

Análisis multivariado para la predicción de fluidos a partir de la mineralogía de yacimientos no convencionales

Por **Diego Gallart** (Practia), **Alberto Cesar Ortiz**, **Carolina Bernhardt** y **Damián Hryb** (YPF); **Julio César Rodríguez Martino** (Practia) y **Gabriel Horowitz** (Y-TEC).

Este trabajo fue presentado en las *VI Jornadas de Geotecnología*, en el marco del *10° Congreso de Exploración y Desarrollo de Hidrocarburos del IAPG* (Noviembre de 2018, Mendoza).

La importancia relativa de las variables en la predicción en algunos casos coinciden con los modelos físicos existentes y, en otros casos, se replantean suposiciones de esos modelos que aportan nuevo conocimiento a este tipo de análisis.



Metodología

Origen de datos

Para este estudio se dispone de un total de 15 variables medidas a distintas profundidades para un pozo NOC perteneciente a YPF.

Las variables analizadas corresponden a la mineralogía expresada en volúmenes obtenida a partir de un modelo petrofísico y los volúmenes poro/fluido de NMR modo T1T2 (Tabla 1).

Grupo	Descripción
Profundidad	Delta 15 cm
M (Mineralogía)	10 variables normalizadas junto con fluidos
VPGM (Volúmenes porales T1T2 agrupados x mapa), VP (Volúmenes porales) y FL (Fluidos)	5 variables

Tabla 1. Grupos de variables.

El set de datos comprende 1296 observaciones tomadas a distintas profundidades con un delta de 15 centímetros.

Análisis exploratorio

A continuación, se muestran los análisis de los diferentes volúmenes poro/fluidos tomados del procesamiento de la resonancia magnética nuclear (NMR) modo T1T2.

Volumen Poro/Fluido T1/T2_WAT

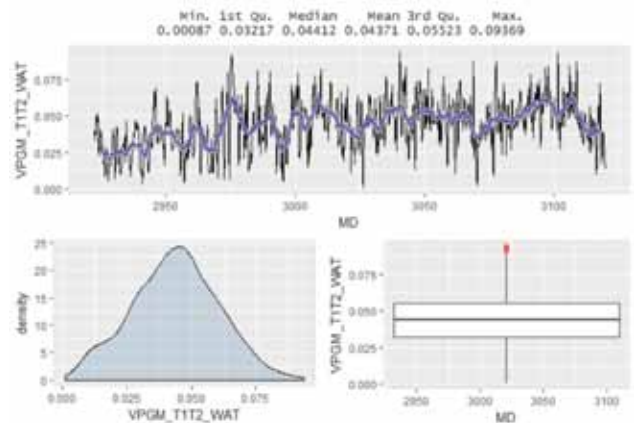


Figura 1. Descripción variable T1T2_WAT.

En la figura 1 vemos los valores de la variable medidos a lo largo de la profundidad del pozo. La línea azul representa el promedio móvil suavizado.

Abajo a la izquierda, el gráfico de distribución de densidad para la variable. Abajo a la derecha el box-plot que muestra puntos rojos en la parte superior. Esos son valores alejados del normal de la población. Son los tres picos que pueden observarse en el gráfico superior y son valores medidos válidos.

Volumen Poro/Fluido T1/T2_HC1

En la figura 2 vemos los valores de la variable medidos a lo largo de la profundidad del pozo. La línea azul representa el promedio móvil suavizado.

La caracterización de los espacios porales y volúmenes de fluidos a partir de datos de T1T2 NMR en un reservorio no convencional permite el entendimiento de sus características a escala nanométrica. Para que estas señales adquieran sentido y relevancia en la predicción de *sweet spots* o lugares de interés hidrocarbúfero es necesario analizarlas en relación con la mineralogía y el contenido de materia orgánica (kerógeno) de la roca.

Las técnicas de análisis multivariado utilizadas se han orientado en esta dirección, lo que ha permitido tener en cuenta la influencia de cada parámetro en los demás, revelando correlaciones que serían muy difícil descubrir y analizar con los métodos tradicionales.

Este trabajo se realizó en distintas etapas: se comenzó por estudiar las correlaciones entre diferentes variables mineralógicas, materia orgánica, de fluidos y porosidad NMR T1T2 en un pozo NOC perteneciente a YPF; se estudió el grado de correlación entre variables, eliminando aquellas redundantes y se utilizaron cuatro modelos de aprendizaje automático para predecir las variables objetivo donde el menor error obtenido en la predicción fue del 10%.

A continuación, se analizó la importancia relativa de las variables en la predicción, que en algunos casos coinciden con los modelos físicos existentes y, en otros, replantean algunas suposiciones y aportan nuevo conocimiento a este tipo de análisis.

Los modelos creados con los datos del pozo analizado pueden ser probados con datos de otros pozos para medir su eficacia y de esta manera tornarse en una herramienta predictiva de las características porales y de fluidos de reservorios tipo *shale*.

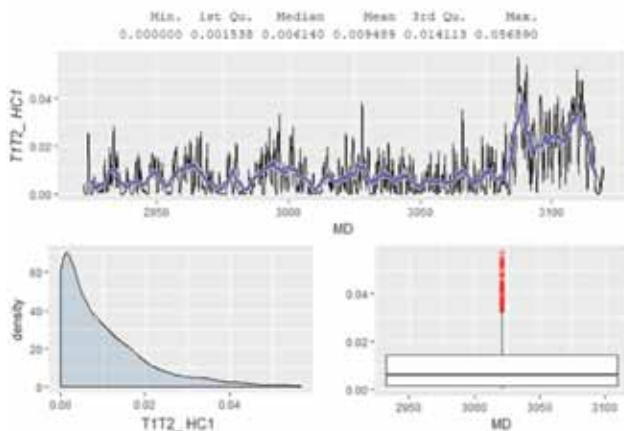


Figura 2. Descripción variable T1T2_HC1.

Abajo a la izquierda, el gráfico de distribución de densidad para la variable. Abajo a la derecha, el box-plot que muestra puntos rojos en la parte superior. Esos son valores alejados del normal de la población. En el gráfico superior, esos puntos rojos, corresponden a los valores de mayor profundidad que se incrementan saliéndose de lo normal a menor profundidad.

Volumen Poro/Fluido T1/T2_IPW

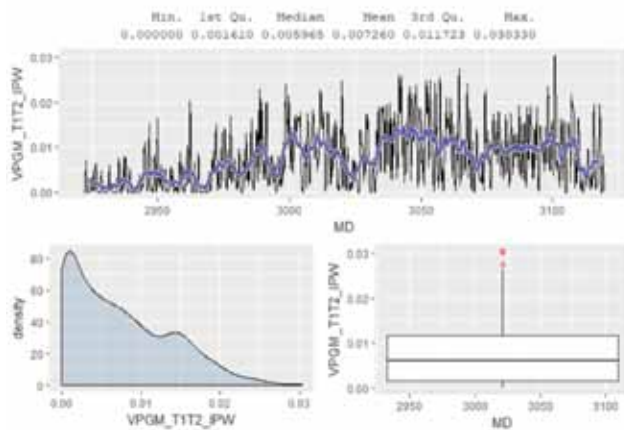


Figura 3. Descripción variable VPGM_T1T2_IPW.

En la figura 3 vemos los valores de la variable medidos a lo largo de la profundidad del pozo. La línea azul representa el promedio móvil suavizado.

Abajo a la izquierda, el gráfico de distribución de densidad para la variable. Abajo a la derecha, el box-plot que muestra puntos rojos en la parte superior. Esos son valores alejados del normal de la población. Son los picos que pueden observarse en el gráfico superior por sobre 0,025 y son valores medidos válidos.

Volumen Poro/Fluido T1/T2_HC3

En la figura 4 vemos los valores de la variable medidos a lo largo de la profundidad del pozo. La línea azul representa el promedio móvil suavizado.

Abajo a la izquierda, el gráfico de distribución de densidad para la variable. Abajo a la derecha, el box-plot. Vemos que, aunque el gráfico superior tiene picos, no llegan a ser valores atípicos.

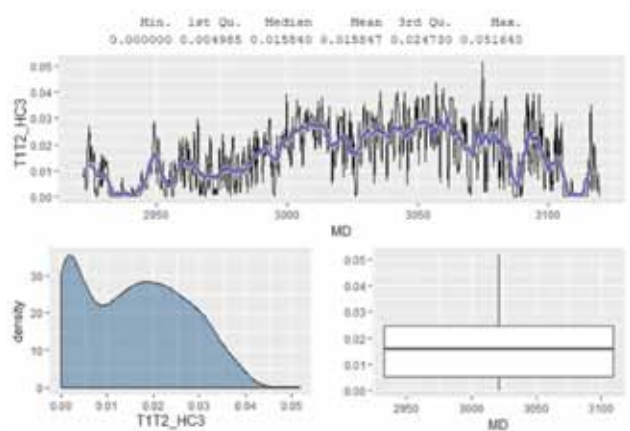


Figura 4. Descripción variable T1T2_HC3.

Volumen fluidos Interpretación – WATER

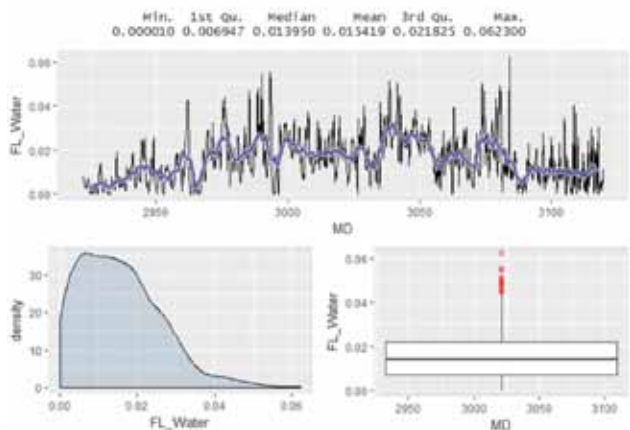


Figura 5. Descripción variable WATER.

En la figura 5 vemos los valores de la variable medidos a lo largo de la profundidad del pozo. La línea azul representa el promedio móvil suavizado.

Abajo a la izquierda el gráfico de distribución de densidad para la variable. Abajo a la derecha el box-plot que nos muestra puntos rojos en la parte superior. Esos son valores alejados del normal de la población. Son los picos que pueden observarse en el gráfico superior por sobre 0,045 y son valores medidos válidos.

Análisis multivariado

Mineralogía

Entre las variables de mineralogía vemos que calcita, illita y cuarzo se encuentran altamente correlacionadas. Calcita correlaciona negativamente con Illita y cuarzo con el mismo coeficiente (-0,85).

Para visualizar las relaciones entre las variables de mineralogía, aplicaremos componentes principales y graficaremos sobre las dimensiones resultantes (Figura 6).

La Figura 7 muestra las correlaciones entre las variables y su contribución respecto de la variabilidad de las componentes.

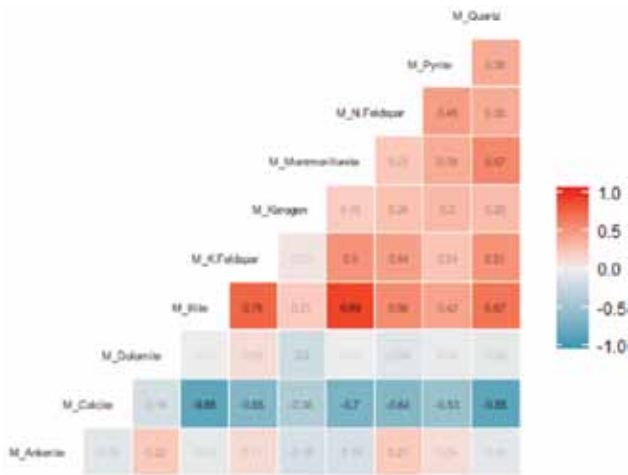


Figura 6. Correlaciones Mineralogía (Pearson).

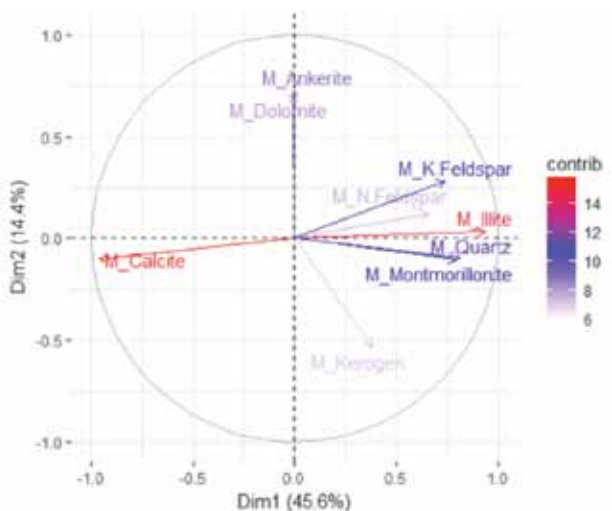


Figura 7. Componentes principales, mineralogía (Dim1- Dim2).

En cuanto a las correlaciones, el ángulo entre los vectores de las variables indica su correlación. Si describen un ángulo de 90°, no están correlacionadas. Si el ángulo es cercano a los 180°, están correlacionadas negativamente.

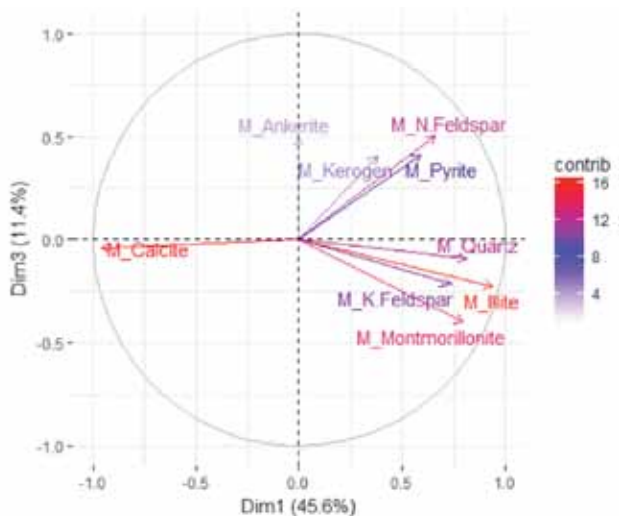


Figura 8. Componentes principales, mineralogía (Dim1- Dim3).

Si el ángulo es cercano a 0°, entonces están positivamente correlacionados.

La distancia al origen (o cercanía a la circunferencia exterior) del gráfico indica cómo son representadas las variables por estas componentes (Dim1 y Dim2). Cuánto más lejos del origen mejor se representadas.

Las dos dimensiones del gráfico explican el 45,6% + 14,4% = 59,5% de la variabilidad de los datos originales.

Al observar las proyecciones sobre la Dim1 vemos que calcita está correlacionada negativamente con la Dim1 mientras que K. Feldspar, illita, cuarzo y montmorillonita correlacionan positivamente entre ellas y con la Dim1. Al ver las mismas proyecciones notamos que estas variables son las que contribuyen mayormente a la Dim1 y que Ankerite, Dolomite y Kerogen son las que más contribuyen a la Dim2 (Figura 8).

La Dim3 aporta el 11,4% de explicación de variabilidad que sumado a lo anterior (un 59,5%) da un 70,9%.

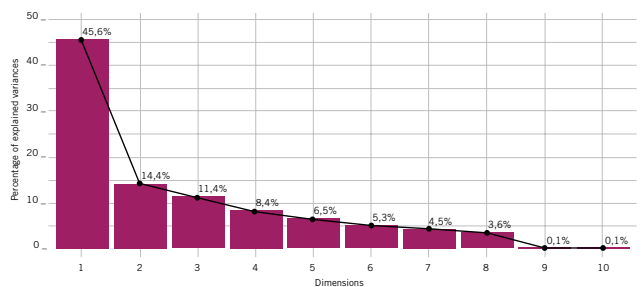


Figura 9. Varianza explicada por cada Dim de PCA.

En la figura 9 vemos que los primeros tres componentes del PCA (Dim1, Dim2, Dim3) explican el 71,4% de la varianza (45,6% + 14,4% + 11,4%).

Análisis de correlaciones entre variables objetivo y las variables de mineralogía (M)

La variable VPGM_T1T2_WAT correlaciona negativamente con Calcite (-0,61), positivamente con Illite (0,62), luego positivamente con Montmorillonite y Quartz (0,55).

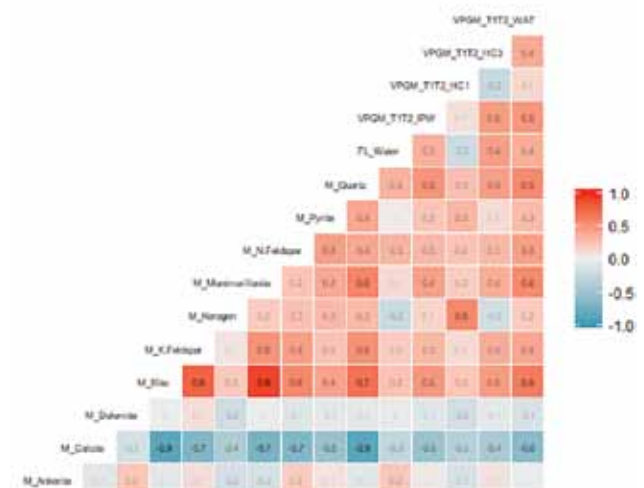


Figura 10. Correlaciones entre las variables objetivo y las variables de mineralogía.

La variable objetivo VPGM_T1T2_HC1 correlaciona negativamente con Calcite (-0,35), positivamente con Kerogen (0,57), luego positivamente con Pyrite y Quartz (0,33 y 0,29, respectivamente).

La variable objetivo VPGM_T1T2_IPW correlaciona negativamente con Calcite (-0,51), positivamente con Illite (0,47), luego positivamente con Montmotillonite y Quartz (0,42 y 0,49, respectivamente).

La variable objetivo VPGM_T1T2_HC3 correlaciona negativamente con Calcite (-0,45), positivamente con Illite (0,46), luego positivamente con K.Feldspar, Montmotillonite y Quartz (0,36, 0,36 y 0,44, respectivamente).

La variable objetivo FL_Water correlaciona negativamente con Calcite (-0,28) y Kerogen (-0,22), positivamente con Illite (0,25), luego positivamente con N. Feldspar y Quartz (0,32 y 0,33, respectivamente) (Figura 10).

Modelos predictores

Para encontrar el mejor modelo predictor y las variables más importantes dividiremos el set de datos (observaciones) disponibles en dos lotes: el 80% de los datos se utilizará para entrenar el modelo y el 20% restante para validarlo.

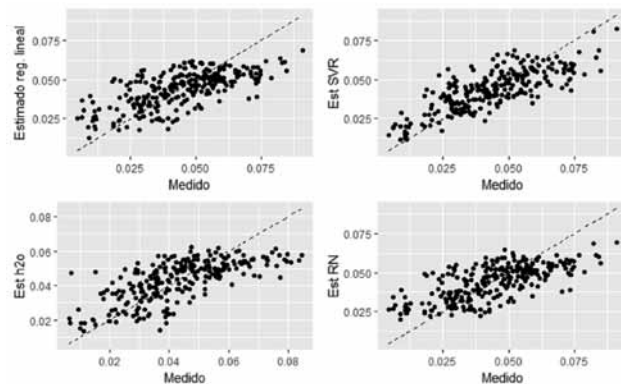
Como modelo más básico utilizaremos regresión lineal múltiple. Luego realizaremos modelos *Support Vector Regression* y redes neuronales. Finalmente, probaremos la performance de diferentes modelos predictivos combinados de forma óptima en una técnica de aprendizaje automático. El resultado será un conjunto de modelos agrupados llamado "ensamble apilado" (*stacked ensemble*).

Resultados

En la figura 11 se muestra el resultado de cada modelo en contraste con los demás. El valor estimado *versus* el valor real medido de la variable. La línea punteada representa la correspondencia sin error de estimado y real. En la tabla se detalla el RMSE de cada modelo (Raíz del error cuadrático medio) que está en la unidad de la variable y el porcentaje o proporción del rango de la variable que eso representa

$$\frac{\text{RMSE}}{(\text{VALmax}-\text{VALmin})}$$

Predecir T1T2_WAT a partir de los volúmenes mineralógicos



RMSE			
Reg. Lineal	SVR	RN	S. Ensemble
0,0129	0,0102	0,0127	0,0117
13,89%	11,06%	13,77%	12,70%

Figura 11. Comparativa de modelos predictivos T1T2_WAT.

Support Vector Regression es el modelo más certero de los construidos.

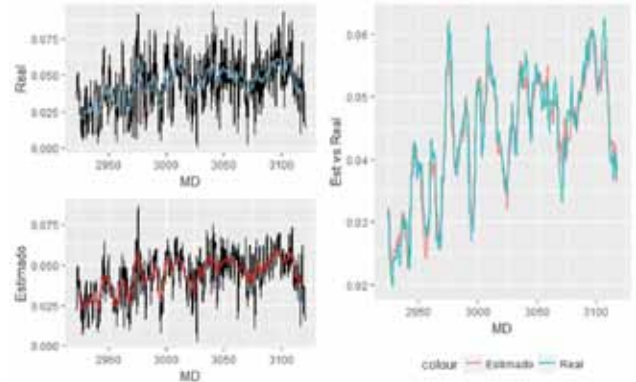
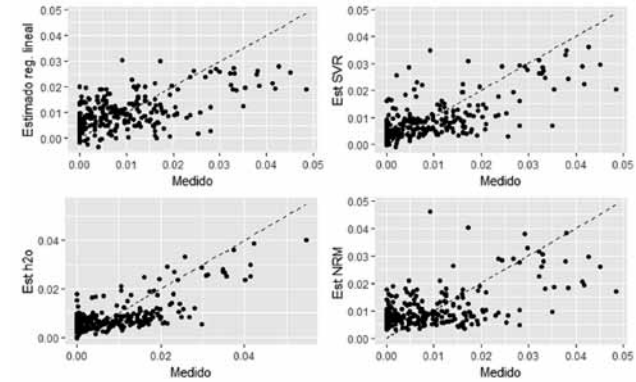


Figura 12. Estimado versus real (SVR - T1T2_WAT).

En la figura 12 se ven las gráficas de la variable según la profundidad tanto para los valores reales como para los

Predecir T1T2_HC1 a partir de los volúmenes mineralógicos



RMSE			
Reg. Lineal	SVR	RN	S. Ensemble
0,0085	0,0074	0,0086	0,0064
15,08%	13,10%	15,18%	11,33%

Figura 13. Comparativa modelos predictivos T1T2_HC1.

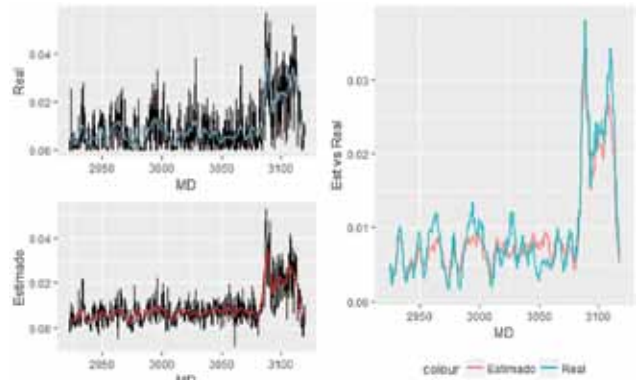
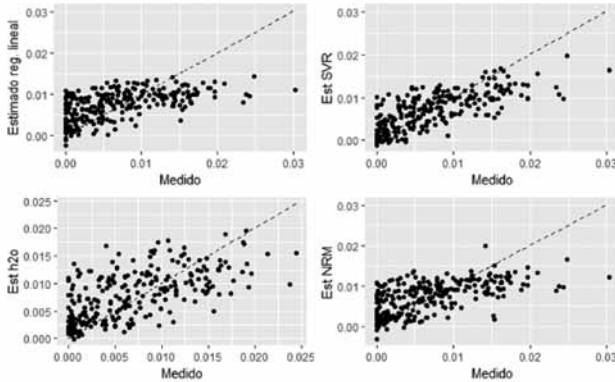


Figura 14. Estimado versus real (Stacked ensemble - T1T2_HC1).

estimados por el modelo. Las líneas roja y celeste son el promedio móvil suavizado de cada caso.

En la figura 14 se ven las gráficas de las variables según la profundidad tanto para los valores reales como para los estimados por el modelo. Las líneas roja y celeste son el promedio móvil suavizado de cada caso.

Predecir T1T2_IPW a partir de los volúmenes mineralógicos



RMSE	Reg. Lineal	SVR	RN	S. Ensemble
0,0049	0,0049	0,0040	0,0048	0,0045
16,33%	16,33%	13,51%	16,05%	15,09%

Figura 15. Comparativa modelos predictivos T1T2_IPW.

Support Vector Regression es el modelo más certero de los construidos.

En la figura 16 se ven las gráficas de la variable según la profundidad tanto para los valores reales como para los estimados por el modelo. Las líneas roja y celeste son el promedio móvil suavizado de cada caso.

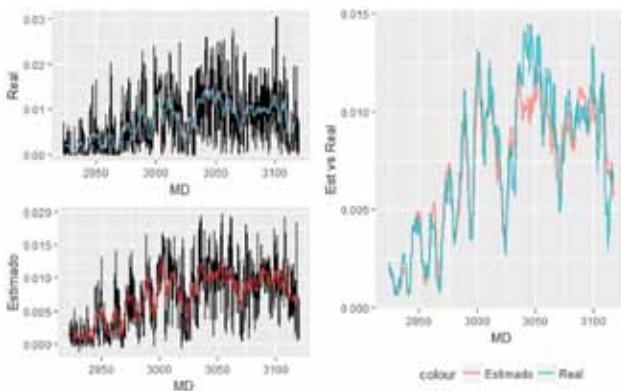
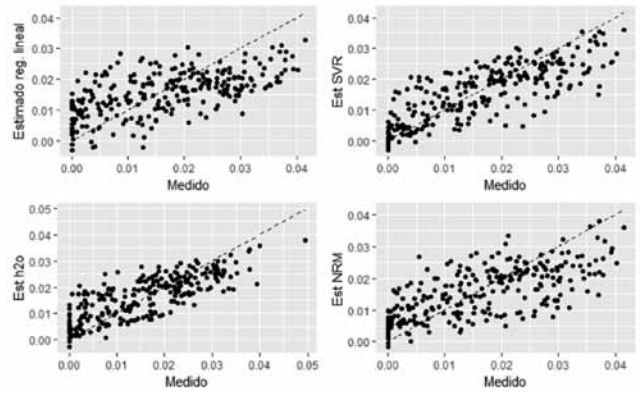


Figura 16. Comparativa modelos predictivos T1T2_IPW.

Predecir T1T2_HC3 a partir de los volúmenes mineralógicos

En la figura 17 se ven las gráficas de la variable según la profundidad tanto para los valores reales como para los estimados por el modelo. Las líneas roja y celeste son el promedio móvil suavizado de cada caso.



RMSE	Reg. Lineal	SVR	RN	S. Ensemble
0,0091	0,0091	0,0066	0,0083	0,0066
17,62%	17,62%	12,93%	16,09%	12,97%

Figura 17. Comparativa modelos predictivos T1T2_HC3.

Support Vector Regression es el modelo más certero de los construidos.

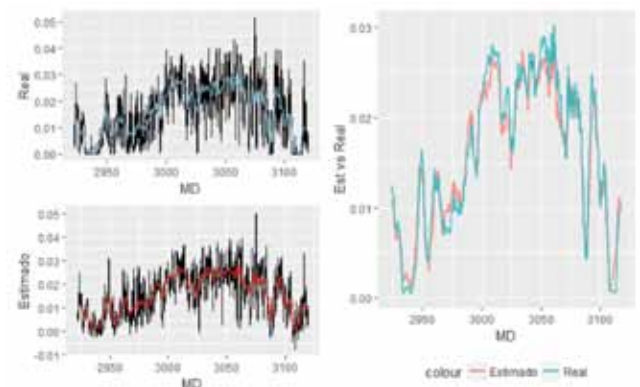
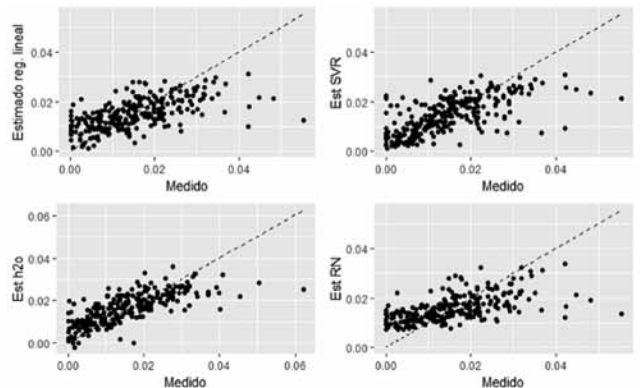


Figura 18. Estimado versus real (SVR - T1T2_HC3).

En la figura se ven las gráficas de la variable según la profundidad tanto para los valores reales como para los

Predecir WATER a partir de los volúmenes mineralógicos



RMSE	Reg. Lineal	SVR	RN	S. Ensemble
0,008	0,008	0,0064	0,0081	0,0072
12,86%	12,86%	10,35%	13,06%	11,52%

Figura 19. Comparativa modelos predictivos WATER.

Support Vector Regression es el modelo más certero de los construidos.

estimados por el modelo. Las líneas roja y celeste son el promedio móvil suavizado de cada caso.

En la figura 20 se ven las gráficas de la variable según la profundidad tanto para los valores reales como para los estimados por el modelo. Las líneas roja y celeste son el promedio móvil suavizado de cada caso.

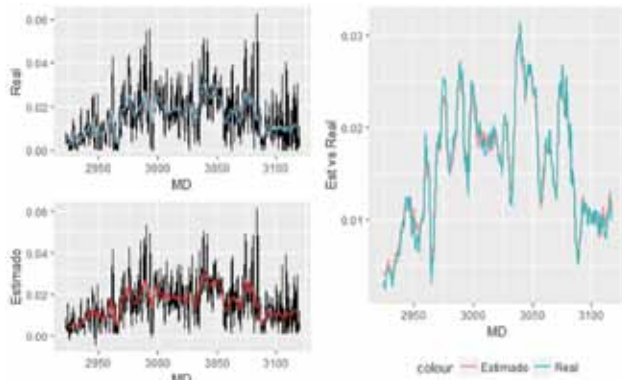


Figura 20. Estimado vs Real (SVR - WATER)

Conclusiones

Dentro de las variables de mineralogía se observan correlaciones altas. Calcita, illita y cuarzo muestran cierta colinealidad. Esto se traduce en que, si fuera costoso obtener las mediciones de alguno de ellos, se podría intentar predecir en base a los otros.

Se construyeron modelos predictores para las variables que van desde un 10% a un 14% de precisión. Esto es que el valor objetivo con respecto al rango de la variable queda predicho con un error medio de +/- ese porcentaje.

Aunque en la mayoría de los casos SVR fue el modelo que mejor resultados dio. El modelo creado con aprendizaje automático (*stacked ensemble*) puede dar mejores resultados si se lo deja procesando alternativas por un tiempo mayor. En este trabajo se dejó correr el algoritmo por una hora para cada variable.

Los modelos creados con los datos del pozo en estudio pueden ser probados con datos de otros pozos para medir su eficacia. ■