

# Identificación de fracturas potencialmente abiertas en carbonatos mediante el análisis topológico de datos, (TDA)

*Luis Ahindel Castellanos Bassoult*  
*Pemex Exploración y Producción, STEP*

Artículo recibido en junio de 2023-evaluado-correcto y aceptado en marzo de 2024

## Resumen

Los yacimientos con mayor aporte de producción de hidrocarburos en México, son los naturalmente fracturados; existen registros geofísicos y metodologías que permiten identificar fracturas de forma directa o indirecta (registro de imagen, láminas delgadas, modelo fractal, espectroscopias de rayos gamma, factor fotoeléctrico, análisis de esfuerzos, etc.), el problema para identificar fracturas resulta cuando se necesita proponer intervalos a disparar y no se cuentan con estos registros y/o estudios en los pozos objetivo. La importancia de transformar los datos en información es ayudar a generar modelos que permitan disminuir el riesgo en la toma de decisiones de intervalos que impactan en la producción de hidrocarburos. El análisis topológico de datos (TDA) genera estructura de datos, es decir, se puede utilizar un modelo que mejore la eficiencia en las predicciones y de indicadores. Para este trabajo se utilizó la información de 10 pozos en yacimientos naturalmente fracturados de doble porosidad (intervalos productores en caliza y dolomía); los datos analizados fueron los siguientes: Núcleos, registros de imagen, pruebas de presión, datos de producción, registros de producción, registro sísmico dipolar y relación de aspecto de poro. El enfoque principal de este estudio es el de generar un modelo predictivo de fracturas potencialmente abiertas, utilizando cuatro componentes (relación de aspecto de poro, registro sísmico compresional, relación de Poisson y registro de densidad), para aplicarlo en pozos carentes de registro de imagen y/o estudios especiales, que permita seleccionar intervalos a producción con menor riesgo en yacimientos naturalmente fracturados de calizas y dolomías.

**Palabras clave:** Análisis topológico de datos, producción de hidrocarburos en México, yacimientos naturalmente fracturados de doble porosidad.

## Potentially open fractures identification in carbonates through topological data analysis (TDA)

### Abstract

The most hydrocarbon-productive reservoirs in Mexico are naturally fractured; there are geophysical logs and methodologies that allow the direct or indirect identification of fractures (image logging, thin sections, fractal models, gamma-ray spectroscopies, photoelectric factor, stress analysis, etc.). The problem arises when it is necessary to propose intervals to be stimulated and there are no such logs and/or studies in the target wells. The importance of transforming data into information is to help generate models that reduce the risk in decision-making regarding intervals that impact hydrocarbon production. Topological data analysis (TDA) generates data structures, meaning that a model can be used to improve efficiency in predictions and indicators. For this work, information from 10 wells in naturally fractured double-porosity reservoirs (producing intervals in limestone and dolomite) was used. The analyzed data included: cores, image logs, nuclear magnetic resonance logs, pressure tests, production data, production logs, dipolar sonic log, and pore aspect

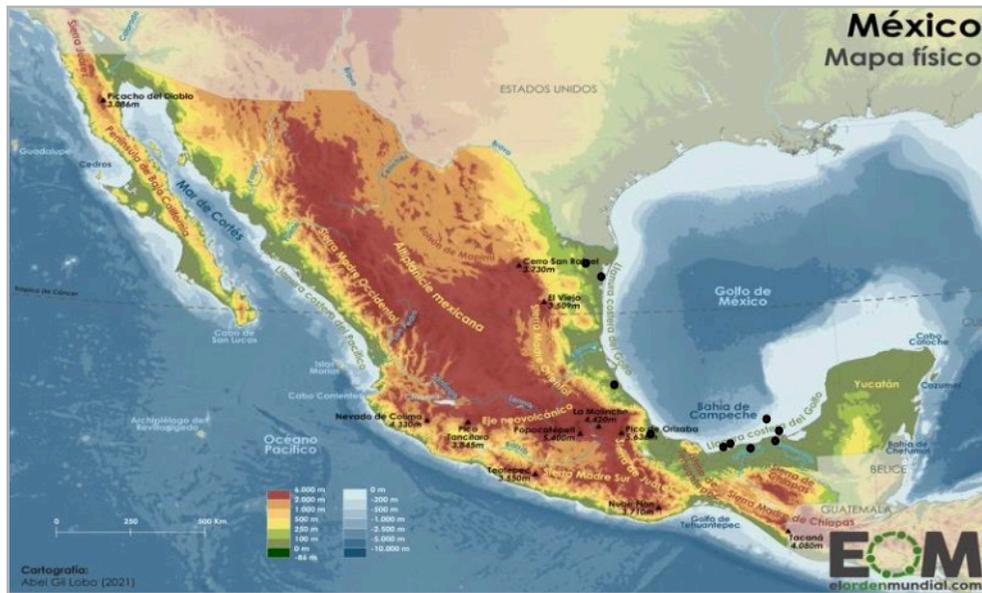
ratio. The focus of this study is to generate a predictive model of potentially open fractures, using four components (pore aspect ratio, compressional sonic log, Poisson ratio, and density log), to apply it in wells lacking image logs and/or special studies, allowing for the selection of production intervals with lower risk in naturally fractured limestone and dolomite reservoirs.

**Keywords:** Topological data analysis, hydrocarbon production in Mexico, naturally fractured double porosity reservoirs.

## Introducción

La distribución espacial tridimensional de las propiedades petrofísicas está controlada por los procesos deposicionales y diagenéticos. La diagénesis típicamente reduce la

porosidad, redistribuye el espacio poroso y altera la permeabilidad y características de capilaridad. La porosidad de los yacimientos de carbonatos en México oscila entre 1 y 17%, **Figura 1**. El proceso diagenético estudiado en este trabajo es el fracturamiento.



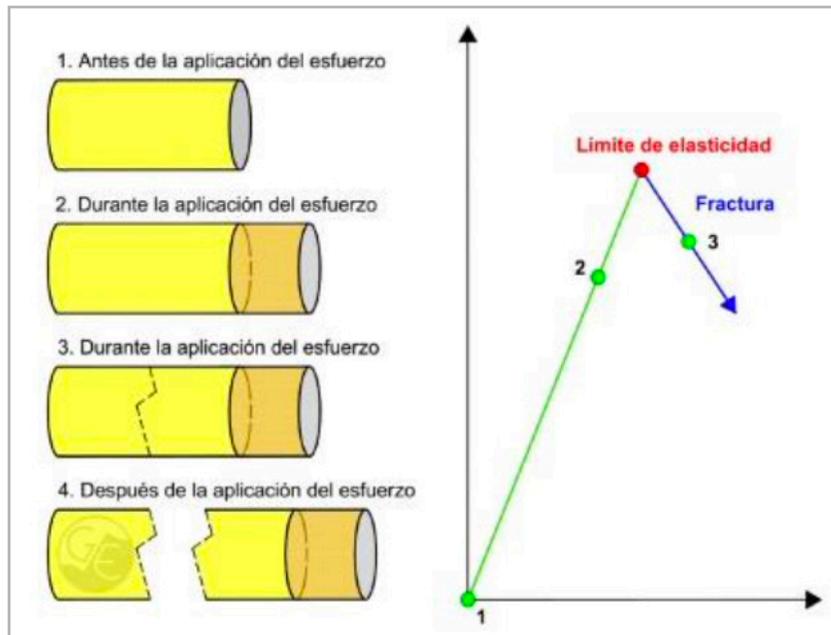
**Figura 1.** Distribución de los yacimientos analizados en carbonatos para este estudio (puntos en color negro). Fuente EOM.

Actualmente existen muchas metodologías que intentan predecir zonas de fractura sin la necesidad de tomar algún registro especial y/o corte de núcleo; la diferencia principal de este estudio radica en el uso del análisis topológico de datos (TDA) y de yacimientos productores naturalmente fracturados del país. Los campos estudiados presentan información y producción en carbonatos fracturados (caliza y dolomía). Es decir, sólo se seleccionaron pozos productores con información especial que de manera directa indicara fracturas potencialmente abiertas y con un ángulo mayor a 80°.

Es importante entender los conceptos por los cuales se rige el fracturamiento, primero conociendo los elementos que afectan la porosidad en función de la textura.

Una fractura es cualquier superficie de discontinuidad producida por la rotura de una masa rocosa (fragilidad). Existen dos tipos principales de fracturas: las fallas y las diaclasas; las fallas son fracturas con un desplazamiento relativo de los bloques que separan, mientras en el caso de las diaclasas no existe un desplazamiento apreciable.

Un material se comporta de forma elástica siempre y cuando cumpla la ley de Hooke, es decir, el material se deforma proporcionalmente al esfuerzo aplicado hasta un punto llamado límite elástico (Popov & Balan, 2000), a partir del cual, el material se fractura. (Ver **Figura 2**).



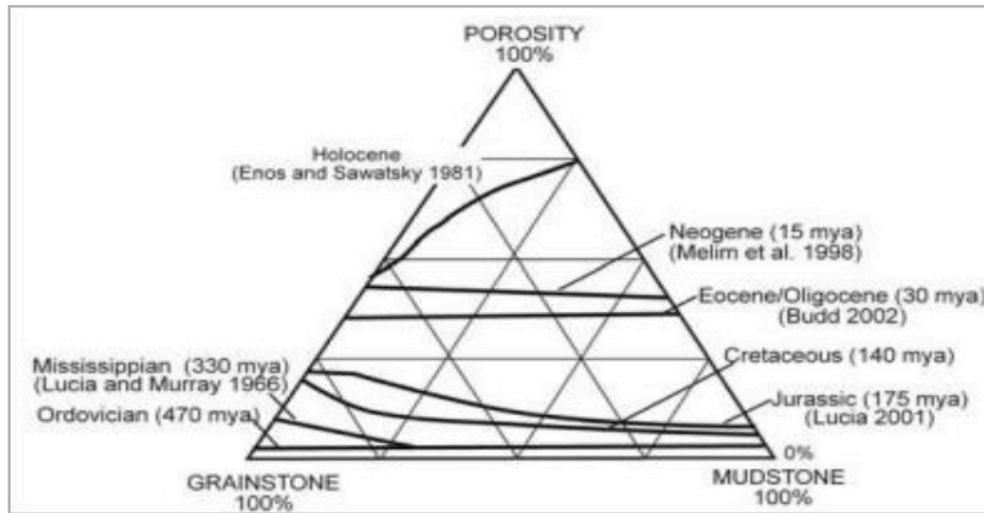
**Figura 2.** Descripción de límite de elasticidad y fractura.

Si el material es sometido a un esfuerzo que no sobrepasa su límite elástico, al cesar la aplicación, el material recuperará su forma inicial sin deformación evidente. Lo anterior es particularmente importante dado que las estructuras geológicas que se observan son aquellas resultantes de esfuerzos que pudieron efectivamente ocasionar una deformación permanente en las rocas; aquellos esfuerzos que ocasionaron deformación elástica en la roca, sin sobrepasar su límite de elasticidad.

Ahora es importante asociar, la porosidad y textura de la roca con el fracturamiento, para ilustrar la pérdida gradual de porosidad con el tiempo, se muestra la **Figura 3** en la

que se incluye un diagrama ternario. La porosidad del 70% que se encuentra en el lodo dominado por sedimentos se reduce después de unos 10-20 millones de años de compactación y cementación, pero todavía son comunes valores de porosidad entre 30 y 40%.

Después de unos 150 millones de años, las calizas dominadas por granos del Jurásico conservan 25% de porosidad mientras que la porosidad de los sedimentos dominados por lodo se ha reducido significativamente. En el Paleozoico Superior, solo predominan lo suficiente para retener porosidad; las calizas paleozoicas son normalmente densas.



**Figura 3.** La relación entre porosidad, textura deposicional y edad geológica. El ritmo de pérdida de porosidad en texturas dominadas por lodo, es más rápido que en texturas dominadas por granos.

La dolomitización se identifica por la presencia del mineral dolomita, y la mineralización de evaporita por la presencia de minerales de evaporita como anhidrita, yeso y halita. La disolución masiva se refiere a la formación de un espacio poroso relativamente grande, sin tener en cuenta la estructura de la roca, y los vacíos pueden presentar una magnitud importante como para colapsar, formando una brecha de colapso y patrones de fractura asociados. Muchos factores relevantes pueden controlar las fracturas de los carbonatos, tales como la deformación regional y local, la inherencia de las fallas más cercanas, la litología, porosidad, estratigrafía mecánica y diagénesis (Milad, 2017; Milad, Ghosh and Slatt, 2018). Los controles estructurales están directamente relacionados con la evolución cinemática de los pliegues.

La diferencia principal de este análisis respecto a otros para identificar fracturas potencialmente abiertas es que se usaron datos de registros especiales en zonas con estabilidad de agujero; se consideró la información de registros convencionales (sónico y densidad) en intervalos productores calibrando con registro de imágenes, resonancia magnética, plt, sónico dipolar, datos de núcleos, entre otros, para identificar formas en los datos y asociarlas

a las zonas fracturadas. Este enfoque permitió establecer una relación entre los datos y las fracturas potencialmente abiertas en yacimientos de carbonatos.

### Fundamentos teóricos para determinar fracturas a partir del análisis topológico de datos (TDA)

#### Análisis de datos

Por el principio físico de cada herramienta es probable identificar zonas fracturadas; son muy evidentes cuando se trata de fracturas con una apertura grande, esto también se observa con la apertura en el diámetro del caliper (washouts) y/o tirones en las lecturas de las herramientas. Se cuenta con una gran información de registros de imagen, núcleos y PLT en intervalos productores de hidrocarburos. Por lo anterior, es importante realizar un análisis de los datos en estas zonas y generar un código que permita poder predecir zonas de fracturas con un riesgo bajo.

Se hizo uso de la metodología de la ciencia de datos aplicada por Rollins, director de IBM, para mejorar cada uno de los criterios en selección de datos dentro del yacimiento, **Figura 4.**

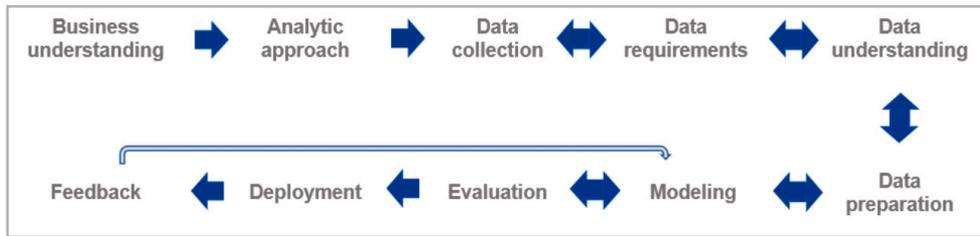


Figura 4. Metodología de la ciencia de datos de Rollins.

Los modelos no se crean una vez, se implementan y se dejan tal como están; en cambio, están constantemente mejorando y adaptándose a las condiciones cambiantes a través de comentarios, ajustes y redespiegues. Tomando en consideración esto, se tiene que entrenar y desplegar el modelo en pozos con información e identificar el porcentaje de error en la predicción.

La mayor cantidad de datos analizados se enfocaron en los registros sísmico compresional (DTC), relación de Poisson (PR), densidad (RHOB) y relación de aspecto de poro (AR), en

las zonas de fracturas abiertas. A continuación, se describen cada uno de estos elementos:

- Registro de densidad (rhob): Es una herramienta de contacto con una profundidad de investigación de 6-9 in. La cual utiliza Cesium radioactivo como fuente, **Figura 5**, que presenta un decaimiento descrito por la reacción siguiente:

Mide la densidad de electrones ( $\text{g}/\text{cm}^3$ ).

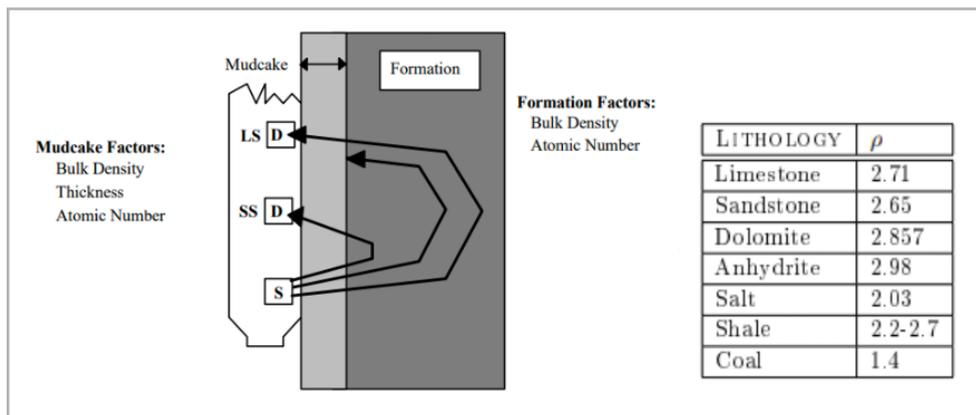
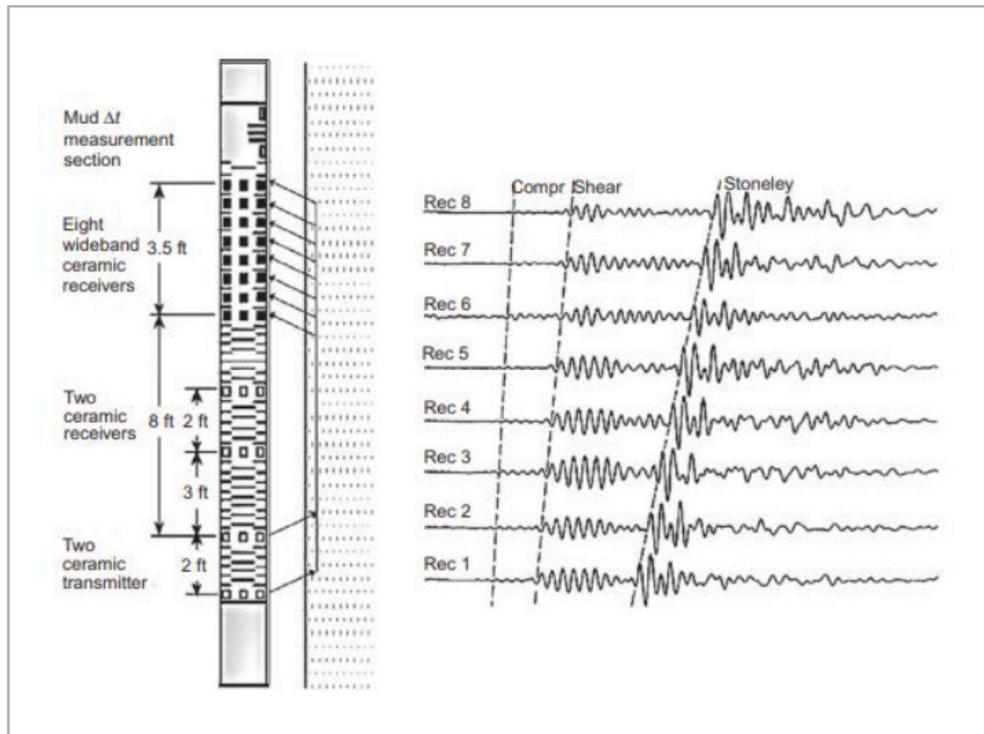


Figura 5. Factores que afectan las medidas del registro de densidad.

- Registro sísmico (dtc): Herramienta centralizada, el uso principal de la herramienta ha sido determinar un valor para la velocidad de compresión de la formación ( $V_p$ ) y la inversa,  $1/V_p$ , que se denomina lentitud. El subíndice “p” se refiere a la onda “primaria” (de compresión). Asimismo, un subíndice “s” denotaría la onda “secundaria” (de

corte), **Figura 6**. Otros términos para la lentitud son delta-t ( $D_{tp}$ ), “tiempo de tránsito de intervalo”, y simplemente “tiempo de tránsito”. La lentitud tiene la unidad de microsegundos/ft o microsegundos/m. La lentitud ( $D_{tp}$ ) es una función de la porosidad de la formación, lentitud del fluido poroso y lentitud de la matriz de formación.



**Figura 6.** Configuración de herramientas sónicas y formas de onda registradas de diferentes receptores (derecha) en una formación rápida.

- La onda en modo compresional viaja en propagación paralela a la dirección del desplazamiento de las partículas. Después de que se desplaza una distancia muy pequeña, cada partícula tiende a volver a su posición inicial debido a la elasticidad del medio. Sin embargo, si la inercia de las partículas hace que se excedan de su posición de reposo, se crean zonas de rarefacción. La propagación de esta compresión y rarefacción constituye una onda de compresión.
- La relación de Poisson ( $\nu$ ) es una función decreciente del parámetro de densidad de fracturas, es la relación entre la deformación relativa de contracción, o deformación transversal (normal a la carga aplicada), y la deformación relativa de extensión, o deformación axial (en la dirección de la carga aplicada). El cual está en función del módulo de Bulk ( $K$ ) y Shear ( $G$ ), se calculan:

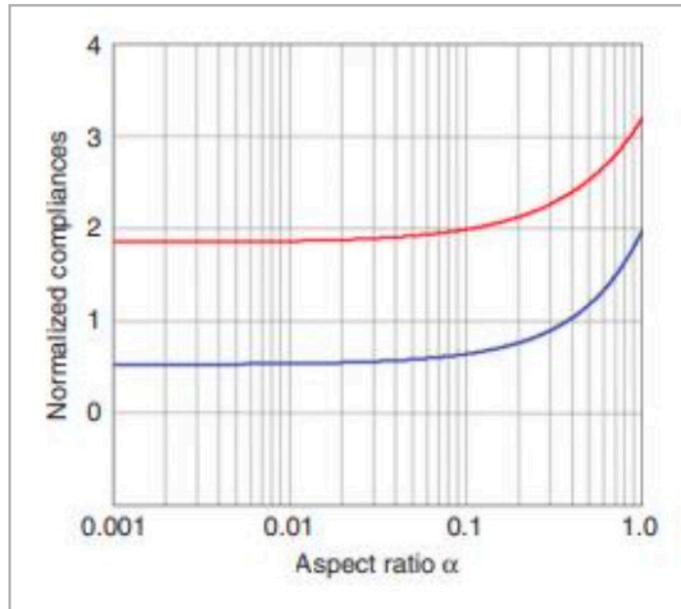
$$G_{dyn} = \frac{\rho_b}{DTCO^2}$$

$$K_{dyn} = \frac{\rho_b}{DTCO^2} - \frac{4}{3}G_{dyn}$$

$$\nu_{dyn} = \frac{3K_{dyn} - 2G_{dyn}}{6K_{dyn} + 2G_{dyn}}$$

- Relación de aspecto de poro (AR). El criterio que una inclusión sea “débil” depende de su forma o relación de aspecto, así como de los módulos relativos del material de inclusión y matriz. La forma correcta de

examinar la dependencia de la relación de aspecto es mantener fijos los diámetros de las grietas, y no su volumen, como se cambia la relación de aspecto, (como en la **Figura 7**).



**Figura 7.** Las curvas se vuelven casi horizontales en relaciones de aspecto inferiores a 0.10, lo que indica insensibilidad de cumplimiento de poros a  $\alpha$  en este rango.

- Los poros en las rocas suelen tener formas muy irregulares, no elipsoidales, que pueden causar que los resultados basados en elipsoides sean inaplicables o engañosos. El factor de concavidad tiene un papel importante: las formas cóncavas producen contribuciones de aporte mucho mayores que las convexas del mismo orden.

también es útil en otros casos, como series temporales. Una serie de tiempo es una secuencia de datos observados, medidos o recopilados en intervalos regulares a lo largo del tiempo. Un registro geofísico se puede asumir como una serie de tiempo, analizar estas series de tiempo puede proporcionar información importante sobre las propiedades del subsuelo y los procesos geológicos que ocurren en un área determinada.

### Análisis topológico de datos (TDA)

El objetivo del análisis topológico de datos (TDA) es proporcionar métodos matemáticos, estadísticos y algorítmicos bien fundamentados, para explotar las estructuras geométricas topológicas y subyacentes en los datos, **Figura 8**. Por lo general, es adecuado para datos tridimensionales, pero la experiencia muestra que TDA

La información es un conjunto de datos ya procesados, la cual ayuda a obtener conclusiones y tomar decisiones; los datos no. Por eso la importancia de analizar estos datos y llevarlos de un punto a otro. TDA es una colección de métodos de análisis de datos que encuentran la estructura en los datos. Estos métodos incluyen PCA, reducción de dimensión no lineal entre otros. Ver Figura 8.

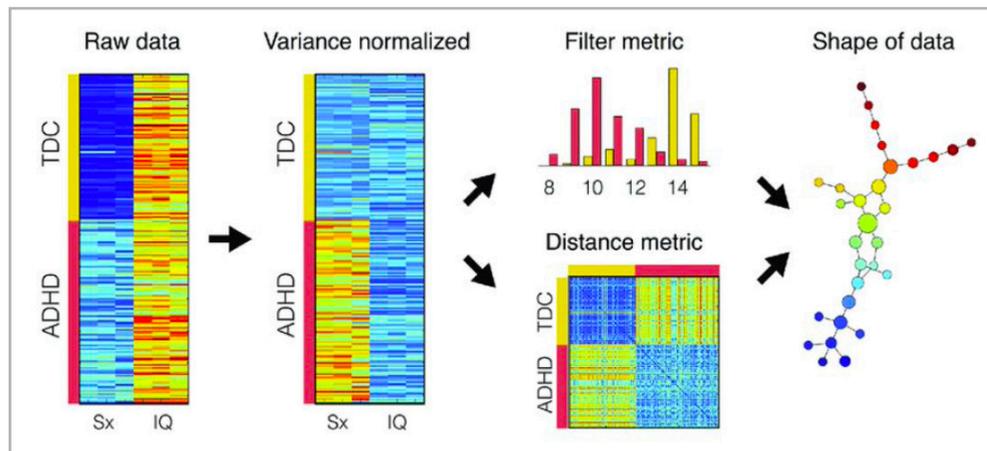


Figura 8. Flujo de trabajo del TDA para obtener información.

TDA usa herramientas algebraicas y captura la información dando forma a los datos. Es decir, se puede utilizar un modelo que mejore la eficiencia en las predicciones. Una de las ventajas principales es que, en el espacio de puntos, estas señales no se distorsionan por información irrelevante. Si considera una figura y se colocan dos puntos cualesquiera sobre ella, esos puntos van a estar a cierta distancia uno del otro. Por ejemplo, la distancia del trabajo a casa es un ejemplo de posicionar dos puntos sobre una esfera (la tierra). Dependiendo de la transformación, esa distancia puede hacerse más grande o pequeña. Por ejemplo, si la Tierra fuera de plastilina, y se le convirtiera en otra figura, entonces la distancia entre la casa y el trabajo evidentemente cambiaría, para un topólogo, un círculo es lo mismo que una elipse, una esfera es lo mismo que un cubo, y se aplica un razonamiento similar para otras figuras.

Muchos métodos de clasificación utilizan la “distancia” como forma de afirmar si existe una similitud fuerte o no entre varias observaciones. Por lo tanto, si se deforma el plano de forma correcta, es posible que puedan encontrar patrones en los datos que de otra forma estarían ocultos para nosotros y para los algoritmos. Entonces, estas nubes enormes de datos pueden ser subdivididas en clústeres más pequeños y cada uno, unirse con sus vecinos en relación con la distancia que exista entre ellos, formando así, lo que en topología se conoce como complejos simples.

## Generación de funciones

El modo de extracción de características encuentra las correlaciones para describir los patrones de entrada. Por lo

tanto, se selecciona un conjunto reducido de características que discrimina diferentes grupos, en este caso los datos de registros de las zonas con fracturas.

La clasificación utiliza los grupos definidos en intervalos de entrenamiento para asignar un punto de datos desconocido a los datos más cercanos al grupo. Por ello, primero se identifican aquellos datos de  $\rho_{ob}$ ,  $pr$ ,  $d_{tc}$  y  $ar$ , que discriminan mejor. Para esto, se deben encontrar características que mantienen los grupos de datos lo más compactos posible y en una distancia máxima entre sí en zona de fractura. Es decir, de todos los campos analizados se extrajo únicamente la zona en donde habían fracturas potencialmente abiertas.

Primero, se necesita una manera de transformar un conjunto discreto de puntos en un espacio topológico más rico para tener invariantes topológicas interesantes para calcular. Segundo, se debe tener en cuenta la escala de las características de los datos; es decir, se necesita determinar la relación entre el tamaño de las características geométricas significativas de los datos y las distancias entre los puntos muestreados.

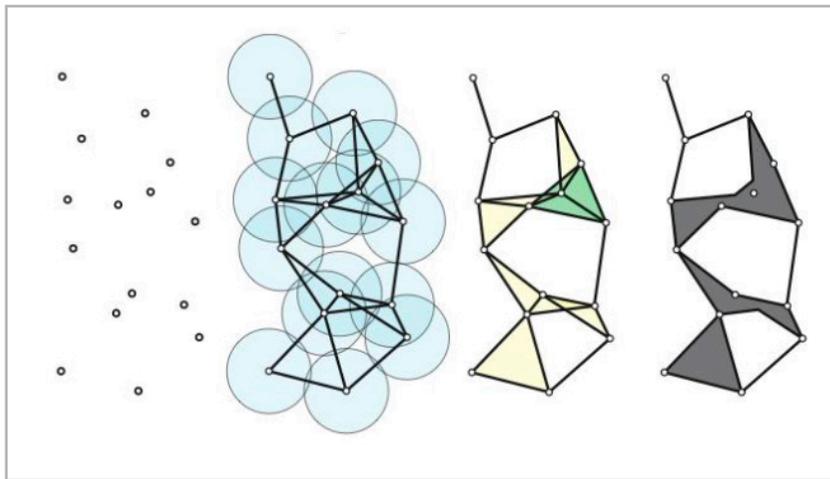
A los registros se les aplica la homología persistente (básicamente, un gráfico formado por un conjunto de puntos y sus relaciones, también conocido como líneas y triángulos en un espacio bidimensional, **Figura 9**), para identificar patrones topológicos en los datos. Se muestra un ejemplo simplificado (registro de densidad):

- Primero se necesita convertir los datos en un formato adecuado para el análisis topológico.
- Preprocesamiento de datos: se puede considerar una serie de tiempo bidimensional donde la profundidad es una dimensión y la densidad es la otra dimensión.
- Creación de complejo simplicial: utilizando estos puntos como vértices, se puede construir un complejo simplicial que capture la conectividad entre los puntos. Por ejemplo, se puede conectar cada punto con sus vecinos más cercanos para formar un complejo simplicial.
- Cálculo de homología persistente: aplica algoritmos de homología persistente para identificar los

componentes conectados y los agujeros en el complejo simplicial. Esto dará información sobre la estructura topológica de los datos geofísicos en función de la profundidad y la densidad.

En Python utilizando la biblioteca ripser para el cálculo de homología persistente y matplotlib para visualización.

Generalmente, la construcción de esta filtración se basa en la definición de una función de distancia, cuyos valores se utilizan para indexar los complejos de la familia.



**Figura 9.** Construcción visual del complejo simplicial.

Visualmente, en dos dimensiones, se considera cada punto de forma independiente: se dibuja un círculo de radio creciente alrededor de esos puntos hasta que se encuentran algunas intersecciones, (densidad). Se persigue el dibujo hasta que se cubren algunos puntos y se destruyen algunas de las estructuras creadas previamente (fracturas). Finalmente, se termina con componentes (densidad y fractura), que se representan a través de ese diagrama de persistencia y código de barras.

El problema de la falta de uniformidad de los vectores extraídos de los diagramas de persistencia es un verdadero inconveniente para las aplicaciones de aprendizaje automático. Para lidiar con esto sin profundizar necesariamente en el nuevo tipo de redes neuronales, se tienen que transformar las

representaciones de persistencia anteriores en algo que recopile de alguna manera la misma información, que puede ser considerable. Para el aprendizaje automático, la extracción de los puntos de persistencia periféricos podría representar funciones excelentes, pero parte de la información subyacente contenida en el diagrama podría perderse a través de ese proceso.

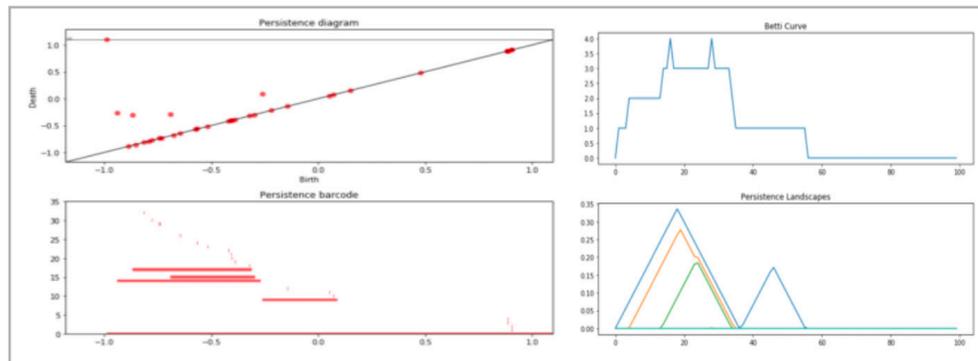
Aquí se subrayan tres posibilidades: las curvas de Betti, los paisajes de persistencia y las imágenes de persistencia, (todos los cálculos se realizaron empleando el paquete Gudhi, desarrollado por el equipo francés DataShape de INRIA). Para utilizar Gudhi en Python, primero se debe instalar la biblioteca. Se puede hacer mediante pip: pip install gudhi.

De manera resumida:

- **Curvas de Betti:** gráficos que muestran cómo varían los números de agujeros (o cavidades) de diferentes dimensiones en un conjunto de datos a medida que se aumenta un parámetro de filtrado.
- **Paisajes de persistencia:** Los paisajes de persistencia son representaciones visuales de la homología persistente en forma de funciones escalares bidimensionales. Muestran cómo cambian las características topológicas de los datos en función de dos parámetros: el tiempo de vida de la característica y la escala en la que aparece.
- **Imágenes de persistencia:** Las imágenes de persistencia son representaciones visuales de la

homología persistente en forma de matrices de colores. Cada píxel de la imagen corresponde a una característica topológica y su color representa el tiempo de vida de esa característica.

Primero, se observa gráficamente a través del diagrama de persistencia y el código de barras, que hay cuatro componentes resaltados, los registros dtc, pr, ar y rhob. Sin embargo, uno de los componentes se coloca en el infinito: esto corresponde al primer punto encontrado mientras se consideran los niveles superiores de la filtración, automáticamente se le asigna un valor infinito. Como consecuencia, sus paisajes de persistencia solo representan claramente los tres registros. Ver **Figura 10**.



**Figura 10.** Se muestra curva de Betti, paisajes de persistencia, código de barras e imagen de persistencia.

Se debe notar el punto siguiente: Incluso si el número de registros aparece claramente (correspondientes a los puntos alejados de la diagonal), permanece una noción de ruido topológico (puntos cercanos a la diagonal). Como consecuencia, los paisajes de persistencia son generalmente de mejor uso, ya que se mantienen más o menos robustos, ya que describen preferentemente los objetos más persistentes del conjunto de datos. No obstante, el ruido topológico es a veces información útil, típicamente cuando se trabaja con series de tiempo cuya desviación estándar es de gran importancia para fines de clasificación.

### Flujo de trabajo

El poder correlacionar los registros geofísicos con metodologías estadísticas, permitió desarrollar un modelo predictivo de zonas con fracturas potencialmente abiertas en yacimientos compuestos principalmente por caliza y dolomía, según fuera el caso; para ello se analizaron los diferentes tipos de registros convencionales y especiales con lo que contaban cada uno de los diez yacimientos estudiados. Para identificar las fracturas potencialmente abiertas se utilizó el diagrama mostrado en la **Figura 11**.

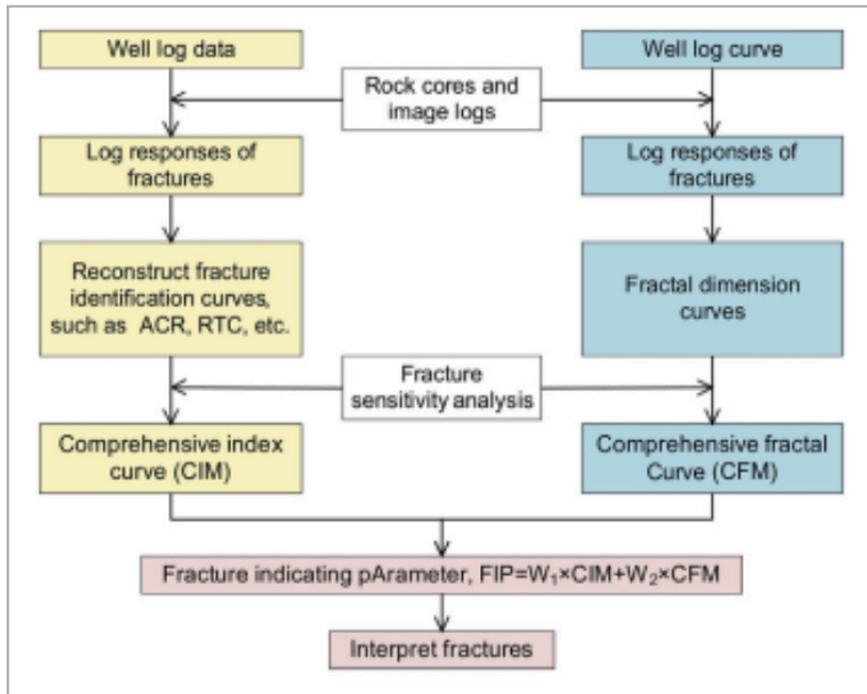


Figura 11. Interpretación de fracturas utilizando registros de pozo.

Al tener sólo los datos deseados en las zonas de fracturas y ordenados por pozo, se hizo uso de la plataforma GUDHI, para llevar a cabo el análisis topológico de los datos. Ver Figura 12.

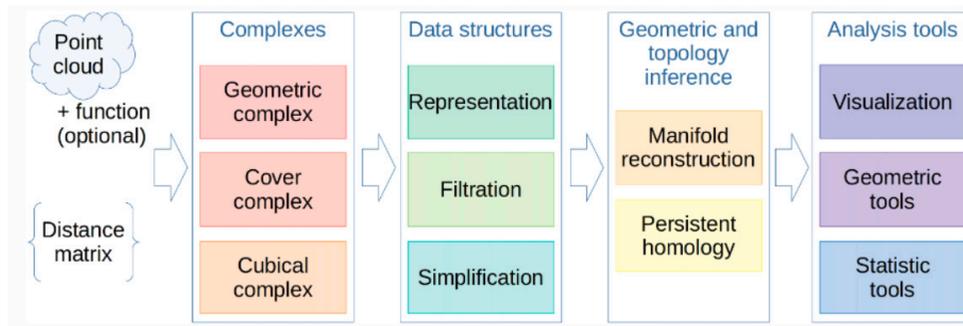
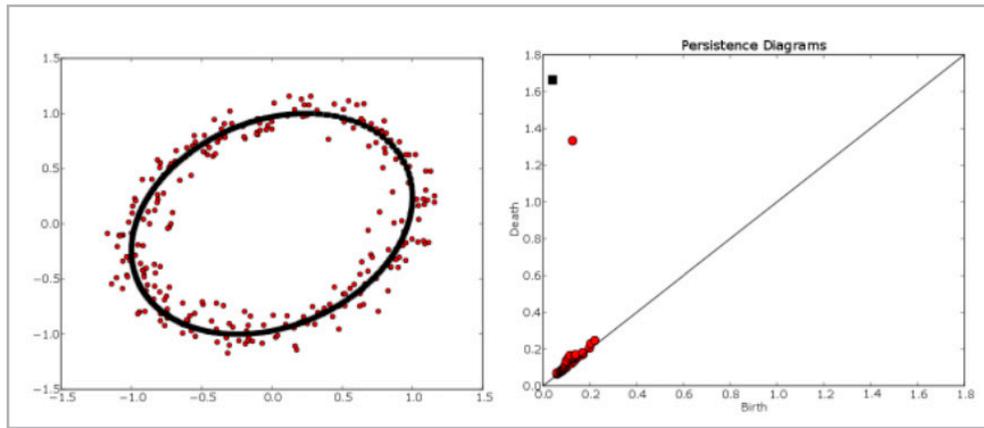


Figura 12. La biblioteca Gudhi es una biblioteca de código abierto para la Topología Computacional y el Análisis Topológico de Datos (TDA).

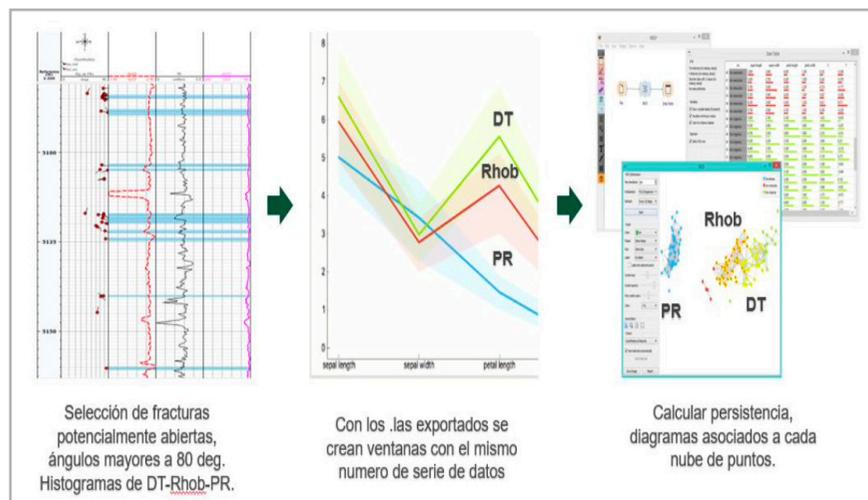
En la plataforma se cargan los datos en donde se obtiene una colección de datos (nube de puntos) de los diferentes registros para definir la distancia (por tipo de yacimiento). El objetivo es identificar coincidencias dentro de la nube de datos. Ver **Figura 13**.



**Figura 13.** Comportamiento de registro en zona fractura.

Los datos reales siempre son finitos, por lo que su estudio requiere que se tenga en cuenta la estocasticidad. El análisis estadístico nos da la capacidad de separar las características verdaderas de los datos de los artefactos introducidos por el ruido aleatorio. La homología persistente no tiene un mecanismo inherente para distinguir entre características de baja probabilidad y características de alta probabilidad.

En la **Figura 14**, se puede observar el flujo de trabajo por las 3 variables resultantes Dtc, Rhob y PR para identificar la estructura topológica por variable y los límites de éstas.



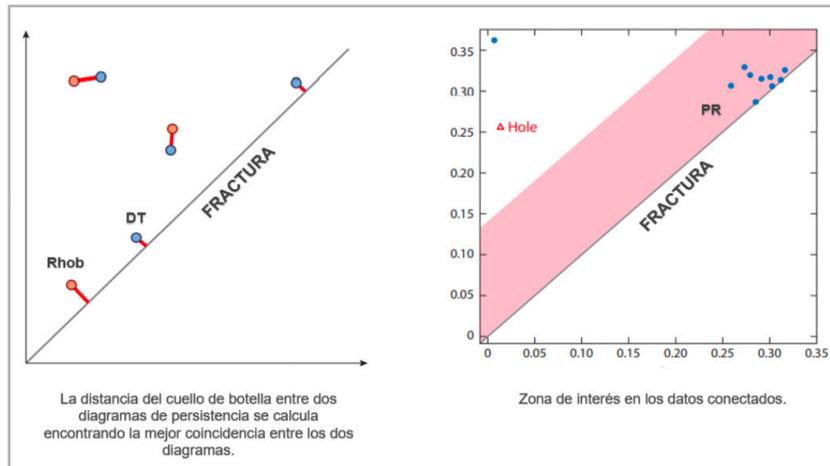


Figura 14. Análisis topológico por variable analizada.

Gran parte del Análisis Topológico de Datos se basa en la noción de que existe una idea de proximidad entre estos puntos de datos. Por ejemplo, si cada punto de datos consiste en valores numéricos  $x = \{x_1, \dots, x_n\}$ , se tiene una definición fácil de proximidad que proviene de la distancia euclidiana estándar: ésta es la generalización de la distancia estándar en el plano  $d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$ .

La distancia euclidiana proporciona un buen punto de partida intuitivo para los requisitos de una distancia generalizada en el sentido matemático. Nótese que la distancia euclidiana asume que las entradas son numéricas. Sin embargo, tener datos completamente numéricos no es un requisito para definir una distancia generalizada. También se pueden definir muchas distancias diferentes cuando los datos son categóricos en lugar de numéricos. Esto podría hacerse simplemente mirando las coincidencias (definir la distancia entre puntos de datos por el número de entradas que son iguales), o incluyendo una vista más matizada de las entradas categóricas.

Complementando el TDA y haciendo uso de los diferentes tipos de registros de imagen (base aceite) se realizó un análisis más detallado en los corredores de fracturas posibles, los cuales son grupos de fracturas relacionadas con fallas que atraviesan todo el yacimiento verticalmente, extendiéndose por varias decenas o cientos de metros lateralmente, de los 10 pozos.

- La permeabilidad en estos corredores puede oscilar muy por encima de los 10 darcys.

- Fracturas con ángulo de inclinación mayores a 80 grados.

Adicional, se programó un exponente de cementación “m”, el cual se ve afectado por la forma, clasificación y empaquetamiento del sistema de partículas, configuración y tamaño de los poros, tortuosidad, tipo de sistema poroso (intergranular, intercrystalino, vóculos y fracturas), compactación debido a la presión de sobrecarga, la presencia de minerales arcillosos y la temperatura del yacimiento.

Dada su importancia dentro del modelo petrofísico y como está relacionado con las fracturas se procedió a incluir un modelo de “m” basado en la textura de la roca, con los siguientes códigos:

La generación de “m variable”, se compone principalmente de wackestone-packstone. Los poros entre partículas generalmente están bien conectados y contribuyen de manera efectiva al flujo de fluidos.

## Resultados

Al tener definidas las estructuras de datos e información de los límites de nuestras variables de entrada de los 10 pozos estudiados, se procedió a desplegar el modelo en nuevos casos de estudio. Es necesario monitorear qué tan bien funciona. Para ello se corrió en Python la integración del lenguaje que permitiera extrapolar el análisis de datos (rango TDA) en pozos que no contaran con información de análisis de fracturas. Ver **Figura 15**.

```

# Estimacion del Fracturamiento >90 deg a partir de registros convencionales #
# Basado en trabajo de datos Topologicos #
#.....PEMEX 2022.....#
LOOP:
  if dtc != MissingValue and pr != MissingValue and rhob != MissingValue:
    #..... Fracture From PR .....#
    if pr > .24 and pr <= .33:
      f_pr=1
    else:
      f_pr=0
    #.....Fracture From DTC.....#
    if dtc > 48 and dtc <= 49.9:
      f_dtc=1
    else:
      f_dtc=0
    #..... Fracture From RHOB .....#
    if rhob > 2.71 and rhob <=2.75:
      f_rhob=1
    else:
      f_rhob=0
    #..... Fracture from Rugosity .....#
    if f_pr=1 and f_dtc=1 and f_rhob=1 :
      f_rugosity= "Fracturas"
    else:
      f_rugosity= "Compacto"
    #..... para dolomias.....#
    if pore != MissingValue:
      mdolo= (2.38-(pore)/(1.048+.001))
# Estimacion del Fracturamiento a partir de registros convencionales #
# Basado en trabajo de datos Topologicos #
#.....PEMEX 2022.....#
LOOP:
  if dtc != MissingValue and pr != MissingValue and rhob != MissingValue:
    #..... Fracture From PR .....#
    if pr > .20 and pr <= .27:
      f_pr=1
    else:
      f_pr=0
    #.....Fracture From DTC.....#
    if dtc > 50.8 and dtc <= 57:
      f_dtc=1
    else:
      f_dtc=0
    #..... Fracture From RHOB .....#
    if rhob > 2.60 and rhob <=2.65:
      f_rhob=1
    else:
      f_rhob=0
    #..... Fracture from Rugosity .....#
    if f_pr=1 and f_dtc=1 and f_rhob=1 :
      f_rugosity= "Fracturas"
    else:
      f_rugosity= "Compacto"
  print "Prediccion de fractura"
    
```

**Figura 15.** Script para dolomías y calizas (cualquier resultado, interpretación, cálculo, dato, investigación o recomendación derivados del uso de este script será por cuenta exclusiva y responsabilidad de quien lo utilice y no tendrá derecho a reclamar).

La **Tabla 1** señala los limites encontrados por el análisis topológico de datos para las variables densidad, relación de Poisson y sónico compresional.

Limites encontrados en la estructura de datos por TDA			
Matriz (80-20)	PR	DTC (us/ft)	RHOB(g/cc)
Caliza	.20-.27	50.8-57	2.60-2.65
Dolomia	.24-.33	48-49.9	2.71-2.75

**Tabla 1.** Limites de registros convencionales en estructura de datos TDA.

La **Figura 16** muestra la aplicación de TDA y python en 3 de los 10 pozos analizados, se observa en el carril la discretización rojo-azul (donde rojo es fractura potencialmente abierta

mayor a 80°, seguido por el carril de interpretación de fracturas por el registro de imagen) vs carril de fracturas, una buena correlación entre lo entrenado y lo descrito.

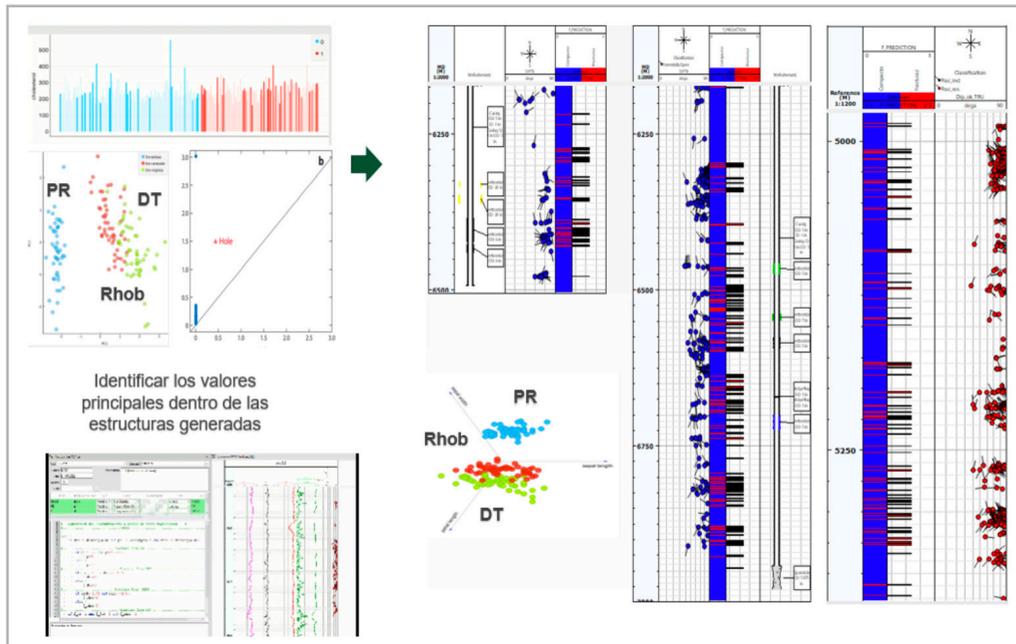


Figura 16. Resultado de aplicar algoritmo TDA en pozos con información.

La Figura 17 muestra en el último carril el script realizado para el cálculo de la “m” únicamente para una dolomía textural.

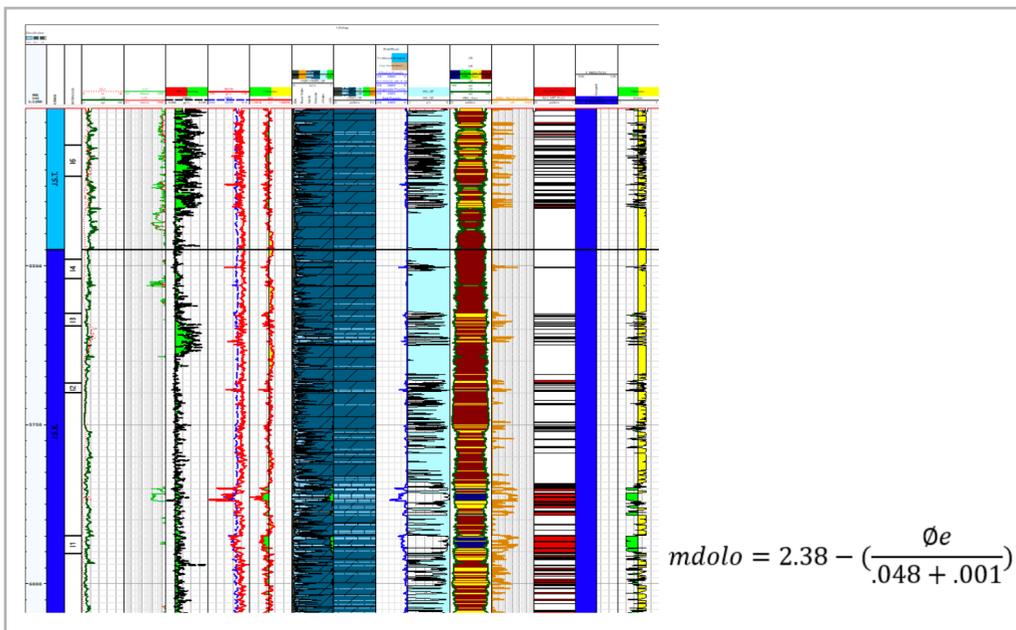
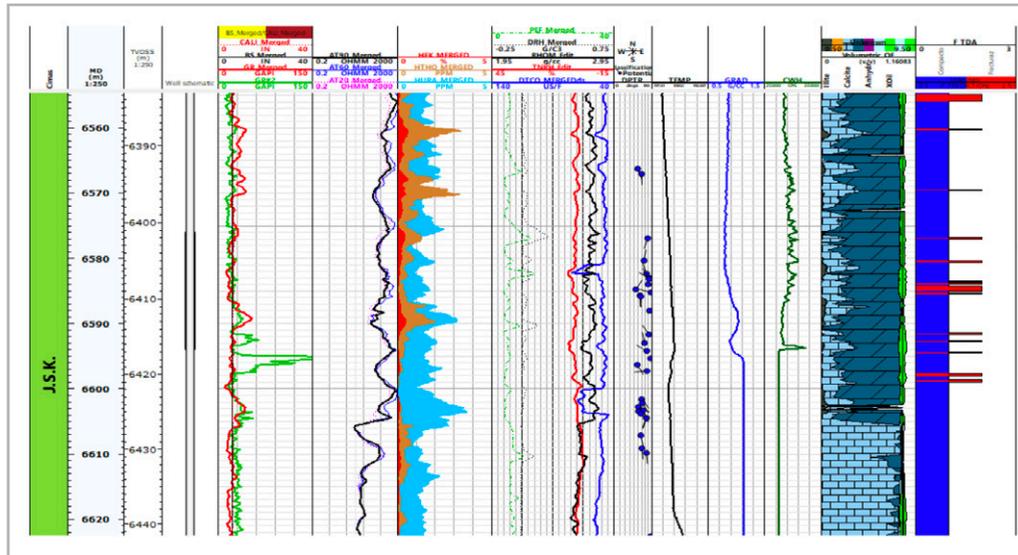


Figura 17. Generación de “m variable” en yacimiento dolomitizado.

La **Figura 18**, muestra la correlación existente entre los datos del PLT y la zona fracturada de aporte con el entrenamiento del TDA.



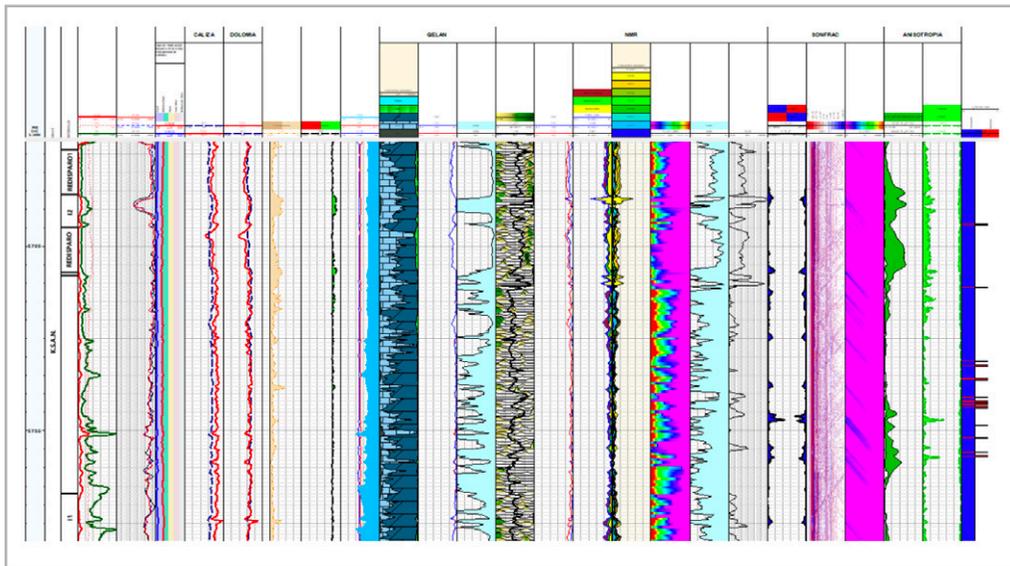
**Figura 18.** Despliegue del modelo TDA (último carril, flag en color rojo indica las fracturas potencialmente abiertas mayores a  $80^\circ$ ) en pozo con información de registro PLT e interpretación de fracturas por registro de imagen resistiva.

Continuando con los resultados obtenidos, la mejor opción es dividir los datos en dos conjuntos: el conjunto de entrenamiento y el conjunto de prueba. Como estos nombres lo indican, se entrena el modelo usando el conjunto de entrenamiento y lo prueba usando el conjunto de prueba. La tasa de error en casos nuevos se denomina error de generalización (o error fuera de la muestra), y al evaluar el modelo en el conjunto de prueba, obtiene una estimación de este error. Este valor indica qué tan bien funcionará el modelo en datos para los cuales no se ha aplicado.

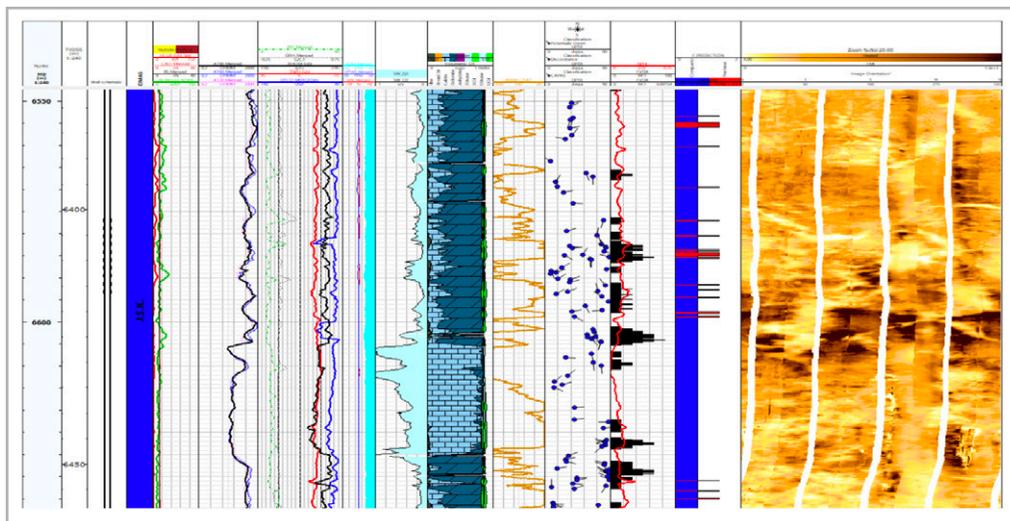
Si el error de entrenamiento es bajo (es decir, su modelo comete pocos errores en el conjunto de entrenamiento),

pero el error de generalización es alto, significa que su modelo está sobre ajustando los datos de entrenamiento.

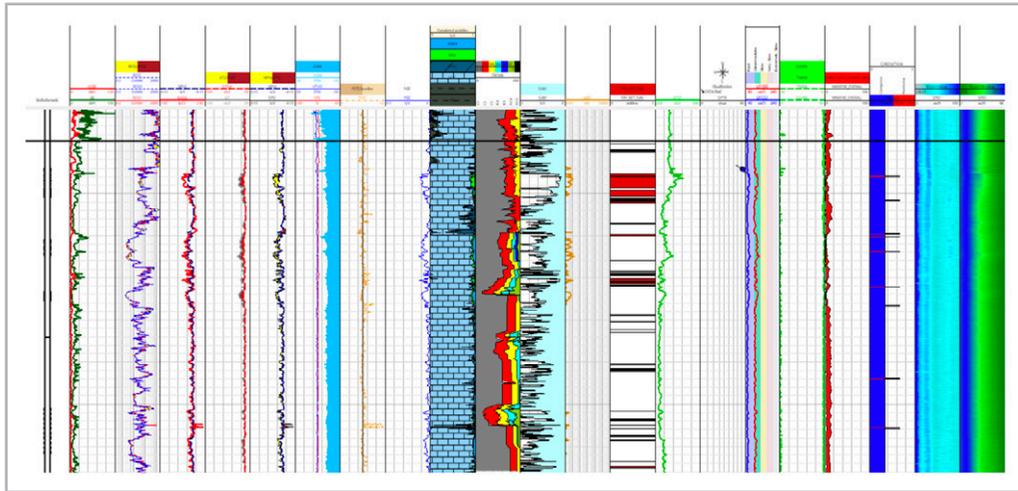
Ahora, en el conjunto de prueba se utilizaron 4 pozos en YNF, los cuales no venían en el grupo de entrenamiento, se muestra el resultado de aplicar TDA. En la **Figura 19** se compara con la interpretación del registro dipolar y resonancia magnética, en la **Figura 20** con el registro de imágenes resistivas, en la **Figura 21** con el sísmico dipolar y en la **Figura 22** con NMR, registro dipolar y resistivo de imagen.



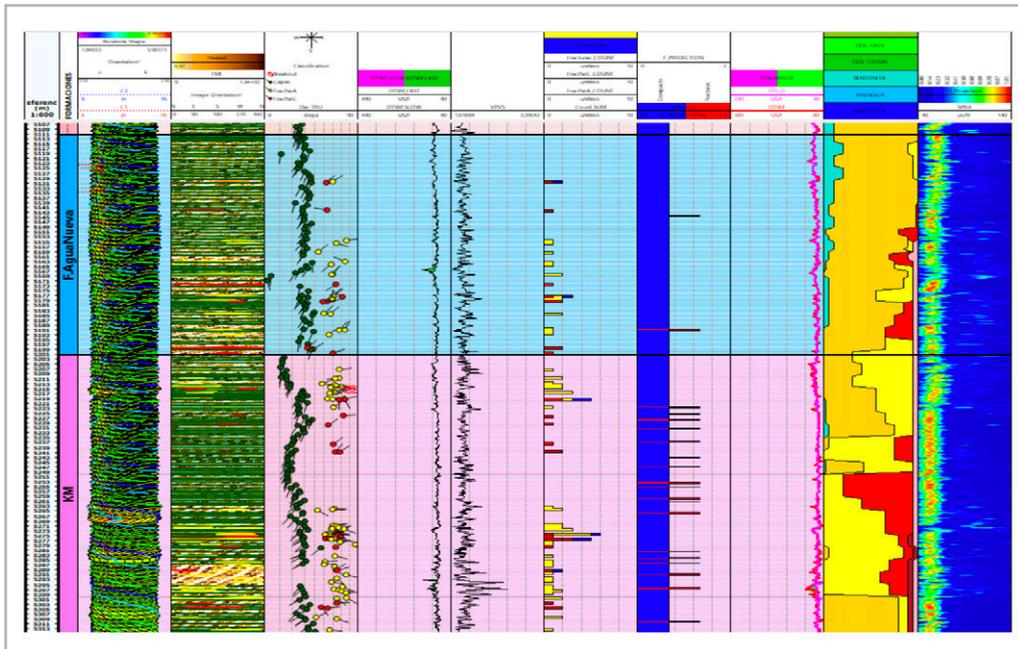
**Figura 19.** El último carril presenta el análisis TDA, en donde flag de color rojo es la fractura potencialmente abierta mayor a  $80^\circ$ . Se correlaciona con la predicción de fracturas observada por el sónico dipolar y bins de porosidad del NMR. Se tiene un error del 5-20 % comparado con datos medidos, (registros especiales).



**Figura 20.** El penúltimo carril presenta el análisis TDA, en donde flag de color rojo es la fractura potencialmente abierta mayor a  $80^\circ$ . Se correlaciona con la predicción de fracturas observada por el registro de imagen. Se tiene un error del 5-20 % comparado con datos medidos, (registro especial).



**Figura 21.** El antepenúltimo carril presenta el análisis TDA, en donde flag de color rojo es la fractura potencialmente abierta mayor a  $80^\circ$ . Se correlaciona con la predicción de fracturas observada por el registro dipolar. Se tiene un error del 5-20 % comparado con datos medidos, (registro especial).



**Figura 22.** El carril del análisis TDA, en donde flag de color rojo es la fractura potencialmente abierta mayor a  $80^\circ$ . Se correlaciona con la interpretación de fracturas observada por el registro de imagen y sónico dipolar. Se tiene un error del 5-20 % comparado con datos medidos, (registro especial).

## Conclusión

A partir del análisis topológico de datos (TDA), con información obtenida de registros geofísicos (registros convencionales y especiales), datos de núcleo, registros PLT, entre otros, y el entrenamiento de redes neuronales, se logró generar un script que permite identificar fracturas (con un margen de incertidumbre del 5-20 %) en yacimientos de carbonatos. Es importante señalar que este trabajo es un apoyo en el caso de yacimientos carentes de registros especiales para identificar fracturas, y por ningún motivo sustituye la toma de información en pozos. Su aplicabilidad se enfocó en yacimientos del país, tanto terrestres como marinos. Este estudio permite disminuir la incertidumbre en la identificación de fracturas para intervalos prospectivos de carbonatos. Este trabajo solo es aplicable para yacimientos naturalmente fracturados de doble porosidad, matriz-fractura.

## Nomenclatura

AR	Relación de aspecto de radio
Dtc	Registro sísmico compresional
NMR	Nuclear Magnetic Resonance
PCA	Principal Component Analysis
PLT	Production Log Tool
PR	Relación de Poisson
RHOB	Registro de densidad
TDA	Topologic Data Analysis
YNF	Yacimiento Naturalmente Fracturado

## Referencias

- Aguilera, R. 1995. *Naturally Fractured Reservoirs*, second edition. Tulsa, Oklahoma: PennWell Books.
- Back, W. 1963. Preliminary Results of a Study of Calcium Carbonate Saturation of a Groundwater in Central Florida. *International Association of Scientific Hydrology. Bulletin* **8** (3): 43-51. <https://doi.org/10.1080/02626666309493336>.
- Bahukhandi, Y., Kar, S., Majumdar, C. et al. 2011. Flow Unit Characterization aided by Quantified Secondary Porosity Result from High Resolution LWD Images in Heterogeneous Carbonate Reservoir: A Novel Approach. Artículo presentado en International Petroleum Technology Conference, Bangkok, Tailandia, noviembre 15-17. IPTC-14772-MS. <https://doi.org/10.2523/IPTC-14772-MS>.
- Ballard, C., Rollins, J., Ramos, J. et al. 2007. *Dynamic Warehousing: Data Mining Made Easy*. Poughkeepsie, NY: IBM Redbook. SG24-7418-00.
- Carlsson, G. 2009. Topology and data. *Bulletin of the American Mathematical Society* **46** (2): 255–308. <https://dx.doi.org/10.1090/S0273-0979-09-01249-X>.
- Carlsson, G., Ishkhanov, T., De Silva, V. et al. 2008. On the Local Behavior of Spaces of Natural Images. *International Journal of Computer Vision* **76**: 1–12. <http://dx.doi.org/10.1007/s11263-007-0056-x>.
- LeCun, Y., Jackel, L. D., Boser, B. et al. 1989. Handwritten Digit Recognition: Applications of Neural Network Chips and Automatic Learning. *IEEE Communications Magazine* **27** (11): 41–46. <https://doi.org/10.1109/35.41400>.
- Long, P. M. and Servedio, R. A. 2010. Restricted Boltzmann Machines are Hard to Approximately Evaluate or Simulate. Proceedings of the 27th International Conference on Machine Learning (ICML'10), Haifa, Israel, junio 21-24.
- Newberry, B. M., Grace, L. M. y Stief, D. D. 1996. Analysis of Carbonate Dual Porosity Systems from Borehole Electrical Images. Artículo presentado en Permian Basin Oil and Gas Recovery Conference, Midland, Texas, marzo 27-29. SPE-35158-MS. <https://doi.org/10.2118/35158-MS>.

## Semblanza del autor

### Luis Ahindel Castellanos Bassoult

Ingeniero Petrolero por la Universidad Nacional Autónoma de México, con Maestría en Ciencias.

Se ha desempeñado en el área de productividad de pozos, análisis estratégico de negocios y actualmente trabaja desarrollando y evaluando modelos petrofísicos y de física de rocas como sísmico petrofísico en Pemex.

Autor de diversos artículos a nivel nacional (CMP-AIPM) e internacional (SPE-SPWLA). Certificado en Python for data science por IBM.