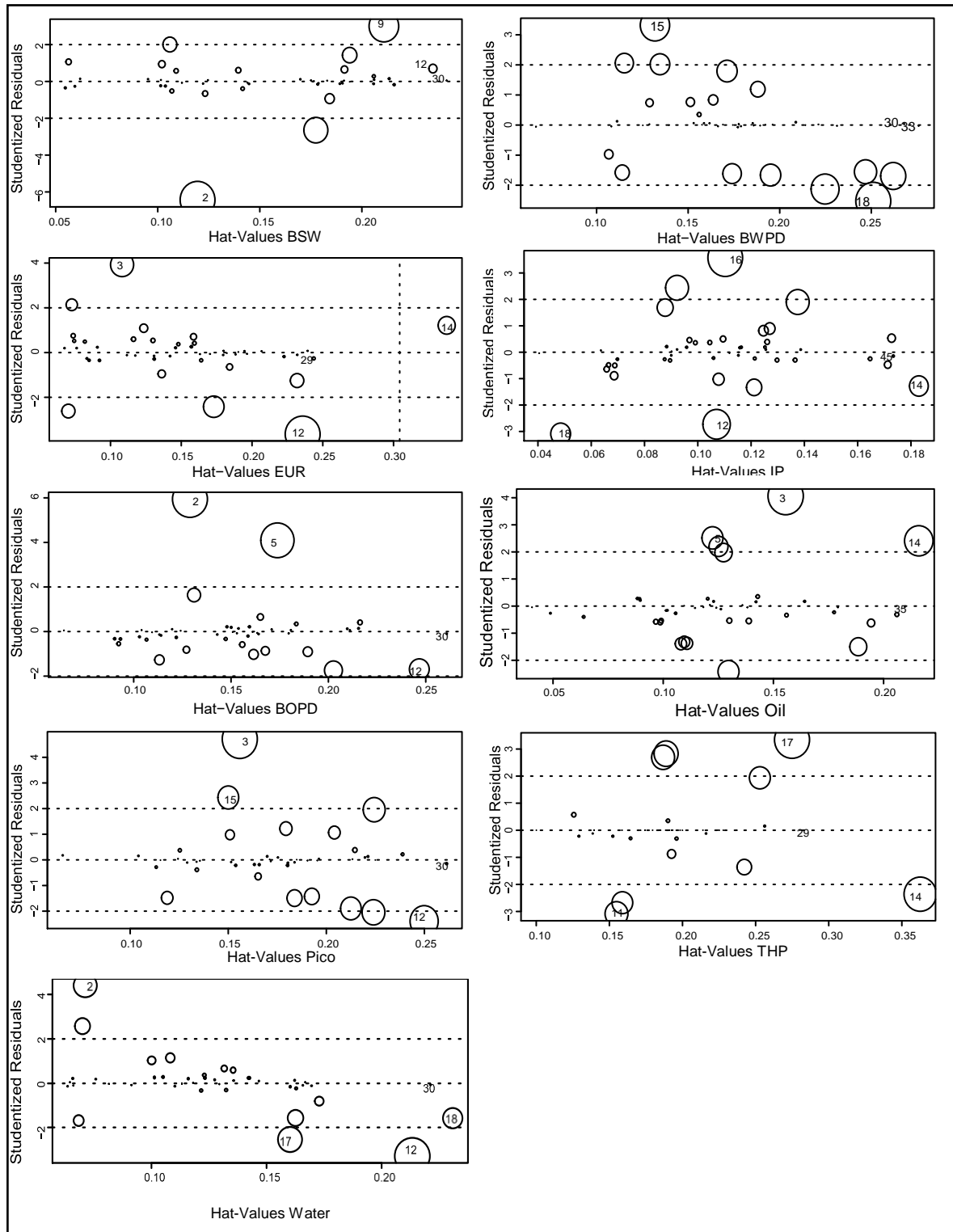
Las condiciones que debe cumplir una observación para ser influyentes son las siguientes.

Si el valor de "*Leverages (hat)*" (ecuación 15) son superiores " $2.5((p+1)/n)$ " siendo p el número de predictores y n el número de observaciones.

Si el valor de "*Distancia Cook*" (ecuación 16) es superior a uno (1).

Con el fin de evidenciar los resultados de una forma gráfica se empleó un diagrama de figuras entre los residuos del modelo y los valores de "*Leverages (hat)*", entre más cerca se encuentren las observaciones a la horizontal (eje x), su influencia y el tamaño del circulo es menor y de forma análoga entre más se aleje la observación de la horizontal, su influencia será mayor. Estos diagramas se pueden visualizar en la figura 25.

**Figura 25.** Residuos frente a los valores “Leverages (*hat*)” del modelo para los parámetros de producción



Fuente: elaboración propia

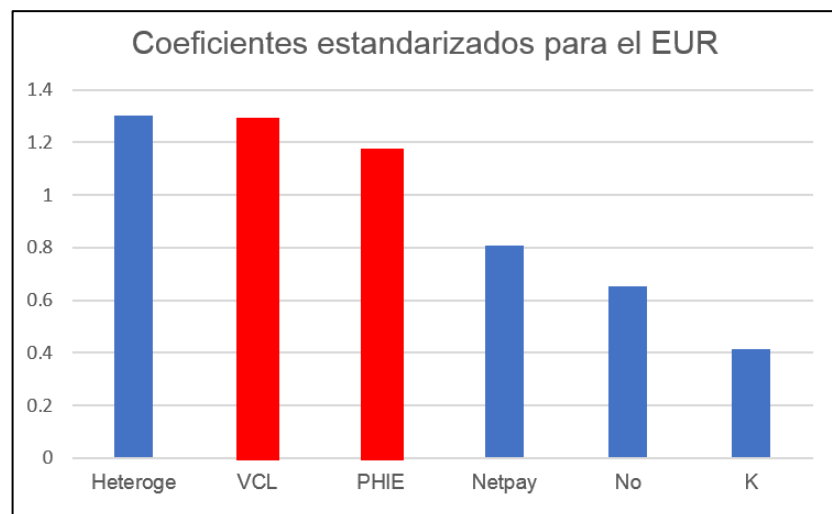
Al haber ingresado las observaciones del modelo bajo la misma numeración de los pozos, se puede identificar por medio de la figura 25 cuáles de ellos son los más influyentes para cada parámetro de producción. Estas observaciones se pueden sustraer para volver a correr el modelo y contrastar cuales son los resultados, sin embargo esta operación no fue realizada debido a la gran extensión que conllevaría generar un nuevo modelo.

### 3.9 BETAS ESTANDARIZADOS PARA EL MODELO PROMEDIO CON MEJORES PREDICTORES

Los modelos de regresión lineal múltiple pueden cuantificar el impacto de forma directa que tiene una variable sobre otra, es decir por la variación de un parámetro independiente se puede cuantificar el cambio en el dependiente por medio de los betas que obtiene el modelo, estos últimos guardan las unidades de la variable dependiente en el numerador y las unidades de la independiente en el denominador. Por esta incompatibilidad de unidades entre los betas de una regresión múltiple, estos no son comparables entre sí, estos deben ser estandarizados para poder cuantificar la magnitud de la influencia de cada variable independiente.

Bajo la intención de determinar en el proyecto cuales son los parámetro petrofísicos con mayor influencia sobre los parámetros productivos ,se procedió a calcular los betas estandarizados para cada uno de los parámetro petrofísicos del modelo promedio con mejores resultados por medio de la ecuación 17. Los valores de los betas estandarizados se pueden observar entre la gráfica 22 a la gráfica 30 donde las barras de color azul indican un relación directamente proporcional entre variables y el color rojo una relación inversamente proporcional.

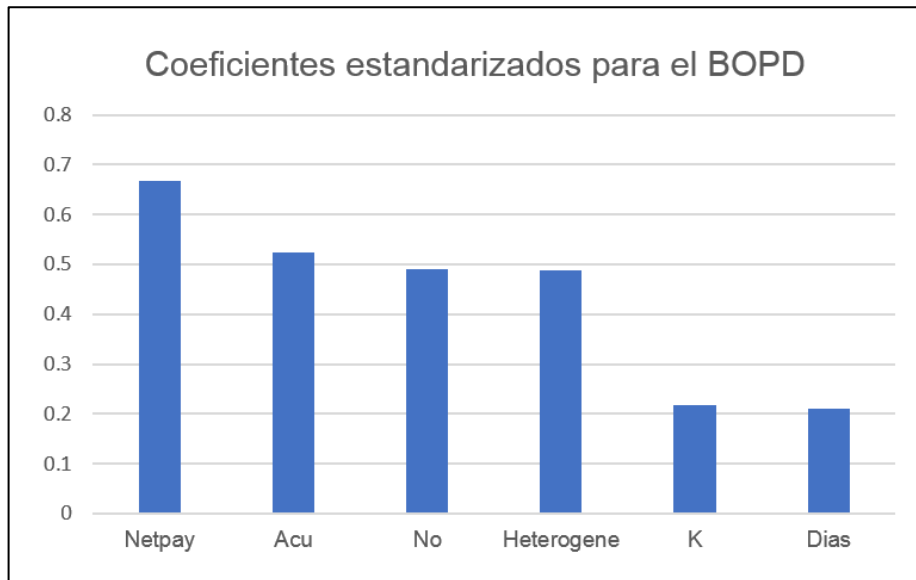
**Gráfica 22.** Betas estandarizados para EUR



**Fuente:** elaboración propia

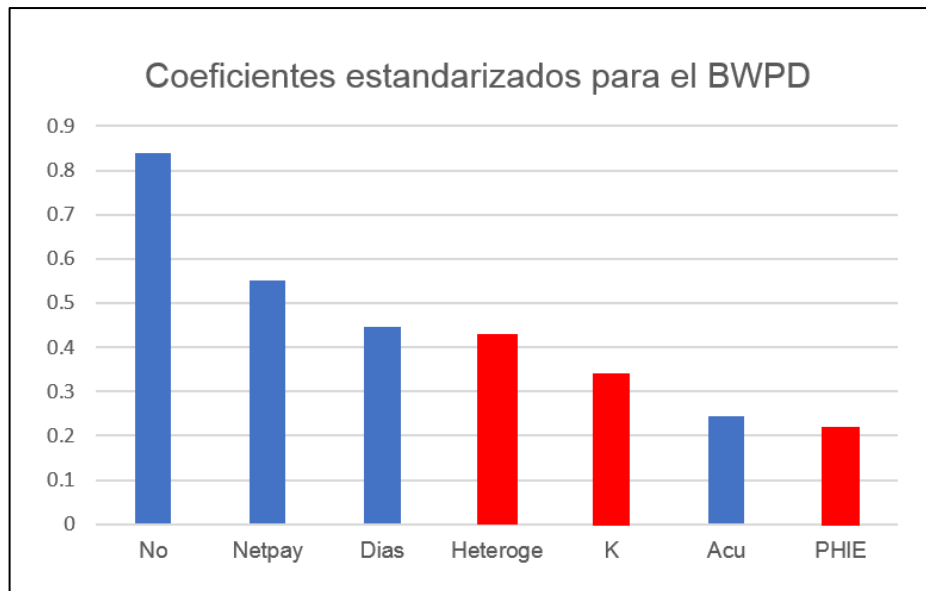


**Gráfica 23.** Betas estandarizados para BOPD



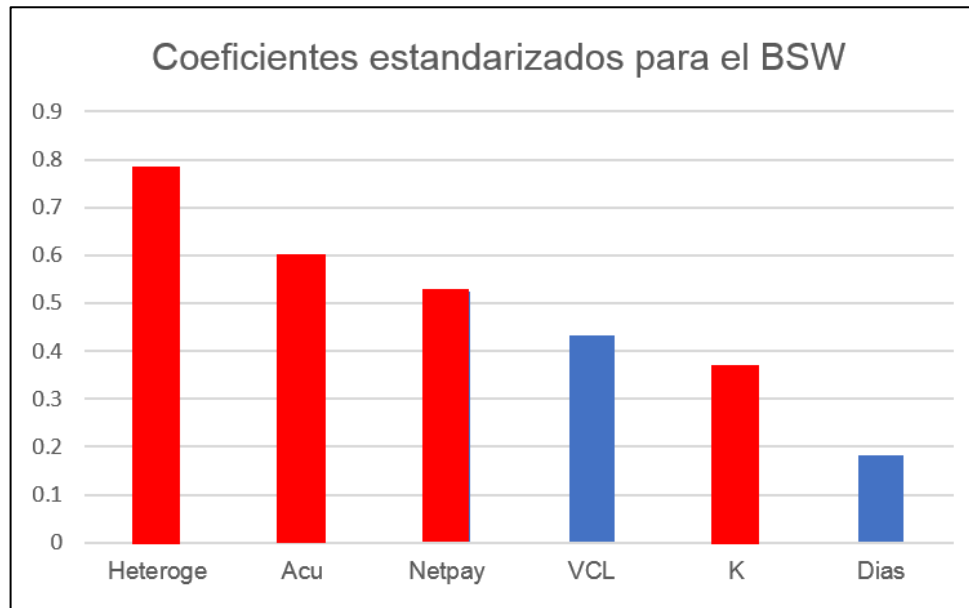
**Fuente:** elaboración propia

**Gráfica 24.** Betas estandarizados para BWPD



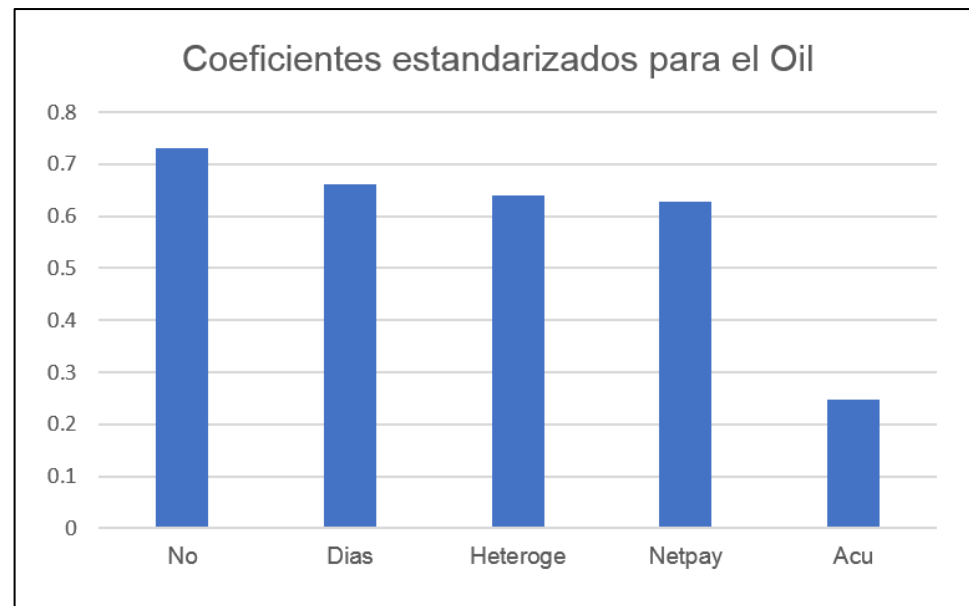
**Fuente:** elaboración propia

**Gráfica 25.** Betas estandarizados para BSW



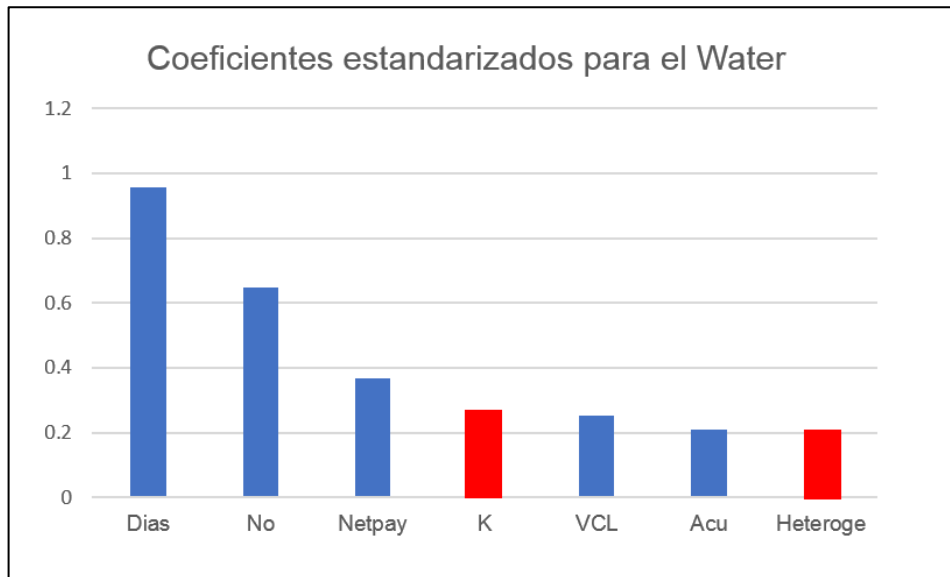
Fuente: elaboración propia

**Gráfica 26.** Betas estandarizados para Oil



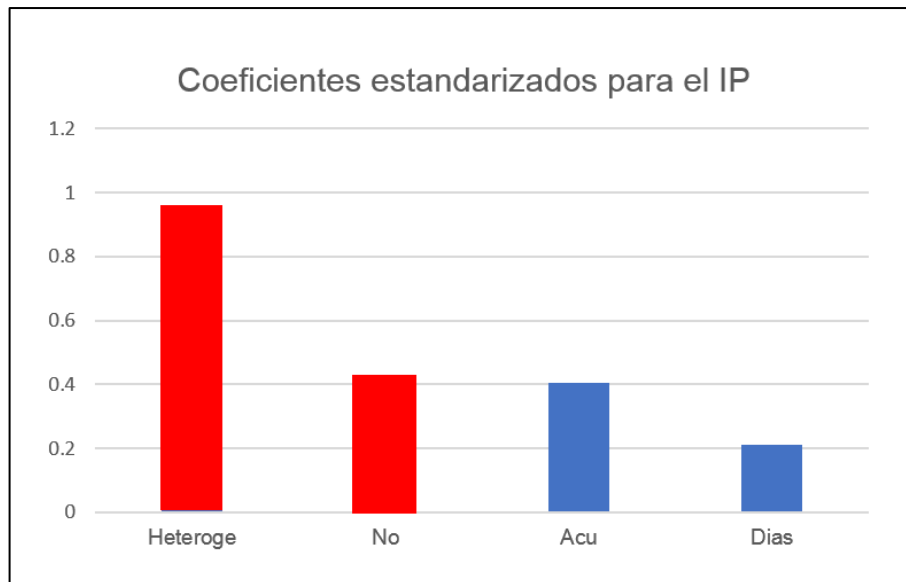
Fuente: elaboración propia

**Gráfica 27.** Betas estandarizados para Water



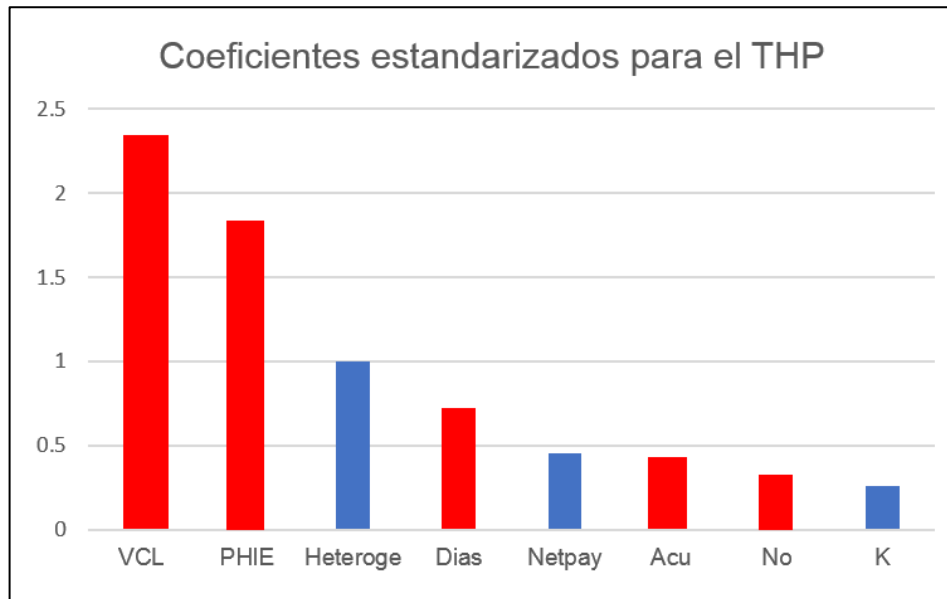
Fuente: elaboración propia

**Gráfica 28.** Betas estandarizados para IP



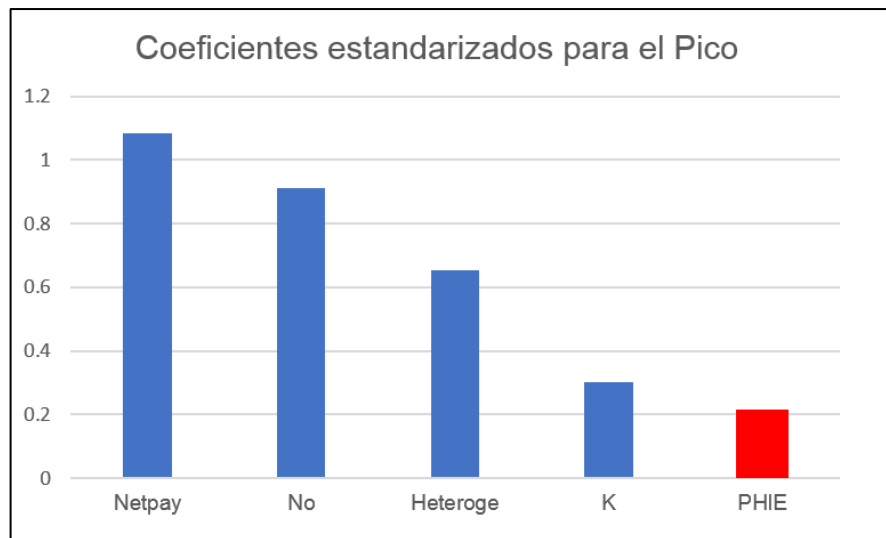
Fuente: elaboración propia

**Gráfica 29.** Betas estandarizados para THP



Fuente: elaboración propia

**Gráfica 30.** Betas estandarizados para Pico



Fuente: elaboración propia

Al observar de la gráfica 22 a la gráfica 30 se puede determinar cuáles son los parámetros petrofísicos con mayor incidencia sobre los parámetros de producción en el modelo promedio de mejores predictores, haciendo la comparación entre los betas estandarizados. Se evidencia que los parámetros de heterogeneidad, espesor de explotación y número de pozos por cluster son los más significativos para predecir los parámetros de producción, debido alguno de estos es el parámetro más importante en siete (7) de los nueve (9) casos.

Es importante resaltar que para los parámetros productivos de producción acumulada de petróleo y barriles de petróleo producidos por día, todos los parámetros petrofísicos de los modelos son directamente proporcionales, es decir que un aumento en cualquiera de los predictores favorece a una mayor producción de petróleo.

Uno de los mayores inconvenientes dentro del campo Imperial es la producción de agua en los pozos, debido a que el sistema de levantamiento primario del yacimiento es un acuífero activo, por medio de estos resultados se podría afirmar que las variables que más influyen sobre la producción de agua son los días de producción y el número de pozos por cluster, lo que quiere decir que entre más tiempo produzca un pozo y mayor número de pozos por cluster haya, la producción de agua también aumenta y finalmente, se establece que a mayor capacidad de tratar el agua en el campo, mayor recobro por pozo.

Para el parámetro de producción “estimado de recobro final” (gráfica 22), los parámetros de petrofísica con mayor incidencia sobre el son la heterogeneidad, el volumen de arcilla y la porosidad con una capacidad explicativa del 30.16, 29.65 y 27 por ciento respectivamente. Para este parámetro la heterogeneidad, espesor para explotación, número de pozos por cluster y permeabilidad tiene una relación directamente proporcional y el volumen de arcilla y la porosidad una relación inversamente proporcional.

Para el parámetro de producción “producción de petróleo diaria” (gráfica 23), los parámetros de petrofísica con mayor incidencia sobre el son el espesor para explotación, distancia con el acuífero, número de pozos por cluster y la heterogeneidad con una capacidad explicativa del 25.72, 20.15 y 18.78 por ciento respectivamente. Para este parámetro la heterogeneidad, espesor para explotación, la distancia con el acuífero, número de pozos por cluster, permeabilidad y días de producción tiene una relación directamente proporcional.

Para el parámetro de producción “producción de agua diaria” (gráfica 24), el parámetro de petrofísica con mayor incidencia sobre él es el número de pozos por cluster con una capacidad explicativa del 27.34 por ciento. Para este parámetro el espesor para explotación, número de pozos por cluster, días de producción y distancia con el acuífero tiene una relación directamente proporcional y la heterogeneidad, permeabilidad y la porosidad una relación inversamente proporcional.

Para el parámetro de producción “corte de agua” (gráfica 25), los parámetros de petrofísica con mayor incidencia sobre el son la heterogeneidad y la distancia con el acuífero con una capacidad explicativa del 27.2 y 22.74 por ciento respectivamente. Para este parámetro el volumen de arcilla y días de producción tiene una relación directamente proporcional y la heterogeneidad, el espesor para explotación, distancia con el acuífero y permeabilidad una relación inversamente proporcional.

Para el parámetro de producción “producción de petróleo acumulada” (gráfica 26), los parámetros de petrofísica con mayor incidencia sobre el son, el número de pozos por cluster, los días de producción, la heterogeneidad y el espesor para explotación con una capacidad explicativa del 25.18, 22.74, 21.97 y 21.61 por ciento respectivamente. Para este parámetro la heterogeneidad, espesor para explotación, la distancia con el acuífero, numero de pozos por cluster y días de producción tiene una relación directamente proporcional.

Para el parámetro de producción “producción de agua acumulada” (gráfica 27), el parámetro de petrofísica con mayor incidencia sobre el son los días de producción con una capacidad explicativa del 32.93 por ciento. Para este parámetro el espesor para explotación, volumen de arcilla, distancia con el acuífero, numero de pozos por cluster y días de producción tiene una relación directamente proporcional y la heterogeneidad y permeabilidad una relación inversamente proporcional.

Para el parámetro de producción “índice de productividad” (gráfica 28), el parámetro de petrofísica con mayor incidencia sobre él es la heterogeneidad con una capacidad explicativa del 47.98 por ciento respectivamente. Para este parámetro la distancia con el acuífero y días de producción tiene una relación directamente proporcional y numero de pozos por cluster y heterogeneidad una relación inversamente proporcional.

Para el parámetro de producción “presión en cabeza de pozo” (gráfica 29), los parámetros de petrofísica con mayor incidencia sobre el son el volumen de arcilla y la porosidad con una capacidad explicativa del 32.03 y 24.87 por ciento respectivamente. Para este parámetro el espesor para explotación, heterogeneidad y permeabilidad tiene una relación directamente proporcional y el volumen de arcilla, porosidad, distancia con el acuífero, numero de pozos por cluster y días de producción una relación inversamente proporcional.

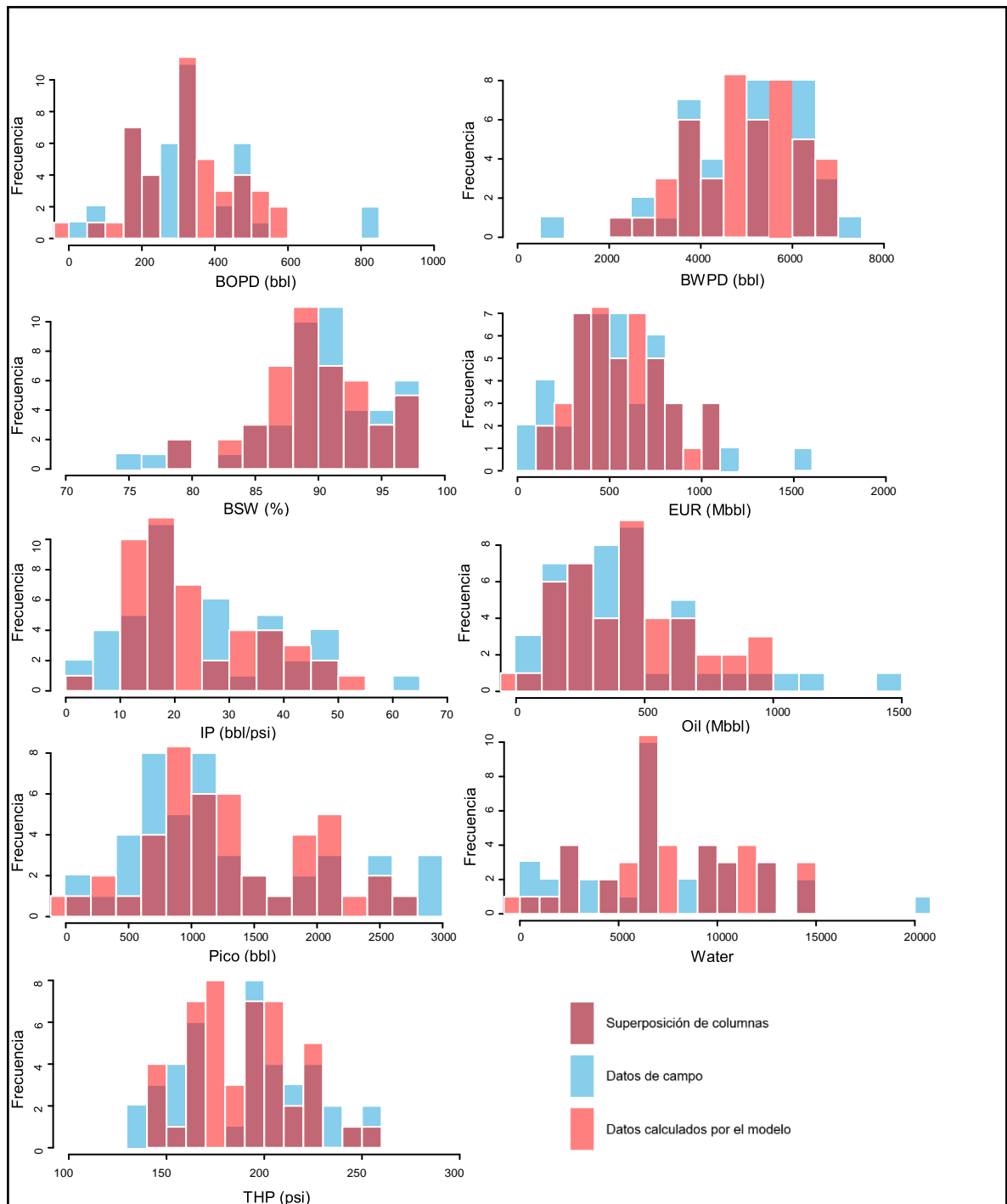
Para el parámetro de producción “pico máximo de producción” (gráfica 30), los parámetros de petrofísica con mayor incidencia sobre el son el espesor para explotación y el número de pozos por cluster con una capacidad explicativa del 34.33 y 28.94 por ciento respectivamente. Para este parámetro el espesor para explotación, heterogeneidad, numero de pozos por cluster y permeabilidad tiene una relación directamente proporcional y la porosidad una relación inversamente proporcional.

### **3.10 RESULTADOS DEL MODELO PROMEDIO CON MEJORES PREDICTORES**

Después de recopilar la información, crear los modelos, seleccionar el mejor de estos, obtener los mejores predictores y evaluar los posibles problemas sobre las bases de datos y sus resultados, se procede a trabajar con la muestra de validación, compuesta por los cincuenta (50) pozos de la muestra de estudio y los cinco (5) pozos adicionales pertenecientes al campo.

A partir de las ecuaciones obtenidas por el modelo promedio con mejores predictores, se calcularon los nueve (9) parámetro de producción para los cincuenta (50) pozos de la muestra de estudio con base a los valores de los parámetros petrofísicos de la base de datos de promedio. Los resultados de las ecuaciones fueron comparados con los valores de los parámetros de producción de campo por medio de la superposición de sus dos (2) histogramas, estos se pueden visualizar en la figura 26.

**Figura 26.** Histogramas de modelo promedio con mejores predictores y datos de campo



**Fuente:** elaboración propia



A partir de los resultados presentes en la figura 26 se pueden comparar los resultados del modelo promedio con mejores predictores contra los parámetros de producción de campo, donde la superposición de los dos (2) histogramas indica que los valores predichos se asemejan a los valores reales, el color azul indica que para esos valores de campo el modelo no los pudo reproducir con exactitud y el color rojo evidencia que el modelo calculo valores erróneos.

Como se mencionó previamente en el apartado de datos atípicos, en algunos parámetros de producción como el BOPD, BWPD, EUR, Oil, IP, Pico y Water se logra apreciar columnas azules en sus valores más altos, lo que ratifica el hecho de que el modelo no tiene una buena capacidad predictiva sobre escenarios productivos de altos valores, en la zona que existe una mayor superposición de los histogramas de forma general es sobre los valores medios de los parámetros.

Se destacan las gráficas de EUR y BSW las cuales tiene una gran área de superposición de histogramas lo cual indica que estos parámetros de producción son predichos por el modelo con una alta exactitud. De forma análoga las gráficas de IP y BOPD presentan una importante carea de no superposición, evidenciando la complejidad predictiva de estos parámetros.

En los histogramas de Water se observa una distribución normal en sus datos, lo cual se atribuye a que un valor entre seis (6) y siete (7) millones de barriles de agua acumulada para un pozo, suele ser un valor común para su cierre, razón por la cual se presenta ese pico. Aquellos pozos que presentan un bajo corte de agua para esas instancias logran superar este acumulado, siendo estos una minoría.

### **3.11 VALIDACIÓN DEL MODELO PROMEDIO CON MEJORES PREDICTORES**

Para lograr comparar y validar los resultados obtenidos por el modelo promedio con mejores predictores se graficaron los datos de producción de campo contra los valores calculados por las ecuaciones del modelo. Dentro de este grafico la recta que tiene como pendiente uno (1) y parte desde el origen, se denomina la línea ideal que representa el infinito conjunto de puntos que el modelo predice con una exactitud del cien (100) por ciento. El gráfico también cuenta con un intervalo de confianza del noventa y cinco (95) por ciento y un intervalo de predicción del noventa y cinco (95) por ciento.

Los gráficos están divididos por parámetro de producción y dentro de él se plasmaron los pozos pertenecientes a la muestra de validación (un total de cincuenta y cinco pozos) haciendo la acotación de que los puntos negros son los pertenecientes a la muestra de estudio y los rojo son aquellos adicionales pero ubicados en el campo Imperial. Los gráficos se pueden observar en la figura 27.

**Figura 27.** Datos de producción de campo frente a los valores calculados por las ecuaciones del modelo

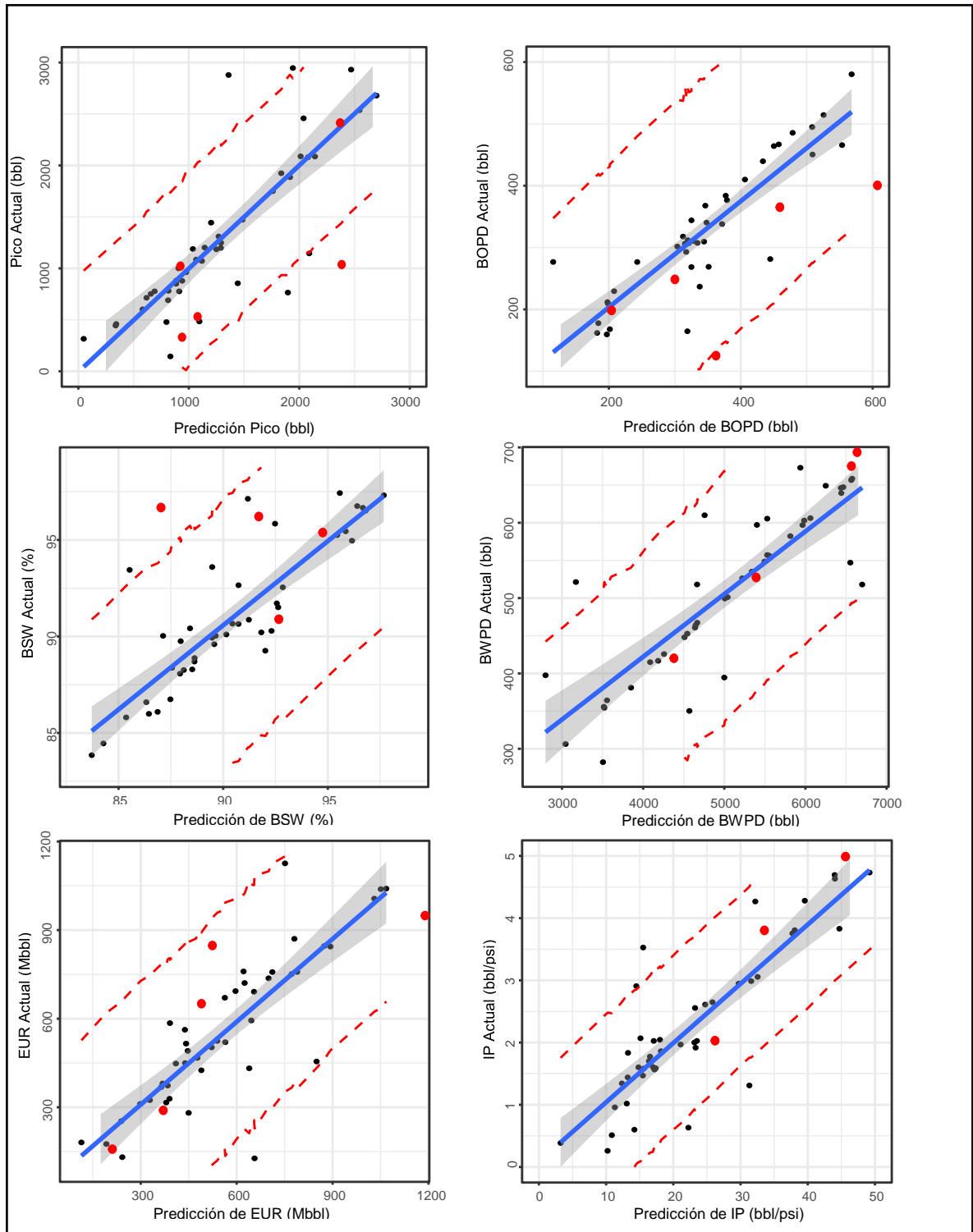
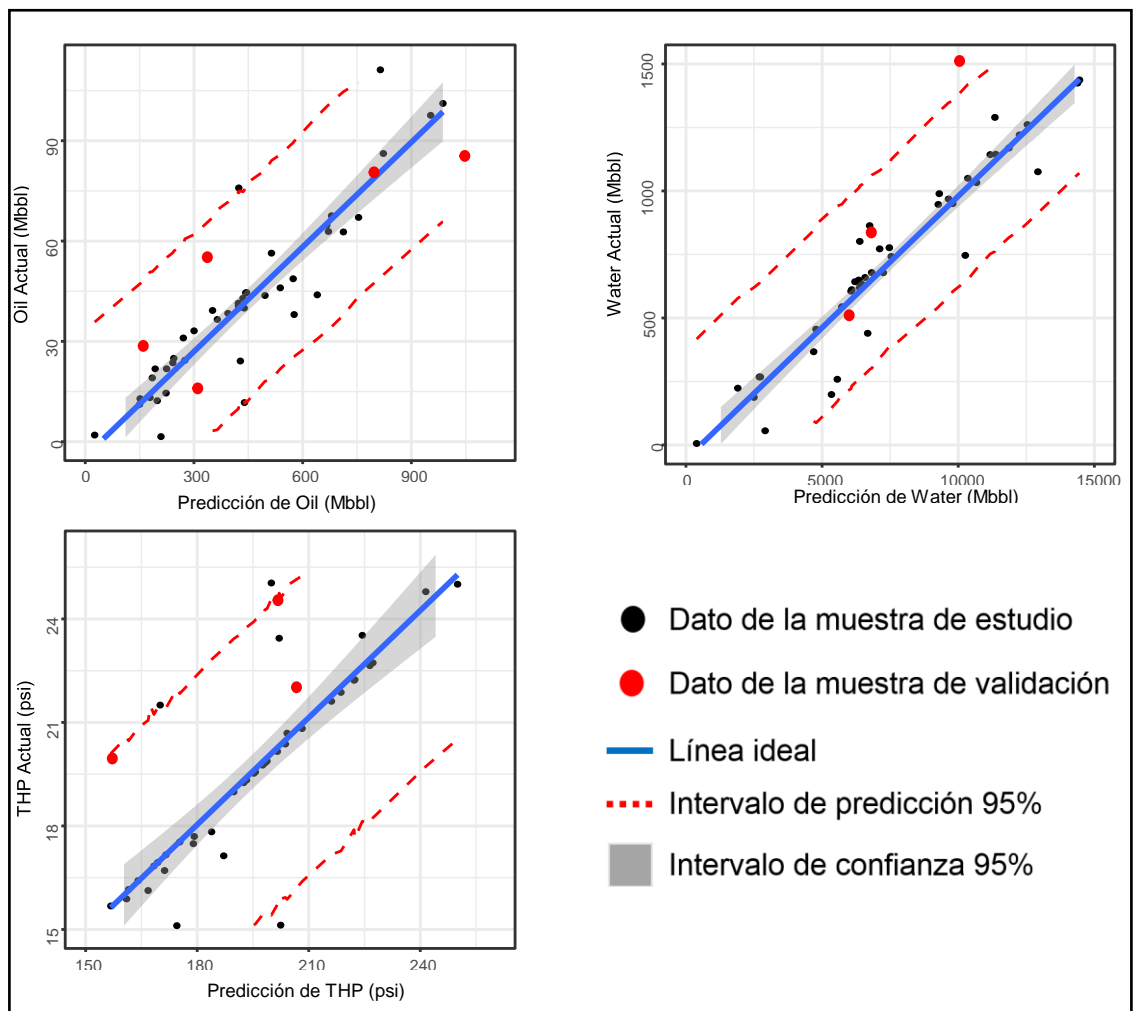


Figura 27. (Continuación)



Fuente: elaboración propia

Al analizar los resultados de la figura 27 se puede observar cómo se comportan los resultados del modelo promedio con mejores predictores frente a los parámetros de producción de campo para los pozos de la muestra de validación. En primera instancia se observa un acople general del modelo a los valores reales con un grado de dispersión que varía entre parámetros.

Para el parámetro pico de producción se evidencia una gran exactitud entre los valores de quinientos (500) y mil quinientos (1500) barriles de petróleo, presentando una alta dispersión de ciertos pozos por encima de este rango. Con respecto a los pozos por fuera de la muestra de estudio se evidencia una predicción aceptable debido a que cuatro (4) de los cinco (5) pozos se encuentran dentro del intervalo de predicción sin presentar una alta exactitud.

Para el parámetro producción de petróleo diaria se evidencia una buena exactitud a lo largo de rango de valores de la variable, presentando una dispersión media general. Con respecto a los pozos por fuera de la muestra de estudio se evidencia una predicción aceptable debido a que cuatro (4) de los cinco (5) pozos se encuentran dentro del intervalo de predicción sin presentar una alta exactitud.

Para el parámetro corte de agua se evidencia una buena exactitud entre los valores de ochenta (80) y noventa y dos (92) por ciento, presentando una dispersión media de ciertos pozos por encima de este rango. Con respecto a los pozos por fuera de la muestra de estudio se evidencia una predicción media debido a que tres (3) de los cinco (5) pozos se encuentran dentro del intervalo de predicción sin presentar una alta exactitud.

Para el parámetro producción de agua diaria se evidencia una buena exactitud a lo largo de rango de valores de la variable, presentando una dispersión baja en general. Con respecto a los pozos por fuera de la muestra de estudio se evidencia una predicción buena debido a que los cinco (5) pozos se encuentran dentro del intervalo de predicción sin presentar una alta exactitud.

Para el parámetro estimado de recobro final se evidencia una media exactitud a lo largo de rango de valores de la variable, presentando una dispersión media en general. Con respecto a los pozos por fuera de la muestra de estudio se evidencia una predicción buena debido a que los cinco (5) pozos se encuentran dentro del intervalo de predicción sin presentar una alta exactitud.

Para el parámetro índice de productividad se evidencia una buena exactitud entre los valores de doce (12) y treinta y dos (32) barriles sobre psi, presentando una dispersión media de ciertos pozos por fuera de este rango. Con respecto a los pozos por fuera de la muestra de estudio se evidencia una predicción media debido a que tres (3) de los cinco (5) pozos se encuentran dentro del intervalo de predicción con una exactitud media.

Para el parámetro producción de petróleo acumulada se evidencia una buena exactitud a lo largo de rango de valores de la variable, presentando una dispersión media general. Con respecto a los pozos por fuera de la muestra de estudio se evidencia una predicción aceptable debido a que los cinco (5) pozos se encuentran dentro del intervalo de predicción con una exactitud media.

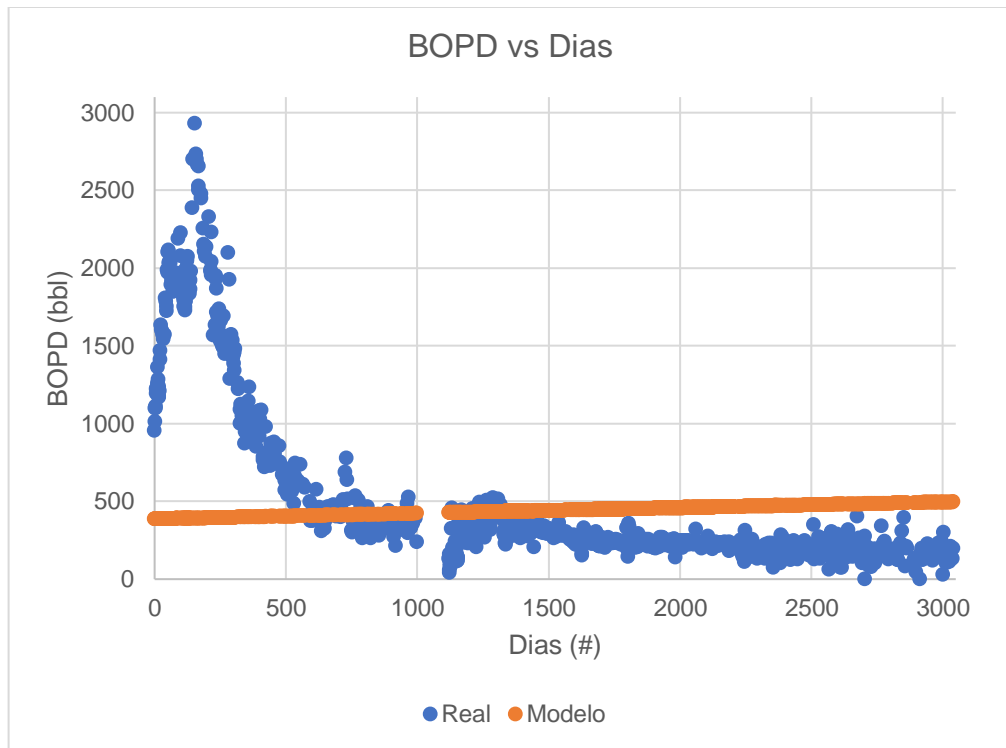
Para el parámetro producción de agua acumulada se evidencia una buena exactitud a lo largo de rango de valores de la variable, presentando una dispersión media entre los valores de seis mil (6000) y ocho mil (8000) barriles. Con respecto a los pozos por fuera de la muestra de estudio se evidencia una predicción media debido a que tres (3) de los cinco (5) pozos se encuentran dentro del intervalo de predicción con una exactitud media.

Para el parámetro de presión de cabeza de pozo se evidencia una alta exactitud en la predicción de pozos de la muestra de estudio a lo largo de rango de valores de la variable y baja predicción para los cinco (5) pozos adicionales.

### 3.12 COMPORTAMIENTO TEMPORAL DEL MODELO

Debido a que el modelo promedio con mejores predictores cuenta con la variable días de producción, se puede visualizar como es el comportamiento de las variables en función del tiempo según los resultados obtenidos. Esto se realiza a partir de la información petrofísica del pozo y tomando como variable los días de producción, como ejemplo se tomará el pozo Imperio 3 el cual tiene un total de tres mil cuarenta (3040) días de producción y los resultados se pueden apreciar en la gráfica 31.

**Gráfica 31.** BOPD para el pozo Impero 3



**Fuente:** elaboración propia

Como se puede apreciar en la gráfica 31, el comportamiento real de la producción de petróleo diaria tiene una tendencia ascendente al inicio y una fase de caída posterior con una estabilización a lo largo del tiempo, mientras que el comportamiento de este parámetro por el modelo es el de una línea recta, debido a que todos los parámetros petrofísicos se mantienen constantes a excepción del tiempo. Se aprecia que el modelo no simula los datos de producción en una fase temprana, tiene a simular los datos de forma más precisa, una vez la producción se empieza a estabilizar (después de los 500 días de producción). Para los datos modelados se calcularon los percentiles de diez (10), cincuenta (50) y noventa (90) por ciento (397, 448 y 479 barriles respectivamente).

#### 4. CONCLUSIONES

- Para los pozos del Campo Imperial existen correlaciones entre los parámetros petrofísicos y de producción, lo cual se evidencia por los valores del R cuadrado del modelo promedio con mejores predictores. Los parámetros de producción que tienen mayor capacidad explicativa son la producción de agua acumulada, el índice de productividad y la producción de petróleo acumulada con los porcentajes de 86.23, 78.79 y 78.11 por ciento respectivamente.
- El parámetro petrofísico con mayor capacidad explicativa general sobre los parámetros de producción es la heterogeneidad, debido a que permanece en todos los modelos del promedio con mejores predictores y ser el parámetro con el beta estandarizado más alto para tres (3) de los nueve (9) parámetros de producción. Se evidencia que la heterogeneidad de un pozo es un factor indispensable para poder predecir el comportamiento productivo de un pozo.
- El porcentaje de pozos que se encuentra por dentro del intervalo de predicción para la muestra de estudio y la muestra de validación es 96.62 y 88.89 por ciento respectivamente, por lo cual los resultados de las ecuaciones obtenidas tienen una mayor exactitud en los pozos pertenecientes a la muestra de estudio que a los pozos que están por fuera de esta.
- El modelo planteado tiene una buena capacidad predictiva sobre el parámetro de producción BSW en un rango entre 80 y 92 por ciento, intervalo donde comúnmente inicia el corte de agua para un pozo del campo a la fecha de hoy. El modelo puede ser usado para anticipar el volumen de agua a tratar en superficie de nuevos pozos perforados, como un valor de referencia.
- Debido a que los valores de los R cuadrados en el modelo promedio con mejores predictores en ninguno de los parámetros de producción en uno (1), se afirma que el rendimiento productivo de un pozo es un comportamiento complejo de predecir y no es únicamente dependiente de las propiedades de la roca y los fluidos.

## 5. RECOMENDACIONES

- Tomar un mayor número de pozos como la muestra de estudio para tener con menor incertidumbre una base de datos normalmente distribuida.
- Añadir parámetros predictores (variables independientes) de otras ramas como la perforación para mejorar la exactitud del modelo obtenido.
- Modificar matemáticamente las variables de entrada para mejorar la exactitud del modelo obtenido.
- Sustraer los parámetros petrofísicos que evidencian una alta colinealidad, para observar el comportamiento del modelo.
- Sustraer las observaciones con una gran influencia en el modelo promedio con mejores predictores, con el fin de observar el comportamiento del modelo.
- Evaluar como variable independiente la permeabilidad vertical y longitud de la sección horizontal de los pozos de la muestra de estudio, para observar el comportamiento del modelo.

## BIBLIOGRAFÍA

ACEROS GARCÍA, Savieth. Diseño del módulo de selección de conectividad para el manual general de completamiento de pozos de Ecopetrol s.a. [en línea] Tesis. Fundación Universidad De América, Bogotá D.C.: 2017. Disponible en: <https://repository.uamerica.edu.co/bitstream/20.500.11839/896/1/5121333-2017-1-IP.pdf>

ARACEL, Tatiana. JAVITA, Revelo. Actualización de reservas y ubicación de pozos de desarrollo en el campo Tapi-Tetete. Tesis. Quito, 2007, Escuela politécnica nacional.

ASOCIACIÓN COLOMBIA DE PETRÓLEO, Todo sobre fracking.[Sitio WEB]. Colombia [28, octubre, 2019] Disponible en: <https://acp.com.co/web2017/en/todo-sobre-el-fracking/955-que-son-los-hidrocarburos-en-reservorios-no-convencionales>

BROWN, Kermit. The technology of artificial lift methods. Denver: PennWell Books, 2007. 1087 p. ISBN 0878141375

BRYAN, Jennifer. readxl: Read Excel Files. R package version 1.3.1. 2019. <https://CRAN.R-project.org/package=readxl>

BUJ, O. GISBERT, J. Petrophysical characterization of three commercial varieties of miocene sandstones from the Ebro valley. Zaragoza: Universidad de Zaragoza, 2007. ISSN: 0465-2746.

CUNNINGHAM, Charlie. COOLEY, Lisa. WOZNIAK, Gregory. PANCAKE, Jim. Using Multiple Linear Regression to Model EURs of Horizontal Marcellus Shale Wells Pennsylvania: SPE International, 2012. 161343.

ECUACIÓN DE DIFUSIVIDAD. (10, octubre, 2019: Bogotá D.C.) RAMOS, Jorge Luis. Bogotá D.C. Universidad América, 2019. 10 p.

ELLIS, Shannon. skimr: Compact and Flexible Summaries of Data. R package version 2.1. 2020. <https://CRAN.R-project.org/package=skimr>

FUQEN, Jaime. Geología de las planchas 98 - durania y 99 – villa. Bogotá, 2011. Instituto colombiano de geología y minería ingeominas.

GEOLOGICAL SURVEY, Payr determination. [Sitio WEB] . Kansas [25, octubre, 2019] Disponible en: <http://www.kgs.ku.edu/software/PfEFFER-java/HELP/PfEFFER/Pfeffer-theory5.html>



HOTHORN, Torsten. Diagnostic Checking in Regression Relationships. R News 2(3), 7-10. 2002. URL <https://CRAN.R-project.org/doc/Rnews/>

INSTITUTO COLOMBIANO DE NORMAS TÉCNICAS Y CERTIFICACIÓN. Compendio de normas para trabajos escritos. NTC-1486-6166. Bogotá D.C. El instituto, 2018 ISBN 9789588585673 153 p.

MICHELENA, Reinaldo. JIMÉNEZ, Juan. Estimation of Petrophysical Properties Using Multiple Attributes: Generalizing Linear Regressions. Texas: SEG Exposition an annual meeting. 2001.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

RESEARCH GATE, Liquid-Liquid Separation [Sitio WEB]. Palermo [22, octubre, 2019] Disponible en: [https://www.researchgate.net/figure/MBT-results-for-a-water-in-oil-emulsion-at-30-BSW-Dots-experimental-results\\_fig5\\_254534820](https://www.researchgate.net/figure/MBT-results-for-a-water-in-oil-emulsion-at-30-BSW-Dots-experimental-results_fig5_254534820)

REVELLE, W. psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, 2019. <https://CRAN.R-project.org/package=psych> Version =1.9.12.

RODRIGO ,Amat. Introducción a la Regresión Lineal Múltiple. . [Sitio WEB]. [Julio, 2016] Disponible en: [https://rpubs.com/Joaquin\\_AR/226291](https://rpubs.com/Joaquin_AR/226291)

ROJAS, Pedro. SEIFERT, Douglas. Field Evaluation of LWD Resistivity Logs in Highly Deviated and Horizontal Wells in Saudi Arabia. Saudi Arabia: SPE International, 2013, Halliburton.

SERRA, Oberto. Fundamentals of well-log interpretation. Francia: Elsevier, 1984. 423p. ISBN 0444421327

SCHLUMBERGER, Interactive Petrophysics. Manual. 2018.

VADULLI, Rolly Buccioni. Estimación del coeficiente de concentración de Gini a partir de la curva estimada de Lorenz. 2012. Chile. Revista Chilena de Economía y Sociedad, 5(1-2).

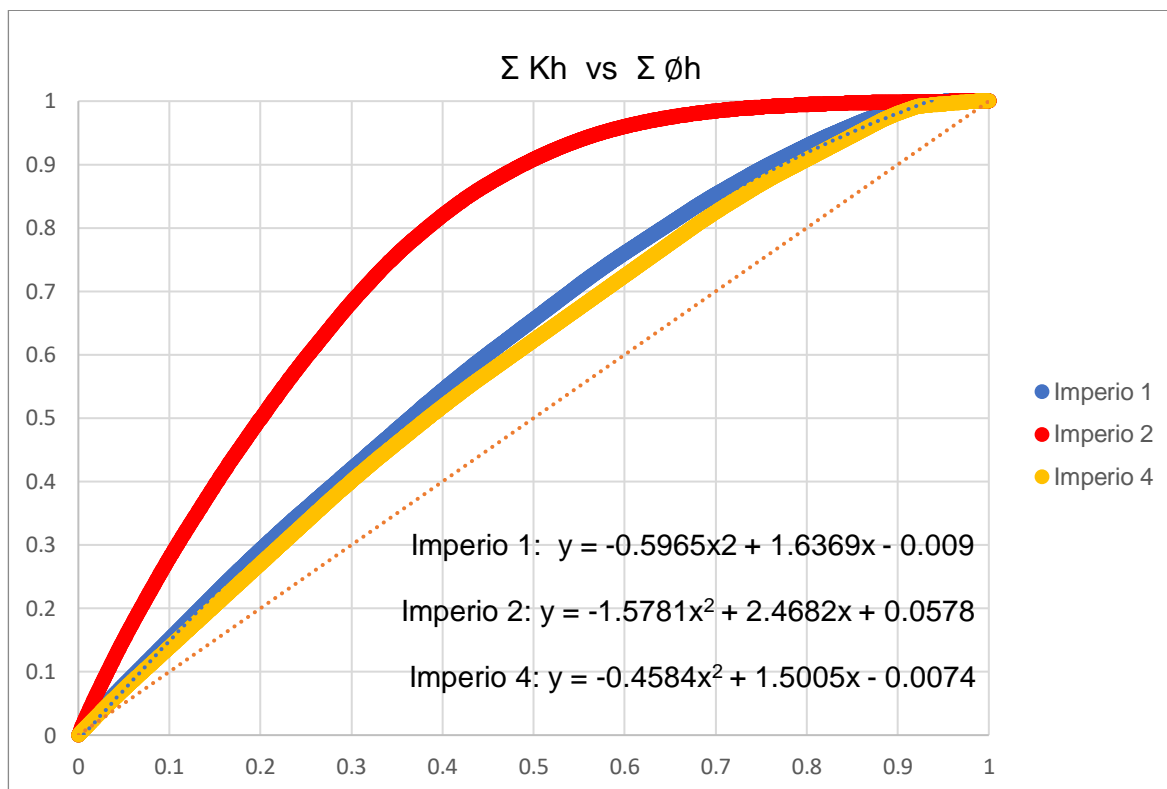
WALPOLE, Myers. Probabilidad y estadística. México: Pearson, 2012. 816 p. ISBN 9786073214179

WICKHAM, Hadley. dplyr: A Grammar of Data Manipulation. R package version 0.8.5. 2020. <https://CRAN.R-project.org/package=dplyr>

## **ANEXOS**

## ANEXO A

### CAPACIDAD ACUMULADA Y NORMALIZADA DE LA PERMEABILIDAD CONTRA LA CAPACIDAD ACUMULADA Y NORMALIZADA DE LA POROSIDAD



Fuente: elaboración propia

## ANEXO B

### ECUACIONES DEL MODELO TABLA COMPLETA

#### Modelo tabla completa

$$EUR = -2638 + 0.1745 \text{ PHIE} + 100.2 \text{ VCL} + 0.007875 \text{ K} - 1.494 \text{ Acu} + 1.391 \text{ Netpay} + 177.1 \text{ Heteroge} + 3919 \text{ No}$$

$$\text{BOPD} = -890.2 + 0.1319 \text{ PHIE} - 40.330 \text{ VCL} + 0.003023 \text{ K} + 2.007 \text{ Acu} + 0.6648 \text{ Netpay} + 67.99 \text{ Heteroge} + 1647 \text{ No} - 0.2741 \text{ Dias}$$

$$\text{BWPD} = 3184 - 0.8738 \text{ PHIE} + 628.9 \text{ VCL} - 0.00211 \text{ K} + 19.67 \text{ Acu} + 1.186 \text{ Netpay} + 227 \text{ Heteroge} - 3167 \text{ No} + 0.9897 \text{ Dias}$$

$$\text{BSW} = 125.4 - 0.005401 \text{ PHIE} + 5.23 \text{ VCL} - 0.00000367 \text{ K} - 0.05713 \text{ Acu} - 0.02161 \text{ Netpay} - 1.89 \text{ Heteroge} - 53.56 \text{ No} + 0.01199 \text{ Dias}$$

$$\text{Oil} = -1664 + 0.0696 \text{ PHIE} + 192.4 \text{ VCL} + 0.005646 \text{ K} - 1.73 \text{ Acu} + 0.7892 \text{ Netpay} + 91.9 \text{ Heteroge} + 2026 \text{ No} + 0.3057 \text{ Dias}$$

$$\text{Water} = -4606 - 0.1937 \text{ PHIE} + 1737 \text{ VCL} + 0.03529 \text{ K} - 4.226 \text{ Acu} + 2.173 \text{ Netpay} + 319.7 \text{ Heteroge} + 2304 \text{ No} + 5.328 \text{ Dias}$$

$$\text{IP} = 60 - 0.003714 \text{ PHIE} + 6.484 \text{ VCL} + 0.0002586 \text{ K} + 0.2811 \text{ Acu} + 0.005863 \text{ Netpay} - 3.472 \text{ Heteroge} - 95.55 \text{ No} + 0.001772 \text{ Dias}$$

$$\text{THP} = 266.7 + 0.0174 \text{ PHIE} - 36.13 \text{ VCL} - 0.001325 \text{ K} + 0.003615 \text{ Acu} - 0.02907 \text{ Netpay} - 3.704 \text{ Heteroge} - 109.2 \text{ No} - 0.003948 \text{ Dias}$$

$$\text{Pico} = -4667 + 0.2029 \text{ PHIE} + 263.1 \text{ VCL} + 0.01463 \text{ K} - 6.039 \text{ Acu} + 2.83 \text{ Netpay} + 400.7 \text{ Heteroge} + 6072 \text{ No}$$

**Fuente:** elaboración propia

## ANEXO C

### ECUACIONES DEL MODELO TABLA MEDIANA

#### Modelo mediana

$$EUR = 626.4 - 6021 \text{ PHIE} - 6400 \text{ VCL} + 0.011630 \text{ K} + 3.658 \text{ Acu} + 0.670 \text{ Netpay} + 95.5 \text{ Heteroge} + 3064 \text{ No}$$

$$\text{BOPD} = -411.6 + 858.5 \text{ PHIE} - 218.2 \text{ VCL} + 0.000514 \text{ K} + 5.632 \text{ Acu} + 0.1367 \text{ Netpay} + 17.14 \text{ Heteroge} + 312 \text{ No} - 0.0096 \text{ Dias}$$

$$\text{BWPD} = 6963 - 10420 \text{ PHIE} - 30150 \text{ VCL} - 0.19020 \text{ K} + 56.62 \text{ Acu} + 1.208 \text{ Netpay} + 475 \text{ Heteroge} - 1115 \text{ No} + 0.6687 \text{ Dias}$$

$$\text{BSW} = 108.7 - 23.251 \text{ PHIE} - 7.37 \text{ VCL} - 0.00003921 \text{ K} - 0.06880 \text{ Acu} - 0.00185 \text{ Netpay} - 0.03 \text{ Heteroge} - 8.96 \text{ No} + 0.00039 \text{ Dias}$$

$$\text{Oil} = -2031 + 2132 \text{ PHIE} + 494.9 \text{ VCL} - 0.008964 \text{ K} + 6.27 \text{ Acu} + 0.4464 \text{ Netpay} + 101.3 \text{ Heteroge} + 1402 \text{ No} + 0.2323 \text{ Dias}$$

$$\text{Water} = -6661 + 6637 \text{ PHIE} - 12050 \text{ VCL} - 0.32040 \text{ K} + 88.63 \text{ Acu} + 2.203 \text{ Netpay} + 1075 \text{ Heteroge} - 2627 \text{ No} + 4.097 \text{ Dias}$$

$$\text{IP} = 162.8 - 239.3 \text{ PHIE} - 710.1 \text{ VCL} - 0.001161 \text{ K} + 0.5836 \text{ Acu} - 0.009141 \text{ Netpay} - 4.357 \text{ Heteroge} + 23.33 \text{ No} + 0.000328 \text{ Dias}$$

$$\text{THP} = 477.5 - 948.9 \text{ PHIE} - 259.8 \text{ VCL} + 0.001883 \text{ K} - 1.32 \text{ Acu} + 0.06156 \text{ Netpay} - 0.357 \text{ Heteroge} + 75.3 \text{ No} - 0.02269 \text{ Dias}$$

$$\text{Pico} = -1053 - 6483 \text{ PHIE} - 13960 \text{ VCL} + 0.02226 \text{ K} + 9.286 \text{ Acu} + 1.84 \text{ Netpay} + 278.7 \text{ Heteroge} + 6591 \text{ No}$$

**Fuente:** elaboración propia

## ANEXO D

### ECUACIONES DEL MODELO PROMEDIO

#### Modelo promedio

$$EUR = 2084 - 10500 \text{ PHIE} - 8238 \text{ VCL} + 0.03821 \text{ K} + 2.11 \text{ Acu} + 0.64 \text{ Netpay} + 91.2 \text{ Heteroge} + 3171 \text{ No}$$

$$\text{BOPD} = -800.1 + 916.4 \text{ PHIE} - 44.59 \text{ VCL} + 0.00806 \text{ K} + 5.889 \text{ Acu} + 0.2661 \text{ Netpay} + 36.49 \text{ Heteroge} + 718 \text{ No} + 0.0373 \text{ Dias}$$

$$\text{BWPD} = 5786 - 12850 \text{ PHIE} - 3409 \text{ VCL} - 0.1272 \text{ K} + 20.12 \text{ Acu} + 1.964 \text{ Netpay} + 522 \text{ Heteroge} - 4176 \text{ No} + 0.6317 \text{ Dias}$$

$$\text{BSW} = 99.6 + 30.87 \text{ PHIE} + 69.76 \text{ VCL} - 0.0005084 \text{ K} - 0.18080 \text{ Acu} - 0.00815 \text{ Netpay} - 0.25 \text{ Heteroge} - 36.30 \text{ No} + 0.00093 \text{ Dias}$$

$$\text{Oil} = -1793 + 796.7 \text{ PHIE} + 1395 \text{ VCL} + 0.001754 \text{ K} + 4.57 \text{ Acu} + 0.5364 \text{ Netpay} + 104.6 \text{ Heteroge} + 1327 \text{ No} + 0.2142 \text{ Dias}$$

$$\text{Water} = -18800 + 29830 \text{ PHIE} + 46160 \text{ VCL} - 0.3637 \text{ K} + 62.62 \text{ Acu} + 4.277 \text{ Netpay} + 1318 \text{ Heteroge} - 9386 \text{ No} + 4.382 \text{ Dias}$$

$$\text{IP} = 31.5 + 73.85 \text{ PHIE} + 9.427 \text{ VCL} - 0.0004796 \text{ K} + 0.3559 \text{ Acu} + 0.003262 \text{ Netpay} - 2.012 \text{ Heteroge} - 96.32 \text{ No} + 0.003272 \text{ Dias}$$

$$\text{THP} = 831 - 1853 \text{ PHIE} - 1626 \text{ VCL} + 0.002307 \text{ K} - 0.8288 \text{ Acu} + 0.03829 \text{ Netpay} - 4.703 \text{ Heteroge} + 252.3 \text{ No} - 0.023650 \text{ Dias}$$

$$\text{Pico} = -975.9 - 8623 \text{ PHIE} - 4709 \text{ VCL} + 0.07494 \text{ K} + 3.136 \text{ Acu} + 2.12 \text{ Netpay} + 315.6 \text{ Heteroge} + 4669 \text{ No}$$

**Fuente:** elaboración propia

## ANEXO E

### ECUACIONES DEL MODELO TABLA PROMEDIO CON LOS MEJORES PREDICTORES

#### Modelo promedio con los mejores predictores

$$EUR = 2455 - 11600 \text{ PHIE} - 8674 \text{ VCL} + 0.03616 \text{ K} + 0.666 \text{ Netpay} + 93.8 \text{ Heteroge} + 3214 \text{ No}$$

$$\text{BOPD} = -525.2 + 0.009952 \text{ K} + 5.279 \text{ Acu} + 0.2887 \text{ Netpay} + 36.86 \text{ Heteroge} + 630 \text{ No} + 0.0362 \text{ Dias}$$

$$\text{BWPD} = 4542 - 9347 \text{ PHIE} - 0.1317 \text{ K} + 20.58 \text{ Acu} + 2.002 \text{ Netpay} + 529 \text{ Heteroge} - 4636 \text{ No} + 0.6438 \text{ Dias}$$

$$\text{BSW} = 107.8 + 49.24 \text{ VCL} - 0.0005314 \text{ K} - 0.195 \text{ Acu} - 0.0073 \text{ Netpay} - 32.58 \text{ No} + 0.001 \text{ Dias}$$

$$\text{Oil} = -1462 + 4.65 \text{ Acu} + 0.5084 \text{ Netpay} + 103 \text{ Heteroge} + 1544 \text{ No} + 0.2123 \text{ Dias}$$

$$\text{Water} = -8536.4 + 23372 \text{ VCL} - 0.3205 \text{ K} + 55.436 \text{ Acu} + 4.168 \text{ Netpay} + 1273.2 \text{ Heteroge} - 6858.56 \text{ No} + 4.293 \text{ Dias}$$

$$\text{IP} = 59.7 + 0.3456 \text{ Acu} - 2.682 \text{ Heteroge} - 104.6 \text{ No} + 0.003076 \text{ Dias}$$

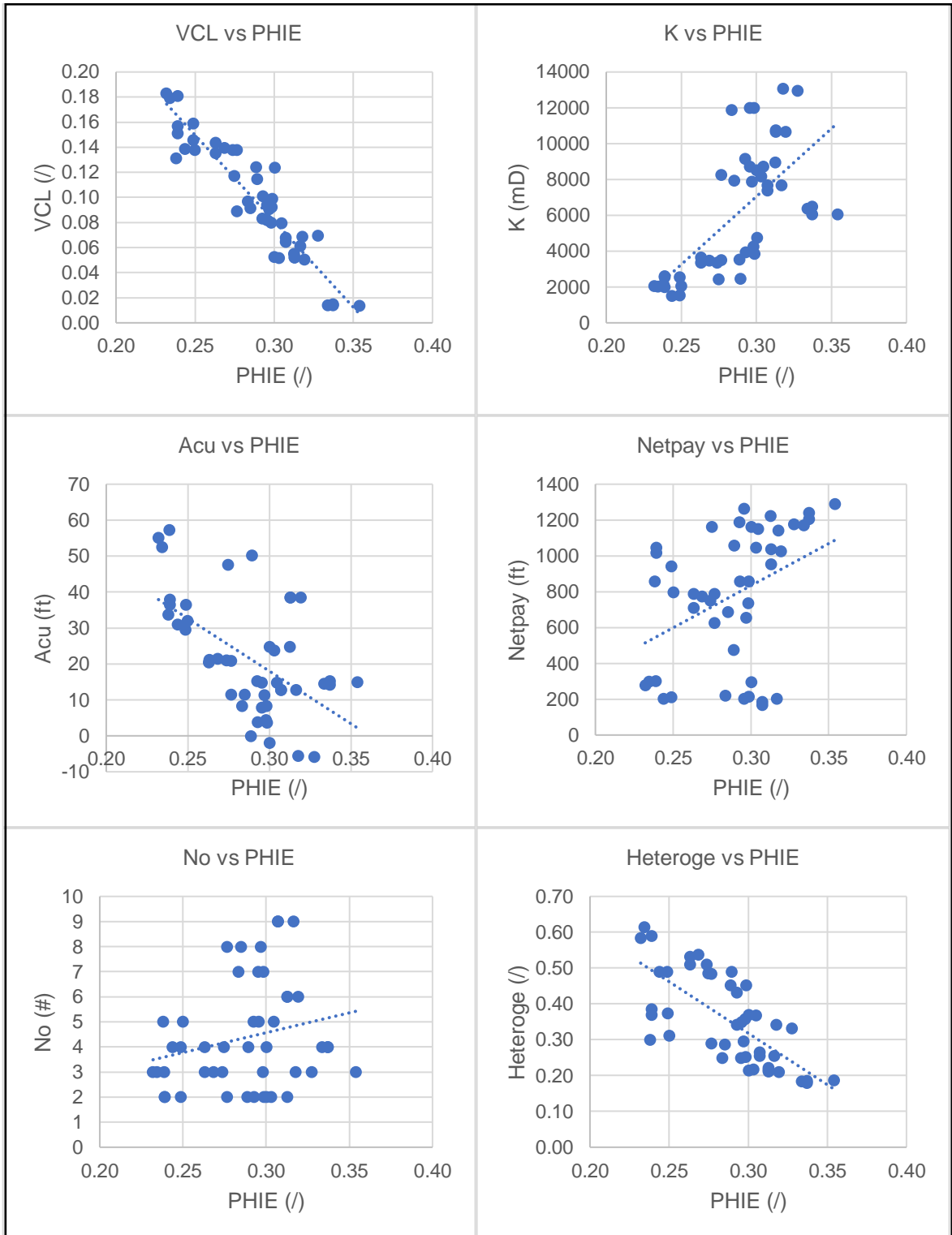
$$\text{THP} = 831 - 1853 \text{ PHIE} - 1626 \text{ VCL} + 0.002307 \text{ K} - 0.8288 \text{ Acu} + 0.03829 \text{ Netpay} - 4.703 \text{ Heteroge} + 252.3 \text{ No} - 0.023650 \text{ Dias}$$

$$\text{Pico} = -2267 - 5040 \text{ PHIE} + 0.06649 \text{ K} + 2.24 \text{ Netpay} + 328.1 \text{ Heteroge} + 4025 \text{ No}$$

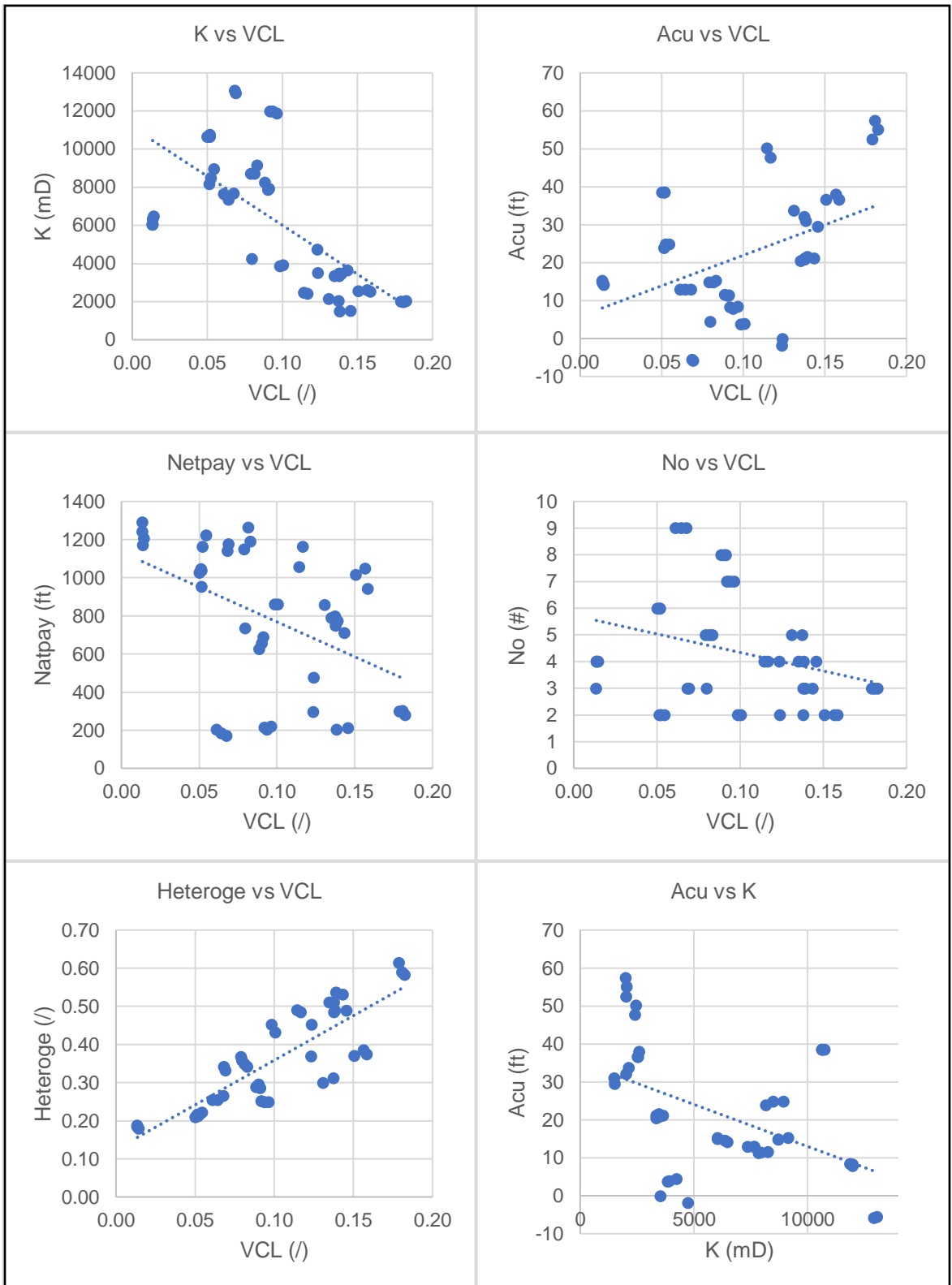
**Fuente:** elaboración propia

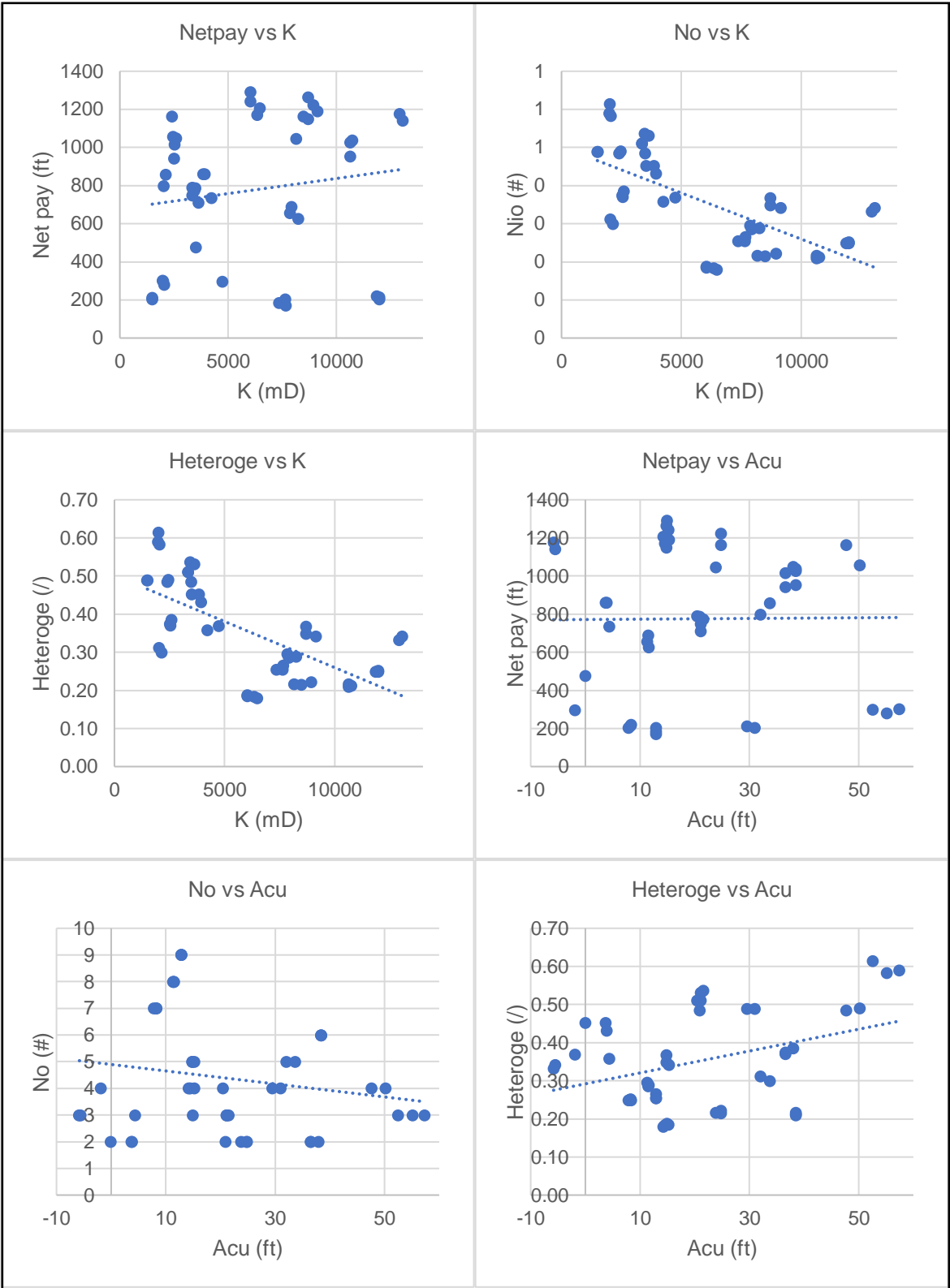
## ANEXO F

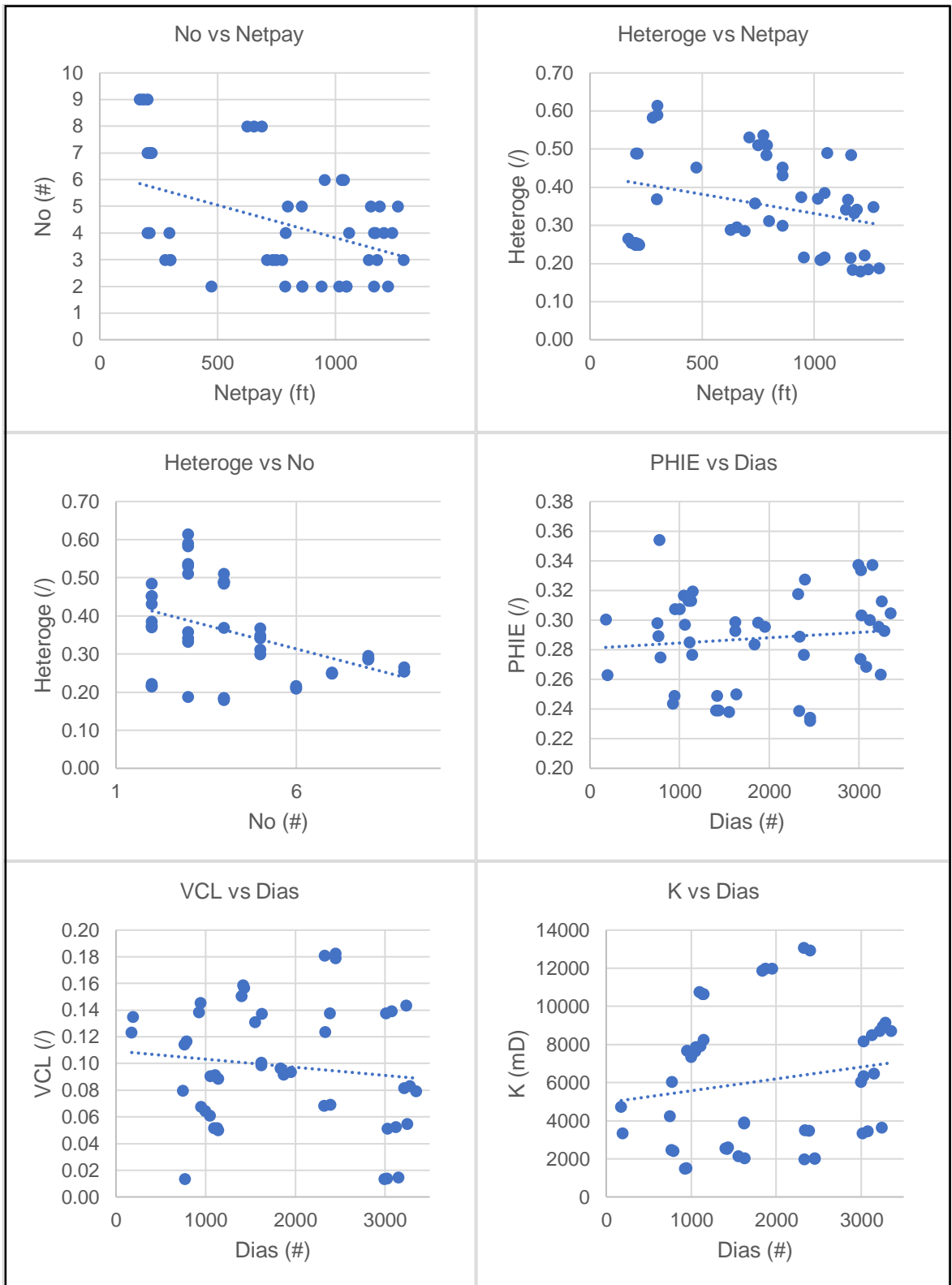
### GRÁFICOS CRUZADOS

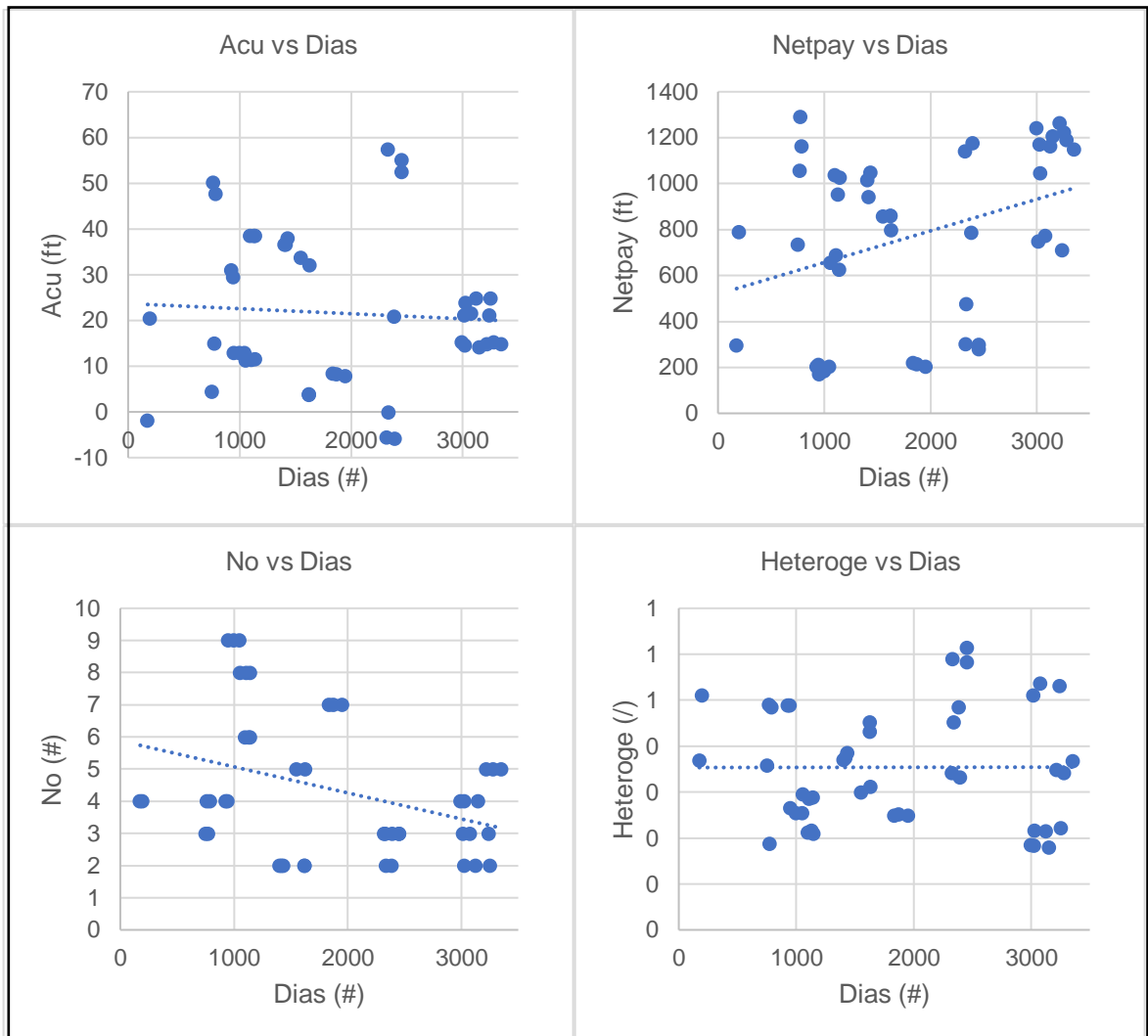












Fuente: elaboración propia