

10º Congreso de Exploración y Desarrollo de Hidrocarburos  
VI Jornadas de Geotecnología: Crear y Administrar Ámbitos Colaborativos  
Subtema: Ciencia de Datos

## ANÁLISIS MULTIVARIADO PARA LA PREDICCIÓN DE FLUIDOS A PARTIR DE LA MINERALOGÍA DE YACIMIENTOS NO CONVENCIONALES

Diego Gallart<sup>1</sup>, Alberto Cesar Ortíz<sup>2</sup>, Carolina Bernhardt<sup>2</sup>, Damián Hryb<sup>2</sup>,

Julio César Rodríguez Martino<sup>1</sup>, Gabriel Horowitz<sup>3</sup>

1: Practia, diego.galart@set.ypf.com, julio.rodriguez.m@set.ypf.com

2: YPF, alberto.ortiz@ypf.com, carolina.bernhardt@ypf.com, damian.hryb@ypf.com

3: Y-TEC, gabriel.horowitz@ypftecnologia.com

Palabras clave: ciencia de datos, correlaciones, mineralogía, no convencional

### ABSTRACT

Multivariate analysis. Prediction of fluids from mineralogy variables on unconventional reservoirs. The characterization of the pore spaces and volumes of fluids from data of T1T2 NMR in an unconventional reservoir allows the understanding of the characteristics of it at nanoscale. For these signals to acquire meaning and relevance in the prediction of “sweet spots” or places of hydrocarbon interest, it is necessary to analyze them in relation to the mineralogy and content of organic matter (kerogen) of the rock.

The multivariate analysis techniques used have been oriented in this direction, which has allowed to consider the influence of each parameter on the others, revealing correlations that would be very difficult to discover and analyze using traditional methods.

The present work was carried out in different stages beginning by studying the correlations between different mineralogical variables, organic matter, fluids and NMR T1T2 porosity in an unconventional reservoir well belonging to YPF. The degree of correlation between variables was studied, eliminating those redundant, and 4 machine learning models were used to predict the objective variables where the smallest error obtained in the prediction was around 10%.

The relative importance of the variables in the prediction was analyzed, which in some cases coincide with the existing physical models and in others they lead to rethinking some assumptions about them, providing new knowledge to this type of analysis.

The models created with the well data analyzed can be tested with data from other wells to measure their efficiency and thus become a predictive tool for poral and fluid characteristics of shale reservoirs.

### INTRODUCCIÓN

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 las características del mismo a escala nanométrica. Para que estas señales adquieran sentido y relevancia en la predicción de “sweet spots” o lugares de interés hidrocarburífero, es necesario analizarlas en relación a la mineralogía y 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 sería muy difícil descubrir y analizar utilizando métodos tradicionales.

El presente trabajo se llevó a cabo en distintas etapas comenzando 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 4 modelos de aprendizaje automático para predecir las variables objetivo donde el menor error obtenido en la predicción fue del orden de 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 llevan a replantear algunas suposiciones de los mismos, aportando 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.

## 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.

| Grupo                                                                                 | Descripción                                 |
|---------------------------------------------------------------------------------------|---------------------------------------------|
| Profundidad                                                                           | Delta 15 centímetros                        |
| 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 extraídos del procesamiento de la resonancia magnética nuclear (NMR) modo T1T2.

### Volumen Poro/FluidoT1/T2\_WAT

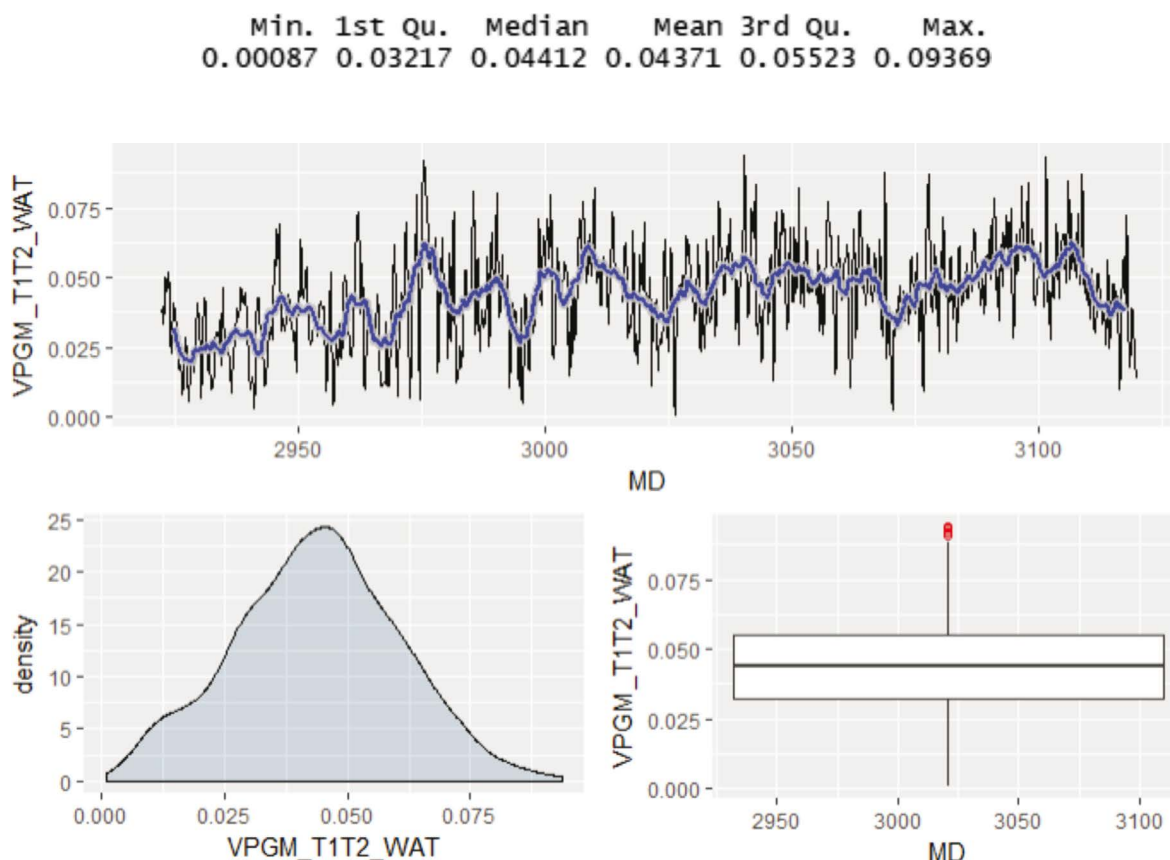


Figura 1. Descripción variable T1T2\_WAT

En el gráfico superior 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 3 picos que pueden observarse en el gráfico superior y son valores medidos válidos.

### VolumenPoro/Fluido T1/T2\_ HC1

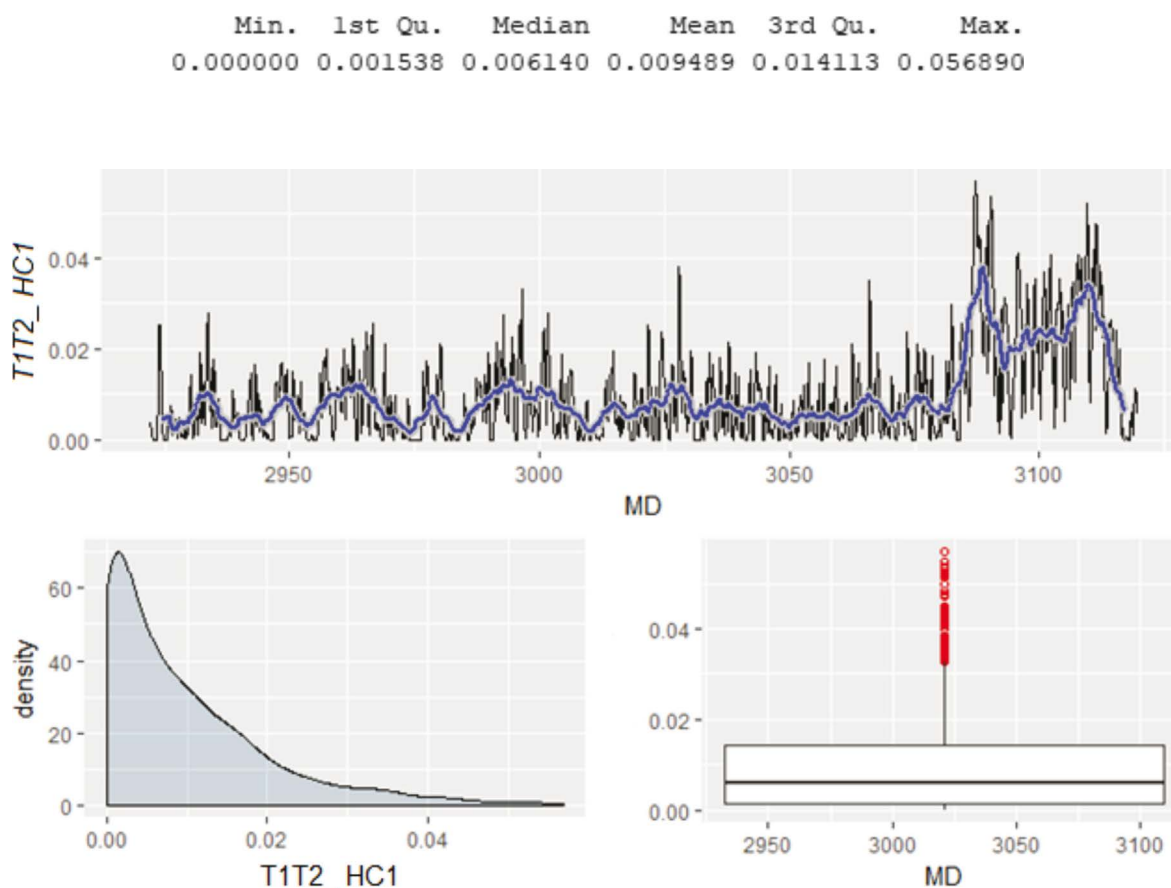


Figura 2. Descripción variable T1T2\_HC1

En el gráfico superior 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. Viendo el gráfico superior, esos puntos rojos, corresponden a los valores de mayor profundidad que se incrementan saliéndose de lo que venía siendo normal a menor profundidad.

### Volumen Poro/Fluido T1/T2 \_IPW

| Min.     | 1st Qu.  | Median   | Mean     | 3rd Qu.  | Max.     |
|----------|----------|----------|----------|----------|----------|
| 0.000000 | 0.001610 | 0.005965 | 0.007260 | 0.011723 | 0.030330 |

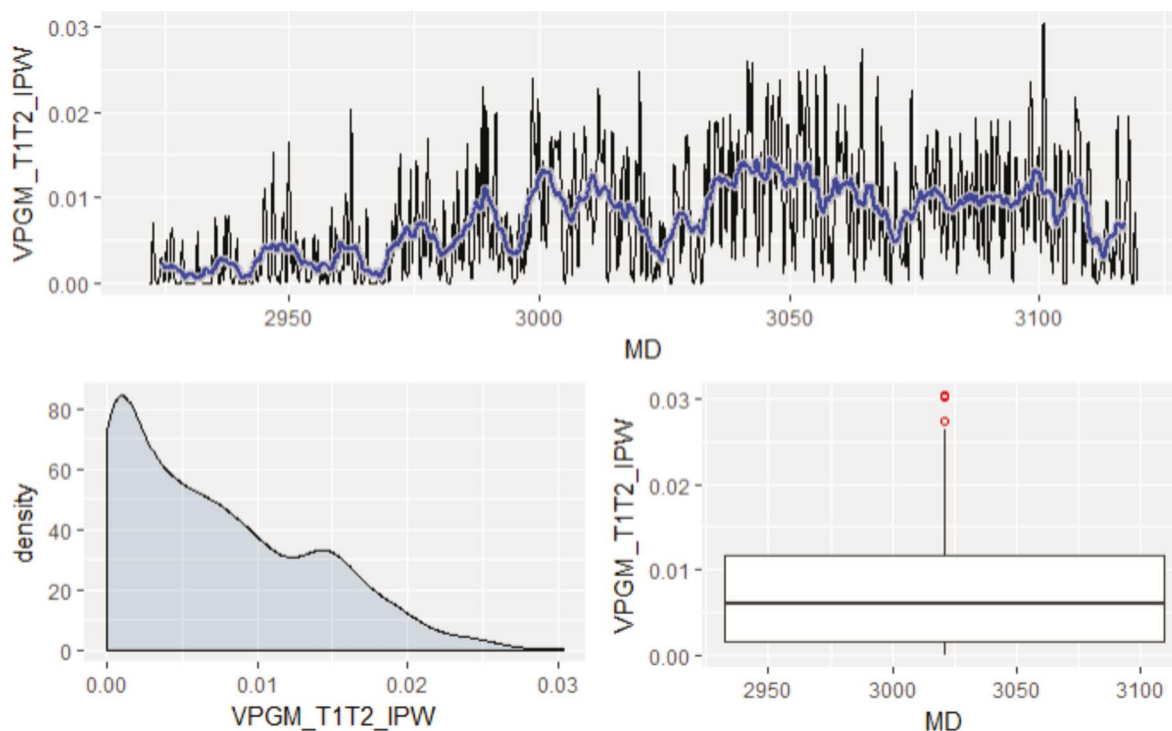


Figura 3. Descripción variable VPGM\_T1T2\_IPW

En el gráfico superior 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 0.025 y son valores medidos válidos.

### Volumen Poro/Fluido T1/T2\_HC3

|          |          |          |          |          |          |
|----------|----------|----------|----------|----------|----------|
| Min.     | 1st Qu.  | Median   | Mean     | 3rd Qu.  | Max.     |
| 0.000000 | 0.004985 | 0.015840 | 0.015847 | 0.024730 | 0.051640 |

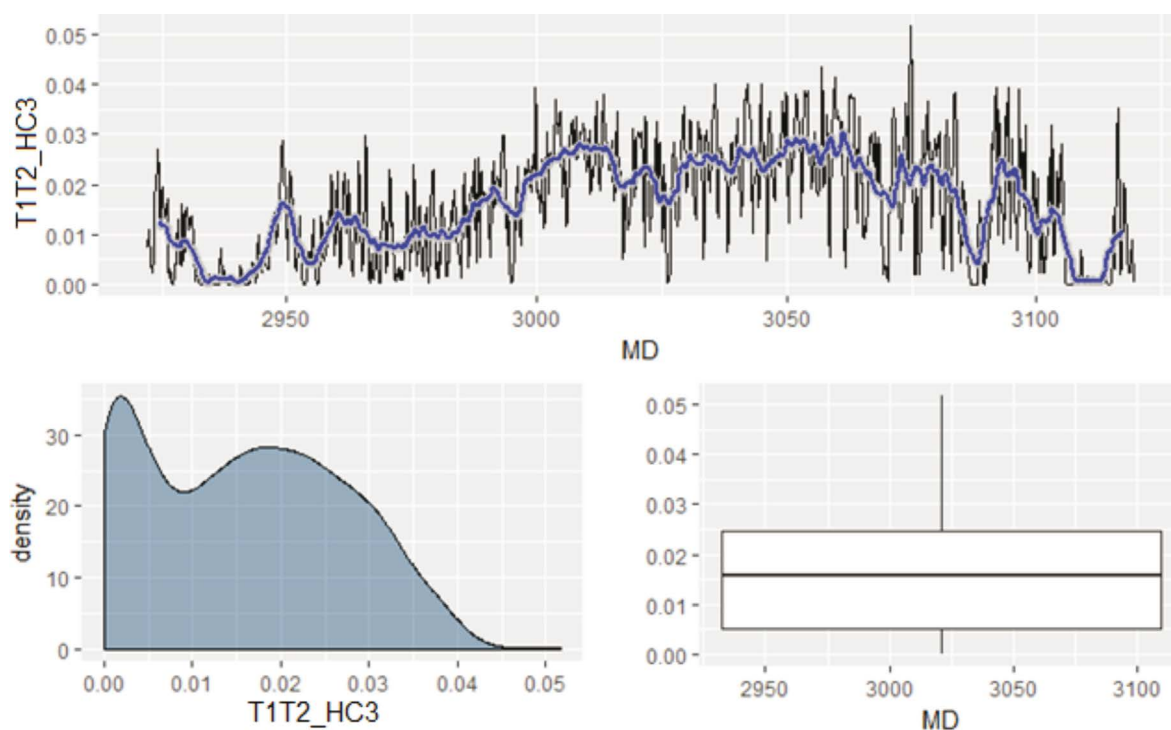


Figura 4. Descripción variable T1T2\_HC3

En el gráfico superior 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, estos no llegan a ser valores atípicos.

### Volumen Fluidos Interpretación – WATER

|          |          |          |          |          |          |
|----------|----------|----------|----------|----------|----------|
| Min.     | 1st Qu.  | Median   | Mean     | 3rd Qu.  | Max.     |
| 0.000010 | 0.006947 | 0.013950 | 0.015419 | 0.021825 | 0.062300 |

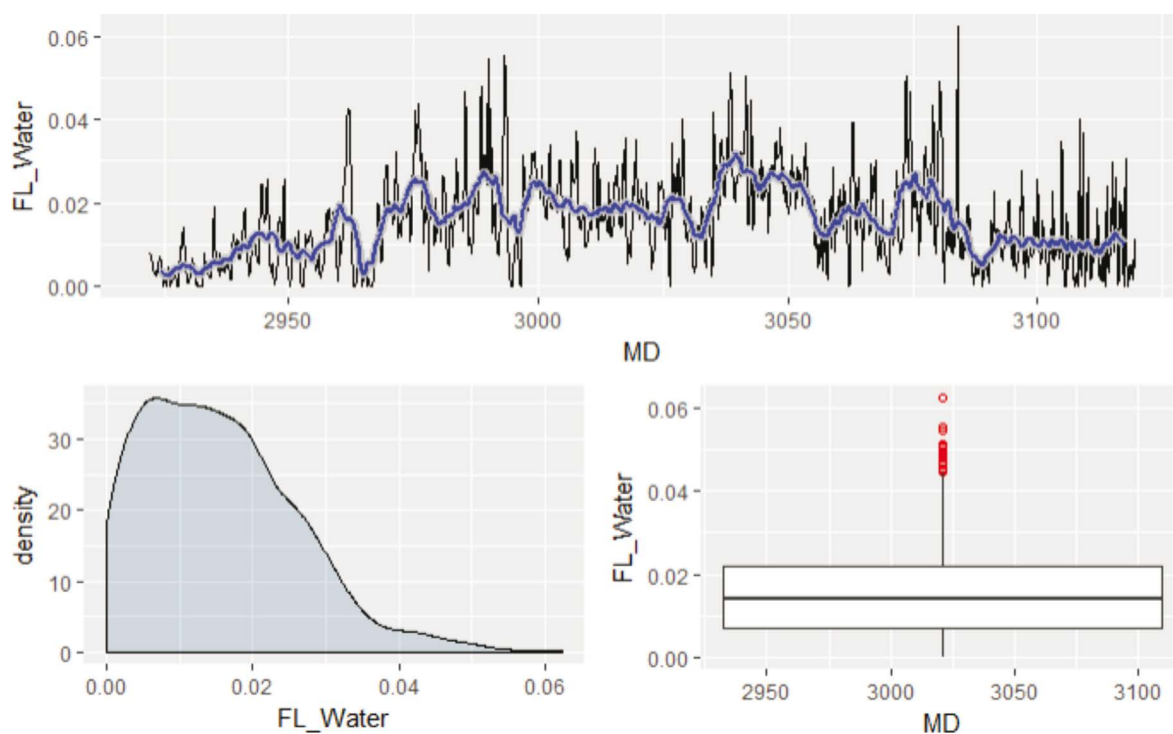


Figura 5. Descripción variable WATER

En el gráfico superior 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

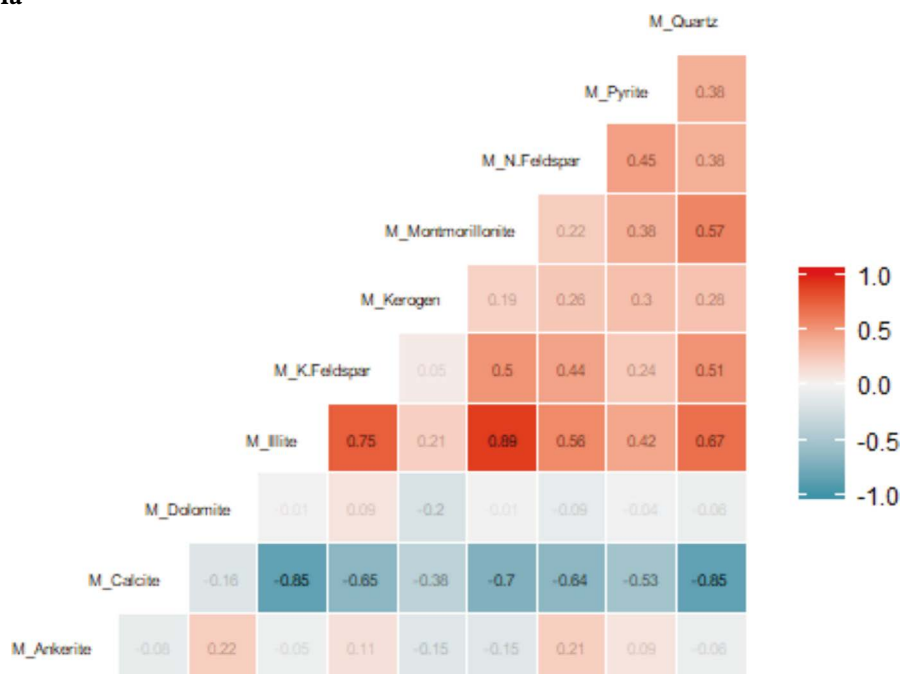


Figura 6. Correlaciones Mineralogía (Pearson)

Entre las variables de mineralogía vemos que Calcite, Illite y Quartz se encuentran altamente correlacionadas. Calcite correlaciona negativamente con Illite y Quartz con el mismo coeficiente (-0.85).

Para visualizar las relaciones entre las variables de mineralogía aplicaremos componentes principales y graficaremos sobre las dimensiones resultantes.

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

En cuanto a las correlaciones, el ángulo entre los vectores de las variables indica su correlación. Si describen un ángulo de 90 grados no están correlacionadas. Si el ángulo es cercano a los 180 grados están correlacionadas negativamente. Si el ángulo es cercano a 0 grados entonces están positivamente correlacionados.

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

Las 2 dimensiones del gráfico explican el  $45.6 + 14.4 = 59.5\%$  de la variabilidad de los datos originales.

Observando las proyecciones sobre la Dim1 vemos que Calcite está correlacionada negativamente con la Dim1 mientras que K.Fekdsopar, Illite, Quarz y Montmorillonte

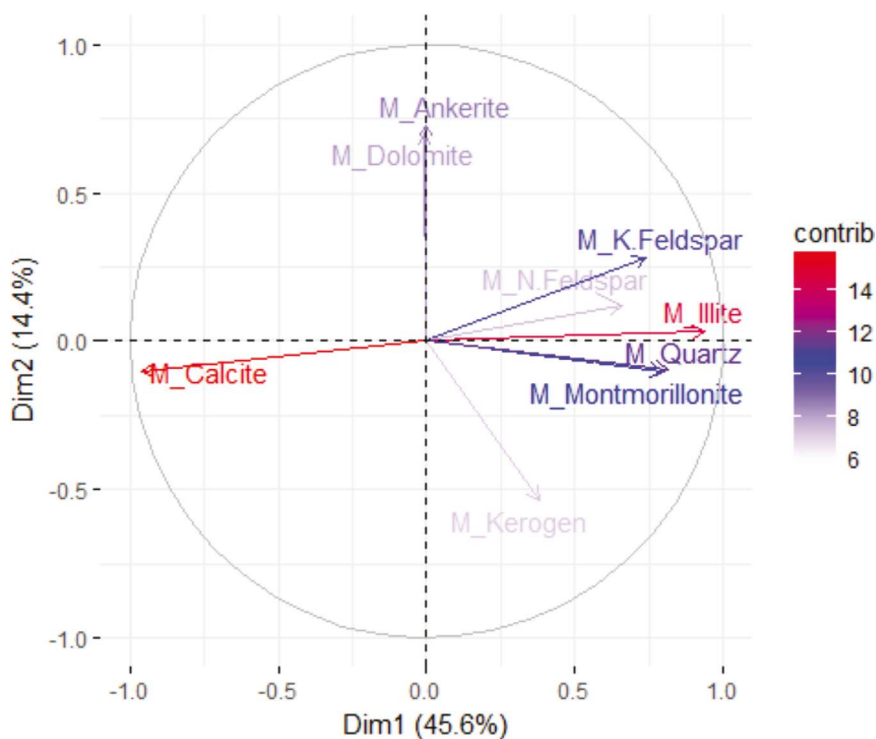


Figura 7 - Componentes Principales Mineralogía (Dim1- Dim2)

correlacionan positivamente entre ellas y con la Dim1. Viendo 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.

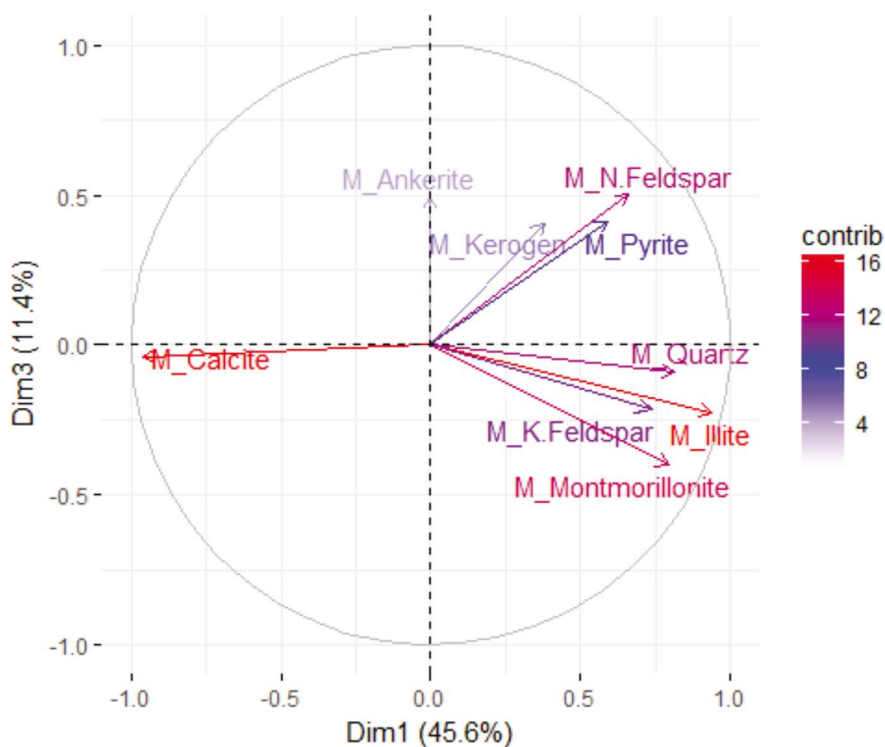


Figura 8 - Componentes Principales Mineralogía (Dim1- Dim3)

La Dim3 aporta 11.4% de explicación de variabilidad que sumado a lo anterior (59.5%) da 70.9%.

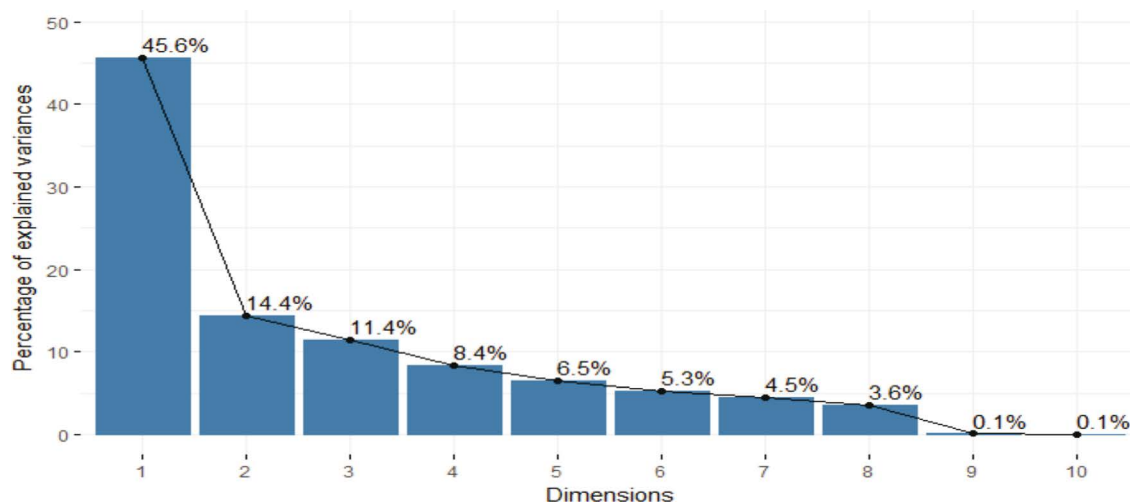


Figura 9. Varianza explicada por cada Dim de PCA

En la Figura 9 vemos que las primeras 3 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).

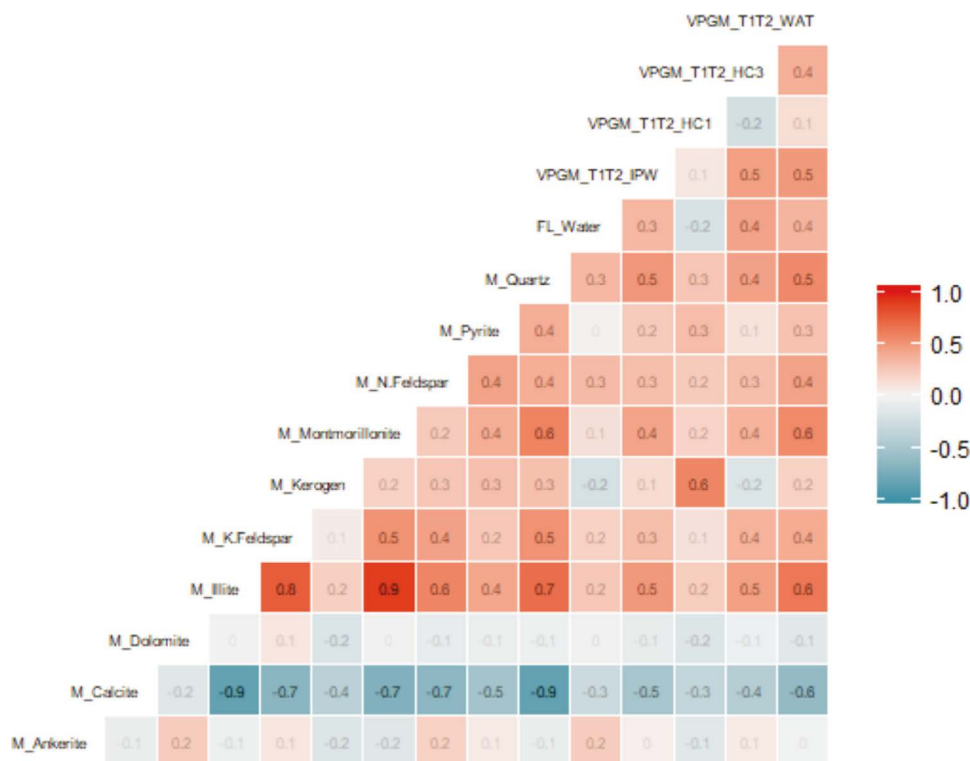


Figura 10 - Correlaciones ente variables objetivo y Mineralogía

La variable VPGM\_T1T2\_WAT correlaciona negativamente con Calcite (-0.61), positivamente con Illite (0.62), luego positivamente con Montmotillonite y Quartz (0.55).

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).

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).

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).

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).

## Modelos Predictores

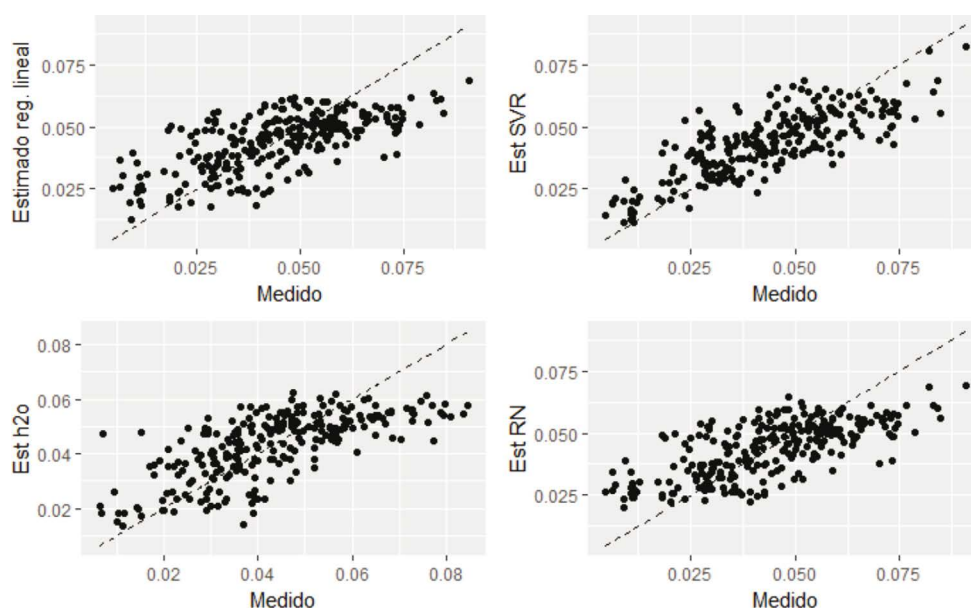
Para encontrar el mejor modelo predictor y las variables más importantes vamos dividir el set de datos (observaciones) disponibles en 2 lotes. El 80% de los datos se utilizará para entrenar el modelo y el 20% restante para validarlo.

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

A continuación se muestra gráficamente 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 debajo 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  $\left(\frac{RMSE}{VAL_{max}-VAL_{min}}\right)$

## 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 modelos predictivos T1T2\_WAT

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

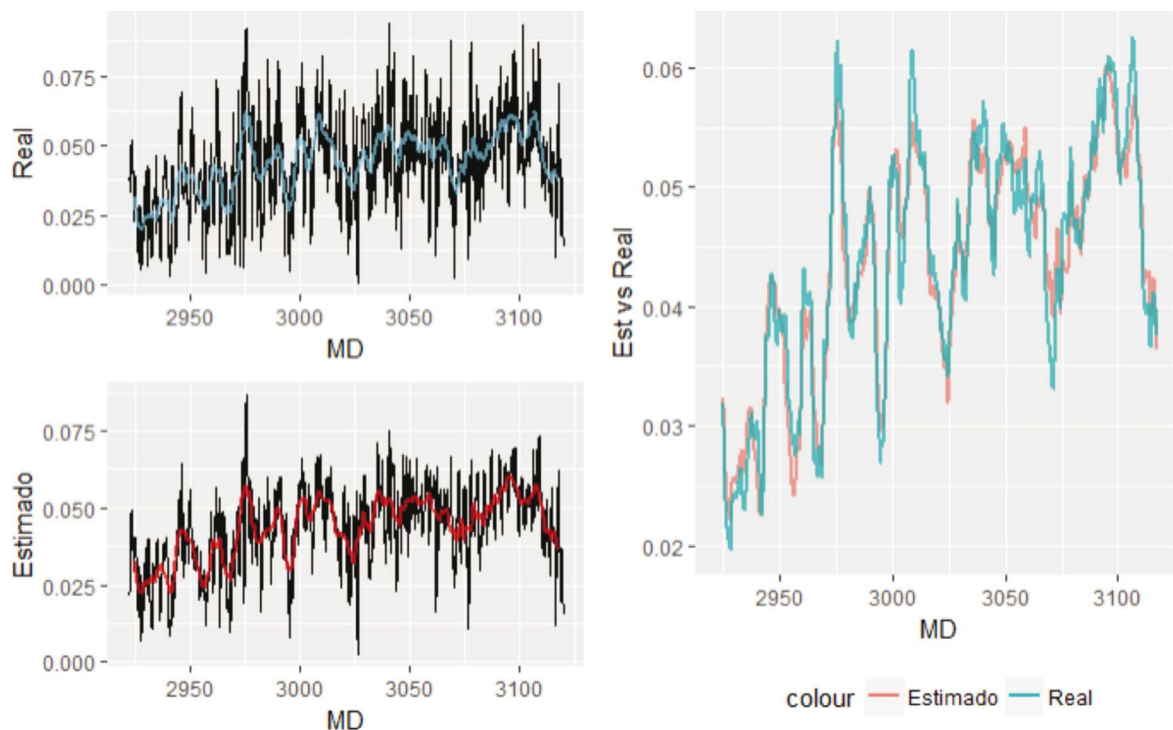
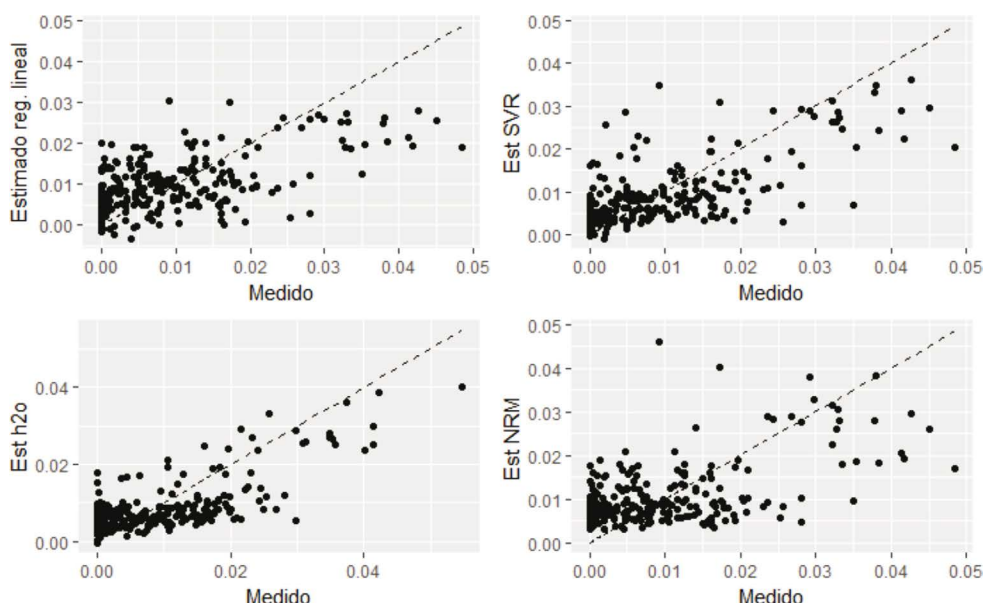


Figura 12. Estimado vs Real (SVR - T1T2\_WAT)

En la ilustración 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.

### Predicir 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

El stacked ensemble es el modelo más certero de los construidos.

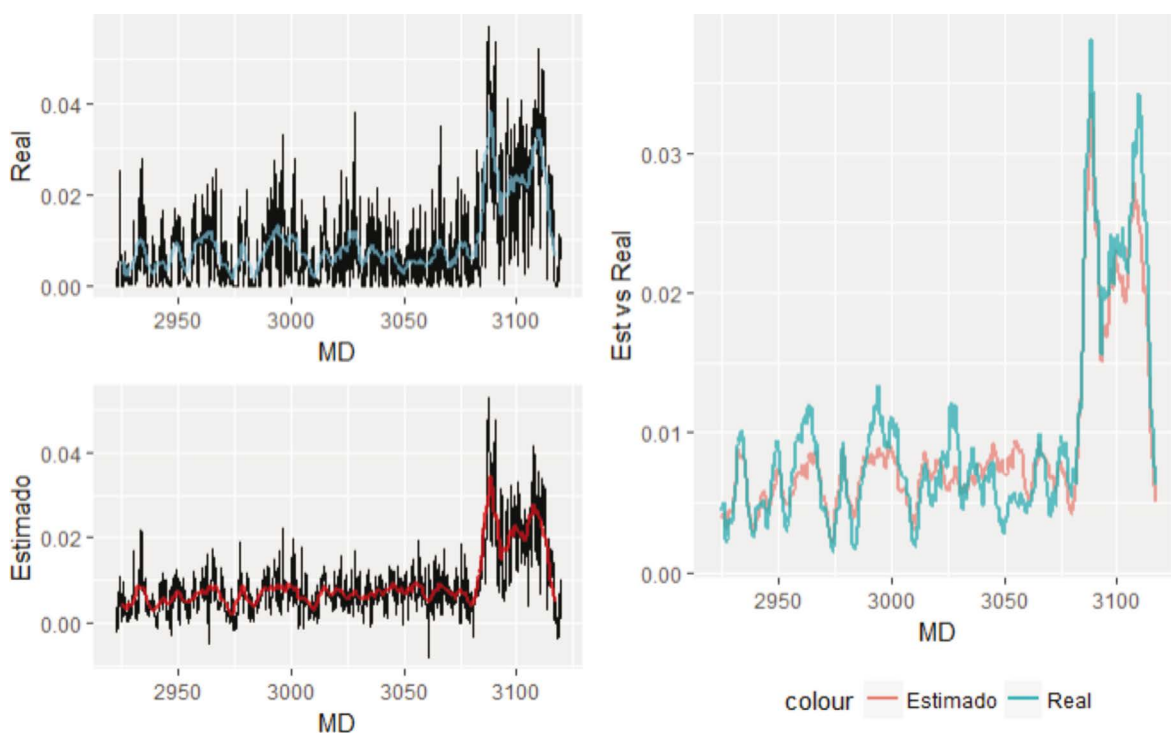
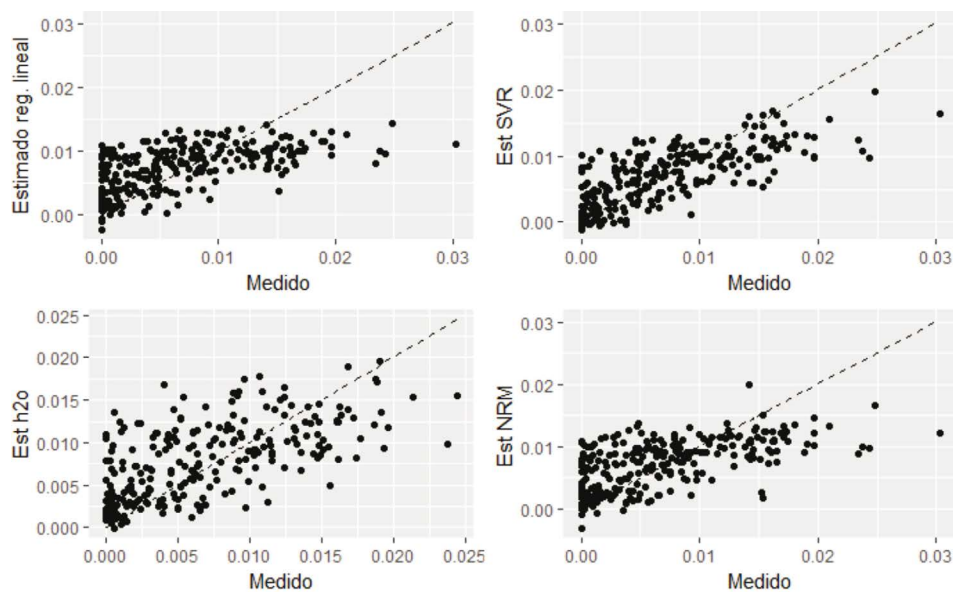


Figura 14. Estimado vs Real (Stacked ensemble – T1T2\_HC1)

En la ilustración 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.

## Predecir T1T2\_IPW a partir de los volúmenes mineralógicos



| RMSE       |        |        |             |
|------------|--------|--------|-------------|
| Reg.Lineal | SVR    | RN     | S. Ensemble |
| 0.0049     | 0.0040 | 0.0048 | 0.0045      |
| 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.

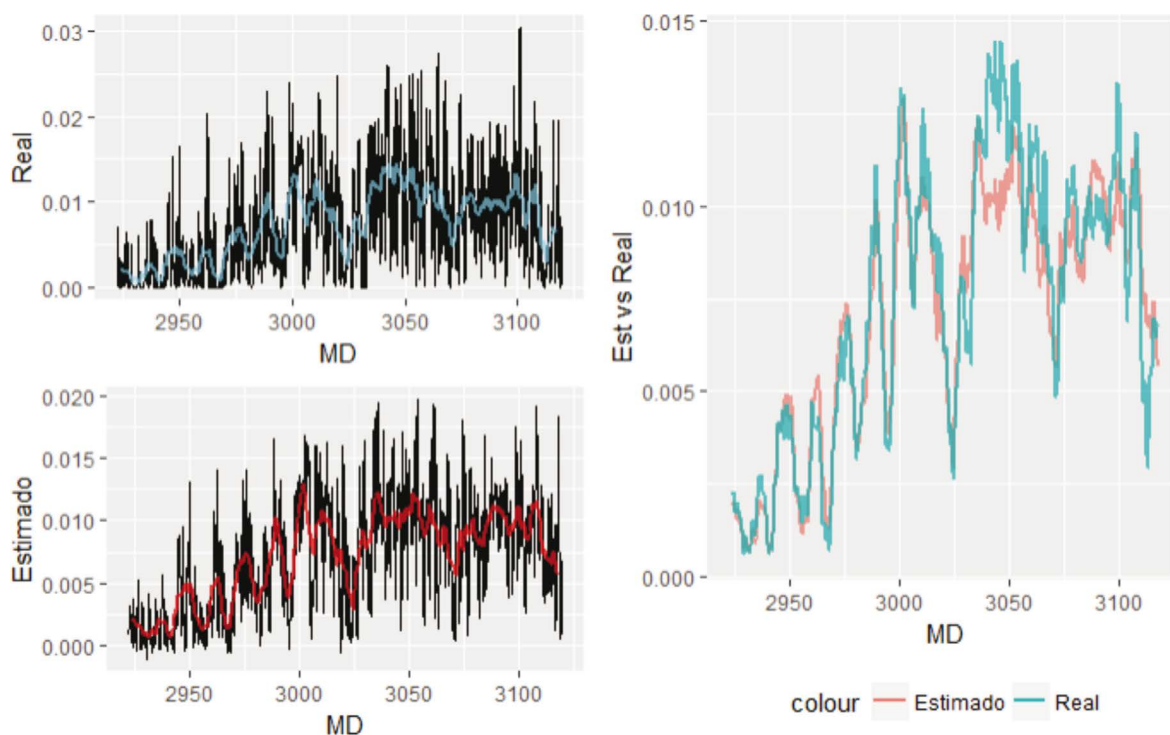
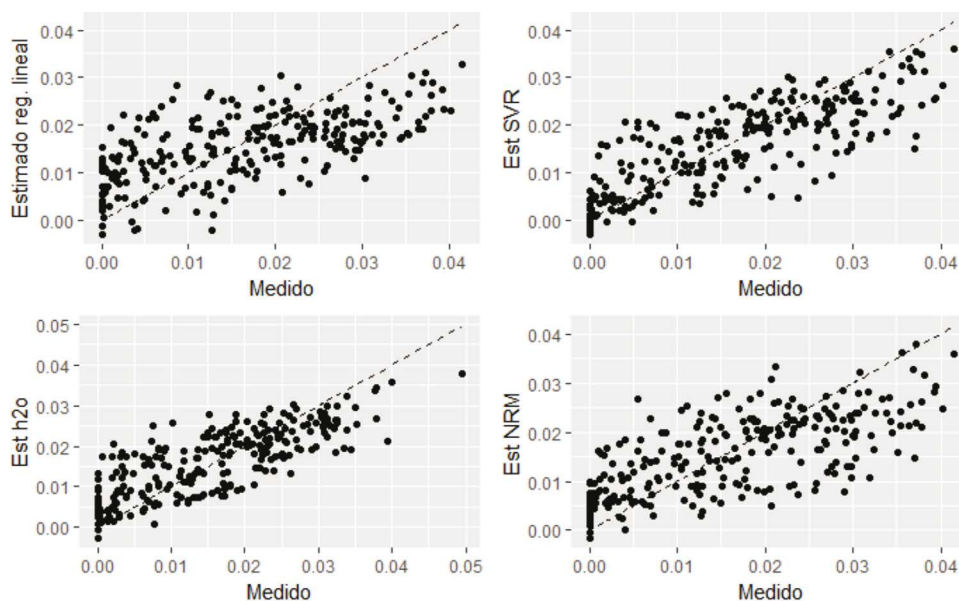


Figura 16. Estimado vs Real (SRV - T1T2\_IPW)

En la ilustración 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.

### Predecir T1T2\_HC3 a partir de los volúmenes mineralógicos



| RMSE       |        |        |             |
|------------|--------|--------|-------------|
| Reg.Lineal | SVR    | RN     | S. Ensemble |
| 0.0091     | 0.0066 | 0.0083 | 0.0066      |
| 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.

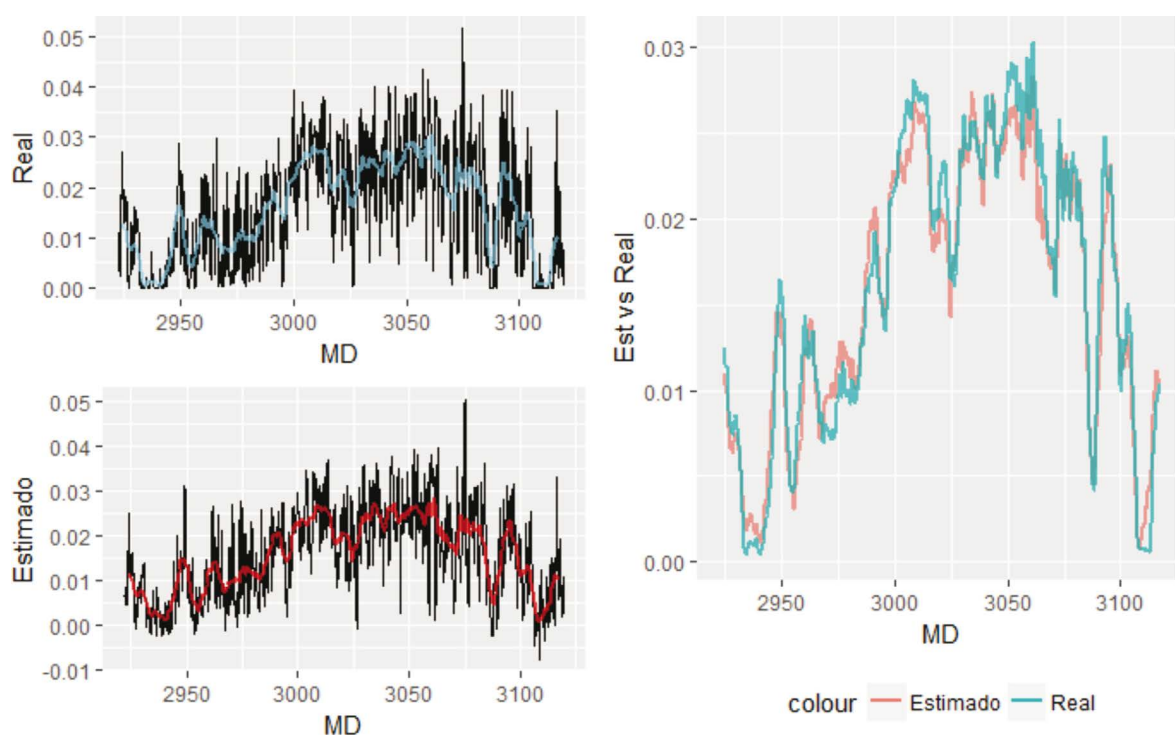
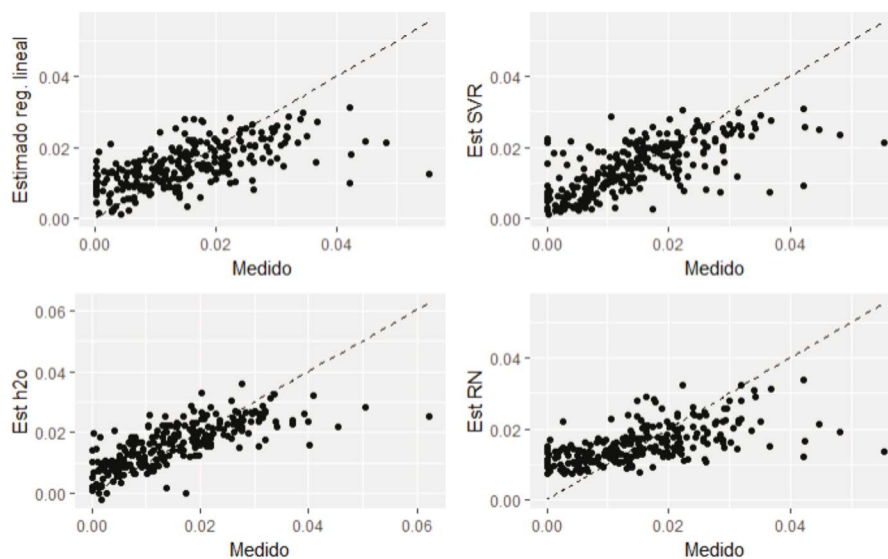


Ilustración 22. Estimado vs Real (SVR - T1T2\_ HC3)

En la ilustración 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.

## Predecir WATER a partir de los volúmenes mineralógicos



| RMSE       |        |        |             |
|------------|--------|--------|-------------|
| Reg.Lineal | SVR    | RN     | S. Ensemble |
| 0.008      | 0.0064 | 0.0081 | 0.0072      |
| 12.86%     | 10.35% | 13.06% | 11.58%      |

Figura 18. Comparativa modelos predictivos WATER

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

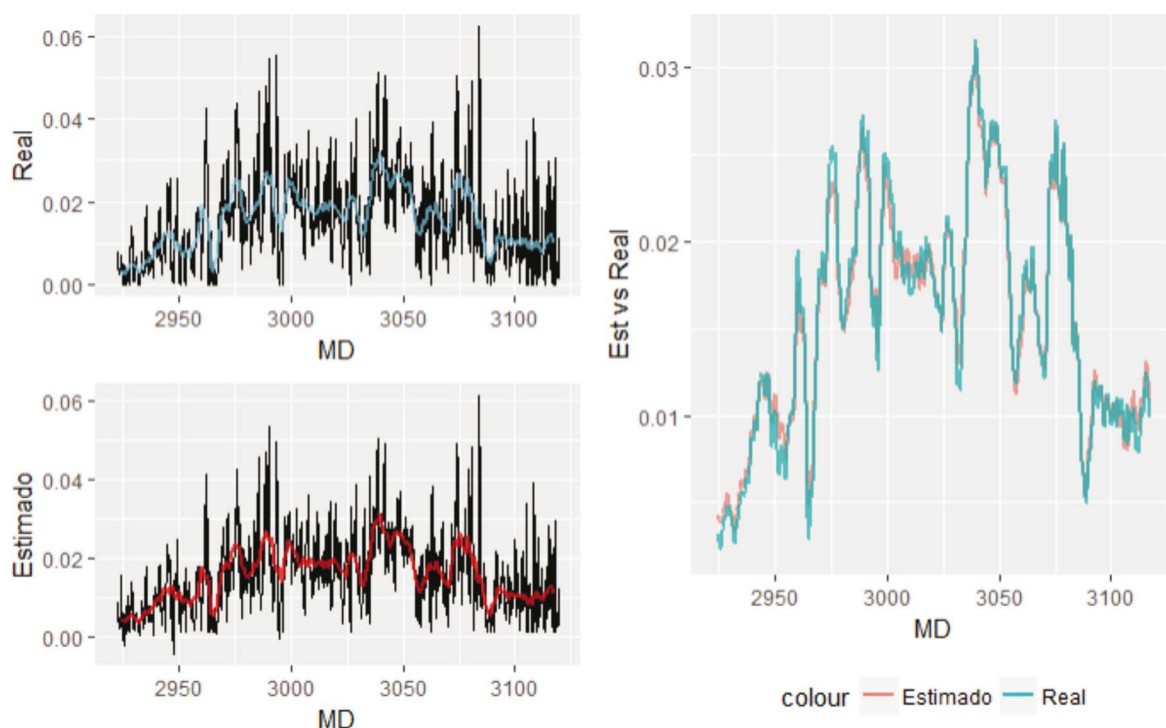


Figura 19. Estimado vs Real (SVR - WATER)

En la ilustración 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.

## CONCLUSIONES

Dentro de las variables de mineralogía se observan correlaciones altas. Calcite, Illite y Quartz 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 arrojar mejores resultados si se lo deja procesando alternativas por un tiempo mayor. En este trabajo se dejó correr el algoritmo por 1 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.



**10° CONGRESO DE  
EXPLORACIÓN Y  
DESARROLLO DE  
HIDROCARBUROS**

**Energía y Sociedad  
aliados inseparables  
Mendoza 2018**