

ESTIMACIÓN DE LA POROSIDAD A PARTIR DE MODELOS PROBABILÍSTICOS Y TÉCNICAS DE MACHINE LEARNING EN LA FORMACIÓN VACA MUERTA

Carolina Bernhardt¹, Diego Gallart¹, Alberto Ortiz²

1: YPF S.A. carolina.bernhardt@ypf.com, diego.gallart@ypf.com

2: NZC Solutions. alberto.ortiz@nzcsolutions.com

Palabras claves: Vaca Muerta, *Shale*, Machine Learning, porosidad

ABSTRACT

In shale reservoirs, quantification of total porosity and water saturation are considered key parameters for calculating the volume of hydrocarbon in place. However, the petrophysical interpretation necessary for the definition of these parameters represents a challenge due to mineralogical complexity of the reservoir as well as the presence of kerogen and different fluids depending on their thermal maturity. Therefore, for a good petrophysical characterization, it is necessary to drill pilot vertical wells to acquire a wide variety of data, both from logging tools and core samples. As a result of this, there is a rise in the budget associated with the acquisition of data in the shale characterization projects. On the other hand, there are many old wells that have been drilled and logged in the Vaca Muerta-Quintuco formation section, due to their position in the lithological column overlying deeper conventional targets.

The purpose of this work is to present a workflow that allows the use of data from these legacy wells with a limited set of logs applying calibrations carried out using machine learning techniques based on probabilistic petrophysical interpretation carried out in wells with a complete set of logs and laboratory data. Once the target log has been defined, an exploratory analysis of the logs available in the aforementioned set of wells is carried out. The logs are studied separately and their relations as a whole, obtaining as a result a set of data with validated quality and with a format suitable for feeding machine learning models. Data is divided into training and validation sets. Then, different models are trained and validated, adjusting them to finally obtain the best result.

The implementation of this technique not only provides relevant information for improving property maps and volumetric calculations, taking into account the uncertainty associated with the results obtained, but also allows to analyze variability of porosity related to other geological characteristics, such as the depositional system and the thermal maturity of the rock. Finally, a better use of the available data allows to reduce costs allocated to the acquisition of new data in areas where the information is available from previously drilled wells.

INTRODUCCIÓN

La determinación de la porosidad total de la roca en un reservorio no convencional tipo “*shale*” es muy importante ya que esta variable generalmente está relacionada entre otras cosas con la calidad de la roca, el sistema deposicional, la capacidad de albergar hidrocarburos líquidos y la madurez térmica.

Es un parámetro crítico que, junto con otros como saturación y permeabilidad, tienen un impacto en la estimación de la productividad del reservorio.

La estimación de porosidad se hace a partir de datos provenientes de registros eléctricos y mediciones en laboratorio. La precisión del cálculo final estará fuertemente condicionada por la cantidad, tipo e integridad de los datos disponibles (Ortiz *et al.*, 2020). En el reservorio Vaca Muerta los registros de Resonancia Magnética, Neutrón y Densidad entre otros, permiten resolver esta propiedad con menor incertidumbre a través de la implementación de modelos petrofísicos determinísticos y/o probabilísticos (Mosse *et al.*, 2014; Rodríguez *et al.*, 2015; Cuervo *et al.*, 2016; Bernhardt *et al.*, 2017).

Sin embargo, pocos pozos tienen datos de perfiles y corona con estudios suficientes como para mantener altos niveles de precisión en los resultados. Es por ello que es necesario la aplicación de otro tipo de metodologías, a fin de determinar esta propiedad con un *set* limitado de perfiles y minimizando o al menos conociendo las imprecisiones en el cálculo final.

El incremento de datos de porosidad como variables de entrada a modelos estáticos y dinámicos permitirá sin dudas una mejor selección y definición de los diferentes niveles de navegación de los pozos horizontales planeados para la etapa de desarrollo del campo.

El objetivo de este trabajo es presentar una nueva metodología y flujo de trabajo que permita estimar porosidad con la máxima precisión posible, utilizando como base de calibración la interpretación petrofísica de tipo probabilística llevada a cabo en pozos con *set* completo de perfiles y datos de laboratorio, a través de técnicas de machine learning siendo las variables de entrada los registros eléctricos básicos.

Para tal fin, se evaluaron diferentes alternativas de análisis y basados en el error de cada una de ellas se eligieron 3 modelos predictivos (lineales y no lineales). La utilización y precisión obtenida de cada uno de ellos dependerá de los datos disponibles en el pozo.

METODOLOGÍA

El presente trabajo atravesó un largo proceso que va desde la creación de un modelo petrofísico probabilístico a detalle de determinación de propiedades, calibradas con datos de laboratorio y utilizando un *software* específico de petrofísica; para luego pasar a una etapa de análisis multivariable, creación de modelos y estimación de la porosidad, a través de herramientas de *machine learning*.

Modelo Probabilístico

A partir de un modelo petrofísico del reservorio basado en datos de registros eléctricos y laboratorio (Bernhardt *et al.*; 2017), se determinaron las propiedades petrofísicas a detalle de un grupo reducido de 26 pozos mediante una interpretación petrofísica de tipo probabilística. Los pozos están bien distribuidos en diferentes ventanas de madurez termal y además atraviesan toda la sección

orgánica de Vaca Muerta. Se puso especial énfasis en la consistencia de los datos según la compañía de perfilaje y en la aplicación de parámetros de adquisición y procesamiento similares.

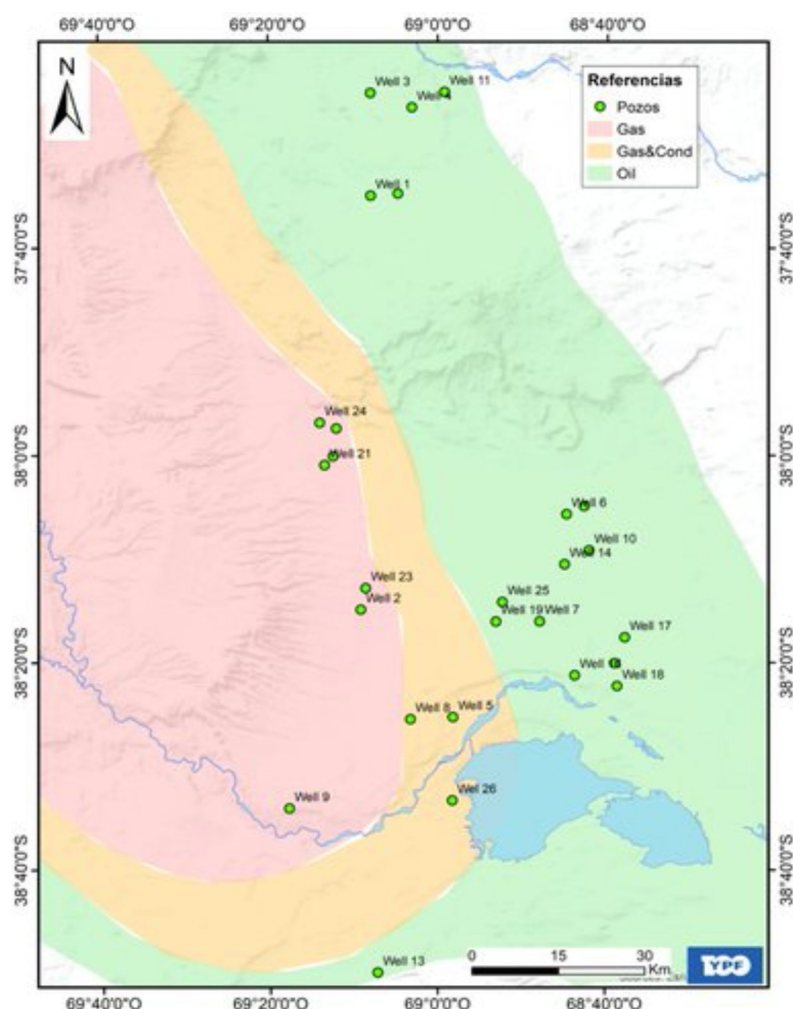


Figura 1. Ubicación de los 26 pozos con registros eléctricos básicos y especiales

Fuente de Datos

Los datos de entrada al modelo son registros eléctricos básicos y especiales (Tabla 1).

Herramienta	Curva	Mnemónico	Unidad
Resonancia Magnética Nuclear	Porosidad Total Resonancia	MRP	m3/m3
Dieléctrico	Porosidad Llena de agua	PWYO	m3/m3
Neutrón	Índice de Hidrógeno	IH	m3/m3
Densidad	Densidad	RHOZ	g/cm3
Espectroscopia captura e inelástica	Peso en Seco elemento (i)	DWelemento (i)	Kg/Kg

Tabla 1. Datos de entrada al modelo probabilístico

Evaluación Petrofísica

La primera parte del cálculo del modelo consistió en la aplicación de correlaciones entre los elementos y minerales obtenidas a partir de una gran cantidad de datos laboratorio de roca. Esta correlación de tipo matricial permite transformar luego el % en peso (% w/w) de elementos determinados con la herramienta de Espectroscopia de captura e inelástica, a % en peso mineral (% w/w). Los datos de corona consisten en mediciones de Fluorescencia de Rayos X (FRX) y Difracción de Rayos X (DRX), en 5 coronas de más de 120 metros de espesor orgánico y extraídas en diferentes ventanas de madurez, realizadas por el mismo laboratorio siguiendo la misma metodología interpretativa. La Figura 2 muestra un pozo en ventana de gas, con la calibración elemental y la estimación de 8 minerales partir de los elementos previamente calibrados.

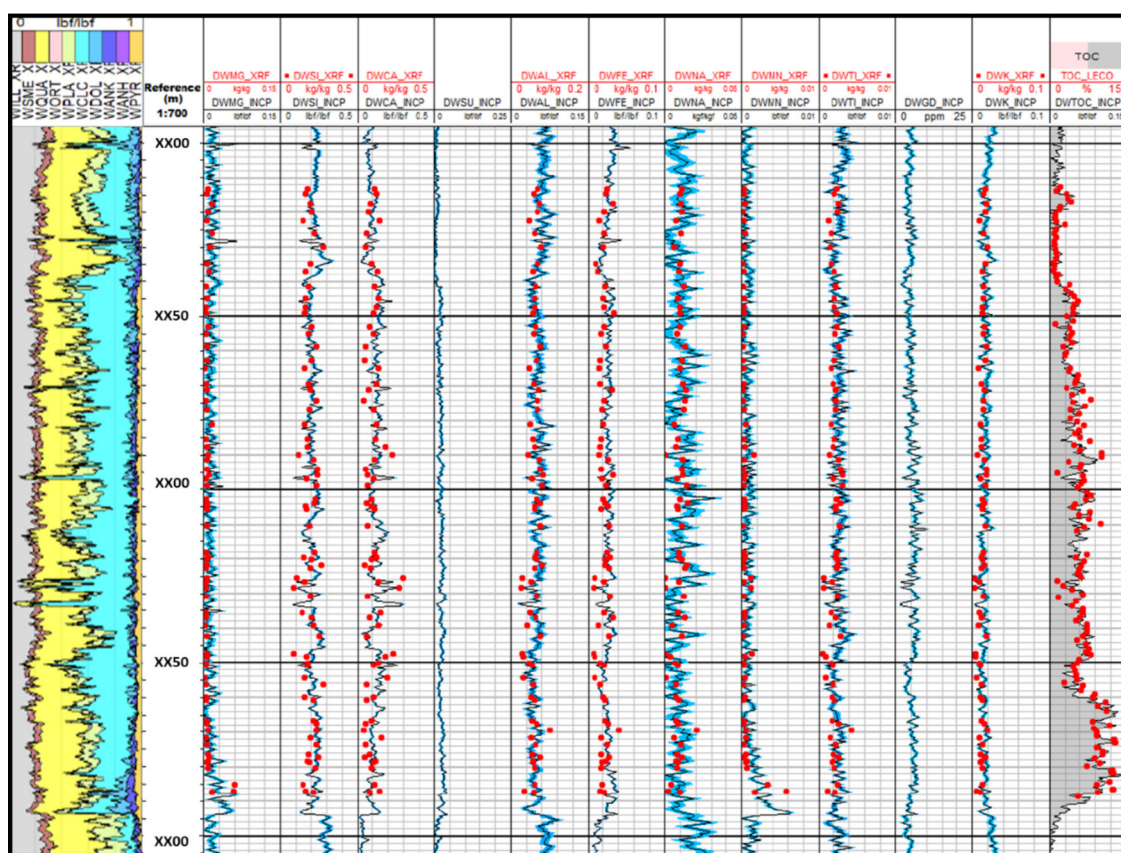


Figura 2. Elementos de laboratorio (FRX-Técnica Perla) vs elementos de Espectroscopia de captura e inelástico. En el track #1 mineralogía en % en peso seco estimada a partir de los elementos calibrados.

Posteriormente se continuó con la determinación de volúmenes minerales, porosidad y fracción fluidos, a través de un modelo probabilístico. Las variables de entrada son los minerales en % w/w de la etapa anterior, la resonancia magnética, densidad, neutrón y dieléctrico. Según la madurez termal en que se encuentra el pozo, se ajustaron los end point de cada componente.

Madurez (Ro)	Light Hydrocarbon				Bitumen				Kerogen				Light Heve Bit	Water		
	RHOZ_EC	HI_LIM_EC	MRP	DW_TC	RHOZ_EC	HI_LIM_EC	MRP	DW_TC	RHOZ_EC	HI_LIM_EC	MRP	DW_TC		RHOZ_EC	HI_LIM_EC	MRP
	Density (Density)	Neutron (HI)	MRP (m3/m3)	End Point DWTOT	Density (Density)	Neutron (HI)	MRP (m3/m3)	End Point DWTOT	Density (Density)	Neutron (HI)	MRP (m3/m3)	End Point DWTOT		Density (Density)	Neutron (HI)	MRP (m3/m3)
<1	0.97	XXxx	XXxx	1.20	XXxx	1.03	1.00	0.95	XXxx	XXxx	0.00	XXxx	1.992	1.13	0.92	1.00
	0.90	XXxx	XXxx	1.10	XXxx	1.03	1.00	0.95	XXxx	XXxx	0.00	XXxx	2.122	1.13	0.92	1.00
	0.83	XXxx	XXxx	1.01	XXxx	1.03	1.00	0.95	XXxx	XXxx	0.00	XXxx	4.321	1.13	0.92	1.00
>1.4	0.77	XXxx	XXxx	0.93	XXxx	1.03	1.00	0.95	XXxx	XXxx	0.00	XXxx	4.789	1.13	0.92	1.00
	0.72	XXxx	XXxx	0.85	XXxx	1.03	1.00	0.95	XXxx	XXxx	0.00	XXxx	5.347	1.13	0.92	1.00

Figura 3. Matriz con parámetros ajustados en función de la madurez termal

Los resultados de las interpretaciones fueron comparados con los datos de corona. La estimación de porosidad (PHIT_QE) se comparó en particular con las mediciones de porosidad obtenidas a partir de la técnica *Dean Stark* (GRI), en el cual, a diferencia del método de la retorta no presenta limitaciones relacionadas a la presencia de bitumen en la roca (Figura 4).

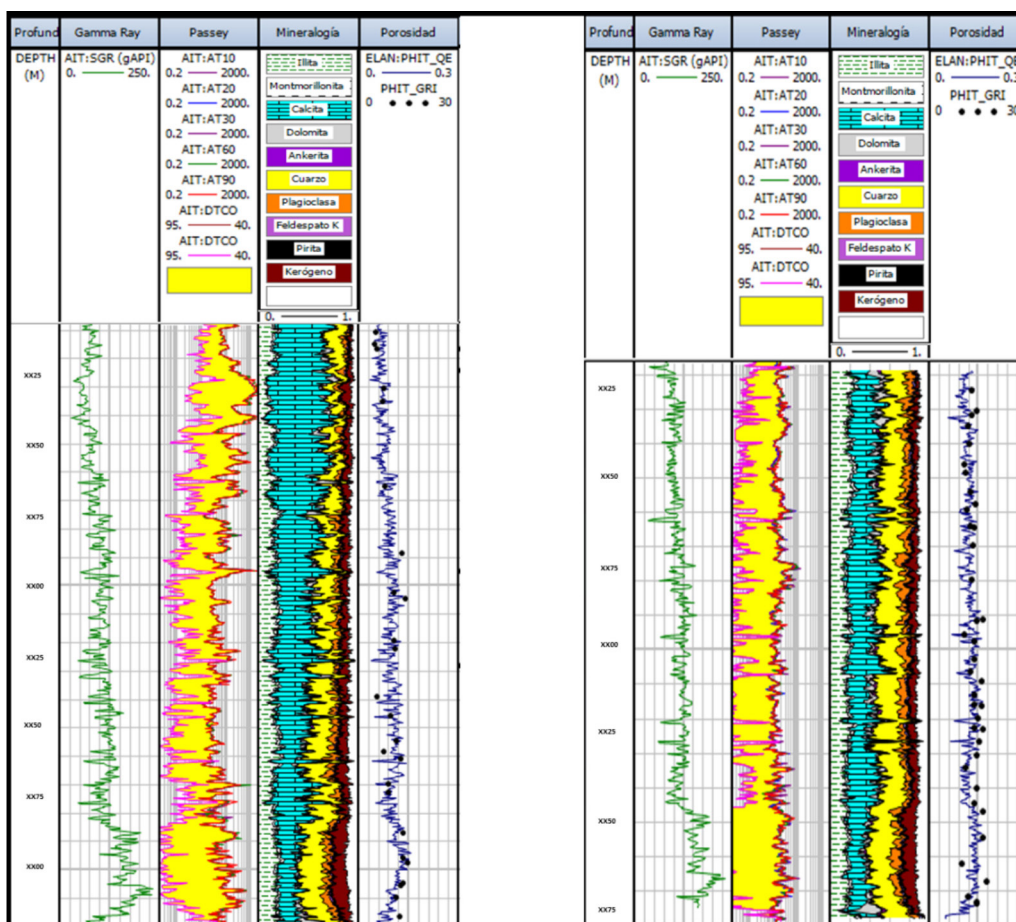


Figura 4. Ejemplo de 2 pozos en diferente ventana de madurez con la interpretación del modelo probabilístico. Track#1 Profundidad, Track#2 *Gamma Ray*, Track#3 DTCO vs Resistividad (cruce de Passey), Track#4 Volúmenes minerales, Track#5 Porosidad Total interpretada vs Porosidad Total de roca (PHIT_GRI)

Técnicas de Machine Learning

Fuente de Datos

El estudio se realiza sobre los datos disponibles para los 26 pozos evaluados con el modelo probabilístico. Las 15 variables consideradas inicialmente y la variable objetivo, curva de porosidad PHIT_QE resultado del modelo probabilístico.

Del *set* inicial de datos y luego de un trabajo de pre -procesamiento, que incluyó definición de topes, *cutoff* de valores, normalización de curva, completitud de los registros entre otros, el *set* final quedó comprendido por 11 variables y 39559 datos (Tabla 2).

Las variables iniciales son de registros eléctricos básicos y otros datos no relacionados con perfiles y conocidos para cualquier pozo (Ej: variable 1 a 5).

Variables Iniciales	Descripción	Unidad	Variables Finales
DEPTH	Profundidad del pozo	m	x
RT	Resistividad	ohm-m	x
Variable 1			x
Variable 2			x
DTCO	Sonico compresional	us/ft	x
NPHI	Neutrón	m3/m3	x
RHOZ	Densidad	g/cm3	x
Variable 3			x
Variable 4			x
Variable 5			x
NMR	Porosidad NMR	v/v	x
GR	Gamma Ray	API	
THOR	Torio	ppm	
URA	Uranio	ppm	
POTA	Potasio	%	
PHIT_QE	Porosidad estimada	m3/m3	x

Tabla 2. *set* de variables iniciales y finales con sus respectivas unidades de medición.

Análisis Exploratorio

Esta etapa del análisis de datos consiste en el estudio de los datos por separado y sus relaciones en conjunto, obteniendo como resultado un *set* de datos con calidad validada y con formato aceptable para alimentar modelos de machine learning.

El objetivo de esta etapa es entender los datos, validar su calidad y dejarlos listos para ser consumidos por el modelo de *machine learning*.

En la Figura 5 se visualizan las distribuciones de las variables que componen el estudio, donde algunos valores son constantes dentro del pozo.

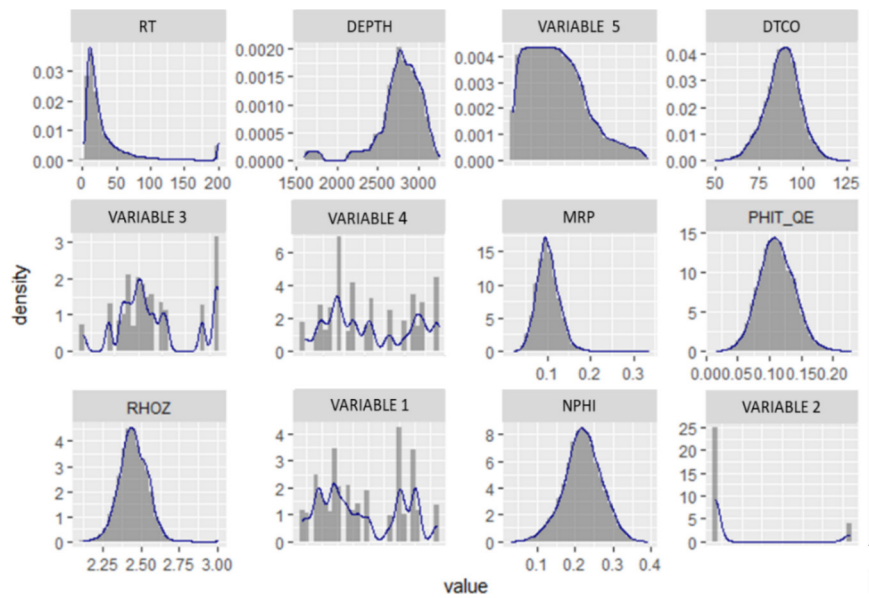


Figura 5. Distribuciones

A partir de la variable 2, que separa dos niveles en la formación VM, se distinguió y analizó la distribución de los 39559 datos por pozo (Figura 6).

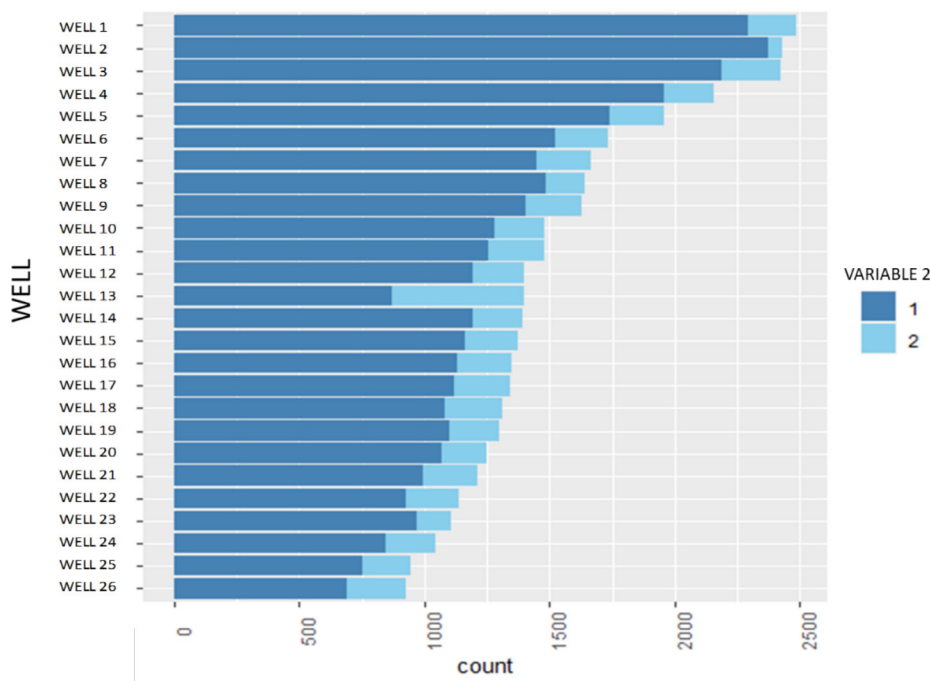


Figura 6. Observaciones por pozo

En la Figura 7 se observa el caso de la variable objetivo en cada nivel VM. La muestra es de mayor tamaño para el nivel 1, que tiene 34056 datos, mientras que para el nivel 2 el volumen de datos es menor con 5503 datos. En el gráfico de la derecha se grafican las medias con un intervalo de confianza de 95% para los dos valores de VM. Para VM=1 la media es 0.108 y para VM=2 es 0.126.

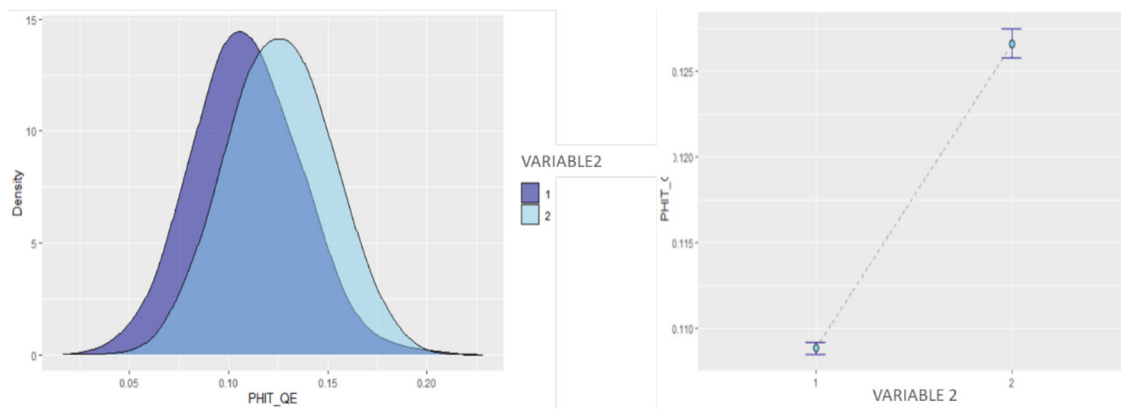


Figura 7. PHIT_QE según nivel de VM

En la Figura 8 se visualizan las distribuciones por pozo de alguna de las variables como PHIT_QE, NMR (porosidad de NMR), DTCO (onda compresional sónica) y RT (resistividad de la roca).

En el caso de las 3 primeras variables se observa que en solo 3 pozos (ubicados en un mismo sector) la media del nivel 2 es levemente menor a la del nivel 1, a diferencia del resto de los pozos.

Para la variable RT a nivel pozo no es tan clara la tendencia general por niveles.

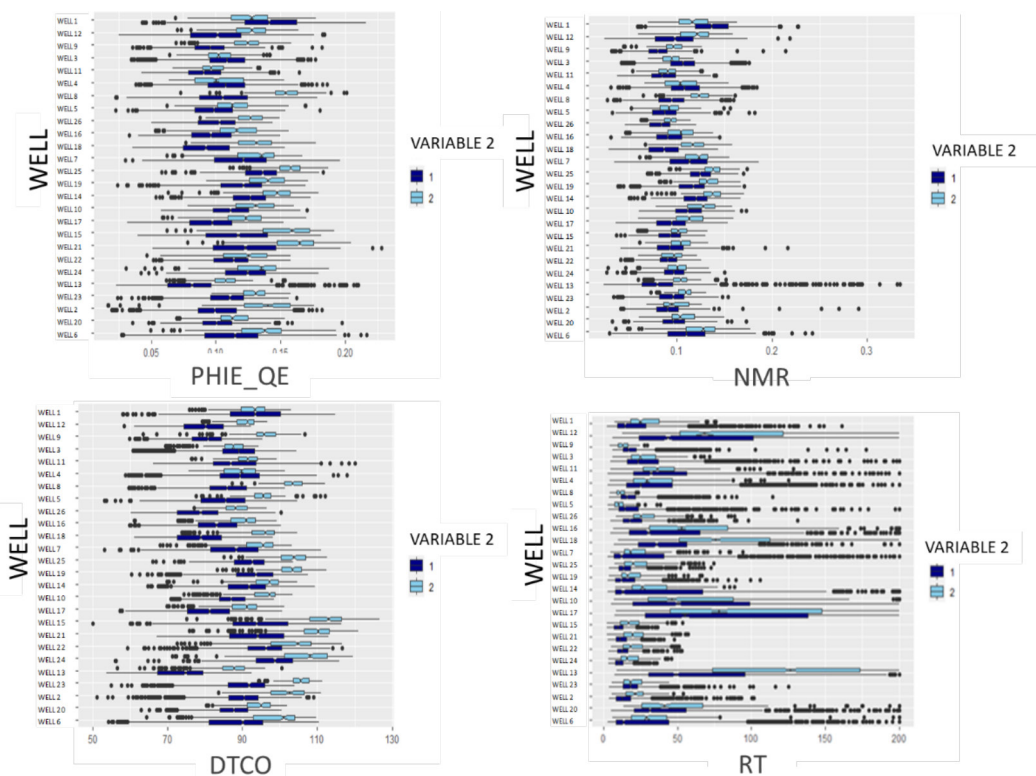


Figura 8. Distribución de PHIT_QE, NMR, DTCO y RT según nivel de VM.

Análisis Multivariable

Luego del análisis exploratorio se avanzó en el análisis multivariado, integrando todas las variables para estudiar sus relaciones y buscando como resultados conjuntos de datos con poder predictivo sobre la variable objetivo porosidad (PHIT_QE).

Según la disponibilidad de los perfiles para un número significativo de pozos y su poder predictivo, se seleccionaron 3 grupos de variables predictoras (Figura 9).

GRUPO 1	GRUPO 1_1	GRUPO 6
VARIABLE 2		
VARIABLE 5		
DEPTH	DEPTH	DEPTH
DTCO	DTCO	DTCO
VARIABLE 3	VARIABLE 3	
VARIABLE 4	VARIABLE 4	
RT	RT	
RHOZ	RHOZ	RHOZ
VARIABLE 1		VARIABLE 1
NPHI		NPHI
		NMR

Figura 9. Grupos de variables analizadas para datos de entrada de los modelos.

Para el Grupo 1 y Grupo 1_1 realizamos un análisis de correlaciones según el coeficiente de Pearson como se observa en la Figura 10. La variable con mayor correlación respecto a PHIT_QE es DTCO. Se observa que en el *set* no hay correlaciones superiores a 0.8 entre las variables. Analizando esas relaciones decidimos avanzar con este *set* para input de la construcción de modelos tanto lineales como no lineales.

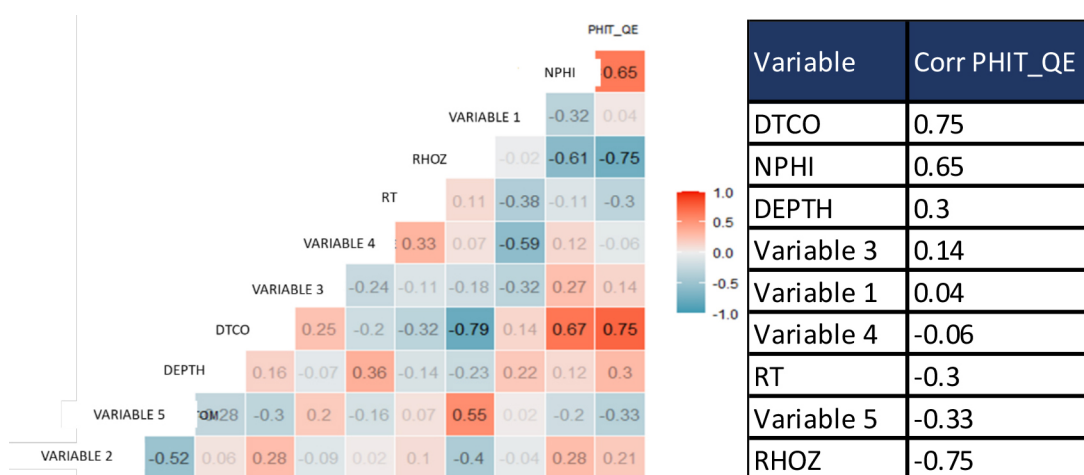


Figura 10. Correlaciones Grupo 1 y Grupo 1_1.

Complementariamente se realizó el mismo análisis, pero por separado para cada pozo en estudio con objetivo de entender las particularidades de cada uno relacionadas con su ubicación geográfica y sus espesores.

En la Figura 11 se visualiza las relaciones de las variables con más correlación con la variable objetivo. Vemos algunas observaciones con RHOZ cercanos a su máximo valor 3.00 que rompen el patrón de relación.

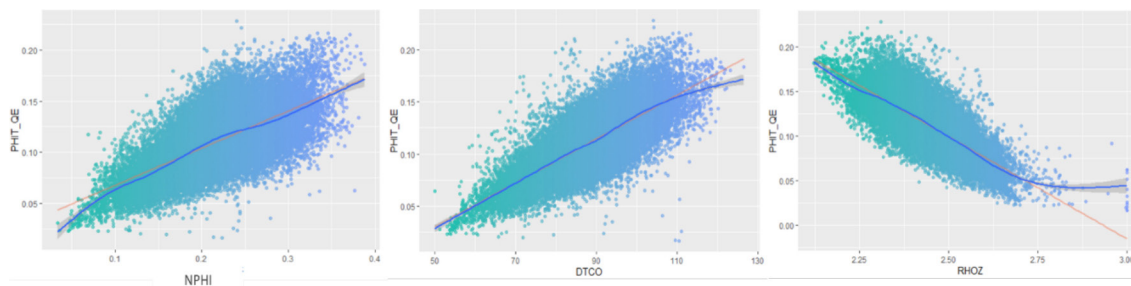


Figura 11. Relaciones entre variables Grupo 1 y Grupo 1_1.

Para entender de forma temprana el potencial predictivo de estos grupos de variables y, dado que el volumen de datos y las capacidades de procesamiento nos lo permite, decidimos realizar una prueba exhaustiva de las diferentes combinaciones de variables en modelos lineales. En la Figura 12 se visualizan para el GRUPO 1_1 el aumento de la medida de score R^2 Ajustado a 0.71, partiendo de un modelo univariado considerando DTCO, y finalmente incluyendo todas las variables disponibles. Las variables que más información aportan son el DTCO, RHOZ, DEPTH.

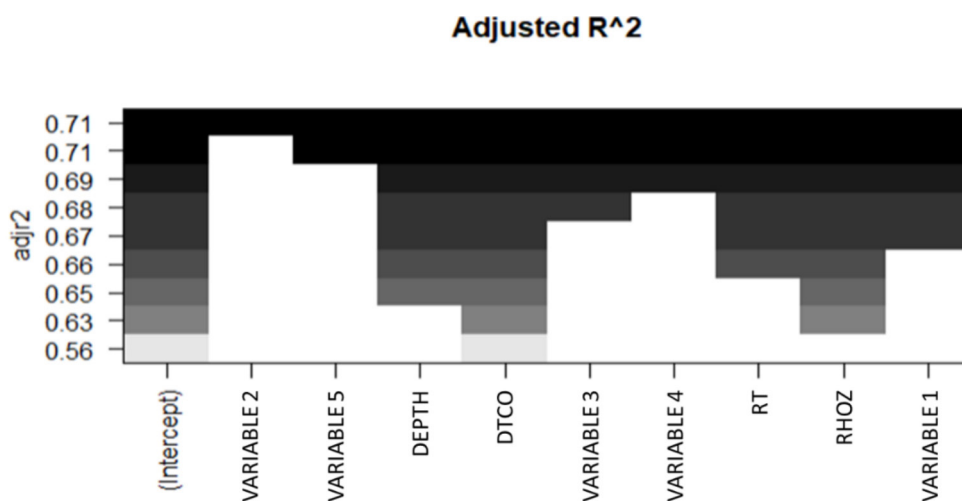


Figura 12. Prueba de combinaciones posibles para modelo lineal

Se realizó un análisis análogo para las variables que componen el GRUPO 6 observando que, como era esperado, NMR es la variable con mayor correlación con la variable objetivo (Figura 13). La variable NMR está disponible para un número menor de pozos.

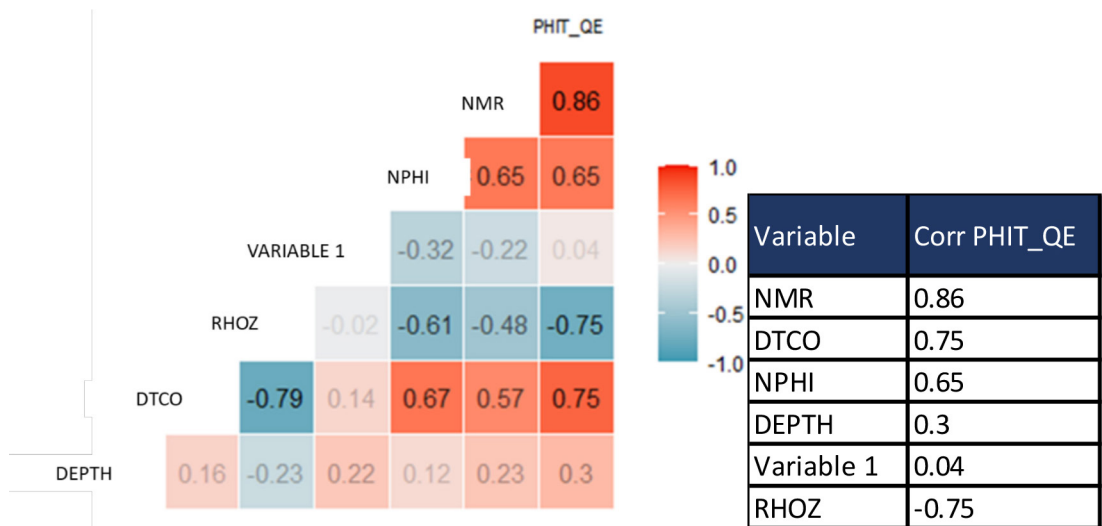


Figura 13. Relaciones entre variables Grupo 6.

En la Figura 14 visualizamos las relaciones de las variables exclusivas del GRUPO 6 más correlacionadas con la variable objetivo. Vemos algunas observaciones con NMR 0.2 y 0.3 que rompen el patrón de relación.

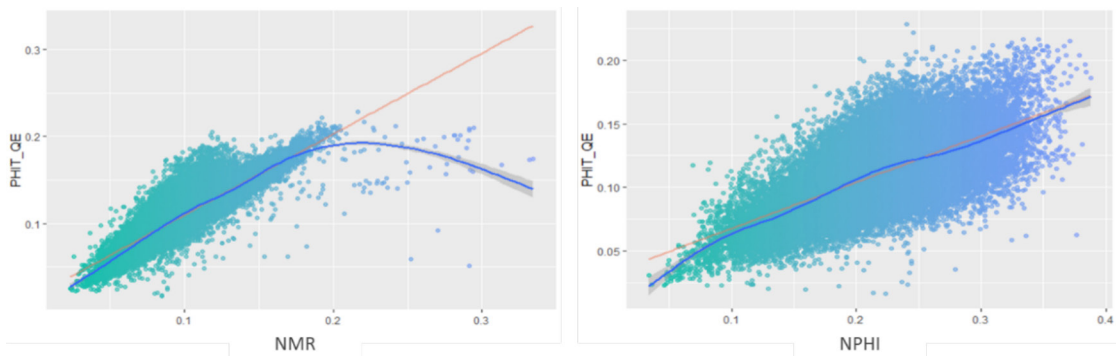


Figura 14. Relaciones entre variables Grupo 6.

Para entender el origen de los valores de NMR en el rango entre 0.2 y 0.3 visualizamos en la Figura 15 las distribuciones de NMR por pozo. Se visualiza claramente que uno de los pozos que mayor aporte hace en este sentido está ubicado en el extremo sur de la zona de estudio.

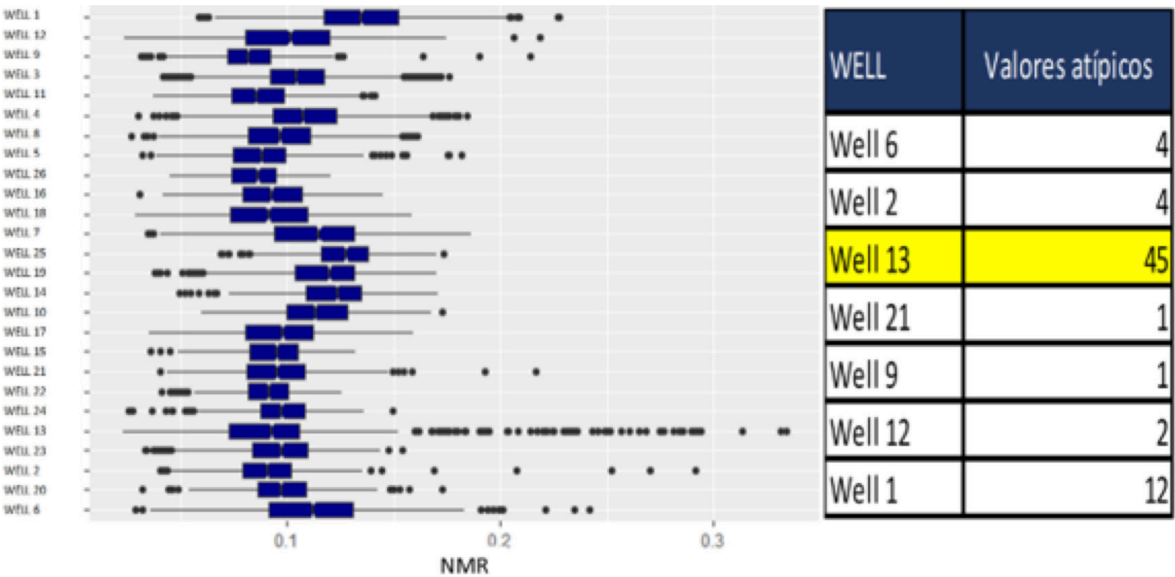


Figura 15. Valores atípicos de NMR.

- Consideración especial de linealidad para AT90

La variable RT muestra una relación log contra la variable objetivo. Por eso para utilizarla en modelos lineales se aplica la función $\log()$ como se muestra en la Figura 16. Esto no se hace para utilizarla en modelos no lineales.



Figura 16. RT y su relación con la variable objetivo.

Modelos Predictores

Una vez finalizado el análisis multivariado y teniendo una mayor comprensión del conjunto de datos se avanzó en la construcción de modelos predictivos para PHIT_QE con los 3 grupos de variables descriptos.

En un primer acercamiento se trabajó con modelos de regresión lineal por su simpleza para ver el efecto de las variables en la predicción. Luego se continuó con modelos más complejos priorizando la precisión por sobre la interpretabilidad. La efectividad de los modelos se mide dividiendo los datos al azar, utilizando el 80% de los datos como *set* de entrenamiento y el 20% restante, nunca observado por el modelo, como *set* de validación.

Como métricas de regresión decidimos utilizar el RMSE y MAPE.

RMSE es la raíz cuadrada del error cuadrático medio, que nos da una idea del error de predicción en la escala de la variable objetivo.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicho_i - Real_i)^2}{N}}$$

MAPE es el error porcentual absoluto medio, nos da una idea del error en proporciones.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Predicho_i - Real_i}{Real_i} \right|$$

Finalmente se realizó una validación del tipo “*leave one out*” sobre los modelos elegidos para los tres grupos con el fin de detectar el poder predictivo para cada pozo por separado, entrenando con el resto de los pozos.

Como resultado, las predicciones puntuales, serán acompañadas de un límite inferior y superior de un intervalo de confianza del 90% calculado con técnicas de estadística robusta.

- Modelos de regresión lineal

Una vez separado el *set* de datos se entrenan modelos lineales multivariable con el 80% de los datos. Se evalúa la performance del modelo con el 20% de observaciones reservadas para validación. En la Figura 17 se observa el contraste entre valores reales y predichos de la variable objetivo para los grupos de variable GRUPO 1 y GRUPO 1_1.

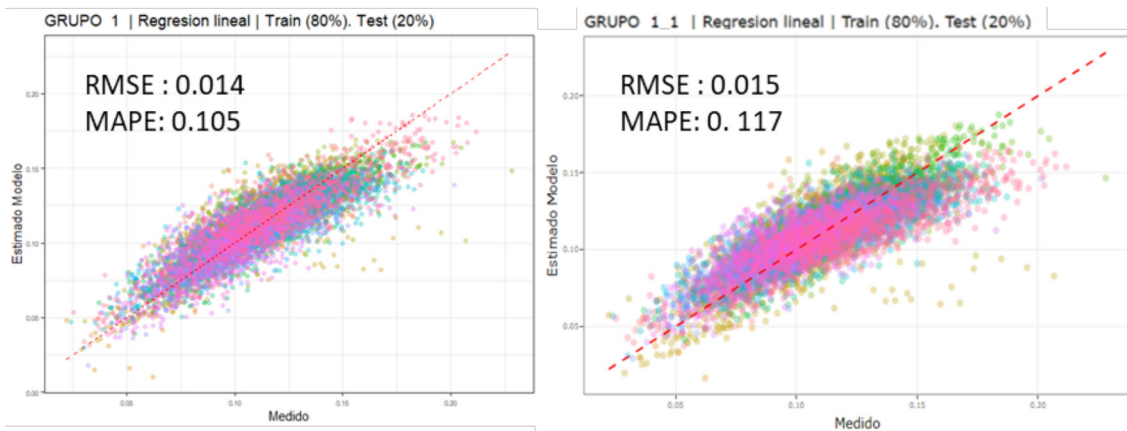


Figura 17. Modelo Lineal para GRUPO 1 & GRUPO 1_1. Valores estimados vs reales.

En la Figura 18 se observa, como era esperado, que el modelo del GRUPO 6 tiene un mayor poder predictivo que los anteriores debido a la variable NMR.

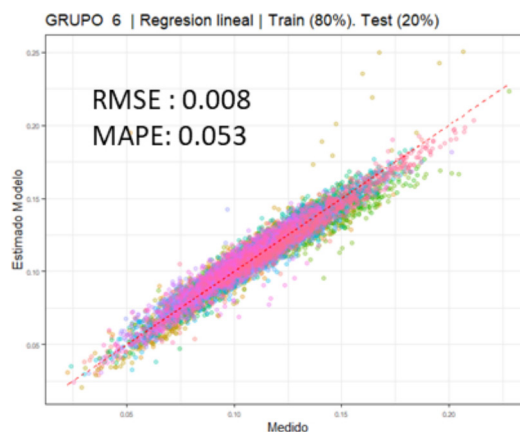


Figura 18. Modelo Lineal para GRUPO 6. Valores estimados vs reales.

Las predicciones con mayor error corresponden al pozo *Well* 13, ubicado más al sur de la zona de estudio. La predicción del modelo del GRUPO 1, el modelo predice por debajo del real con PHIT_QE mayores a 0.17, mientras que para el GRUPO 6 el modelo predice valores mayores al real con PHIT_QE mayores a 0.21.

En la Figura 19, se muestra un gráfico xplot (PHIT_QE vs RHOZ) donde se puede explicar esos errores para el pozo *Well* 13. Si bien se mantiene la regla de correlación inversa entre ambas variables, hay valores de PHIT_QE altos con respecto a la generalidad de los pozos.

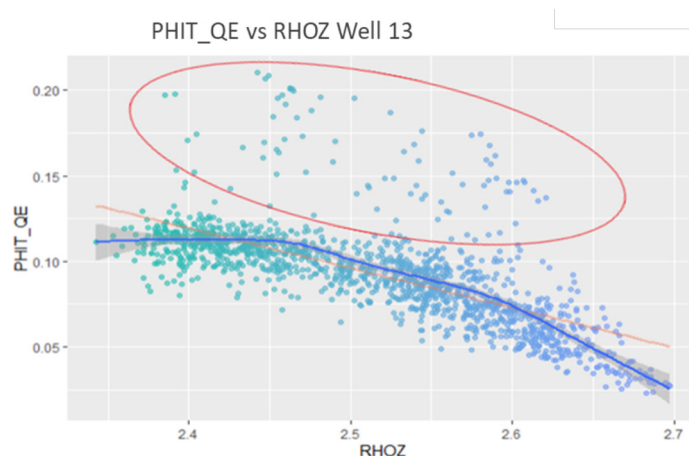


Figura 19. Valores atípicos de PHIT_QE vs RHOZ en el pozo Well 13

Las tablas a continuación muestran las variables que aportan la mayor información para la predicción, en orden de relevancia, en las predicciones.

Variable	Estadístico t	Variable	Estadístico t
RHOZ	84.02523	NMR	350.609631
DEPTH	63.40928	RHOZ	167.875136
log (RT)	57.94175	Variable 1	104.408505
Variable 3	54.5676	DTCO	3.440832
Variable 1	50.97019	NPHI	1.955348
Variable 4	45.1410500		
DTCO	34.9331700		
Variable 5	28.0571100		
NPHI	26.5714500		
Variable 2	17.45014		

Tabla 3. Valores - Importancia de las variables.

- Modelos de regresión no lineal

Se consideraron en este trabajo distintos algoritmos para la construcción de modelo no lineales. Comenzando desde los árboles binarios y avanzando con random forest y técnicas de *boosting*. A medida que se complejiza el modelo se obtiene una ganancia en el poder predictivo. Finalmente, los modelos seleccionados son de tipo ensamble GBM por ser los que lograron mayor precisión.

A continuación, se muestran los gráficos xplot con el contraste entre lo predicho por el modelo (80% de los datos) y lo real (conjunto de test que son el 20 % de las observaciones).

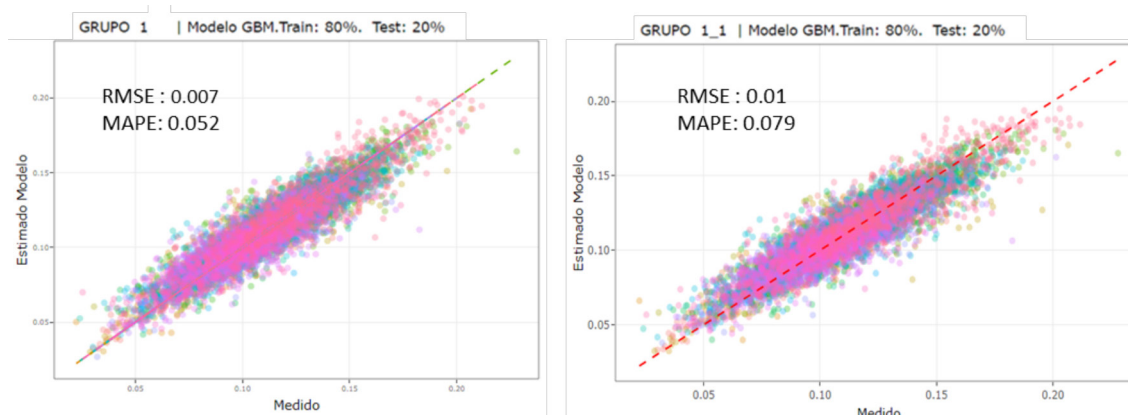


Figura 20. Modelo No Lineal para GRUPO 1 & GRUPO 1_1. Valores estimados vs reales

En la Figura 20 se visualiza que el modelo final para el GRUPO 1 es levemente mejor en cuanto a la precisión que el modelo construido con el GRUPO 1_1. Ambos modelos mejoran lo logrado con el modelo lineal en cada caso.

En la Figura 21 se observa el resultado de evaluar el modelo no lineal construido con el GRUPO 6 en el *set* de Test.

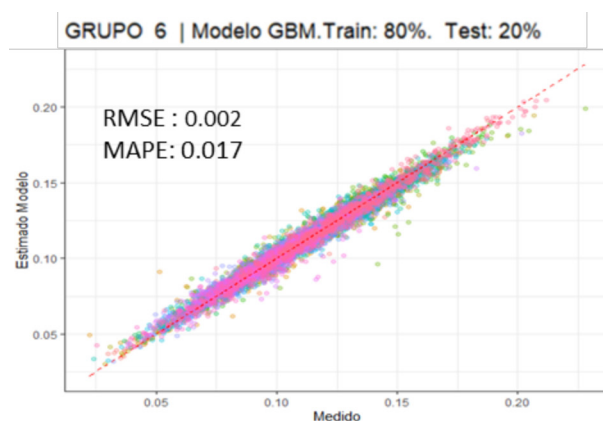


Figura 21. Modelo No Lineal para GRUPO. Valores estimados vs reales.

Una vez construidos los modelos analizamos la importancia de las variables. Dado que este tipo de modelos no son simples de interpretar como los lineales, nos valemos de una técnica proveniente de la teoría de juegos cooperativos, llamada *SHapley Additive exPlanations* (SHAP).

Con esta técnica logramos explicar los aportes de las variables y sus valores a la predicción final según lo aprendido por el modelo. En la Figura 22 vemos un ejemplo de gráfico de tornado, para el modelo del GRUPO 1_1, que nos muestra la importancia de las variables ordenadas de arriba hacia abajo y su efecto en la predicción según valores de la variable.

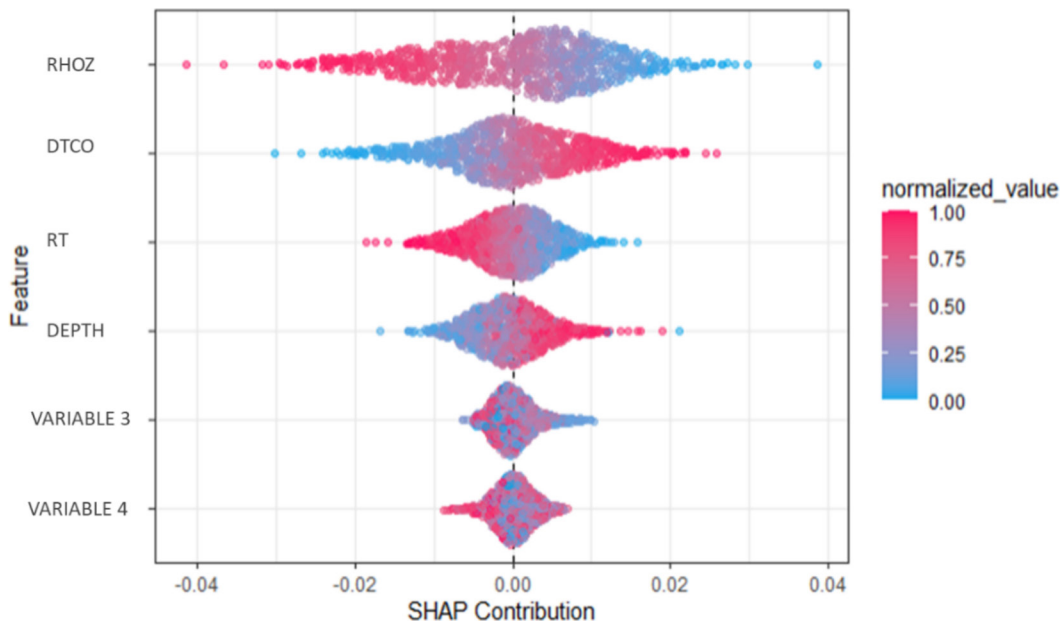


Figura 22. Shap Values para modelo GBM GRUPO 1_1

Así vemos que, por ejemplo, el modelo aprendió que valores de RHOZ bajos contribuyen de forma positiva a la porosidad.

- Validación *Leave one out*

La validación del tipo “*leave one out*” se refiere, en este caso, a entrenar un modelo con 25 pozos de 26 donde tenemos grupos de observaciones y validarlo con el pozo restante que fue dejado fuera del conjunto de entrenamiento. De esta manera, para cada grupo de variables, se construyen 26 modelos y se verifican los resultados. El objetivo es identificar con qué precisión es posible estimar un pozo con los restantes y de esta forma detectar pozos con condiciones particulares.

WELL	RMSE GRUPO 1	RMSE GRUPO 1_1	RMSE GRUPO 6
Well 6	0.015	0.015	0.007
Well 20	0.016	0.016	0.004
Well 2	0.014	0.015	0.008
Well 23	0.014	0.018	0.012
Well 13	0.021	0.020	0.008
Well 24	0.014	0.014	0.008
Well 22	0.014	0.018	0.008
Well 21	0.018	0.015	0.012
Well 15	0.013	0.013	0.006
Well 17	0.013	0.013	0.004
Well 10	0.013	0.013	0.005
Well 14	0.013	0.013	0.004
Well 19	0.016	0.016	0.008
Well 25	0.019	0.014	0.007
Well 7	0.013	0.014	0.005
Well 18	0.012	0.012	0.004
Well 16	0.014	0.014	0.005
Well 26	0.013	0.010	0.007
Well 5	0.015	0.016	0.006
Well 8	0.016	0.014	0.009
Well 4	0.014	0.016	0.005
Well 11	0.017	0.021	0.007
Well 3	0.012	0.013	0.007
Well 9	0.014	0.016	0.006
Well 12	0.017	0.015	0.004
Well 1	0.026	0.017	0.005

Tabla 4. Validación LOO (*Leave one out*)

En la Tabla 4 se visualizan los resultados de la validación *leave one out*. En rojo se resaltan las mayores discrepancias entre variables a predecir y el valor predicho. Para los modelos de GRUPO 1 los pozos peor predichos son un pozo al norte y otro en el extremo sur del área analizada. Esto podría indicar que para este grupo de variables esos pozos son particulares y el muestreo no tiene el peso en el *set* de datos para poder aprender sus características.

Para los modelos del GRUPO 6 los pozos peor predichos son *Well 21* y *Well 23*. Dos pozos del centro del mapa.

Uno de los mejores predichos por los modelos para GRUPO 6 es el pozo *Well 1* que había sido el peor para GRUPO 1.

En promedio, los modelos para GRUPO 6 mejoran a los de GRUPO 1 por $0.008 \text{ m}^3/\text{m}^3$.

Considerando los modelos creados para los dos grupos de variables, el pozo que parece ser más difícil de predecir es *Well* 21, pero esto no quiere decir que el modelo entrenado anteriormente para GRUPO 6 no haya logrado entender este pozo.

Intervalos de confianza

Apoyándonos en técnicas de estadística robusta buscamos calcular un intervalo de confianza para cada predicción.

Esto es, dada una muestra $y_1 \dots y_n$ todas con distribución F , encontrar un intervalo $I = [a, b]$ tal que si y_0 (observado) tiene distribución F entonces $P(a \leq y_0 \leq b) = 1 - \alpha$ donde α es la confianza.

El intervalo de confianza se construye en base a un análisis de los errores de las predicciones para los 26 pozos originales utilizados para construir el modelo.

En la Figura 23 se visualiza el intervalo de confianza para una porción de un pozo predicho con el modelo para GRUPO 6.

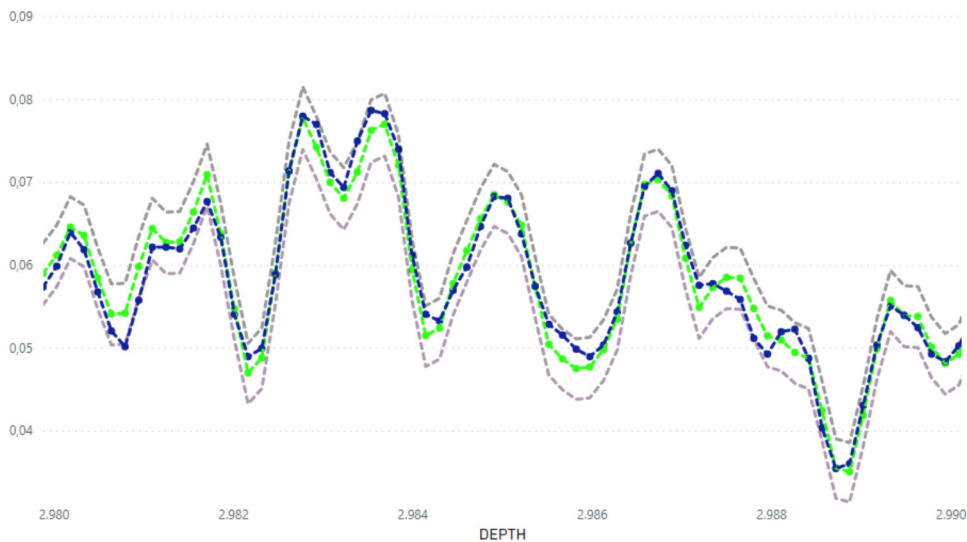


Figura 23. Intervalo de confianza para las predicciones.

CONCLUSIONES

Con técnicas de Machine Learning y un *set* limitado de registros se puede predecir con robustez la variable PHIT.

Como resultado del trabajo se obtuvieron 3 modelos a utilizar según datos disponibles. Cada modelo permite predecir PHIT aún con baja calidad de perfiles adquiridos, sin significativamente perder el poder predictivo, logrando precisión y consistencia en las predicciones.

En el conjunto de validación los modelos mostraron en promedio un error cercano al 5%, para el GRUPO 1, 8% para el GRUPO 1_1 y menor al 2% para GRUPO 6.

Se observaron algunas relaciones entre los valores extremos en las variables y su ubicación geográfica, lo cual es esperable debido a variaciones laterales de condiciones geológica. A consecuencia de esto se logró una menor predicción en aquellos pozos ubicados en los extremos Norte y Sur del área de estudio.

La robustez de los resultados permite adoptar *Data Analytics* como herramienta de mucha utilidad para los especialistas y además reducir los costos destinados a la adquisición de nuevos datos y ampliando de manera significativa la disponibilidad de datos para el mapeo a partir de la inclusión de pozos con *set* limitado de perfiles adquiridos.

AGRADECIMIENTOS

Los autores quieren agradecer a Juan Martín Elorriaga, Gabriel Horowitz y Ricardo Maronna (YPF) por su colaboración en el proyecto. A Pablo Saldungaray (Schlumberger), Mariela Silka, Mauro Sanchez y Lourdes Vera (YPF) por el soporte técnico para validar los resultados. A Guillermina Sagasti por la revisión del trabajo. También a YPF por permitir la publicación del trabajo.

REFERENCIAS CITADAS

- Bernhardt, C., Ortiz, A. y González Tomassini, F. 7-11 de agosto, 2017. Interpretación Petrofísica de detalle en la Formación Vaca Muerta a partir de datos de testigo de roca y perfiles de última generación. El Problema de la calibración roca/perfil. XX Congreso Geológico Argentino.
- Cuervo, S., Lombardo E., Vallejo D., Crousse L., Hernandez C., Mosse L. 2016. Towards a Simplified Petrophysical Model for the Vaca Muerta Formation. In UNCONVENTIONAL Resources Technology Conference: URTeC: 2437257
- Mosse, L., Rodriguez, L. M., Chiapello, E., Lambert, L., Leduc, J.P., Decoster, E., Grau, J., Horkowitz, J., 2014. Accurate lithology and TOC obtained from an induced gamma ray spectroscopy tool. Two major challenges in *shale* reservoirs. In: Simposio Evaluación de Formaciones. XIX Congreso Argentino de Exploración y Desarrollo de Hidrocarburos IAPG. Mendoza, Argentina.
- Ortiz, Alberto César, Luisa Crousse, Carolina Bernhardt, Dolores Vallejo, and Laurent Mosse, 2020, Reservoir properties: Mineralogy, porosity, and fluid types, in Daniel Minisini, Manuel Fantín, Iván Lanusse Noguera, and Héctor Leanza, eds., Integrated geology of unconventional: The case of the Vaca Muerta play, Argentina: AAPG Memoir 121, p. 329–350.
- Rodriguez, L., Chiapello, E., Lambert, L., Leduc, J., Mosse, L., Sanchez, M., 2015. Quantitative and Comparative Evaluation of Mineralogy and TOC Analysis from Cores, Cuttings and Logs in Vaca Muerta Unconventional Play. SPE-178-MS/URTeC:2172972