

## EXPLORACIÓN PRELIMINAR DE NUEVOS RECURSOS ENERGÉTICOS MEDIANTE LA MINERÍA DE DATOS. CASO HIDRÓGENO

### PRELIMINARY EXPLORATION OF NEW ENERGY RESOURCES THROUGH DATA MINING. HYDROGEN CASE

JULIO ERNESTO GÓMEZ HERRERA\*, ZULEMA DOMÍNGUEZ SARDIÑAS, LOURDES JIMÉNEZ DE LA FUENTE

*Centro de Investigación del Petróleo. CUPET-MINEM. Churruca No 481, Cerro, La Habana, Cuba C.P. 12000.*

*\*[juliog@ceinpet.cupet.cu](mailto:juliog@ceinpet.cupet.cu)*

**RESUMEN:** La disminución de las reservas de hidrocarburos, la necesidad energética de los países y la necesaria disminución de la huella de carbono para mitigar el cambio climático, determinan la obligación de buscar nuevas fuentes energéticas o alternativas, para resolver estos problemas. Por tal motivo, ha cobrado una gran fuerza la exploración y explotación del gas natural y otras fuentes alternativas y Cuba no es una excepción. En el año 2013, se comenzó un estudio de exploración de gas en las áreas de la Franja Petrolera Norte Cubana y el área de la Cuenca Central donde se encuentra el yacimiento de Motembo (1881), histórico por sus volúmenes de gas natural y petróleos ligeros entrampados en rocas ofiolíticas. En el muestreo de los gases en esta área se evidenció la existencia de hidrocarburos gaseosos y de altos valores de hidrógeno molecular. A nivel mundial, en países como los EE. UU., Rusia, Francia, Mali, etc. se investiga este gas para incorporarlo a la canasta energética. Este trabajo es una primera mirada para comenzar a estimar este recurso energético alternativo en Cuba. El objetivo es, obtener una metodología para su evaluación con la asociación con otros atributos geólogo - geofísicos, aplicando las técnicas de análisis exploratorio y minería de datos. Este ejercicio preliminar indica que la aplicación de estas técnicas son una importante herramienta para el análisis multivariado y la de relaciones entre los atributos gaseosos y las variables geólogo - geofísicas.

**Palabras Clave:** Hidrógeno molecular, atributos geólogo - geofísicos, minería de datos, energías no convencionales.

**ABSTRACT:** The decrease in hydrocarbon reserves, the energy needs of countries and the necessary reduction in the carbon footprint to mitigate climate change, determine the obligation to search for new or alternative energy sources to solve these problems. For this reason, the exploration and exploitation of natural gas and other alternative sources has gained great strength and Cuba is no exception. In 2013, a gas exploration study began in the areas of the Cuban Northern Oil Belt and the Central Basin area where the Motembo field (1881) is located, historic for its volumes of natural gas and light oils, trapped in ophiolitic rocks. In the gas sampling in this area, the existence of gaseous hydrocarbons and high values of molecular hydrogen were evident. Globally, in countries like the USA, Russia, France, Mali and others, this gas is being investigated to incorporate it into the energy basket. This work is a first look to begin to estimate this energy resource in Cuba, with the objective of obtaining a methodology for its evaluation and association with other geologist-geophysical attributes, applying exploratory analysis and data mining techniques. This preliminary exercise indicates that the application of these techniques is an important tool for multivariate analysis and search for relationships between gas attributes and geologist-geophysical attribute.

**Keywords:** Molecular hydrogen, geologist-geophysical attributes, data mining, unconventional energies.

**Recibido:** 03/04/2024

**Aceptado:** 15/05/2024



Este artículo se encuentra bajo los términos de la licencia Creative Commons Attribution-NonCommercial (CC BY-NC 4.0).  
<https://creativecommons.org/licenses/by-nc/4.0/>



## INTRODUCCIÓN

La disminución de las reservas de hidrocarburos, el aumento la necesidad energética para el desarrollo del país y la necesaria e impostergable disminución de la huella de carbono para mitigar el cambio climático, ha determinado la obligación de buscar nuevas fuentes energéticas o alternativas, para resolver estos problemas. Por tal motivo, ha cobrado una gran fuerza la exploración y explotación del gas natural y otras fuentes alternativas y Cuba no es una excepción. En el año 2013, se comenzó el estudio de exploración de gas convencional y no convencional, en las áreas de la Franja Petrolera Norte Cubana (FPNC) y la Cuenca Central (CC), con énfasis en el área de Motembo, ejemplo histórico de Cuba por la calidad de su gas natural y petróleos ligeros entrampados en el complejo ofiolítico.

El área de estudio de Motembo y su yacimiento homónimo, se localiza al suroeste de la región geográfica de Corralillo y sus propiedades gasopetrolíferas se conocen desde antes de la colonización de la Isla de Cuba. Cuna de las primeras minas de nafta y gas natural conocidas que datan del año 1880, ocupando una superficie de más de medio millón m<sup>2</sup>. Los primeros trabajos de exploración datan del año 1881 y los primeros pozos se perforaron con grandes manifestaciones de nafta y gas entre 244 y 305m de profundidad, estabilizándose la explotación comercial en la década de los años 40.

Aunque ya no se explotan industrialmente los pozos, sí se extraen pequeños volúmenes de nafta de las serpentinitas y se emplea el gas para uso doméstico por los pobladores.

La región se caracteriza por el desarrollo de formaciones cuyas edades oscilan entre el Jurásico Superior al Mioceno, representadas mayormente por secuencias carbonatas, ofiolitas, olistostromas poli-componentes y secuencias arenosas. Su estructura geológica está representada por mantos cabalgados y la superposición de diferentes eventos tectónicos con sus implicaciones en las subsiguientes deformaciones plicativas y disyuntivas que agudizan la complejidad estructural de la zona (Figura 1 y 2).

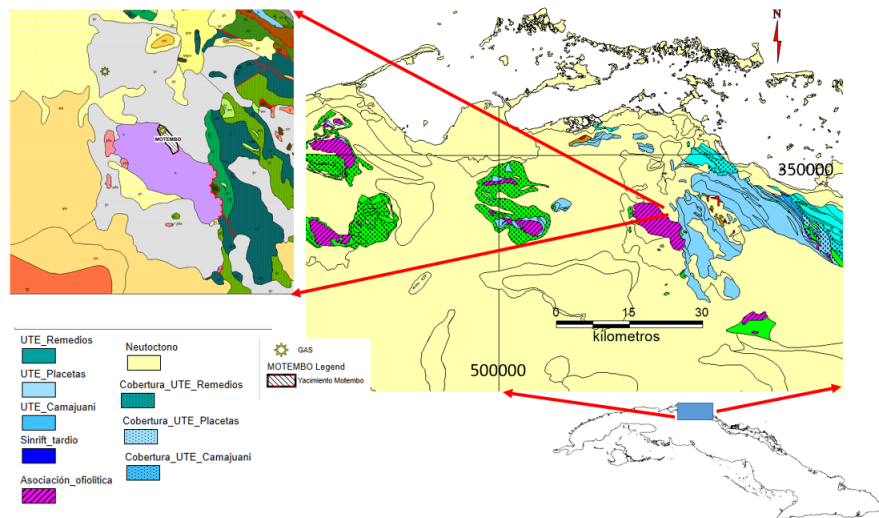


Figura 1. Área de estudio de Motembo

En las disciplinas geológicas, que en forma genérica se denominan "ciencias poco formalizadas", se dan un conjunto de características, que hacen que el problema de la modelación matemática se torne tremendamente complejo. Por eso podemos afirmar que, en la ciencia geológica, la mayoría de los modelos geológicos son verdaderos y falsos a la vez; verdaderos, porque representan a la naturaleza, pero por la incertidumbre de la información, hacen que este se convierta en incompleto, por lo que es falso (Botezatu, 1976).



Otro ejemplo se muestra en la **Figura 4**, donde, con la utilización de interpretaciones de imágenes satelitales y el modelo de elevación digital, se propone la diferenciación de las zonas potencialmente emisoras de  $H_2$ . La similitud geológica y la presencia de numerosas depresiones subcirculares morfológicas, constituyen un buen indicador de las emanaciones de  $H_2$  en Namibia. La vegetación ya observada parece capaz de discriminar estas estructuras (**Moretti, 2022**).



**Figura 3.** Área de Jones Lake Bay en Carolina del Norte, EE UU (Zgonnik, 2015)



**Figura 4.** Imagen en el campo, donde las depresiones no estaban sin vegetación, sino con pasto en lugar de árboles (Moretti, 2022)

Se han descubierto depósitos de hidrógeno en Rusia, Estados Unidos, Brasil, Mali y Francia. Se están llevando a cabo estudios geoquímicos de superficie, consistentes en medir las concentraciones de hidrógeno en el suelo, a alrededor de un metro de profundidad, con el fin de determinar los flujos de emisión hacia la atmósfera.

Una fuente de hidrógeno natural es la asociada a las ofiolitas. Desde el punto de vista geológico, en formaciones ofiolíticas, el hierro ferroso se disuelve en el agua y se produce una reacción de reducción-oxidación; el hierro ferroso se oxida a hierro férrico ( $Fe^{3+}$ ), mientras que el agua se reduce y da lugar a hidrógeno molecular. Se trata de la denominada serpentinización, que responde a procesos químicos.

## MATERIALES Y MÉTODOS

Para el análisis de toda la información se confeccionó una Tabla-Objeto-Propiedad (TOP) para su posterior modelación matemática, mediante algoritmos de reconocimiento de patrones de naturaleza lógico-combinatoria. Los atributos utilizados fueron los análisis geoquímicos (Figura 5), interpretaciones geomorfológicas (Figura 6), sensores remotos (Figura 7), campos potenciales (Figura 8) y otros atributos geólogo - geofísicos.

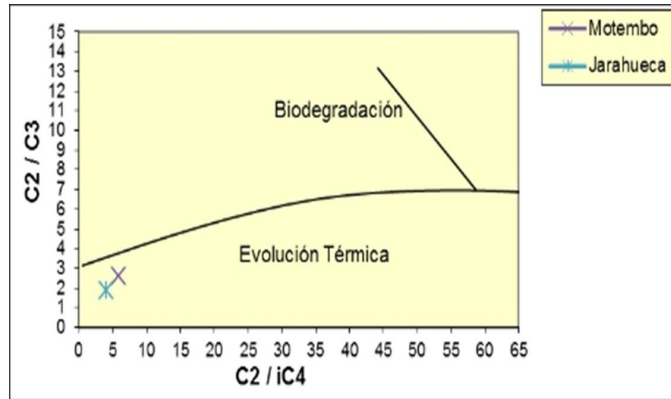


Figura 5. Análisis geoquímico

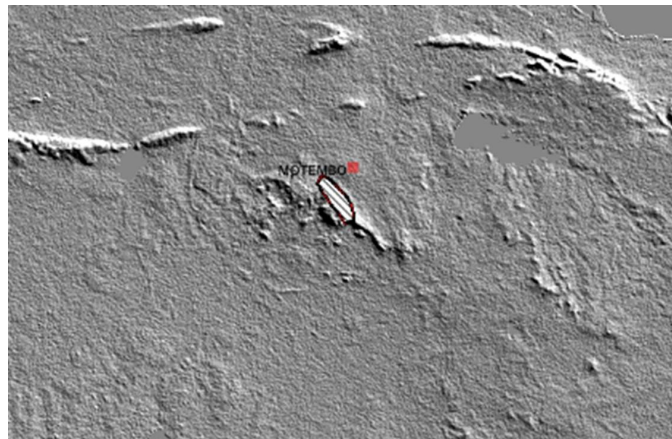


Figura 6. Modelo Digital del terreno

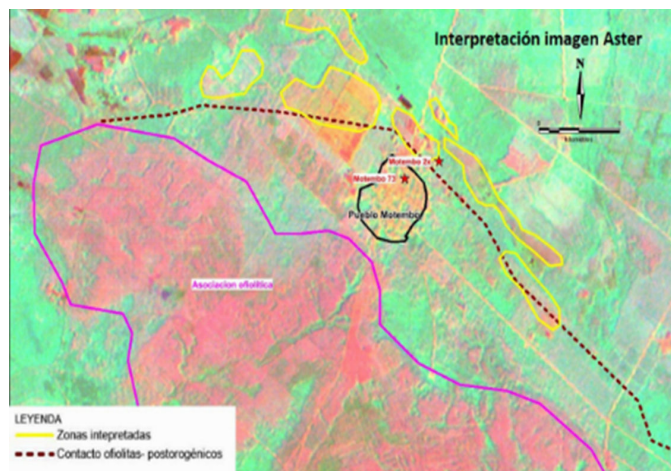


Figura 7. Sensores Remotos. Imagen Aster

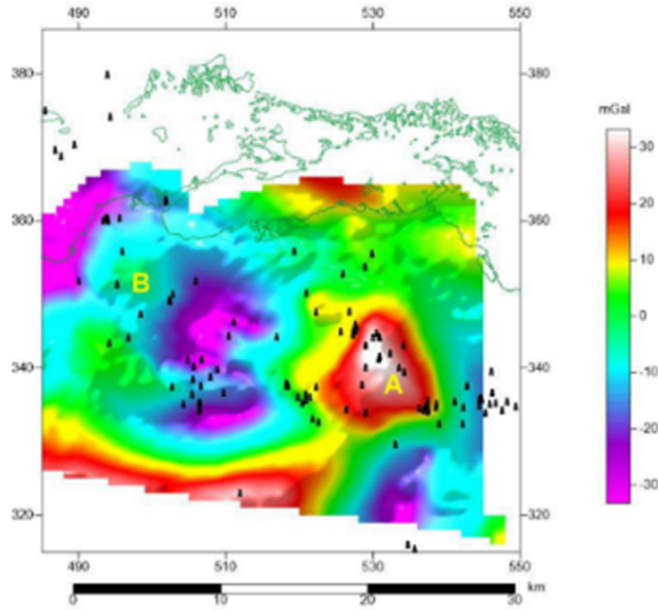


Figura 8. Campos Potenciales y sus derivados

## RESULTADOS Y DISCUSIÓN

Para este análisis preliminar- exploratorio solo se realizó un estudio, con el objetivo de encontrar alguna relación de dependencia entre los componentes gaseosos  $C_nH_n$  y  $H_2$  y dejar para más adelante el análisis complejo entre todos los atributos geólogo - geofísicos.

Para la modelación se caracterizó, como área patrón, la porción de la región donde se realizaron las mediciones de los elementos gaseosos, con buena información en la mayoría de los atributos. (Figura 9).

En la tabla 1 se presenta el análisis exploratorio básico de los atributos. Un elemento llama la atención y son los valores de hidrógeno, los cuales son extraordinariamente grandes con respecto a la cadena de hidrocarburos gaseosos medidos. Esta anomalía fue lo que provocó la concentración del estudio en este fenómeno y la hipótesis de que el hidrógeno molecular debe tratarse como un gas no convencional, con capacidad energética.

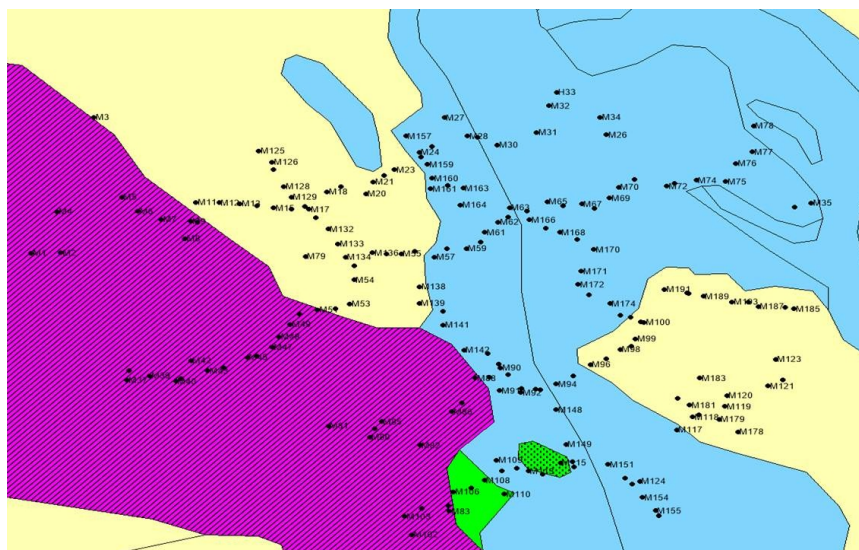


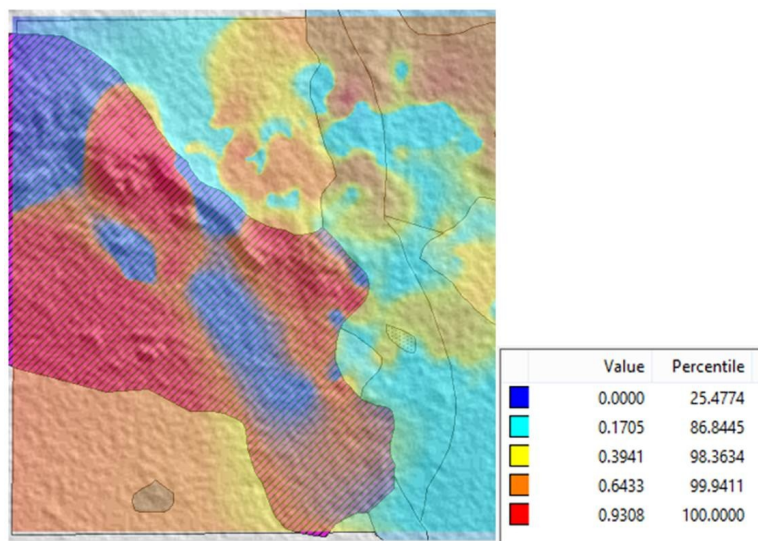
Figura 9. Área patrón con la información geólogo-geofísica y puntos medidos

**Tabla 1.** Análisis exploratorio de la interpretación geoquímica (ppm)

Gases	HeNe	H2	C1/C2	C1/C3	C2/C3	C2/iC4	iC4/nC4	C1/ (C2+C3)
Número de valores	261.00	261.0	261.00	261.00	261.00	261.00	261.00	261.0
Valor Mínimo	11670.00	4654.00	2.55	2.17	0.46	1.51	0.10	1.18
Valor Máximo	37432.00	10911645.00	468.01	1578.09	4.58	49.12	1.44	360.96
Rango	25762.00	10906991.0	465.46	1575.92	4.12	47.60	1.34	359.78
Media	24778.34	558339.17	14.19	32.95	2.01	11.20	0.56	9.56
Mediana	25070.00	115334.00	8.63	19.96	1.98	8.24	0.55	5.96
Primer cuartil	22591.25	47963.0	6.80	10.95	1.36	5.36	0.43	4.36
Tercer cuartil	26965.75	388683.25	12.57	25.10	2.61	14.86	0.67	8.07
Error Standard	258.47	81827.84	2.33	7.71	0.05	0.53	0.01	1.79
95% intervalo de confianza	509.40	161266.30	4.60	15.19	0.10	1.04	0.02	3.53
99% intervalo de confianza	671.28	212515.07	6.06	20.02	0.13	1.37	0.03	4.65
Varianza	17437133.99	1747602430771.58	1421.50	15508.51	0.61	73.11	0.04	836.65
Desviación media	3090.47	685694.57	9.68	27.97	0.65	6.48	0.15	6.97
Desviación Standard	4175.78	1321969.15	37.70	124.53	0.78	8.55	0.20	28.92
Coefficiente de variación	0.17	2.37	2.66	3.78	0.39	0.76	0.35	3.03
Kolmogórov-Smirnov Prueba	0.07	0.34	0.38	0.41	0.06	0.16	0.07	0.39
ANOMALIA GENERAL	28954.0	1880308.0	52.0	157.0	3.0	20.0	1.0	38.0
ANOMALIA Normalizada	0.67	0.17	0.11	0.10	0.57	0.38	0.49	0.10

Como los valores de hidrógeno eran muy desproporcionados con respecto a los otros gases, el análisis exploratorio se realizó con los datos originales y normalizados (valores entre 0 y 1).

En la **Figura 10** se pueden apreciar los altos valores de los gases de hidrógeno sobre las rocas del complejo ofiolítico.



**Figura 10.** Altos valores del hidrógeno sobre las ofiolitas (rojo)

Para realizar este estudio, se utilizó el procesamiento por una herramienta compleja de minería de datos denominado WEKA (*Waikato Environment for Knowledge Analysis*). El paquete Weka contiene una colección de herramientas de visualización y algoritmos para análisis de datos y modelado predictivo, unidos a una interfaz gráfica de usuario para acceder fácilmente a sus funcionalidades. Los resultados se muestran en la **Figura 11**. Este es un análisis predictivo de clasificación con la utilización del algoritmo J48 (técnica de minería de datos de árboles de decisión). Con árboles de decisión, no se suelen obtener modelos tan precisos como con otras técnicas (Redes Neuronales, SVM (*support vector machine*), etc.) pero se puede extraer conocimiento oculto, interpretar qué variables son más importantes y descubrir reglas útiles para la toma de decisiones. Además, son robustos ante los datos incompletos o *missing*, se pueden trabajar con una gran cantidad de variables y son fáciles de programar.

```

=== Summary ===

Correctly Classified Instances      232           88.8889 %
Incorrectly Classified Instances    29            11.1111 %
Kappa statistic                    0.6899
Mean absolute error                0.1846
Root mean squared error            0.3038
Relative absolute error            47.3621 %
Root relative squared error        68.8942 %
Total Number of Instances         261

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.969   0.333   0.890     0.969   0.928     0.702   0.889    0.932    M
                0.667   0.031   0.885     0.667   0.760     0.702   0.889    0.762    T
Weighted Avg.   0.889   0.253   0.889     0.889   0.883     0.702   0.889    0.887

=== Confusion Matrix ===

  a  b  <-- classified as
186  6 |  a = M
 23 46 |  b = T

```

**Figura 11.** Resultados de la clasificación del análisis con Weka

En resumen, de las 261 muestras gaseosas patrones medidas clasificaron correctamente 232 para una efectividad de un 89%, para un error medio cuadrático de un 30,38%. El índice Kappa toma el valor 0,69 por lo que existe un alto grado de concordancia. Con estos resultados se considera que existe una buena correlación entre la familia de los gases y el hidrógeno molecular, ratificado mediante la matriz de confusión donde su diagonal indica un bajo porcentaje de falsos positivos.

En la **Figura 12**, se proyecta el árbol de decisión donde se muestra la relación entre el atributo de los gases  $C_1$  (metano),  $C_2$  (etano) e hidrógeno ( $H_2$ ) (rama de la izquierda).

## CONCLUSIONES

1. Solamente con este simple análisis preliminar se demuestra que, existe una buena relación entre la familia de los hidrocarburos gaseosos y el hidrógeno.
2. El área estudiada de Motembo se comporta de una manera muy parecida a las existentes a nivel mundial, por lo que apunta hacia una nueva dirección de exploración para la futura utilización de esta nueva fuente de energía alternativa, por reportarse buenas anomalías de contenidos de hidrocarburos gaseosos y zonas anómalas asociadas a contenidos significativos de hidrógeno gaseoso, dentro del complejo ofiolítico.
3. La aplicación de las técnicas de minería de datos aporta una importante herramienta para el análisis multivariado y la búsqueda de relaciones entre los atributos gaseosos.

4. Se debe combinar el análisis de los gases con los demás parámetros geólogo - geofísicos utilizando, tanto los algoritmos de minería de datos, como los métodos de reconocimiento de patrones, para encontrar relaciones más complejas.
5. Se deben proseguir las investigaciones sobre este aspecto para poder evaluar las potencialidades de los gases de los hidrocarburos y del hidrógeno molecular, al mismo tiempo.

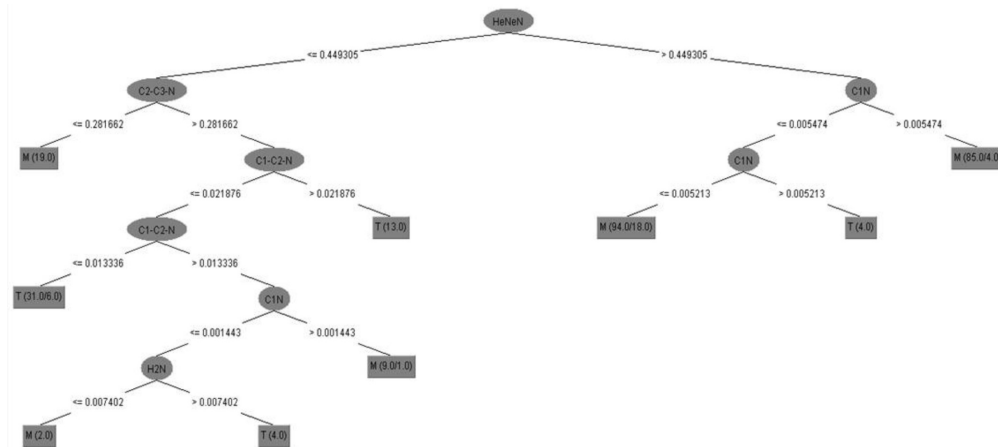


Figura 12. Análisis del árbol y la relación entre el atributo de los gases

## REFERENCIAS

- Botezatu, R. (1976). *Modele geofizice ale alcatuirii geologice a Romanici. Editura Academici, Bucuresti, Romania. 1976.* Bucuresti, Romania. 1976, Romania: Universidad de Bucarest.
- Cheremisina, E. (1989). Sistema automatizado ALISA para la solución de problemas en zonas poco formalizadas del conocimiento. Un ejemplo de la Geología. (ICIMAF, Ed.) *Revista Ciencias Matemáticas*(10), 53-64.
- Moretti, I. (2022). Natural hydrogen emanations in Namibia: Field acquisition and vegetation indexes from multispectral satellite image analysis. (S. Universities, Ed.) *Hydrogen Energy Publications LLC. Published by Elsevier Ltd.*
- Zgonnik, Z. (2015). Evidence for natural molecular hydrogen seepage associated with Carolina bays (surficial, ovoid depressions on the Atlantic Coastal Plain, Province of the USA). *Progress in Earth and Planetary Science*, 2:31. doi: <http://doi.org/10.1186/s40645-015-0062-5>

**Julio Ernesto Gómez Herrera.** Graduado de Ingeniero Geólogo – Geofísico en la Universidad de Bucarest, Rumania (1981). Doctor en Ciencias Técnicas. Ha trabajado para diferentes compañías petroleras fuera de Cuba y actualmente trabaja en el Centro de Investigación del Petróleo (CEINPET) en la Dirección de Geología, Grupo de exploración petrolera. Profesor Auxiliar de la Universidad Tecnológica de La Habana "José Antonio Echeverría" (CUJAE), Facultad de Informática. Perfil ocupacional; Interpretación compleja geólogo – geofísica, modelación matemática, reconocimiento de patrones, informática petrolera. Sistemas de Información Espacial, Procesamiento de imágenes satelitales, Análisis de decisión y riesgo en la Exploración Petrolera, geoestadística y minería de datos, gerencia de Proyectos de Investigación y Desarrollo y gestión de la Información. Ha participado en numerosos eventos científico técnicos y fórum de su especialidad y proyectos de investigación. Ha recibido diferentes reconocimientos estatales y no gubernamentales. Es autor de gran cantidad de publicaciones. Miembro de la Sociedad Cubana de Geología.