

MODELO DE APRENDIZAJE ESTADÍSTICO PARA LA PREDICCIÓN DE PRODUCCIÓN A PARTIR DE INFORMACIÓN DE SUBSUELO Y VARIABLES DE ESTIMULACIÓN

Ariel Kautyian¹, Soledad Lagos¹, Walter Brinkworth¹, Luciana De Marzio¹,
Mauro Weimann¹ y Guillermina Sagasti¹

1: YPF S.A. ariel.kautyian@ypf.com, soledad.lagos@ypf.com, walter.brinkworth@ypf.com, luciana.demarzio@ypf.com, mauro.i.weimann@ypf.com, gsagastic@ypf.com

Palabras clave: Aprendizaje Estadístico, producción, predicción, Vaca Muerta, shale

ABSTRACT

The main objective of this work is to define a statistical-learning model that combines data from multiple sources to predict cumulative production for a set of horizontal wells targeting Vaca Muerta formation. The set of predictive features includes geologic or reservoir information, geo-navigation data and stimulation parameters for each well. Geologic information is represented by seismic attributes of different types extracted along well trajectories. A thorough exploratory analysis is conducted over the input data to understand the distribution and variability of proposed features. Outliers are identified and removed, and the dimensionality of the problem is reduced by evaluating collinearity effects and performing a principal component analysis.

In this first stage of the work, we propose a multiparametric linear model through LASSO method (Least Absolute Shrinkage and Selection Operator), which allows for an optimum feature selection while training the model. Hence, the method contributes to further dimensionality reduction and to avoid overtraining effects, therefore balancing the bias-variance tradeoff. The linearity of the model helps interpreting the effect of each feature on the target variable, and its coefficients represent the importance of variables. We conclude that the trained statistical-learning model can explain a significant fraction of the well production data, while revealing synergy between geological properties and stimulation parameters. One of the advantages of using seismic data as features in the model is that it can be used to construct predictive maps of cumulative production. These maps could be interpreted as an indicator of productivity potential due to the quality of reservoir.

INTRODUCCIÓN

Se conoce como *Statistical Learning* o Aprendizaje Estadístico a un conjunto de herramientas que tienen por objetivo modelar, analizar y entender datos complejos. Es un área en desarrollo dentro de la estadística que combina diversos aspectos de las ciencias de la computación. El objetivo de estos algoritmos o modelos es poder extraer información de los datos con el fin de

entender su estructura, posibles interrelaciones y tomar decisiones a partir de ellos (James *et al.* 2021). En este campo se utilizan diversos algoritmos y métodos, entre los que podemos mencionar los modelos de regresión, árboles de regresión y clasificación y support vector machine, entre otros. En particular, los modelos en los que se busca predecir una variable objetivo o target a partir de los datos o features se denominan métodos de predicción supervisados (James *et al.* 2021). En los últimos años, el uso de este tipo de herramientas ha ido en aumento, considerando este enfoque para analizar problemas de diferente naturaleza y diversas disciplinas.

Una posible aplicación de modelos de Aprendizaje Estadístico supervisado corresponde a la búsqueda de criterios y estrategias para predecir la productividad y optimizar el rendimiento de los campos asociados a reservorios no convencionales. En el trabajo publicado por Wang y Cheng (2019) se implementan algoritmos de *Machine Learning* con el fin de optimizar los parámetros de estimulación y otras variables, considerando como objetivo a la producción acumulada de los pozos en su primer año, en la formación Montney en la cuenca sedimentaria del oeste de Canadá. Un objetivo similar se propone Rice (2014) para un conjunto de pozos horizontales en la cuenca de Delaware, Estados Unidos. Por su parte, en el trabajo de Smith *et al.* (2019) se incorporan además datos geofísicos (coherencia sísmica) y geológicos (Phi y TOC) para predecir la variación temporal de la producción (curvas de declino) en pozos con objetivo en la formación Eagle Ford. En Argentina, un análisis de este tipo fue realizado por Gilardone *et al.* (2021a, 2021b) a partir de datos de acceso público a través de la Secretaría de Energía.

El desarrollo de reservorios no convencionales en Argentina, en continuo aumento desde el año 2015, ha sido liderado por YPF, que cuenta con más del 60% del total de los pozos perforados y fracturados a Vaca Muerta. Este desarrollo es acompañado por la generación de datos que permiten mejorar el entendimiento del reservorio y tomar decisiones que contribuyan a optimizar los recursos.

El presente trabajo tiene por objetivo desarrollar un flujo de trabajo que permita integrar el volumen de información disponible asociada a distintas disciplinas que de manera aislada analizan características específicas de los pozos, la formación objetivo y la eficiencia de los diferentes diseños de fractura. Se busca encontrar una metodología óptima para combinar esta información con el fin de generar un modelo capaz de explicar las observaciones asociadas a los datos de producción, con la suficiente capacidad de generalización que permita predecir, al menos en parte, el rendimiento de otros pozos no considerados para el modelo.

En esta primera etapa, se trabajó en un área de aproximadamente 1760 km² al este de la Provincia de Neuquén, en la Región del Engolfamiento Neuquino. Se trata de un cubo sísmico con una migración pre-stack en tiempo. Además de los datos sísmicos se recopilieron las trayectorias, parámetros de estimulación, datos de geonavegación y producciones acumuladas de más de 200 pozos en la región de estudio. En las sucesivas secciones se describe la información

utilizada para el trabajo, la metodología implementada y los resultados observados en esta etapa del proyecto, y se discuten posibilidades de mejora y limitaciones de los modelos generados.

DATOS

A continuación, se describen los datos utilizados como variables predictivas para el modelo de Aprendizaje Estadístico.

Información de estimulación

Para cada uno de los pozos utilizados se cuenta con el conjunto de parámetros que definen el diseño de estimulación, y que refleja la evolución en el tiempo que han tenido los diseños debido a la incorporación de conocimiento y tecnología. La información disponible está dada por: la longitud estimulada (en metros), la intensidad de fluido y de arena (en bbl/ft), la cantidad de *clusters* por etapa, el espacio entre *clusters* (en metros), el caudal por *cluster* (en bbl/min/#) y el volumen por *cluster* (en bbl/#). Como es de esperar, y por tratarse de variables de diseño, varias de ellas están relacionadas, por lo tanto, existen importantes colinealidades entre los distintos parámetros de estimulación. Este efecto puede ser analizado a partir de la matriz de correlaciones de la Figura 1a. De esta manera, por ejemplo, el número de *clusters* es inversamente proporcional al espaciamiento de *clusters*. La información redundante tiende a ir en detrimento de la eficiencia de los modelos de Aprendizaje Estadístico, por lo que se realizó un análisis de componentes principales sobre este conjunto de datos para identificar la información que efectivamente agrega valor al modelo. Como se puede observar en la Figura 1b, las primeras tres componentes principales calculadas a partir del total de variables de estimulación logran explicar el 90% de la varianza de las mismas.

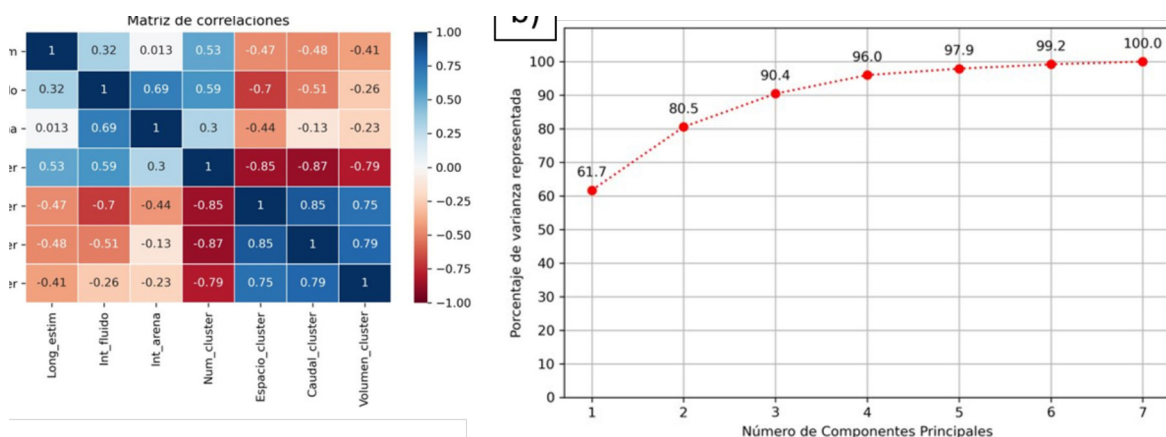


Figura 1. Matriz de correlaciones de los parámetros de estimulación y varianza acumulada para las componentes principales.

Información derivada de la sísmica

La cobertura areal del volumen sísmico utilizado se muestra en la Figura 2a. A partir de este dato, se interpretaron horizontes en toda la extensión del *survey* para identificar los niveles de interés (Fig. 2b), dados por los niveles de navegación de los pozos.

Posteriormente, se calcularon atributos sísmicos de distinta naturaleza para cada uno de los niveles de interés.

La información incorporada al modelo estadístico, proveniente del dato sísmico, está compuesta por atributos *post-stack*. Uno de estos atributos es la impedancia P derivada de un proceso de inversión acústica. Los atributos sísmicos son herramientas que permiten cuantificar propiedades estratigráficas y estructurales del medio. Distintos atributos, obtenidos mediante diferentes algoritmos brindan información de, por ejemplo, la litología, propiedades del reservorio como la porosidad o el contenido de carbono total, la presencia de fallas y fracturas, la posición en el sistema depositacional, etc. Por lo tanto, se asume que las propiedades geológicas del subsuelo pueden ser descriptas a través de los mismos. En la Figura 3 se listan los atributos calculados y se muestran algunos de ellos para el Nivel 1. Una vez obtenido el conjunto de atributos, se genera un dataframe inicial que contiene la unión de los mismos.

Como parte del control de calidad y análisis exploratorio de datos se detectan y remueven valores anómalos. El criterio para la remoción de valores anómalos se basa en el análisis de los rangos intercuartiles de los datos, siempre que estos tengan una distribución afín a la aplicación de este criterio (atributos como la coherencia deben ser analizados de forma individual). Tras eliminar los valores anómalos, que representaban aproximadamente el 1% de los datos, se analizó la posible colinealidad entre atributos, eliminando aquellos que representaban información claramente redundante y reduciendo el número de variables de 18 a 12. Esto corresponde a pares de variables con correlaciones superiores al 50%. En cada caso, se conservó aquella variable con mayor significado físico. En la Figura 4a) se muestra la matriz de correlación de estas variables. A diferencia de lo que se observaba para el caso de las variables de estimulación, estas variables sísmicas muestran una correlación muy baja entre ellas, que también se manifiesta en la curva de varianza acumulada por componente principal de la Figura 4b): se necesitan al menos 11 variables para explicar el 90% de la varianza de los datos. Esta observación muestra que a partir del cálculo de distintos atributos es posible capturar información presente en los datos sísmicos de diferente naturaleza y con independencia entre variables.

A partir de los atributos sísmicos seleccionados, se realiza una extracción a lo largo de la trayectoria de cada uno de los pozos disponibles. Para esto, se asume que la resolución vertical de la sísmica permite promediar los atributos en el intervalo interpretado. Además, dado que las trayectorias de los pozos son predominantemente N-S, la distancia E-O definida para la extrac-

ción del atributo representa la extensión de la zona fracturada en esta dirección, mientras que la variabilidad en la dirección N-S refleja los efectos de la resolución espacial del dato sísmico. Como se ilustra en la Figura 5, para cada punto de la trayectoria horizontal de un pozo, se reali-

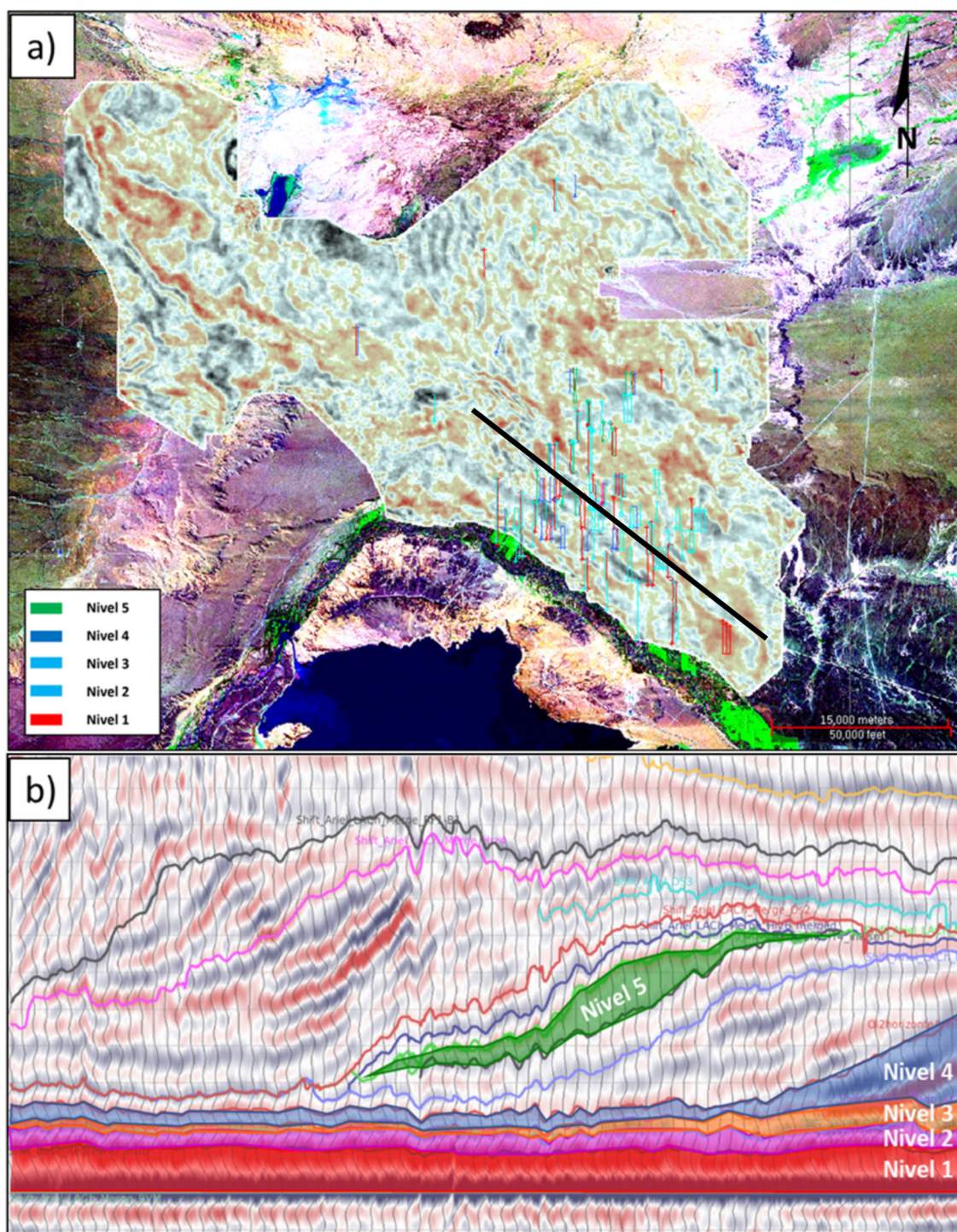


Figura 2. a) Área de cobertura del dato sísmico. Ubicación de pozos y distribución por nivel de navegación. b) Esquema de la interpretación sísmica de los niveles de navegación a lo largo de la sección indicada en a).

za la extracción del atributo asumiendo un largo (x) y un ancho (y) para la zona fracturada, dando como resultado una curva del atributo a lo largo de la trayectoria del mismo. La variabilidad del dato a lo largo de la trayectoria podría ser utilizada, por ejemplo, para explicar diferencias en la producción por etapas que se puede observar en pozos con información de trazadores o PLT's. No obstante, dado que la disponibilidad de dicha información es muy limitada, en esta instancia consideramos el valor promedio del atributo a lo largo de toda la trayectoria como variable a incorporar al modelo. Para tener en cuenta posibles desviaciones de la trayectoria real de cada pozo respecto de la propuesta, que pueden dar lugar a una navegación fuera de nivel en algunos intervalos, se incorporó además información de geonavegación para definir el porcentaje de navegación de cada pozo dentro de cada nivel de interés. De esta manera, el valor final de cada atributo, extraído para cada pozo, es el promedio a lo largo de la curva antes mencionado, ponderado por el porcentaje de navegación por nivel.

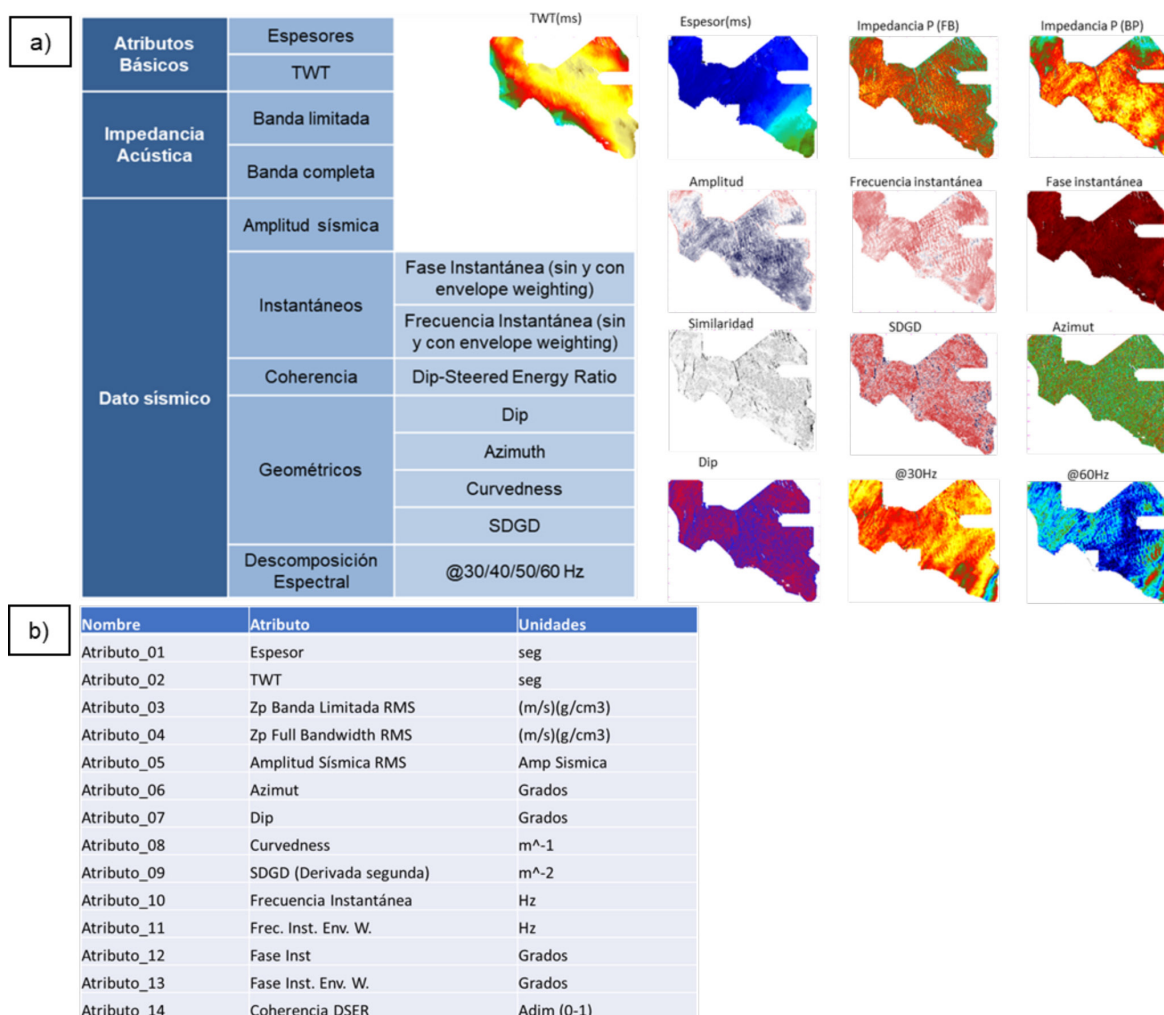


Figura 3. a) Atributos extraídos a partir del cubo sísmico y del cubo de impedancias y b) tabla con la codificación usada para cada atributo sísmico.

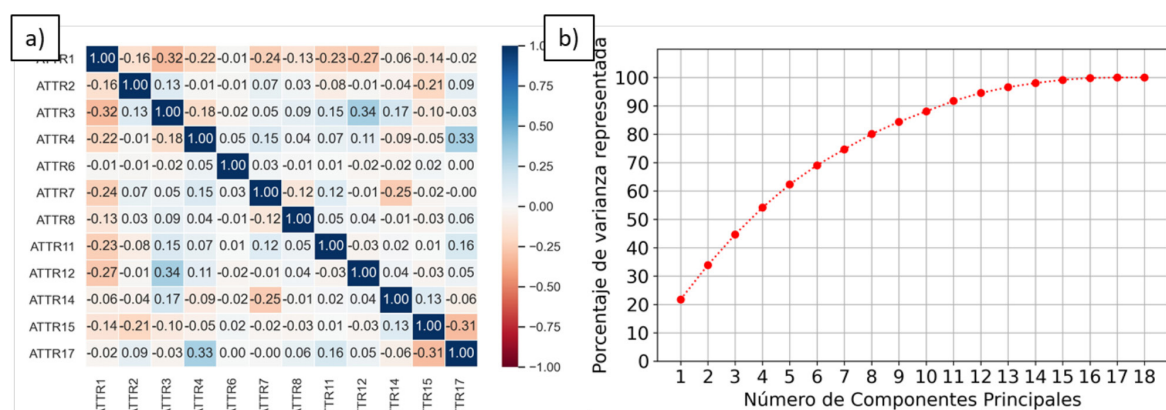


Figura 4. a) Matriz de correlaciones de los atributos sísmicos finales (12) y b) curva de varianza acumulada para las componentes principales calculadas a partir de los atributos sísmicos iniciales (18).

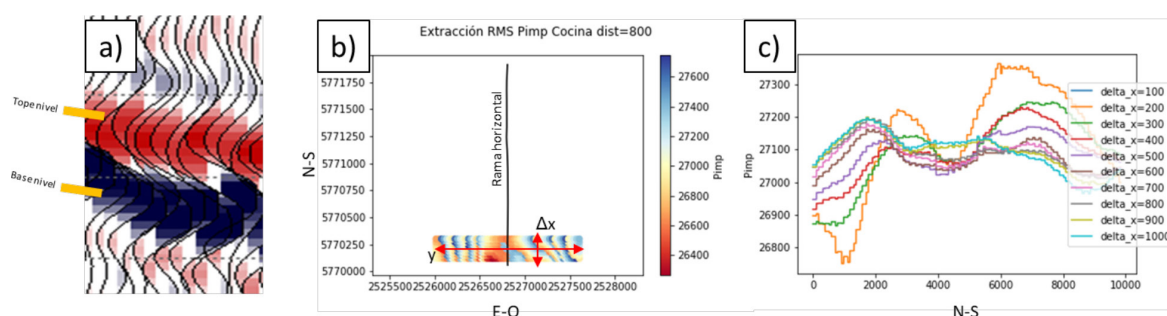


Figura 5. Extracción de atributos por cada nivel a lo largo de la trayectoria del pozo.

Información de producción

Los datos de producción representan la variable objetivo o target del modelo de Aprendizaje Estadístico. En este caso se consideró la producción acumulada de líquido y petróleo en km³ de alrededor de 200 pozos distribuidos en el área cubierta por el *survey* sísmico, con una mayor concentración de ellos hacia la región sudeste. Se contaba con los datos de producción acumulada medidos en distintos momentos: a 120, 180, 270, 365, 545 y 730 días desde el inicio de la misma. En la Figura 6 se muestra la distribución de estas producciones acumuladas por nivel de navegación para el caso de petróleo. En función de la antigüedad de cada pozo, a mayor tiempo de medición, menor cantidad de datos disponibles. Se observa que la distribución de las producciones acumuladas se estabiliza alrededor de los 270 días (no se observan cambios significativos en las distribuciones a partir de allí), por lo que se elige este intervalo de medición como variable target: las producciones a menores tiempos son más dependientes de variables tales como la apertura de cada pozo mientras que a largo plazo predominan las características propias del medio y la extensión alcanzada por la fractura. Se define como variable objetivo o target a la producción acumulada de petróleo a 270 días.

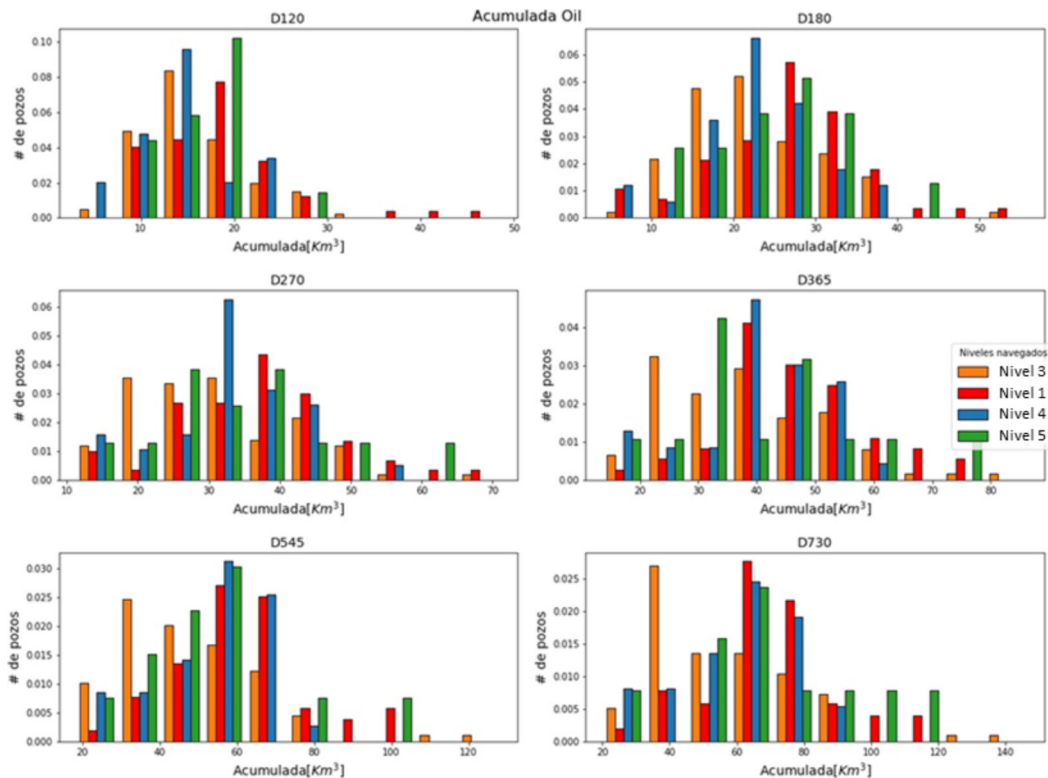


Figura 6. Producción acumulada de petróleo en distintos intervalos de medición para cada nivel de interés.

METODOLOGÍA

Con el conjunto de parámetros de estimulación y el valor promedio de cada atributo sísmico extraído para cada pozo se procedió a generar un único set de datos que constituye la matriz de features con la que se buscará entrenar modelos de Aprendizaje Estadístico. Este data frame está formado, por 12 atributos sísmicos de diferente naturaleza y por 3 componentes principales de los parámetros de estimulación que explican el 90% de la varianza de los mismos. El conjunto de datos se separa en un set de entrenamiento, dado por el 80% del total, con el cual se optimizarán los parámetros del modelo, y un set de testeo, dado por el 20% restante, que se utilizará para evaluar la eficiencia del modelo en la predicción de la variable objetivo: la producción acumulada.

Como modelo de Aprendizaje Estadístico se consideró una regresión lineal múltiple por medio del método LASSO (*Least Absolute Shrinkage and Selection Operator*). Se trata de una alternativa a la clásica regresión lineal por mínimos cuadrados, que incorpora una estrategia de regularización que busca reducir el valor absoluto de los coeficientes del modelo:

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j \cdot x_{ij} \right)^2 + \alpha \sum_{j=1}^p |\beta_j|$$

$$\beta_0, \beta_j: \min \left(\sum_{i=1}^n (y_i - \hat{y}_i)^2 + \alpha \cdot |\beta_j| \right)$$

donde x_{ij} es el j -ésimo feature del modelo en la observación i , β_j son los coeficientes o pesos de cada feature, y_i y $\hat{y}_i$ corresponden a la i -ésima observación de la variable objetivo y a su predicción, respectivamente, y α es el parámetro de regularización del método LASSO, que busca penalizar el módulo de cada coeficiente. Esto último se traduce en un método de selección de los features que efectivamente descartando aquellos que no aportan valor adicional al mismo (casos en los que $\beta_j = 0$). Las cantidades n y p corresponden al número de observaciones y a la cantidad de features, respectivamente. En consecuencia, el método permite reducir la dimensionalidad del problema y minimizar los efectos de sobreentrenamiento.

El valor del hiperparámetro α o parámetro de regularización depende de cada problema y es necesario optimizarlo como parte de la implementación del modelo. Con este fin, se realizó una validación cruzada sobre el set de datos de entrenamiento. Esta consiste en serie de pasos iterativos en los que se realiza una subdivisión del *set* de entrenamiento en dos subconjuntos de entrenamiento y validación y se ajusta el modelo con un valor de α diferente en cada paso. El valor de α con el que se logra el mejor ajuste es considerado el valor final a utilizar en el entrenamiento del modelo.

Existen distintas métricas para evaluar la eficiencia del modelo. En este trabajo consideramos como score del ajuste al coeficiente de determinación o R^2 que se define como

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

donde $\bar{y}$ es la media de las observaciones. R^2 toma valores entre 0 y 1. $R^2 = 1$ corresponde al caso en que el modelo predice exactamente todas las observaciones. En la evaluación de los modelos, consideramos los valores de R^2 obtenidos tanto para el set de entrenamiento como para el de testeo.

RESULTADOS

En primer lugar, se entrenaron modelos que utilizaran la información de estimulación y la información derivada de la sísmica de manera individual, con el fin de analizar la capacidad predictiva de cada tipo de información de manera individual. Posteriormente, se consideró el total de la información para entrenar un modelo único (previa optimización del parámetro de regularización) y analizar el grado de sinergia entre los datos de distinta naturaleza para predecir la producción acumulada de los pozos.

En un primer paso, entonces, se consideró la matriz de features dada por los parámetros de estimulación de cada pozo y se entrenaron modelos de regresión lineal en dos situaciones: con la variable objetivo sin normalizar por la longitud y con la variable objetivo normalizada. Los resultados de ambos casos se muestran en la Figura 7 a) y b), respectivamente. En los paneles superiores se muestra la comparación entre las producciones (curvas nítidas) acumuladas de líquido total (azul) y petróleo (rojo) y los correspondientes datos de testeo (curvas con transparencia). En cada caso, se observa que la curva predicha tiene una semejanza significativa con la de observaciones, sin llegar a predecir los valores más extremos. En el panel inferior en cada caso, se muestra el valor del R^2 obtenido en cada una de las 100 realizaciones (100 modelos entrenados con distintas particiones de entrenamiento y testeo, considerando la acumulada de petróleo como variable target) para el set de testeo (curva azul) y el set de entrenamiento (curva negra). Como se puede observar, en el caso de la variable objetivo sin normalizar por longitud, el modelo alcanza un R^2 de 0.5 en el set de entrenamiento. Es decir, es capaz de predecir un 50% de la información de producción. Por su parte, en el caso de los datos normalizados por longitud, las predicciones no logran representar más que el valor medio de las observaciones, mientras que el entrenamiento del modelo se vuelve inestable, con un R^2 de 0.2 para el set de entrenamiento. Esto refleja que la variable más relevante para predecir la producción de un pozo, en el conjunto de datos estimulación, es la longitud del pozo. Se realizaron modelos de predicción tomando la producción normalizada y sin normalizar, dando como resultado un mejor ajuste para el caso de la producción sin normalizar por longitud estimulada.

Por otro lado, se realizó el ajuste del modelo considerando únicamente la información derivada de la sísmica. Los resultados se muestran en la Figura 8. Los paneles superiores muestran un gráfico de dispersión de producción acumulada de líquido medida versus predicha (izquierda) y los coeficientes ajustados mediante LASSO para cada feature (derecha). El panel central y el panel inferior muestran las observaciones (rojo) y las predicciones (azul) para el set de entrenamiento y testeo, respectivamente. Se puede observar que no se logra un buen ajuste de las predicciones, salvo en su tendencia general, lo cual es verificado por el valor de R^2 que toma el valor 0.27 para el set de entrenamiento y 0.21 para el de testeo. Es decir, la información geológica representada por el conjunto de atributos sísmicos solo es capaz de explicar alrededor del 20% de las producciones de los pozos.

Por último, se incorporó el total de la información al modelo de Aprendizaje Estadístico. En las Figuras 9 a) y b), se muestran los resultados para dos particiones distintas de los datos en entrenamiento y testeo. En los paneles inferiores se observa un importante acople entre las curvas correspondientes a los datos predichos (azul) y observados (rojo), salvo para algunos pozos con valores extremos de producción. El R^2 para estos modelos ronda el valor 0.6, de manera que en conjunto la información de estimulación y la información derivada de la sísmica explican un 60% de las observaciones de los sets de testeo. Esto muestra que existe una importante sinergia

entre ambos tipos de datos, mejorando así la capacidad predictiva cuando son consideradas en conjunto respecto de lo observado en los modelos entrenados con los datos por separado. Podemos notar además que los coeficientes más importantes del modelo corresponden a las componentes principales 1 y 3 de las variables de estimulación. Estas dos componentes principales son las que presentan mayor correlación con la longitud de estimulación. Como ya se mencionó, la longitud estimulada es la variable más importante para predecir la producción acumulada de los pozos.

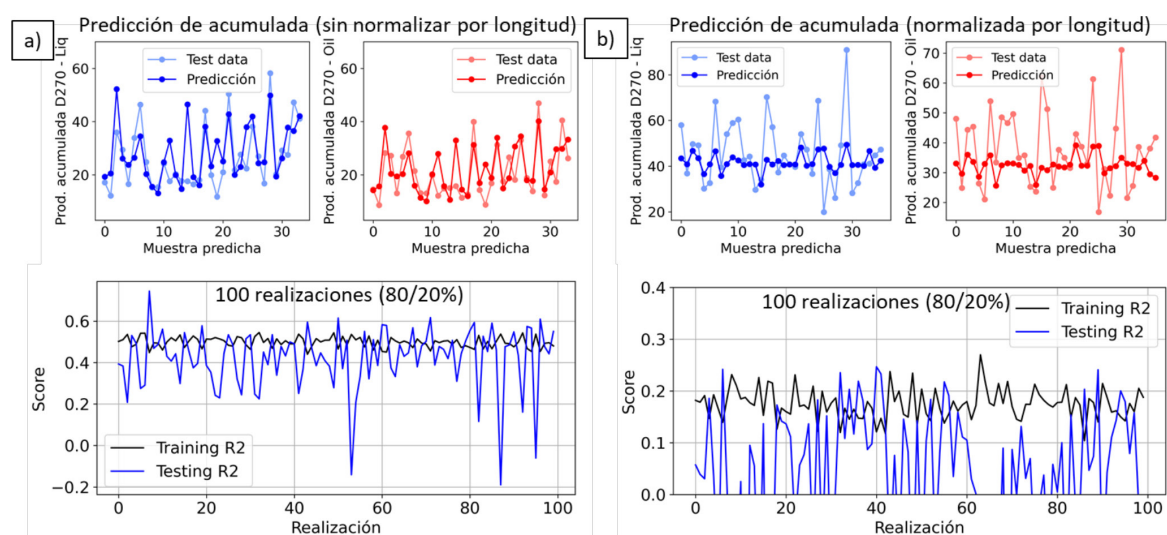


Figura 7. Ajuste de modelos utilizando la información de estimulación asumiendo como variable objetivo a) las producciones sin normalizar por longitud y b) las producciones normalizadas a 2500 m.

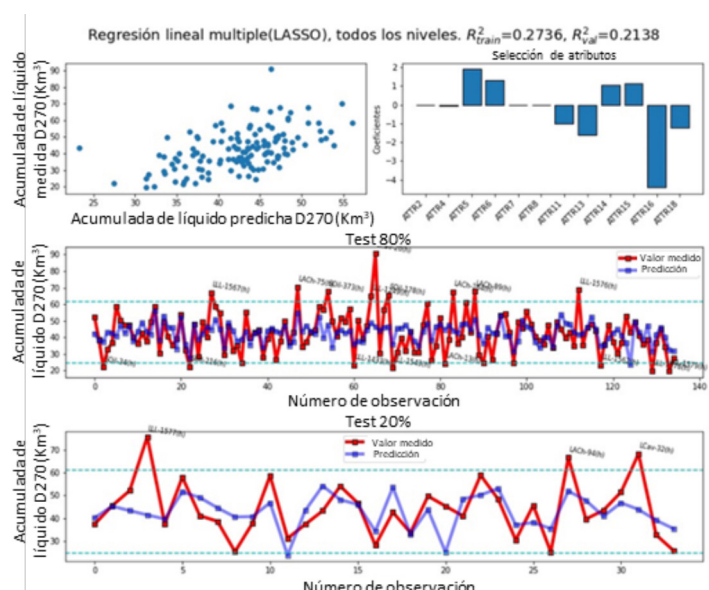


Figura 8. Ajuste de modelo LASSO utilizando únicamente la información derivada de la sísmica.

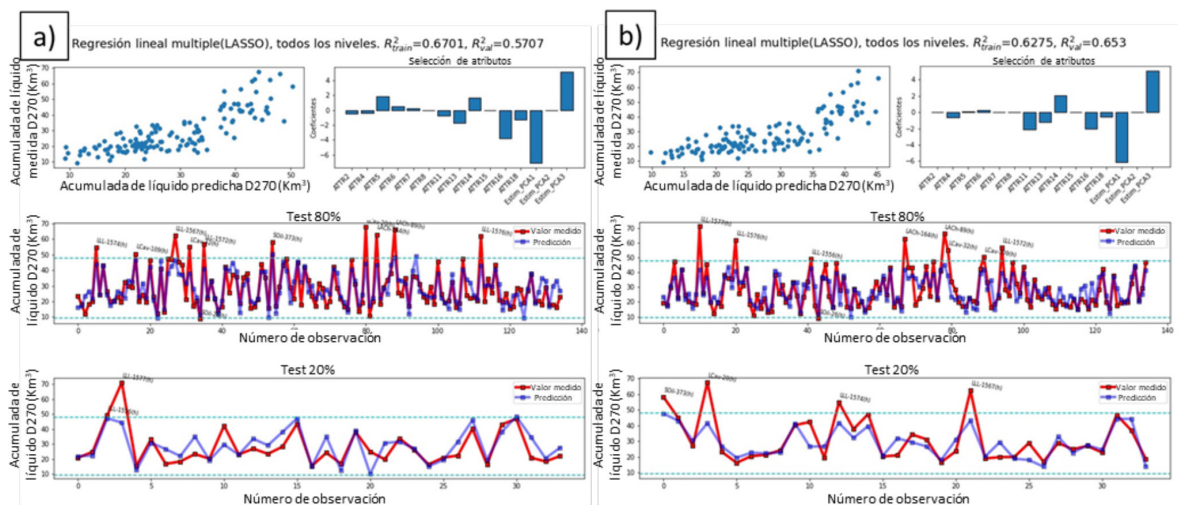


Figura 9. Ajuste de modelo LASSO utilizando conjuntamente la información de estimulación y la información derivada de la sísmica.

DISCUSIÓN

El punto más importante a tener en cuenta al trabajar con modelos de Aprendizaje Estadístico o de *Machine Learning* es la cantidad de datos disponibles, que se asocia directamente a la capacidad de los modelos para generalizar los resultados y evitar caer en el sobreentrenamiento.

En este trabajo se utilizaron alrededor de 200, cuya distribución espacial es heterogénea cubriendo un área muy pequeña en relación al área total definida por la cobertura del *survey* (Figura 2a).

Esto constituye una importante limitación para la construcción de un modelo predictivo con capacidad de generalizar, en consecuencia, se hizo énfasis en la reducción de la dimensionalidad del problema (reducción del número de features que componen el modelo) y se optó por el método LASSO debido a su capacidad de seleccionar variables y reducir la varianza de las predicciones asociada al sobreentrenamiento de los datos. No obstante, es necesario actualizar el modelo en la medida que se cuente con información de producción de pozos con suficiente historia para ser incorporados al análisis. Por otra parte, existe una fracción importante de los datos de producción (alrededor del 40%) que no logra ser explicada por el modelo. Las principales diferencias entre los datos predichos y las observaciones se asocian a pozos con valores de producción altos o bajos en extremo, particularmente aquellos que muestran una significativa variabilidad espacial respecto de pozos cercanos, que no pueden ser explicadas por las variaciones espaciales presentes en el dato sísmico (teniendo en cuenta su resolución espacial) ni por la implementación de diseños de estimulación fuera de los estándares generales. Estos datos de producción particulares deben ser analizados en mayor detalle para descartar errores de medición o cuestiones operativas excepcionales que pudieran generar estas anomalías. Además,

existen otras variables que podrían agregar información y variabilidad al modelo (geomecánica, información de fluidos, anisotropía, presión, etc.) y que se espera puedan ser incorporadas al mismo en la continuidad de este trabajo.

CONCLUSIONES

Se calcularon atributos sísmicos de distinta naturaleza en el área de estudio, bajo la hipótesis de que estos son representaciones de las características geológicas del reservorio. Los atributos fueron calculados para cada nivel de navegación de interés, teniendo en cuenta los datos de geonavegación de cada pozo. Se desarrolló un flujo de trabajo para la extracción de los atributos sísmicos a lo largo de la trayectoria de los pozos horizontales, asumiendo una longitud de fractura y teniendo en cuenta la resolución sísmica para la definición de las dimensiones de extracción. Se trabajó además con información de parámetros de estimulación y se realizaron análisis exploratorios de toda la información disponible. Se observó que la complejidad del dato sísmico es tal que es necesario considerar múltiples atributos para representar la varianza de los datos, mientras que la información de estimulación puede reducirse a 3 componentes principales.

Se testearon modelos de Aprendizaje Estadístico individuales para la información sísmica y la información de estimulación que mostraron una baja capacidad predictiva y un predominio de la longitud de estimulación como variable predictiva. Por último, se entrenaron modelos con ambos tipos de información de manera conjunta mejorando la capacidad predictiva a un 60%, mostrando sinergia entre ellos y un potencial de la metodología para mejorar los modelos incorporando nueva información geológica o de reservorio que no esté capturada en los datos sísmicos.

AGRADECIMIENTOS

Los autores de este trabajo agradecen a YPF S.A. por otorgar los permisos necesarios para mostrar los resultados obtenidos.

REFERENCIAS CITADAS

- | | |
|--|--|
| Gilardone, C., Canel, C.A., Albuquerque, L., Ruiz Benitez, M., Cabello, A., 2021a. Vaca Muerta's Productivity and Economic Performance. 7 years in review. SPE Annual Technical Conference and Exhibition. | Gilardone, C., Albuquerque, L., Ruiz Benitez, M., Cabello, A., Lencinas, S., Dri, J.M. y Sivori, J.I., 2021b. Correlaciones entre los parámetros de producción, petrofísicos y de completación para predecir la productividad de petróleo en la Fm. Vaca |
|--|--|

- Muerta. 2° Jornadas Virtuales Revolución Digital e Innovación Tecnológica para Petróleo y Gas.
- James, G., Witten, D., Hastie, T. y Tibshirani, R., 2021. An Introduction to Statistical Learning with Applications in R. Segunda Edición. Editorial Springer.
- Rice, R., 2014. Identifying the Ideal Stimulation Design via Quantitative Multivariate Analysis: Second Bone Springs Sand, New Mexico. SPE Annual Technical Conference and Exhibition.
- Smith, R., Mukerji, T. y Lupo, T., 2019. Correlating geologic and seismic data with unconventional resource production curves using *Machine Learning*. Geophysics, Vol. 84, P. O39-O47.
- Wang, S. y Chen, S., 2019. Insights to fracture stimulation design in unconventional reservoirs based on *Machine Learning* modeling. Journal of Petroleum Science and Engineering. Vol. 174, P. 682-695.