



Metodología y aplicación de atributos sísmicos en la modelación de facies para zonas potencialmente productoras de gas. Área Santa Cruz-Jibacoa, Cuba

Julio Ernesto Gómez-Herrera¹*; Osvaldo Rodríguez-Morán¹; Yeniley Fajardo-Fernández¹;
José Hernández-León¹; José Alberto Arias del Toro²; Zulema Domínguez-Sardiñas¹;
Lourdes Jiménez de la Fuente¹

doi: <https://doi.org/10.18273/revbol.v42n2-2020004> 

Forma de citar: Gómez-Herrera, J.E.; Rodríguez-Morán, O.; Fajardo-Fernández, Y.; Hernández-León, J.; Arias del Toro, J.A.; Domínguez-Sardiñas, Z.; Jiménez de la Fuente, L. (2020). Metodología y aplicación de atributos sísmicos en la modelación de facies para zonas potencialmente productoras de gas. Área Santa Cruz-Jibacoa, Cuba. *Boletín de Geología*, 42(2), 69-87. doi: 10.18273/revbol.v42n2-2020004.

Resumen

Históricamente, la exploración petrolera en Cuba, se ha apoyado en estudios geoquímicos, petrofísicos y de ingeniería de yacimientos en los pozos para caracterizar los intervalos productores de hidrocarburos y la explotación del gas disuelto. El objetivo de este estudio es detectar en esos reservorios, las acumulaciones comerciales de gases húmedos a partir del uso de los atributos sísmicos con la instrumentación de una metodología de trabajo. Se escogieron dos yacimientos que se encuentran en la Franja Norte Petrolera Cubana; el Santa Cruz del Norte (patrón de aprendizaje) y Jibacoa (patrón predictivo). Como objeto se estudió la Formación Vega Alta, secuencia sinorogénica de edad Paleoceno-Eoceno Medio, con gran complejidad litológica, considerada sello regional, pero paradójicamente, con apreciables manifestaciones de hidrocarburos, principalmente de gas. Se realizó un análisis microfacial, fundamentalmente con la información de los *cuttings* (1921 m), construyéndose un modelo ternario para las rocas clásticas y otro para las carbonatadas y silíceas, con predominio de las primeras. Para obtener información sobre la relación entre las litofacies y los atributos sísmicos, se efectuó un análisis exploratorio de los datos. Se establecieron las firmas sísmicas para cada litotipo y con la aplicación de métodos de minería de datos, se ajustó la capacidad de esos atributos para diferenciar cada litofacies. Fueron discretizados 4130 intervalos patrones en pozos. Para trabajar dentro de un dominio con gran incertidumbre, se empleó un algoritmo geoestadístico de simulación secuencial de indicadores para calcular las probabilidades. Se ejecutó la conversión de atributo sísmico a litofacies, mediante una clasificación supervisada, en una matriz de 350000 celdas (aproximadamente 43750 km³ en total), con los atributos sísmicos seleccionados como significativo y por último se estableció un cálculo de riesgo-recompensa que pronostica el alcance económico de esta estrategia.

Palabras clave: Litofacies; Atributos sísmicos; Franja Norte Petrolera Cubana; Geoestadística; Minería de datos; Riesgo exploratorio.

Methodology and application of seismic attributes in modeling facies for potentially gas producing areas. Santa Cruz-Jibacoa oilfields area, Cuba

Abstract

Historically the Cuban exploration strategy has been characterized by discovering productive oil intervals and to use its associated gas by conventional methods. The objective of this study was to create a methodology to detect reservoirs with free gaseous hydrocarbons contain, by seismic attributes. Were chosen two oilfields located in the Cuban Northern Oil Belt; (main Cuban oil region); Santa Cruz del Norte (learning pattern) and Jibacoa oil fields (predictive pattern). Were studied the Vega Alta Formation, a synogenic sequence of Paleocene-Middle Eocene age, with great lithological complexity and considered as a regional seal, but paradoxically, with significant hydrocarbons presence. A microfacial analysis was carried out, with the information of the cuttings, building a ternary model for clastic rocks and another model for carbonated and siliceous rocks. Were studied 1921 m by cuttings samples, predominantly clastic lithofacies. An exploratory data analysis was carried out, which permitted obtaining information about the unknown relationship between lithofacies and seismic attributes. The seismic signatures of lithofacial classification were determined, and with the assistance of data mining methods were adjusted the seismic attributes to differentiate lithofacies. These were discretized in 4130 standard patterns intervals. Probability values were calculated applying a geostatistics sequential simulation algorithm. The inverse task (conversion of seismic attributes to lithofacies) was executed, through a supervised classification, into a 350000 cells matrix (approximately 43750 km³), with the selected seismic attributes and finally, an exploration risk analysis was calculated for estimate the economic strategy.

Keywords: Lithofacies; Seismic attributes; Cuban Northern Oil Belt; Geostatistics; Data mining; Exploration risk.

¹Centro de Investigación del Petróleo, La Habana, Cuba. (*) julio@ceinpet.cupet.cu; ormoran2016@gmail.com; yeniley@ceinpet.cupet.cu; joseh@ceinpet.cupet.cu; zulds@ceinpet.cupet.cu; lourdes@ceinpet.cupet.cu

²Oficina Nacional de Recursos Minerales, La Habana, Cuba. jose@onrm.minem.cu

Introducción

Hasta este momento en la exploración petrolera en Cuba, la caracterización de los intervalos potencialmente productores de gas en la Franja Norte Petrolera Cubana (FNPC) no ha formado parte de la estrategia en los trabajos de exploración y cuando se ha efectuado, se ha apoyado fundamentalmente en los métodos de ingeniería de yacimientos, criterios geoquímicos y petrofísicos. En esos análisis lo que ha predominado es la evaluación del gas acompañante y no propiamente la de gas húmedo. Muchas de estas manifestaciones de gas se han encontrado en secuencias que no forman parte de los objetivos principales o reservorios, por lo que la información sobre todo de geofísica de pozo, ha sido insuficiente. Existe la limitante de la imposibilidad de evaluar ese potencial en los lugares que no posean pozos.

El objetivo de este estudio es precisamente detectar posibles reservorios (en una determinada formación geológica), que puedan contener hidrocarburos, a partir de poca información geológica directa, pero sí con datos sísmicos, para poder obtener conocimiento utilizando los atributos sísmicos.

El estudio se realizó en las áreas de los yacimientos Santa Cruz del Norte (tomado como patrón de aprendizaje-predictivo) y Jibacoa (patrón predictivo), que se encuentran en la FNPC, principal zona productora de hidrocarburos de la República de Cuba (Figura 1). Como objetivo de trabajo se utilizó la controvertida Formación Vega Alta, considerada oficialmente como el sello regional de la FNPC. Esta es una secuencia sinorogénica que corresponde a depósitos de la cuenca de antepaís en profundidades batiales de edad Paleoceno-Eoceno Inferior. Está constituida por depósitos caóticos policomponentes que contienen bloques y fragmentos de calizas de diferentes edades, serpentinitas, rocas volcánicas y brechas. En la base se observan depósitos turbidíticos distales. Una breve caracterización lito-paleontológica determina que la conforman cuatro facies principales: a) facies arcillosa con clastos conglomeráticos (olistostromas); b) facies conglomerática arcillosa-arenosa; c) facies arcillosa silícea y d) facies arcillosa-carbonatada. En muchos intervalos se producen significativas manifestaciones de petróleo y gas, pero no deben olvidarse las limitadas posibilidades productivas que tienen estas secuencias, debido a la abundante matriz arcillosa que las caracteriza (Morales, 2014).

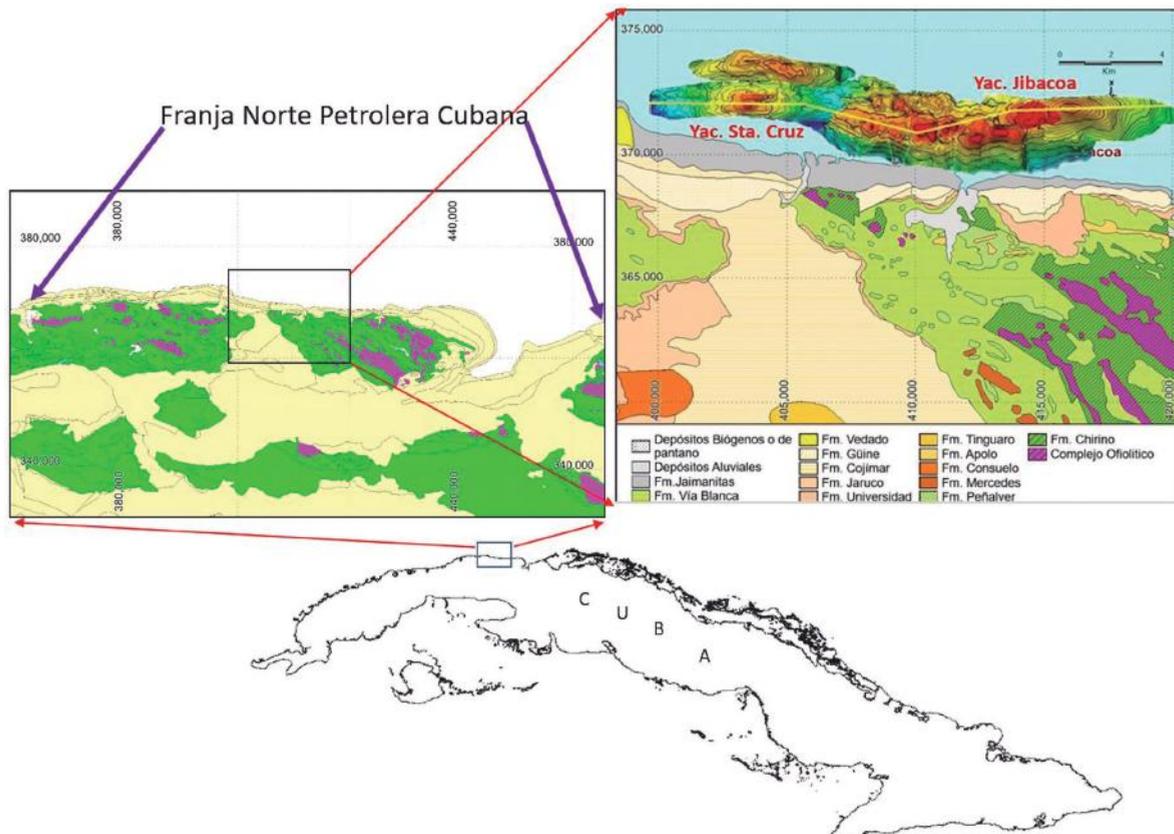


Figura 1. Área de estudio dentro de la Franja Norte Petrolera Cubana (FNPC).

Se instrumentó una metodología de trabajo consistente en tareas de recopilación, transformación, evaluación e interpretación de datos.

Metodología

Para el aprendizaje se seleccionaron intervalos reportados como productores de gas en pozos de estas áreas. Se realizó un análisis microfacial con la información dada por los *cuttings* y se calculó la proporción de la matriz arcillosa (porque esta no se recupera, ya que se disuelve en el fluido de perforación, o es lavada durante el proceso de preparación de las muestras). El volumen de la matriz arcillosa se obtiene, aplicando una metodología donde se calcula el porcentaje de arcilla libre por registros geofísicos radioactivos (Hernández, 2015). Con esta información, se construyó un modelo ternario de microfacies en correspondencia con la distribución de las rocas en esa área, según los siguientes criterios:

- Criterio geoquímico: localización de ciertas anomalías según su espectro de cromatografía gaseosa y registros de hidrocarburo.
- Criterio petrofísico: intervalos con registros geofísicos que de alguna manera indiquen características petrofísicas de ser potenciales reservorios de gas.
- Criterio petrográfico: determinación de los porcentajes litológicos por los *cuttings*.
- Criterio espacial: intervalos con espesores mayores que 10 m.

- Criterio sísmico: Intervalo cubierto por la sísmica 3D.

En este caso se realiza el estudio para posibles acumulaciones de gas húmedo, preferiblemente de tipo gas no asociado (gas libre que no está en contacto con el petróleo, formado por hidrocarburos livianos a intermedios, que no se condensan en el reservorio, pero sí lo hacen en superficie) (Martínez y Valdez, 2011).

En el análisis litofacial se tuvieron en cuenta grupos litológicos muy diversos, siendo en su gran mayoría clastitas (conglomerados, areniscas, limolitas y *claystone* de diversos tipos) y en menor cuantía rocas carbonatadas y silíceas. Las rocas clásticas están relacionadas con formaciones orogénicas, formadas mayormente por olistostromas de ambientes turbidíticos de alta energía, que son muy inmaduras texturalmente, con abundantes cantidades de matriz arcillosa.

Teniendo los valores de granulometría, calculados durante el estudio de las muestras y el porcentaje de matriz arcillosa calculado por registros gamma, se aplican entonces las clasificaciones universales establecidas para estas rocas (Pettijohn, 1975; Folk, 1980), utilizando para ellos gráficos ternarios preparados a tal efecto. Como la mayoría de estos intervalos, pertenecientes a la Formación Vega Alta, se encuentran fuera de los tradicionales reservorios, con un escenario geológico complejo y gran incertidumbre, carecen de una información petrofísica aceptable (Figura 2).

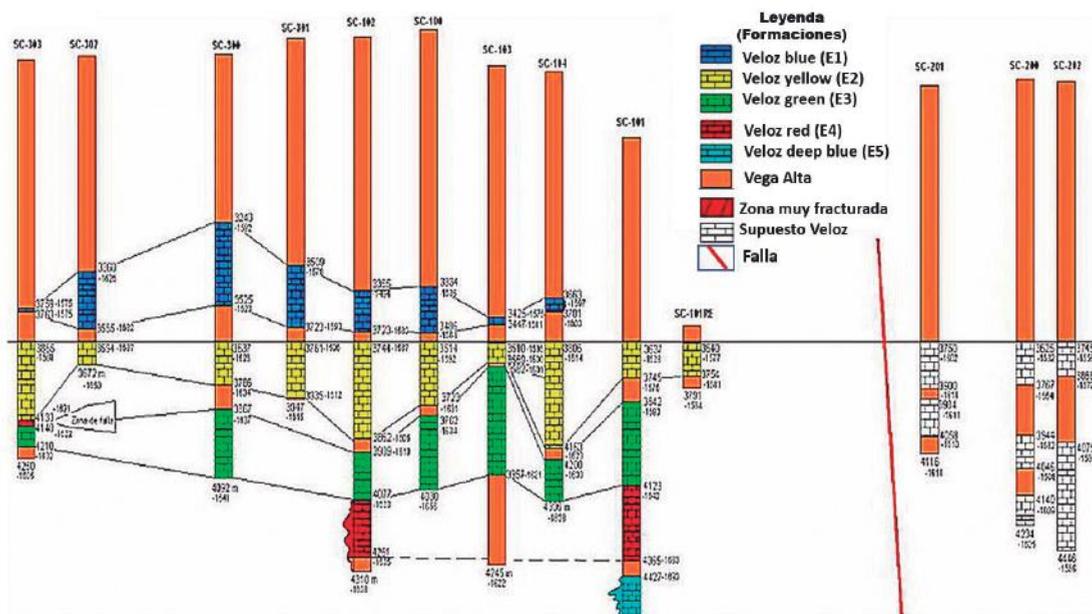


Figura 2. Correlación, donde se observa la información desigual en los intervalos de la Formación Vega Alta y los reservorios subyacentes para el yacimiento patrón Santa Cruz del Norte.

El problema científico radica en que, si se logran caracterizar litológicamente los intervalos gasíferos en los pozos, considerándolos como patrones, y se determina su grado de concordancia con el complejo de atributos sísmicos, entonces se podrá realizar la conversión de los atributos sísmicos a sismofacies, mediante la modelación, y apreciar de

este modo posibles reservorios con características gasíferas, donde solo existe cobertura sísmica y como consecuencia se amplía el volumen de la exploración.

En la Figura 3 se presenta el modelo estructural correspondiente a la Formación Vega Alta en los yacimientos Santa Cruz del Norte - Jibacoa.

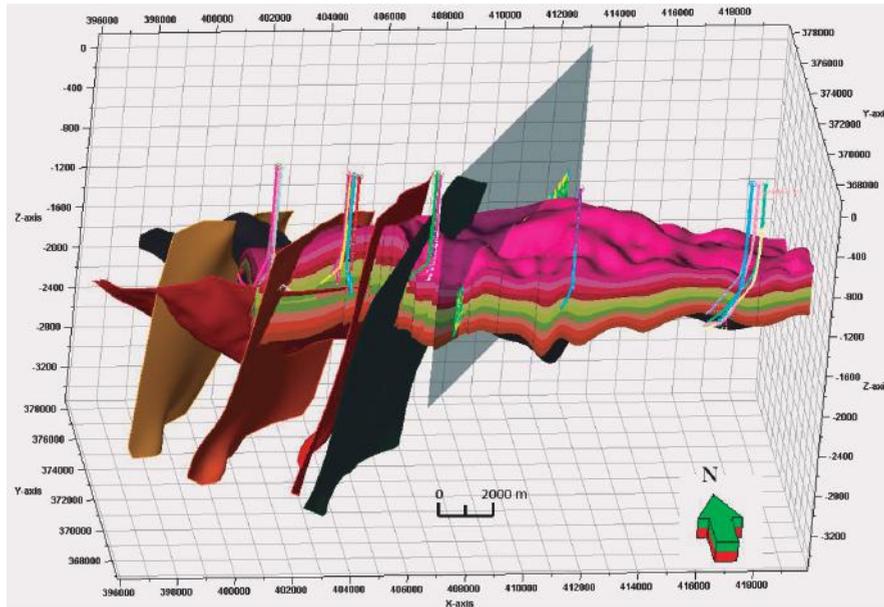


Figura 3. Modelo estructural del área Santa Cruz del Norte - Jibacoa, para la Formación Vega Alta.

Como los pozos están dirigidos hacia los reservorios más profundos (carbonatos plegados del Grupo Veloz, en sus diferentes escamas *yellow*, *green*, *red*, *blue* y *deep blue*, de edad Jurásico Superior- Cretácico Inferior por debajo de la Formación Vega Alta), no están uniformemente distribuidos en el área (prácticamente alineados de oeste a este), haciendo que la información existente esté muy focalizada, como se puede observar en el yacimiento patrón Santa Cruz del Norte (Figura 4).

Se caracterizaron litofacialmente 45 intervalos seleccionados como posibles productores de gas, donde la mayoría de éstos pertenecen a la Formación Vega Alta. Para la determinación de las microfacies terrígenas de grano grueso y medio (gravas y arenas), fue creado un modelo ternario, con arcilla, grava y arena (Figura 5), el cual permite su clasificación con gran facilidad cuando se han determinado con anterioridad los porcentajes de cada componente.

Para la clasificación de las microfacies carbonatadas arcillosas y arcillosas silíceas de grano fino, se compuso

un modelo de doble entrada. Los componentes que se toman en cuenta son arcilla y ópalo (el principal es el componente silíceo, proveniente de los radiolarios). En el caso de las litofacies calcáreas se toma en cuenta el componente calcáreo (calcita) y su mezcla con arcilla (Figura 6).

De modo preliminar, puede afirmarse que los paquetes litológicos seleccionados pueden ser agrupados en las siguientes litofacies (Hernández, 2018):

- Litofacies terrígenas de grano grueso: conglomerados y conglomerados arenosos.
- Litofacies terrígenas de grano medio: arenitas y grauvacas diversas.
- Litofacies terrígenas de grano fino: diversas rocas, tales como limolita, claystone y claystone limoso.
- Litofacies arcilloso-silíceas, claystone silíceo y radiolaritas.
- Litofacies calcáreas: calizas y margas diversas.

El yacimiento Santa Cruz del Norte es el más prolífico en la selección de intervalos potencialmente

productivos de gas y con más pozos perforados, con un total de 30 tramos, escogidos por sus propiedades petrofísicas y a los que se les realizó el análisis facial. En estos predominan las litofacias clásticas de grano

grueso y medio y en menor cuantía las de grano fino y las calcáreas. La expresión gráfica de sus espesores y la distribución en los modelos litofaciales se muestra en las Figuras 7, 8 y 9, respectivamente.

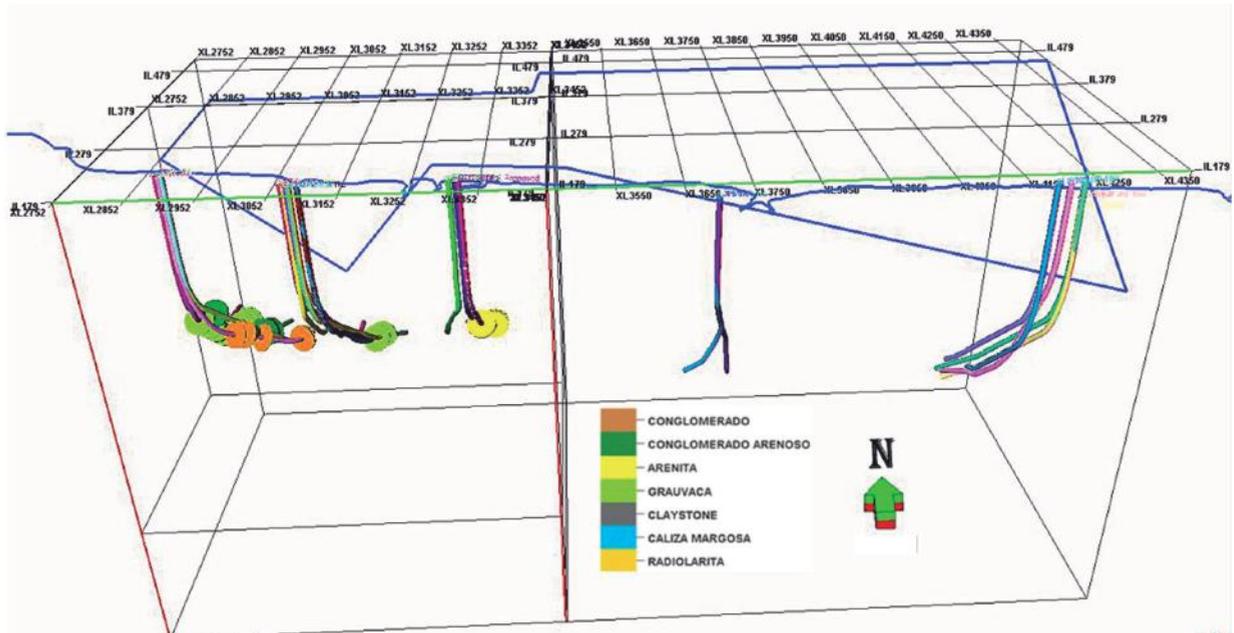


Figura 4. Trayectorias de los pozos en los yacimientos Santa Cruz del Norte y Jibacoa. Se muestran los intervalos con gas y sus respectivas facies (patrones), dentro del cubo sísmico.

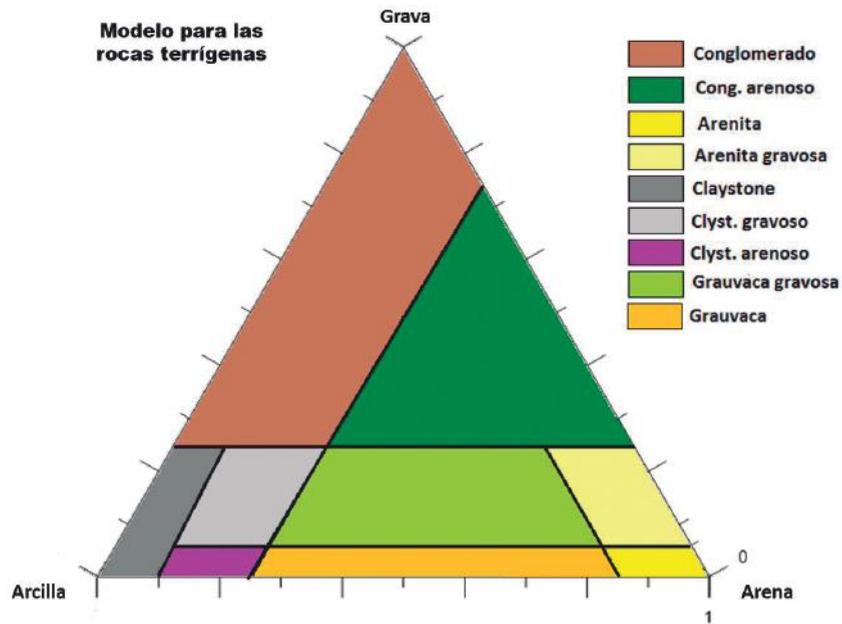


Figura 5. Gráfico ternario para la clasificación de las rocas terrígenas (Hernández, 2018).

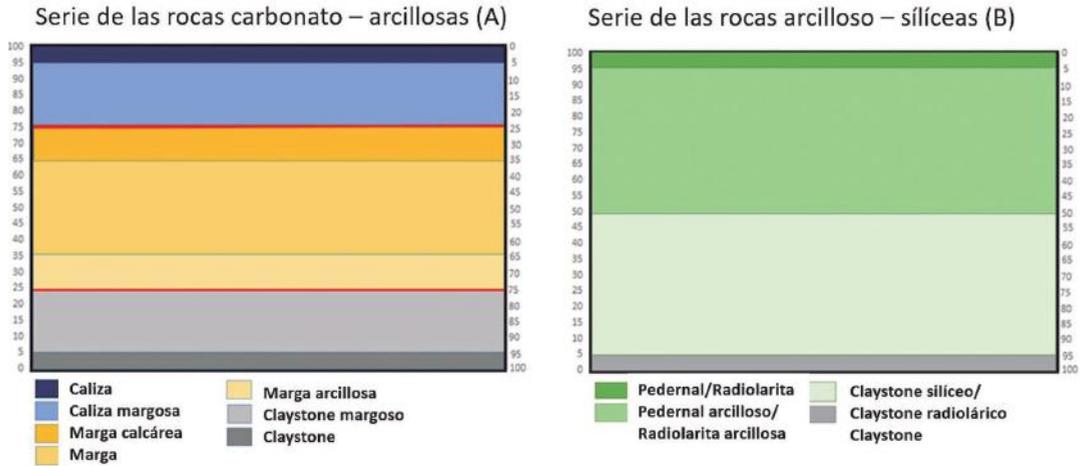


Figura 6. Gráficos de doble entrada para la clasificación de las microfacies. A. Calcáreas. B. Las arcilloso-silíceas (Hernández, 2018).

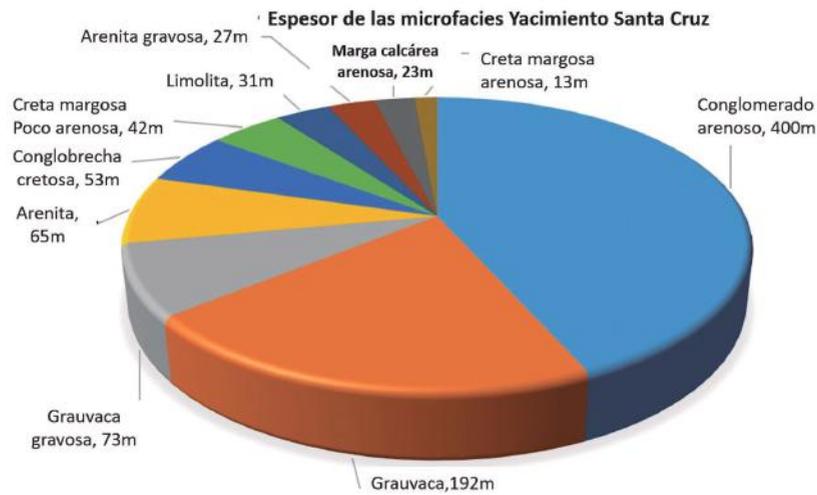


Figura 7. Microfacies del área Santa Cruz del Norte y la expresión gráfica de sus espesores (Hernández, 2018).

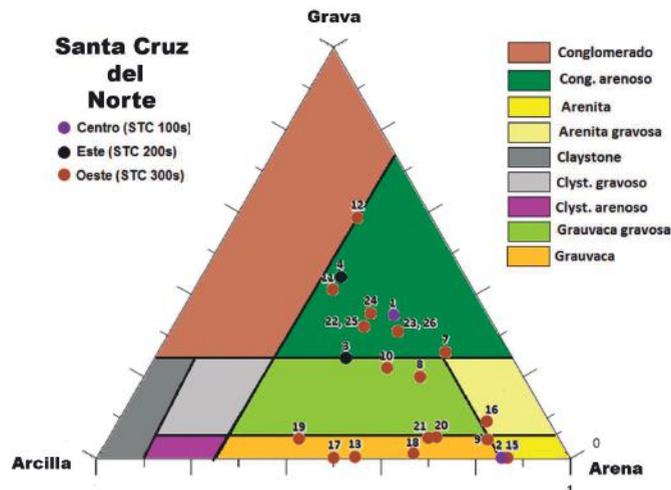


Figura 8. Gráfico ternario para la clasificación de las rocas terrígenas (Hernández, 2018).

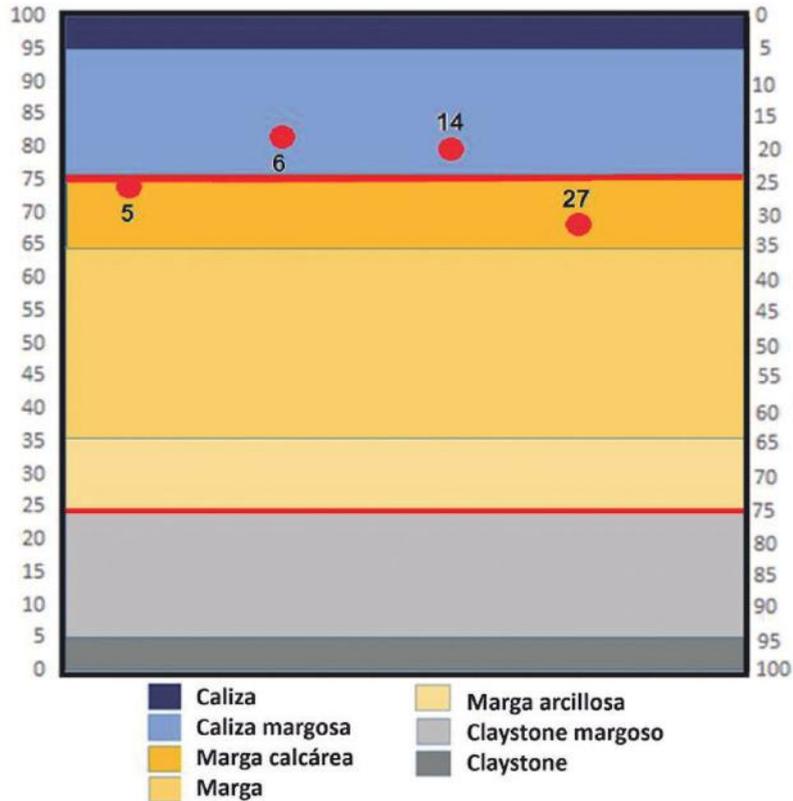


Figura 9. Gráfico de doble entrada para la clasificación de las rocas calcáreas para el área de Santa Cruz del Norte (Hernández, 2018).

Desde el punto de vista del análisis sedimentológico, el área de Santa Cruz del Norte resulta muy prometedora. El 25% del espesor de todos los tramos estudiados (237 m) se consideran reservorios potencialmente

buenos para la producción de gas. Un 37% se evalúan como regulares (341 m) y el 38% restante son de mala calidad (Figura 10).

Calidad de los reservorios y su espesor

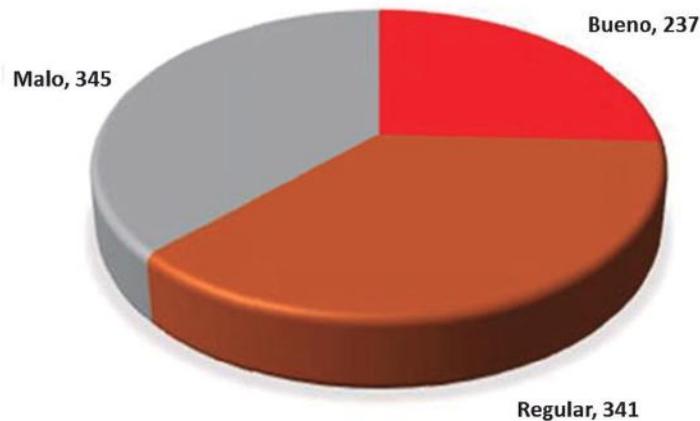


Figura 10. Calidad de los reservorios con potencial de gas en el yacimiento Santa Cruz del Norte (Hernández, 2018).

Los intervalos clasificados como regular (con un contenido de arcilla por encima del límite sedimentológico de 14%), son rocas con una elevada inmadurez textural, que pueden manifestar una permeabilidad aceptable para la producción de gas, por su gran diversidad granulométrica, y ciertos niveles de fracturación, principalmente las grauvacas y conglomerados arenosos.

En función del conocimiento obtenido y la resolución de la sísmica, se confeccionó para el estudio una variante más general de litofacies, la cual permite una

mejor adaptación a las complejidades de las litologías y disminuir la incertidumbre de los resultados de la modelación tridimensional (Hernández, 2018).

De esta manera fueron agrupadas las microfacies en litofacies:

- Modelo ternario para las rocas clásticas que generaron 5 litofacies (Figura 11).
- Modelo lineal para las rocas arcillosas calcáreas y arcillosas – síliceas con 2 litofacies (Figura 12).

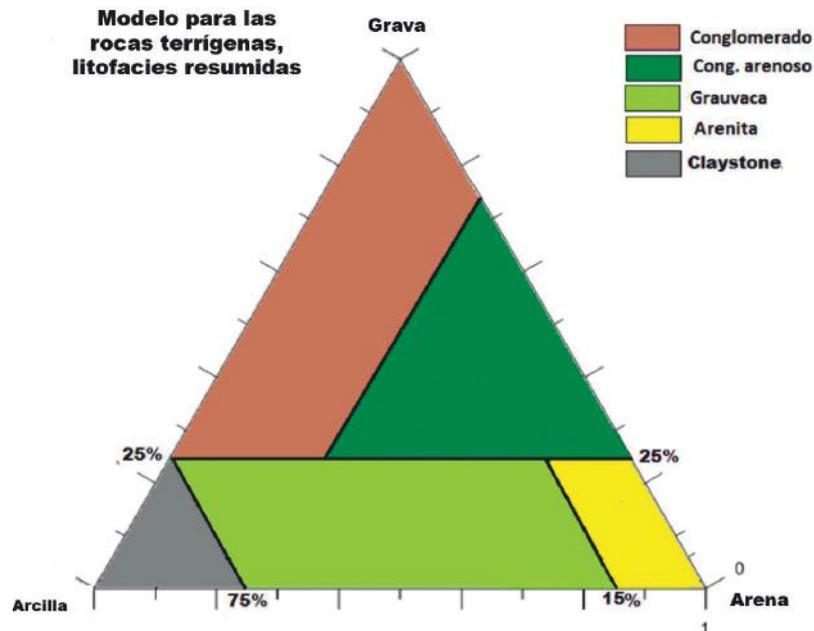


Figura 11. Modelo ternario, para la clasificación de las rocas terrígenas (Hernández, 2018).

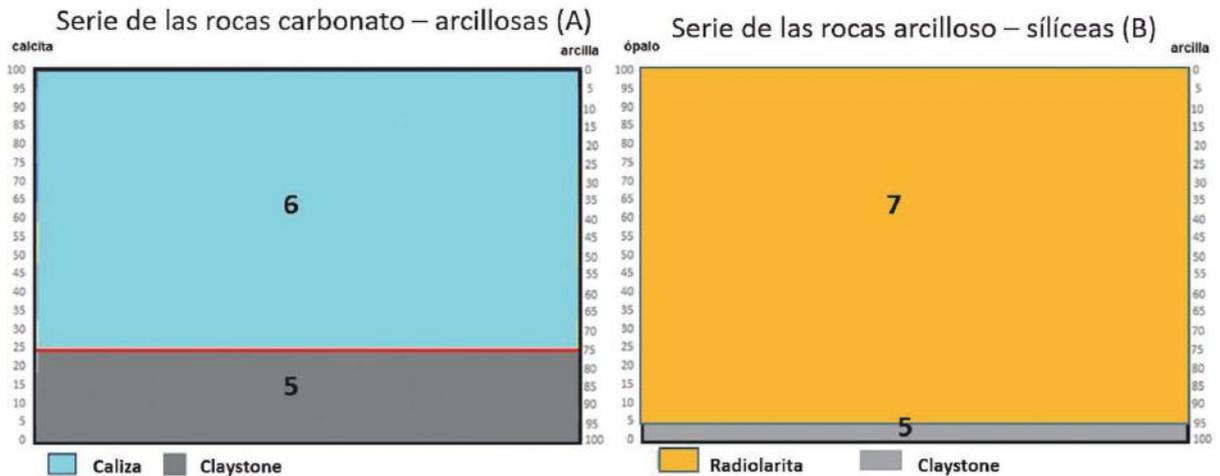


Figura 12. Modelos de doble entrada para la clasificación de las litofacies: A. calcárea; B. arcilloso-síliceas (Hernández, 2018).

Aplicación y calibración de los atributos sísmicos

Se denomina atributo sísmico a toda información obtenida de los datos sísmicos, ya sea por mediciones directas, por razonamiento lógico, o basado en la experiencia (Taner, 2001) o cualquier medida de los datos sísmicos que ayude a mejorar o cuantificar visualmente las características de interés en la interpretación (Chopra y Marfurt, 2007).

Estos pueden ser calculados tanto a partir de datos sísmicos pre o post apilamiento (*pre-stack* o *post-stack*), como antes o después de la migración y se los puede clasificar de diferentes formas.

Posterior a un estudio bibliográfico sobre los atributos y sus características en el sistema de modelación PETREL (Daber y Aqravi, 2010), se decidió experimentar con estos atributos, para la interpretación de las facies, en las difíciles condiciones geológicas existentes en esta zona.

A continuación, se da una breve descripción de los atributos sísmicos seleccionados y de los registros de pozo empleados como apoyo:

- a) Atributo *Impedance*: El atributo impedancia acústica relativa muestra el contraste acústico aparente, indicando los límites de secuencias, facilitando un mejor enlace con la litología y relacionándose con la porosidad.
- b) Atributo Atenuación: Atributo estratigráfico basado en la atenuación de frecuencia que compara las frecuencias por encima y por debajo de los puntos de análisis seleccionados, lo que permite que sea un buen indicador de fracturas abiertas si hay gas presente.
- c) Atributo *Sweetness*: El atributo “dulzura” refleja ligeros cambios de energía. Este atributo frecuentemente se utiliza para discriminar características producidas por contrastes grandes de impedancia, tales como discordancias, fallas, contactos agua - petróleo y para definir límites de comportamiento del yacimiento. Es la combinación del atributo envolvente (*envelope*) y la frecuencia instantánea.
- d) Atributo Coseno: El coseno de fase también llamado amplitud normalizada, funciona con valores de rendimiento o de salida entre -1 y 1. Este atributo es un excelente indicador de continuidad lateral, terminaciones estratigráficas y variaciones de facies sísmicas.

- e) Atributo RMS: Cálculo de la raíz media cuadrática de la amplitud en una traza de una ventana específica. Este atributo puede mostrar contrastes estratigráficos y cambios de densidad en los alrededores del objeto de estudio.
- f) Atributo Isofrecuencia: Atributo estratigráfico de descomposición espectral en frecuencias, proporciona indicaciones sobre el grosor estratigráfico, identifica las características estructurales y estratigráficas sutiles, que lo hacen bueno para la cartografía de facies.
- g) Atributo *Structural smoothing*: Realiza el suavizado de formas, por medio de un promedio ponderado gaussiano. Utiliza como entrada la varianza en la profundidad o caos. Elimina ruido de los datos y puede iluminar contactos de fluidos.
- h) Atributo *Chaos*: Hace la cartografía de la “entropía” de la señal sísmica local. Esto lo hace útil para analizar y discriminar facies sísmicas.
- i) Registro Neutrón: Los registros geofísicos neutrónicos se utilizan principalmente para delinear formaciones porosas.
- j) Registro Gamma (CGR): Son registros geofísicos que permiten correlacionar formaciones, interpretar posibles litologías, determinar zonas permeables e impermeables y estimar el volumen de arcilla.
- k) Registro Gamma (SGR): Registro geofísico que mide la radioactividad natural de la formación. Es un indicador del volumen de arcilla y análisis del ambiente de sedimentación.

Se utilizaron solamente estos registros de pozo, por ser los más comunes, ya que se obtienen en tiempo real, durante la perforación.

Es necesario aclarar que este estudio comprende a dos yacimientos contiguos, con un mismo modelo litofacial.

Descubrimiento de conocimiento en bases de datos

El descubrimiento de conocimiento en bases de datos KDD (del inglés *Knowledge Discovery in Databases*), es un proceso no trivial para identificar patrones válidos, novedosos, potencialmente útiles y en última instancia, comprensibles, a partir de grandes cantidades de datos (*bigdata*), teniendo como objetivo encontrar conocimiento útil, relevante y nuevo sobre un fenómeno o actividad. Este proceso consta de una secuencia iterativa de etapas o fases (Hernández-Orallo *et al.*, 2004):

- a) Recopilación e integración (Bases de datos y datos objetivo). contrasten con la realidad, y la realización del proceso sea inútil.
- b) Selección y limpieza (análisis exploratorio de los datos). Para la gestión de los datos se creó una Tabla de Objeto-Propiedad (TOP) (Tabla 1).
- c) Minería de datos (datos minados).
- d) Evaluación e interpretación (patrones).
- e) Difusión y uso (conocimiento y toma de decisiones). Se analizaron 2785 puntos en los pozos con sus respectivas asignaciones de litofacies según el modelo elegido (Figuras 11 y 12) y que generaron una cantidad de 4130 datos-objetos.

Resultados

El análisis exploratorio de los datos, tiene como objetivo el seleccionar de forma coordinada, los datos más relevantes del proceso, ya que la información recolectada puede contener errores en los datos o valores faltantes. Una mala calidad de los datos trae como consecuencia que los patrones descubiertos

En la Tabla 2 se muestra un análisis estadístico preliminar de los valores de los atributos. Con esa información se procedió a la depuración de la base de datos, de valores extremos o fuera de rango (*outliers*) y se pudo determinar de forma preliminar cuáles son los atributos sísmicos capaces de diferenciar en mayor o menor grado las litofacies (Tabla 3).

Tabla 1. Detalle parcial de la TOP de los datos por pozos, los intervalos muestreados, sus valores de atributos sísmicos y clasificación litofacial.

Pozo	Prof.	Stru	Iso	Aten	Chaos	Imped	Cosen	RMS	Sweet	Facies
JIB-100	2800	0,88	0,72	-0,01	0,07	3,98	0,62	1,92	3,18	Claystone
JIB-100	2800,2	0,88	0,72	-0,01	0,07	3,97	0,62	1,92	3,18	Claystone
JIB-103	3733,84	-0,07	0,82	-0,02	0,13	-0,97	-0,43	0,95	1,28	Claystone
JIB-103	3733,99	-0,07	0,82	-0,02	0,13	-0,97	-0,42	0,95	1,28	Claystone
SC-202	3645	0,54	0,59	0,01	0,12	0,46	0,79	1,1	1,36	CongAren
SC-202	3650	0,49	0,58	0,01	0,11	0,79	0,69	1,08	1,34	CongAren
SC-202	3650	0,49	0,58	0,01	0,11	0,79	0,69	1,08	1,34	CongAren
SC-301	2440	-0,16	0	0	0,71	0	1	0	0	Grauvaca
SC-301	3445	-0,3	0,64	0	0,32	-0,05	-0,97	0,48	0,67	CongAren
SC-301	3450	-0,31	0,57	0,01	0,28	0,02	-0,96	0,5	0,72	CongAren
SC-302	1644	0	0	0	1	0	1	0	0	Grauvaca
SC-302	1644,5	0	0	0	1	0	1	0	0	Grauvaca
SC-303	1281,84	0	0	0	1	0	1	0	0	Grauvaca
SC-303	1281,99	0	0	0	1	0	1	0	0	Grauvaca
SC-304	2275,64	0,22	0,56	-0,03	0,2	-2,43	0,16	1	1,62	CongAren
SC-304	2275,79	0,21	0,56	-0,03	0,2	-2,42	0,17	1	1,62	CongAren

Tabla 2. Análisis estadístico descriptivo básico de los datos primarios.

Variable	Valid N	Media	Mediana	Mínimo	Máximo	Cuartil 25%	Cuartil 75%	Coficiente de variación	Error estándar
Stru	4130	-0,058676	-0,010	-1,14	2,170	-0,450	0,150	-1004,42	0,009171
Iso	4130	0,498964	0,570	-0,14	0,900	0,280	0,690	51,70	0,004014
Aten	4130	-0,011140	0,000	-1,14	0,060	-0,030	0,020	-513,74	0,000891
Chaos	4130	0,258569	0,200	0,03	1,000	0,100	0,270	97,49	0,003923
Imped	4130	0,292797	0,000	-3,89	4,650	-0,650	1,360	504,34	0,022978
Coseno	4130	0,077630	0,155	-1,00	1,000	-0,560	0,750	888,61	0,010734
RMS	4130	0,878266	0,920	0,00	2,480	0,520	1,130	59,05	0,008071
Sweet	4130	1,152121	0,955	0,00	3,970	0,570	1,610	73,47	0,013171

Tabla 3. Análisis estadístico descriptivo después de eliminar los valores fuera de rango.

Variable	Valid N	Media	Mediana	Mínimo	Máximo	Cuartil 25%	Cuartil 75%	Coefficiente de variación	Error estándar
Stru	4130	-0,058676	-0,010	-1,14	2,170	-0,450	0,150	-1004,42	0,009171
Iso	3995	0,493166	0,550	-0,14	0,880	0,280	0,690	52,71	0,004113
Aten	4112	-0,008436	0,000	-1,10	0,060	-0,030	0,020	-404,66	0,000532
Chaos	4125	0,257670	0,200	0,03	1,000	0,100	0,270	97,38	0,003907
Imped	4052	0,278801	0,000	-3,89	4,650	-0,660	1,370	533,25	0,023355
Coseno	4103	0,082939	0,170	-1,00	1,000	-0,560	0,760	830,44	0,010753
RMS	3933	0,809636	0,8910	0,00	1,920	0,520	1,080	52,80	0,006816
Sweet	3951	1,053637	0,920	0,00	3,180	0,530	1,530	68,35	0,011458

Comparando las Tablas 2 y 3 se aprecia una disminución apreciable de la variabilidad de los valores por atributos, que ayuda a la mejor clasificación de los datos. Los valores catalogados como fuera de rango se sustituyeron posteriormente por valores

ausentes (*missing*). Para el análisis (univariado) de la información del comportamiento de los atributos por facies, se emplearon gráficos de tipo caja y bigote. Se mostrarán los más significativos (Figuras 13 a 16).

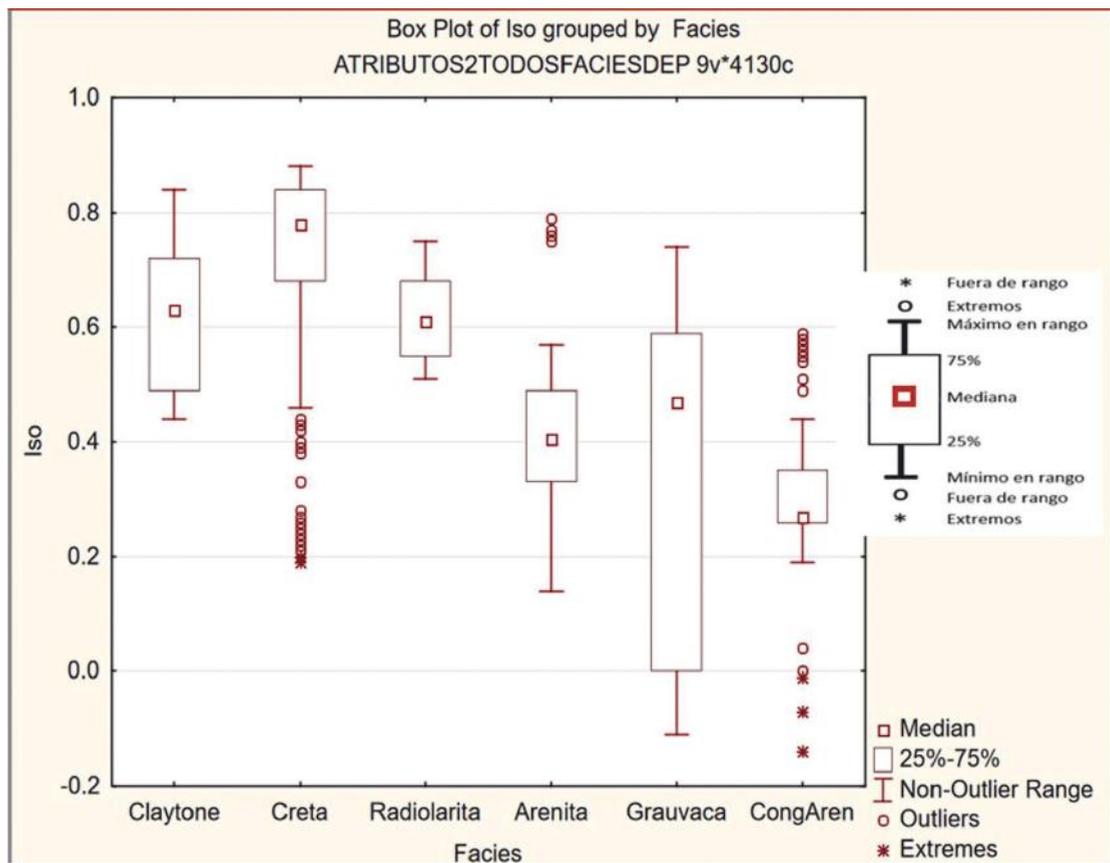


Figura 13. Gráfico de caja y bigote: mediana, percentiles de 25% y 75% y valores máximos y mínimos permitidos (derecha). Atributo isofrecuencia. Atributo con capacidad de diferenciar las litofacies no clásticas (creta, claystone y radiolarita) de las litofacies clásticas.

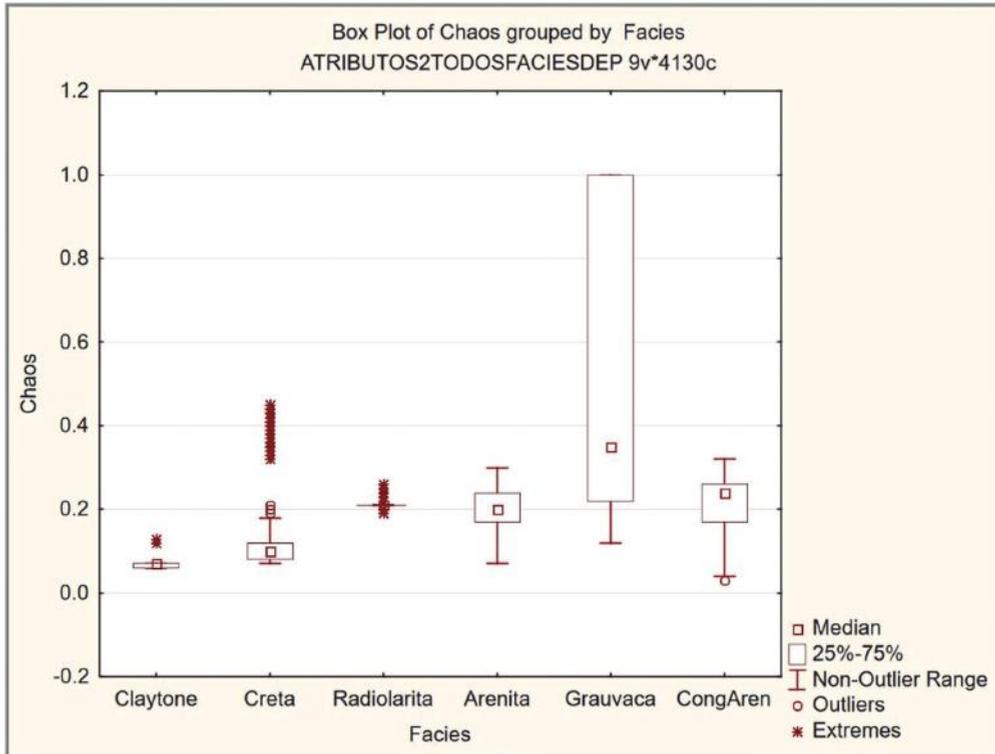


Figura 14. Atributo *Chaos*. Atributo capaz de diferenciar la mayoría de las litofacies agrupando a la arenita y conglomerado arenoso.

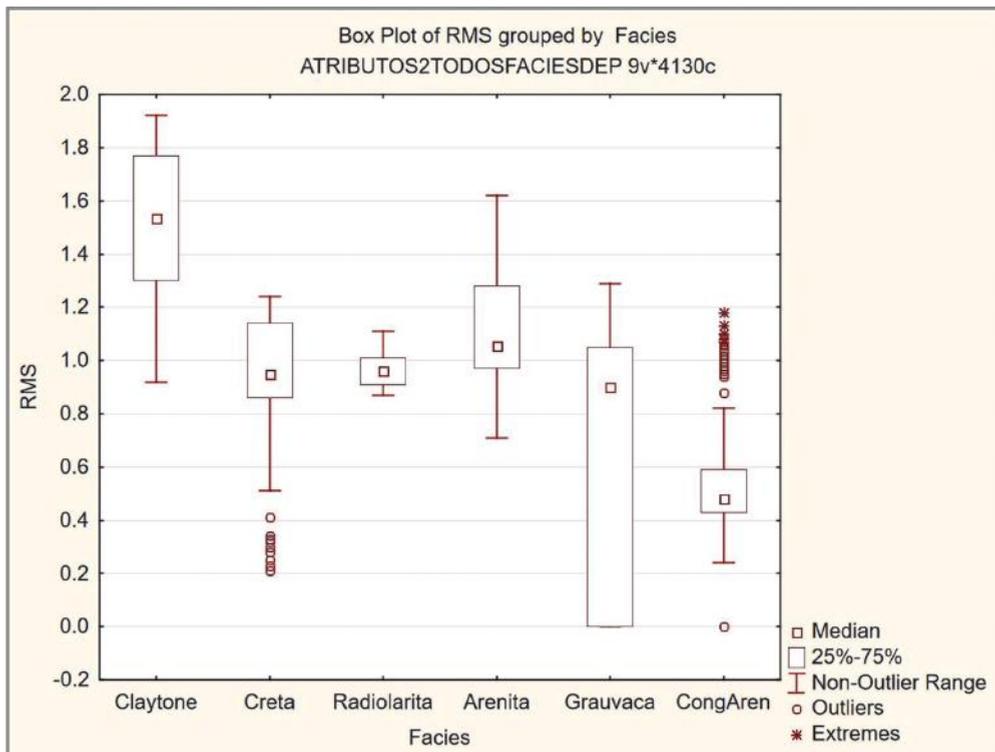


Figura 15. Atributo RMS. Presenta una buena diferenciación entre la litofacies *claystone* de las demás.

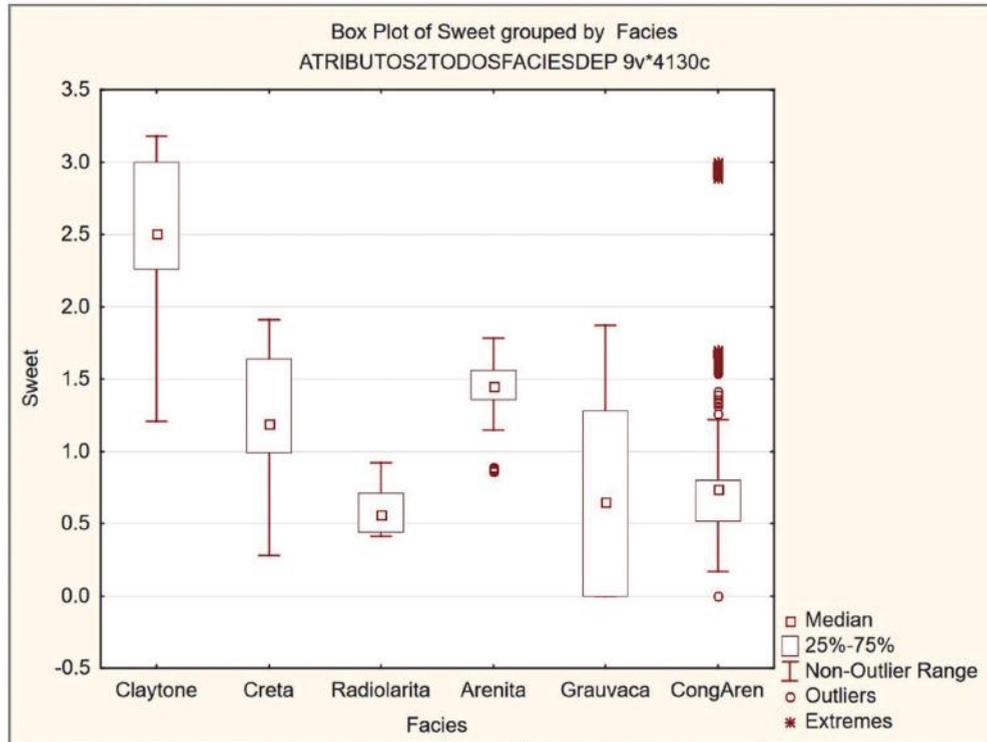


Figura 16. Atributo *Sweetness*. Se puede considerar que posee el mejor comportamiento discriminatorio entre las litofacies.

Después de realizado el análisis exploratorio (univariado) de los datos, se procede a la fase de minería de datos MD (multivariada), que constituye el centro del proceso KDD. Uno de los principales problemas que se enfrentan en la clasificación de objetos geológicos es la precedencia de la incertidumbre, ya que no se cuenta con el conocimiento suficiente de los fenómenos y los parámetros que se evalúan.

Discusión

Con esta información previa sobre el comportamiento de los atributos se puede pasar a la etapa de la utilización de metodologías multivariadas y clasificación supervisada. Se utilizó el *software* Weka 3.9.3 (Weka, 2008), que se define como una colección de algoritmos de aprendizaje por computadora (en librerías JAVA), para tareas de minería de datos y que contiene herramientas para la preparación, clasificación regresión, agrupamiento, reglas de asociación y visualización de datos.

En la Figura 17 se presentan los datos patrones según las litofacies: *claystone* 330 (azul oscuro, 7,99%), *creta* 1047 (rojo, 25,35%), *radiolarita* 303 (azul claro, 7,33%), *arenita* 154 (verde oscuro, 3,72%), *grauvaca*

1351 (ocre, 32,71%) y conglomerado 945 (verde claro, 22,88%) para un total de 4130 muestras con 8 atributos a clasificar.

Para este análisis del conocimiento se determinó manejar los clasificadores J48 (C4.5), *Naive Bayes* y redes neuronales artificiales.

El algoritmo J48 es de tipo árbol de decisión predictivo y es muy utilizado en MD. Permite realizar particiones tanto numéricas como nominales. Es muy robusto frente a datos ausentes. La técnica bayesiana (*Naive Bayes*) tiene entre sus principales características el uso explícito de la teoría de la probabilidad para cuantificar la incertidumbre. Es un modelo que permite un doble uso: descriptivo y predictivo. En los modelos descriptivos, los algoritmos de aprendizaje de redes bayesianas se centran en el descubrimiento de relaciones de interdependencia y/o relevancia entre sus variables.

Por su parte, las redes neuronales artificiales (en su variante *multilayer perceptron*) son modelos matemáticos capaces de adaptar su comportamiento en respuesta al medio de manera supervisada o no supervisada, y hacer una generalización una vez entrenada. Este clasificador realiza el aprendizaje mediante el ajuste multicapas y

sus pesos direccionales (*backpropagation*) (Hernández *et al.*, 2004).

Para el trabajo con estos algoritmos se empleó la opción de trabajo de validación cruzada (*crossvalidation*), que calcula el porcentaje esperado de aciertos mediante una validación de k evaluaciones.

Para la evaluación de los clasificadores en Weka se contabilizaron (para simplificar el análisis):

- a. Cantidad de instancias clasificadas correctamente (*Correctly Classified Instances CC*) (% de verdaderamente positivos o negativos y/o coincidentes entre la clasificación prefijada por el experto o patrón y la clasificación *a posteriori* dada por el algoritmo).
- b. Cantidad de instancias clasificadas incorrectamente % de (*Incorrectly Classified Instances IC*) (falsos positivos o negativos).
- c. Estadígrafo Kappa (*Kappa statistic*): se usa para

medir la correspondencia entre la clase predicha y la clase real, corrigiendo las correspondencias que ocurren por azar. Máximo valor 1 o 100%.

- d. El error cuadrático medio (*Root Mean Squared error, RMS*) es el error cuadrático normalizado por el error cuadrático total de un predictor por defecto en %.
- e. Matriz de confusión (*confusion matrix*) es la matriz o tabla de contingencia, está formada por tantas filas y columnas como clases existan. La cantidad de instancias clasificadas correctamente es la suma de la diagonal (o traza) de la matriz y el resto están clasificadas de forma incorrecta.

A continuación, se presenta en la Tabla 4 un resumen de los parámetros resultantes de cada análisis, con los algoritmos empleados.

En la Tabla 5 se presentan las matrices de confusión para cada algoritmo que incluyen las litofacies.

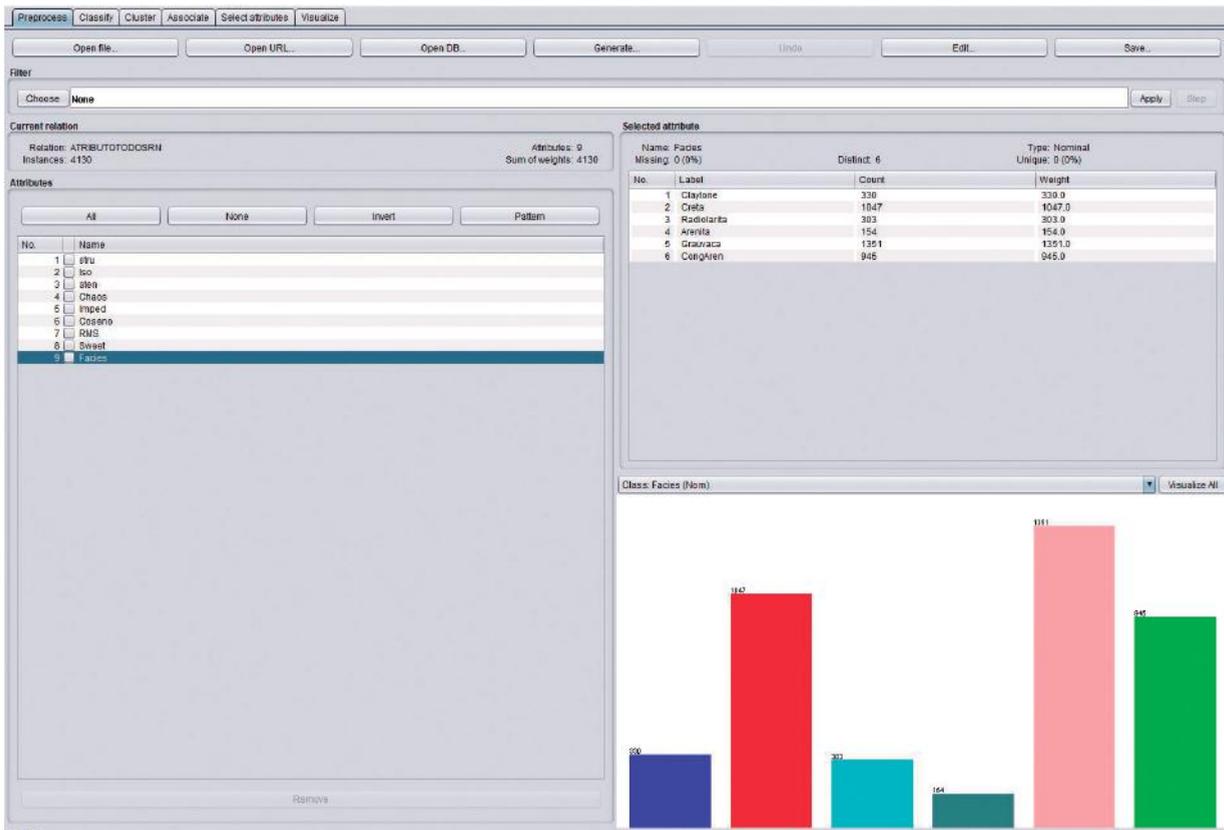


Figura 17. Entrada de datos en Weka, con el análisis preliminar de atributos y objetos.

Todos los algoritmos de clasificación utilizados se comportaron muy eficientes, para poder identificar las litofacies patrones según los atributos sísmicos, aunque el J48 se destaca como ligeramente superior en este estudio (Tabla 4 y 5).

Con el objetivo de profundizar en esta investigación, se aborda el tema de la selección de atributos (eliminación de atributos que pueden resultar redundantes e irrelevantes). Se procedió a determinar el peso informacional de estos,

con la utilización del algoritmo *WrapperSubSetEval* (Weka, 2018). Este algoritmo evalúa los conjuntos de atributos utilizando un esquema de aprendizaje. La validación cruzada se utiliza para estimar la precisión del esquema de aprendizaje, utilizando el algoritmo J48, con el cual se obtuvieron los mejores resultados. Para complementar esta evaluación, se empleó el método de algoritmo *BestFirst* (Weka, 2018) que busca el peso informacional en la combinación de los subconjuntos de atributos.

Tabla 4. Resultados de los algoritmos utilizados en porcentos.

Algoritmo	Objetos	CC%	IC%	Kappa%	RMS%
Naive Bayes	4130	87,53	12,46	83,5	17,8
J48	4130	99,44	0,55	99,27	3,88
Red Neural (1000 iteraciones)	4130	98,11	1,88	97,52	6,97
Red Neural (5000 iteraciones)	4130	98,45	1,54	97,97	6,46

Tabla 5. Matrices de confusión resultantes de los algoritmos utilizados. a) *Claystone*, b) *Creta*, c) *Radiolaritas*, d) *Arenita*, e) *Grauvaca*, f) *Conglomerado arenoso*.

a	b	c	d	e	f	
251	79	0	0	0	0	a
0	980	21	0	25	21	b
0	0	292	0	11	0	c
1	20	0	117	10	6	d
0	114	0	0	1237	0	e
0	65	0	17	125	738	f

Algoritmo *Naive Bayes*

a	b	c	d	e	f	
330	0	0	0	0	0	a
0	1045	1	0	0	0	b
0	1	300	1	0	1	c
0	2	1	147	0	4	d
0	0	0	0	1349	2	e
0	0	0	1	8	936	f

Algoritmo J48

a	b	c	d	e	f	
313	17	0	0	0	0	a
1	1019	3	4	8	12	b
0	3	299	1	0	0	c
0	3	6	125	7	13	d
0	0	0	0	1351	0	e
0	0	0	0	0	945	f

Algoritmo Red Neural (1000 iteraciones)

a	b	c	d	e	f	
321	9	0	0	0	0	a
2	1019	2	4	6	14	b
0	1	301	1	0	0	c
0	4	5	131	5	9	d
0	0	0	0	1351	0	e
0	0	0	2	0	943	f

Algoritmo Red Neural (5000 iteraciones)

Se determinan los atributos por su orden de significancia:

1. *Chaos* (100%)
2. RMS (90%)
3. Atenuación (80%)
4. *Structural smoothing* (70%)
5. Isofrecuencia (60%)
6. Coseno (60%)
7. *Impedance* (50%)
8. *Sweetness* (40%)

Respetando el conocimiento adquirido, tanto en el análisis univariado y multivariado se toma la decisión que todos los atributos utilizados poseen un nivel aceptable de discriminación cuando se desarrollan en un ambiente multivariado y ante la alta incertidumbre que presenta el modelo litofacial, se optó por considerarlos a todos para la clasificación.

Se procedió a la clasificación del modelo litofacial del yacimiento Santa Cruz del Norte y Jibacoa, para la Formación Vega Alta. Para realizar el pronóstico de toda

el área, se creó una matriz (escalado) de una malla con tamaño de celda de 50m x 50m x 50m (0,125 km³), como promedio general en el modelo geológico, para clasificar un total de 349852 celdas (43731,5 km³), que contiene los atributos sísmicos seleccionados, pero sin información litofacial.

Para el cálculo de las probabilidades de semejanza con los patrones se utilizó el sistema *Stanford Geostatistical Modeling Software* (SGeMS). Se basa en la geoestadística; o sea, el estudio estadístico espacial de fenómenos físicos y estimación de fenómenos naturales (Pyrz y Deutsch, 2014). Se aplicó del software SGeMS, el algoritmo geoestadístico de la Simulación Secuencial de Indicadores (SISIM), el cual permite la obtención de modelos simulados equi-probables para variables categóricas (o denominadas discretas desde el punto de vista numérico), dentro de un ambiente probabilístico (Arias del Toro y Pérez-Campos, 2015).

En la Figura 18 se muestra el modelo litofacial producto de la clasificación por atributos sísmicos estudiados con el algoritmo J48.

Para disminuir en cierta medida la entropía presente entre las litofacies y homogeneizar el modelo, a la matriz con dimensión de celdas de 50m x 50m x 50m, se le aplicó

un filtro de promedios móviles (para el cálculo de la probabilidad de pertenencia a cada facies), resultando una nueva matriz con celdas de 250m x 250m x 50m. Ya con las probabilidades calculadas por cada celda, se realizó una nueva depuración que incluye solamente aquellas celdas con un valor de pertenencia de similitud (probabilidad) con las litofacies patrones $\geq 70\%$ (Figura 19), a partir del modelo litofacial de la Figura 18.

Se puede apreciar el predominio de las litofacies de grauvaca y en menor medida de las arenitas y conglomerados. Este comportamiento (aunque modelado) se acerca al concepto estudiado en el modelo de microfacies y corresponde a la tendencia de los procesos turbidíticos ocurridos.

Se procedió a aplicar la misma metodología de clasificación, pero utilizando el algoritmo *multilayer perceptron*. El modelo resultante no se correspondió con el modelo de microfacies patrón, por lo que se descarta por ahora y queda sujeto a futuros estudios.

A partir del modelo obtenido se procedió al cálculo del riesgo para la exploración y potencial de recursos (especulativo) para gas húmedo mediante la modelación de Monte Carlo, según la metodología establecida (Gómez-Herrera y Rodríguez-Morán, 2013) (Figura 20).

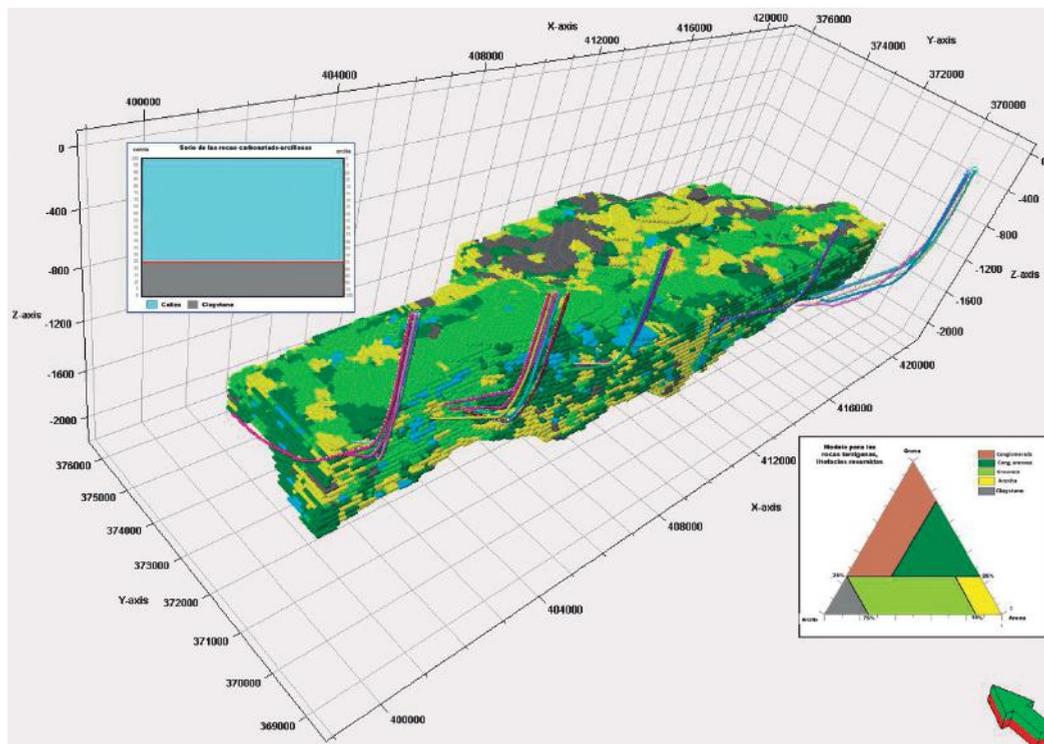


Figura 18. Modelo litofacial clasificado del área Santa Cruz-Jibacoa dentro de la Formación Vega Alta.

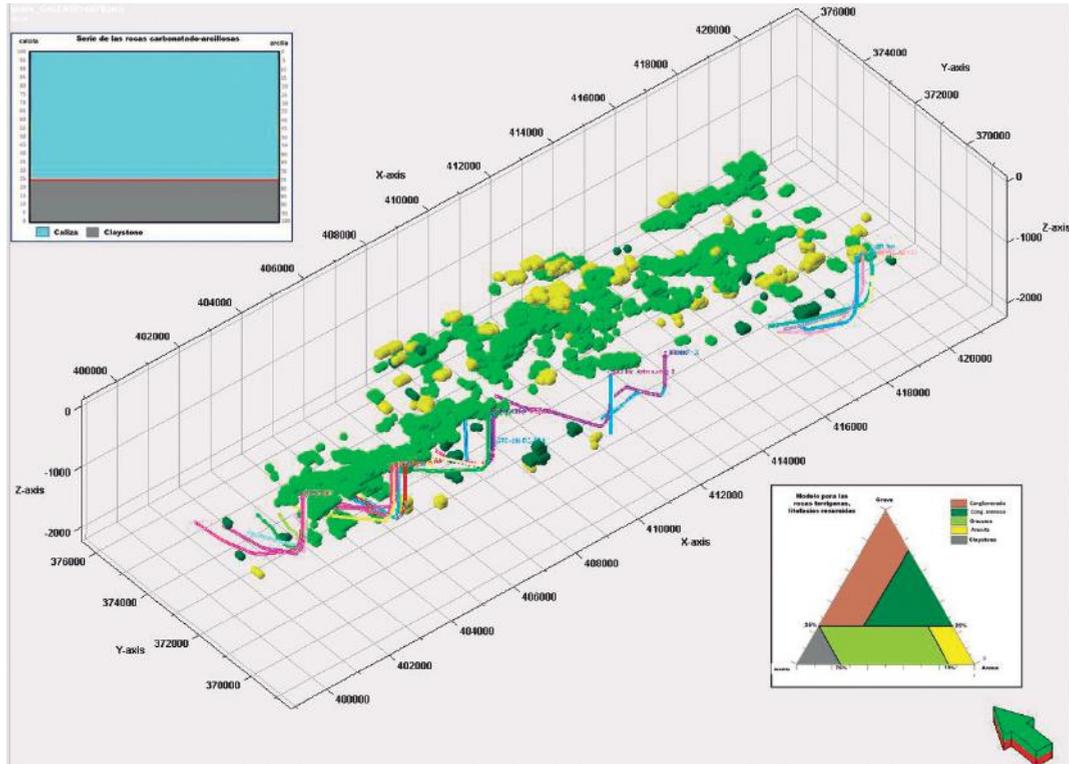


Figura 19. Volumen de reservorio con una confiabilidad (probabilidad) $\geq 70\%$, según modelo litofacial de 5 elementos (parte superior izquierda).

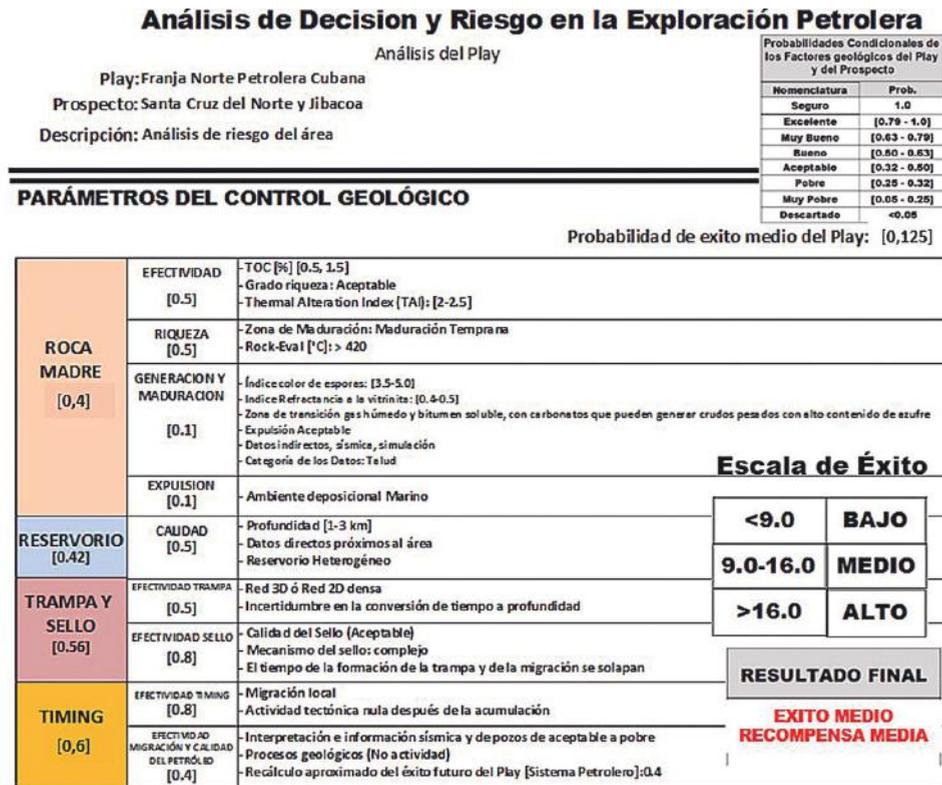


Figura 20. Detalle del cálculo de riesgo geológico para el play donde se ubican los yacimientos de Santa Cruz del Norte y Jibacoa.

Conclusiones

Los resultados obtenidos en el proceso de modelación de las litofacies partiendo del análisis de los atributos sísmicos presentan gran similitud y consistencia en su distribución, con el análisis microfacial de los intervalos seleccionados en los pozos (patrones con presencia de gas) del área Santa Cruz del Norte, derivados del estudio directo.

La utilización de técnicas estadísticas y de minería de datos en el proceso KDD, resultó ser muy satisfactoria y útil en la investigación de ambientes con insuficiente información geológica, alta incertidumbre y estén cubiertos por la sísmica, como es el caso de los reservorios gasíferos en la Formación Vega Alta.

Las litofacies de grauvaca y arenita, con alto grado de heterogeneidad, son las indicadas según el modelo como las más perspectivas para contener depósitos comercializables de gas húmedo.

En esta etapa de estudio, el algoritmo de clasificación supervisada de árboles de decisión J48 fue muy eficiente en estas condiciones de gran variabilidad e incertidumbre entre las litofacies y los atributos en el área de estudio.

Es necesario realizar un estudio más profundo sobre el mal resultado obtenido en la aplicación del algoritmo de redes neuronales en la clasificación. Es posible que esto se deba a un fenómeno de sobreajuste (*overfitting*) o desajuste (*underfitting*), que pueda estar vinculado con una insuficiente caracterización en la complejidad del modelo.

Se recomienda continuar la aplicación de esta metodología en otras áreas de la FNCP, sobre todo en aquellas donde esté planificada la perforación, para la verificación del modelo (en el área de Jibacoa en este momento ya se están perforando dos pozos exploratorios).

Agradecimientos

Los autores agradecen a los especialistas del Centro de Investigaciones del Petróleo, específicamente al Departamento de Petrofísica, por el trabajo científico desarrollado en la creación de los modelos, a la MSc. Dolores González Despaigne, a la división

de interpretación geofísica, al MSc. Carlos Veiga Bravo y a los especialistas del grupo de Gestión del conocimiento, MSc. Lorenza Mejías Rodríguez y MSc. Ramón Pico Peña.

Referencias

- Arias del Toro, J.A.; Pérez-Campos, M.A. (2015). Aplicación del software SGEMS para la modelación geoestadística y la estimación de los recursos minerales. Caso de estudio, depósito arcillas "Moja Hueva". *XII Congreso Cubano de Informática y Geociencias*. La Habana, Cuba.
- Chopra, S.; Marfurt, K.J. (2007). *Seismic Attributes for Prospect Identification and Reservoir Characterization*. Society of Exploration Geophysicists. Geophysical Development Series. No. 11.
- Daber, R.; Aqrabi, A. (2010). *Interpreter's Guide to Seismic Attributes*. Houston: Schlumberger Limited.
- Folk, R. (1980). *Petrology of sedimentary rocks*. Austin: Hemphill Publishing Company.
- Gómez-Herrera, J.; Rodríguez-Morán, O. (2013). Metodología para el análisis del riesgo petrolero y toma de decisiones. Caso de estudio Franja Norte de Crudos Pesados, sector Guanabo-Seboruco, Cuba. *Anuario de la Sociedad Cubana de Geología*, 1, 23-31.
- Hernández, J. (2015). Proyecto 7043. Etapa 3. Bioestratigrafía de los depósitos de los sellos en Cuba occidental. Archivo Técnico del Centro de Investigaciones del Petróleo. La Habana, Cuba.
- Hernández, J. (2018). Proyecto 7082. Exploración de gas en Cuba. Archivo Técnico del Centro de Investigaciones del Petróleo. La Habana, Cuba.
- Hernández-Orallo, J.; Ramírez-Quintana, M.; Ferrer-Ramírez, M.J. (2004). *Introducción a la Minería de Datos*. Madrid: Pearson Educación, S.A.
- Martínez, R.; Valdez, J. (2011). Cálculo de reservas en yacimientos de gas. Tesis, Universidad Nacional Autónoma de México, México.

- Morales, C. (2014). Proyecto 7078. Exploración en el sector Morro - Boca de Jaruco en la Franja Norte de Crudos Pesados. Archivo Técnico del Centro de Investigaciones del Petróleo. La Habana, Cuba.
- Pettijohn, F.J. (1975). *Sedimentary rocks*. 3ra edición. New York: Harper & Row.
- Pyrcz, M.J.; Deutsch, C.V. (2014). *Geostatistical reservoir modeling*. Oxford University Press. 2nd Edition.
- Taner, T. (2001). Seismic Attributes. *CSEG Recorder*, 26(7), 48-56.
- Weka. (2018). Waikato environment for knowledge analysis. Version 3.9.3. Waikato University. New Zealand. <https://www.cs.waikato.ac.nz/ml/weka/>

Julio Ernesto Gómez-Herrera
ORCID: 0000-0001-5184-4148

Osvaldo Rodríguez-Morán
ORCID: 0000-0001-9153-4603

Yeniley Fajardo-Fernández
ORCID: 0000-0002-5686-6490

José Hernández-León
ORCID: 0000-0002-1135-2971

José Alberto Arias del Toro
ORCID: 0000-0001-6879-4626

Zulema Domínguez-Sardiñas
ORCID: 0000-0002-2539-5673

Lourdes Jiménez de la Fuente
ORCID: 0000-0003-3735-442X

Trabajo recibido: junio 06 de 2019
Trabajo aceptado: abril 24 de 2020