

The logo for the Revista Cubana de Ciencias Informáticas (RCCI) is displayed in a blue, bold, sans-serif font. The letters 'RCCI' are prominent, with a small 'i' at the end. The logo is contained within a white rectangular box with rounded corners.

Revista Cubana de Ciencias Informáticas

ISSN: 1994-1536

ISSN: 2227-1899

Editorial Ediciones Futuro

Valcarce Ortega, Rosa María; Rodríguez Miranda, Willy Roberto
Reconocimiento de Patrones aplicado a la identificación litológica del corte geológico.
Revista Cubana de Ciencias Informáticas, vol. 12, núm. 3, 2018, Julio-Septiembre, pp. 93-103
Editorial Ediciones Futuro

Disponible en: <https://www.redalyc.org/articulo.oa?id=378365832007>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UAEM The logo for redalyc.org, featuring the text 'redalyc.org' in a stylized font with a red dot above the 'i'.

Sistema de Información Científica Redalyc
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

Tipo de artículo: Artículo original
Temática: Reconocimiento de Patrones
Recibido: 18/10/2017 | Aceptado: 06/02/2018

Reconocimiento de Patrones aplicado a la identificación litológica del corte geológico

Recognition of Patterns applied to the lithological identification of the geological profile

Rosa María Valcarce Ortega ^{1*}, Willy Roberto Rodríguez Miranda ²

¹ Ingeniera Geofísica, Profesora del Departamento de Geociencias de la Universidad Tecnológica de La Habana “José Antonio Echeverría”, CUJAE, Calle 114 # 11901 E/ Rotonda y Ciclovía, Marianao, La Habana, Cuba. e-mail: rosy@tesla.cujae.edu.cu

² Ingeniero Geofísico, Profesora del Departamento de Geociencias de la Universidad Tecnológica de La Habana “José Antonio Echeverría”, CUJAE, Calle 114 # 11901 E/ Rotonda y Ciclovía, Marianao, La Habana, Cuba. e-mail: willy@civil.cujae.edu.cu

* Autor para correspondencia: rosy@tesla.cujae.edu.cu

Resumen

En la Geociencia las técnicas de Reconocimiento de Patrones encuentran un amplio campo de aplicación. El objetivo de la presente investigación fue identificar las litologías presentes en los acuíferos de la Cuenca Sur de la provincia Artemisa y Mayabeque, y en la Cuenca M-IV-2 de la provincia Matanzas, aplicando la técnica del Análisis Euclidiano de Agrupación sobre parámetros geofísicos de pozo con apoyo muy limitado de descripciones de núcleos. Fue aplicado también el Análisis de Componentes principales y el Análisis de Factores para estudiar la estructura de correlación del sistema de atributos originales. Los resultados obtenidos demuestran el cumplimiento del objetivo trazado y la efectividad de la metodología aplicada, logrando definir tres grupos litológicos básicos diferenciados por su desarrollo cársico y arcillosidad.

Palabras clave: reconocimiento de patrones, identificación litológica

Abstract

Pattern recognition techniques find a broad field of application in Geoscience. The objective of the present investigation was to identify the lithologies present in the aquifers of the southern basin of the province of Artemisa and Mayabeque and in the basin M-IV-2 of the province of Matanzas applying the technique of Cluster Analysis Euclidian with geophysical parameters very limited support of the basic descriptions. Principal Component Analysis and Factor Analysis were also applied to study the original attribute system correlation structure. The results obtained demonstrate the fulfillment of the proposed objective and the effectiveness of the applied methodology, by defining three basic lithological groups differentiated by their karst development and arcillosity.

Keywords: lithological identification, recognition of Patterns.

Introducción

Durante las investigaciones geológicas la identificación de las litologías presentes es una de las tareas básicas que deben ser resueltas por los métodos geofísicos y en este sentido el empleo de las técnicas de reconocimiento de patrones supervisadas y no supervisadas ha demostrado su alto poder resolutivo.

El objetivo de esta investigación es identificar los tipos de rocas presentes en los acuíferos desarrollados en la Cuenca Sur de las provincias Artemisa y Mayabeque, y en la Cuenca M-IV-2 de la provincia de Matanzas, a partir de la información que aportan los registros geofísicos de pozo y las escasas descripciones de núcleos disponibles.

En ambas cuencas los acuíferos están desarrollados fundamentalmente en la Formación Güines en la que se describe una gran variedad de calizas con elevado grado de carsificación y en ocasiones lentes finos de arcillas. La alta carsificación de estas rocas las convierte en un excelente acuífero que almacena los mayores recursos de agua subterránea en estas provincias. Las Figuras 1 y 2 presentan la ubicación de las áreas estudiadas.

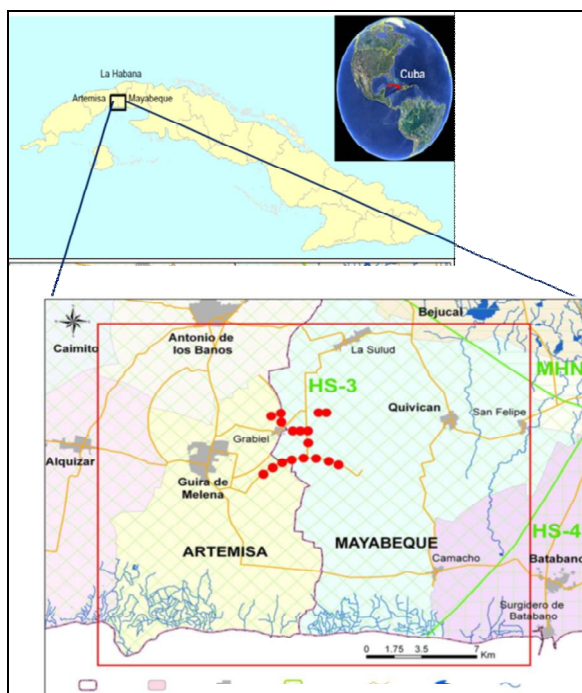


Figura 1. Ubicación del área de estudio en la Cuenca Sur de Artemisa y Mayabeque.

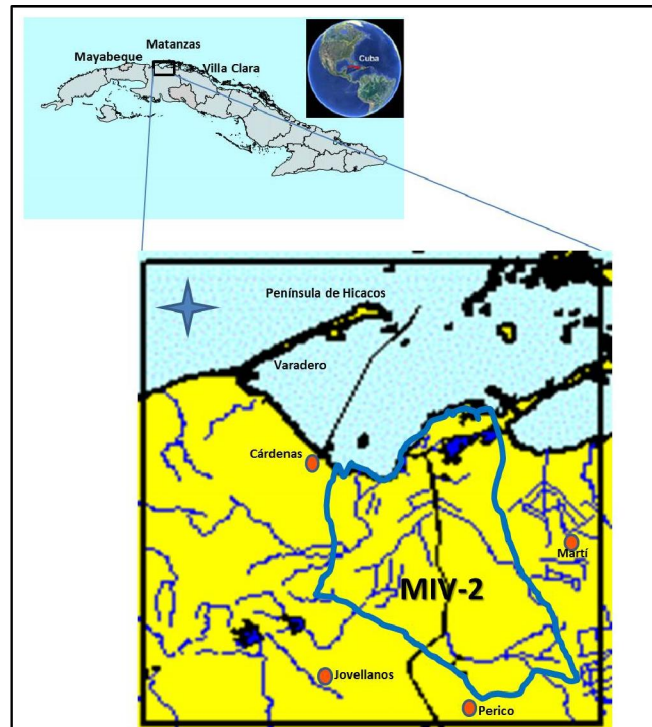


Figura 2. Ubicación de la Cuenca M-IV-2 de la provincia de Matanzas.

Métodos y metodología aplicada

En la Cuenca Sur de las provincias Artemisa y Mayabeque, se dispuso de 33 calas en las que fueron medidos registros de resistividad eléctrica del corte geológico y del agua subterránea, así como de la radiactividad gamma natural de las rocas. En 10 pozos de la Cuenca M-IV-2 se contó además de registros de intensidad neutrónica.

A partir de los registros geofísicos de pozo en las 33 calas de la Cuenca Sur de las provincias Artemisa y Mayabeque, fueron delimitadas las capas geológicas y cada una de ellas fueron caracterizadas por sus valores de radiactividad gamma natural (I_g) y factor de formación (F), obtenido éste como el cociente de la resistividad eléctrica de la roca 100% saturada de agua y la resistividad del agua que la satura. Debe destacarse que incrementos del factor de formación se asocian a bajos valores de la porosidad de las rocas e incrementos de la radiactividad gamma natural a elevados niveles de arcillosidad o a la presencia de dolomitización en estas formaciones geológicas.

En las 10 calas de la Cuenca M-IV-2, a partir de los registros geofísicos medidos, fueron delimitadas las capas geológicas presentes y caracterizadas por sus valores de intensidad gamma natural, factor de formación e intensidad

neutrónica (Inn). Las rocas con alta porosidad poseen bajos valores de la intensidad neutrónica, y viceversa (Valcarce, 1998; Kirsch, 2006).

Para dar respuesta al objetivo planteado, identificar las litologías presentes en el corte geológico a partir de la información que aportan los registros geofísicos de pozo y las escasas descripciones de núcleos disponibles, fueron aplicadas técnicas de clasificación estadística.

En las últimas décadas se ha producido un incremento notable con respecto al tratamiento de la información geológica por métodos lógico – numéricos, motivado por diferentes razones, tal vez la más importante sea el reconocimiento de que el alto grado de variación de los atributos estudiados exige, en mayor o menor extensión, el empleo de los métodos estadísticos de diagnóstico matemático y de modelación, a pesar de las limitaciones prácticas que en ocasiones impone el insuficiente muestreo. Otra razón importante, es la necesidad de realizar generalizaciones y predicciones a partir de un número limitado de datos, en muchos casos debido a la imposibilidad de acceso directo a los objetivos investigados (Díaz y Morales, 2012).

En general en las disciplinas de las Geociencias, al estudiar un sistema se utilizan parámetros cuyo comportamiento es el resultado de la superposición de muchas causas, posiblemente no simultáneas, y el conocimiento de estas causas es incompleto. Ello exige que el parámetro estudiado deba ser descrito empleando métodos de la estadística aplicada con un enfoque estocástico ya que en general adoptan un valor para cada punto del espacio donde pueden ser observados dos aspectos complementarios y aparentemente contradictorios:

- a) Un comportamiento aleatorio, asociado con las variaciones erráticas e impredecibles de estas variables.
- b) Un comportamiento estructurado que refleja las características más generales del sistema bajo estudio.

Es por ello que los problemas que se estudian en las Geociencias difícilmente pueden abordarse por métodos determinísticos y se aplican técnicas de reconocimiento de patrones para dar solución a diferentes tareas. A continuación se resumen los aspectos teóricos de la técnica estadística de clasificación empleada en esta investigación: el Análisis Euclidiano de Agrupación (AEA).

Las técnicas estadísticas de clasificación permiten identificar grupos de elementos de composición relativamente homogénea, estudiando las regularidades del conjunto de atributos usados para su descripción. A partir de la información “a priori” disponible, estas técnicas pueden dividirse en supervisadas y no supervisadas.

Las técnicas supervisadas emplean una muestra de aprendizaje o patrón, que representa a los diferentes objetos geológicos de interés y que permitirá clasificar elementos de origen desconocido, según determinada regla de decisión. En este caso, el resultado de la clasificación tiene un sentido geológico aportado por la muestra de aprendizaje.

Las técnicas no supervisadas, no emplean muestra de aprendizaje, ellas se utilizan cuando no existe información “a priori” sobre los objetos geológicos, o cuando esta información es muy reducida. En estas técnicas la muestra a clasificar se subdivide en grupos a partir del grado de semejanza de los elementos, y el resultado de la clasificación no posee un significado geológico directo: se requiere realizar la correlación geológica de dichos resultados posteriormente. En esta investigación fue empleado el Análisis Euclidiano de Agrupación (AEA) como técnica de clasificación no supervisada. Esta es una de las técnicas de clasificación no supervisadas más simple y efectiva. En ella se asume que puede existir un número finito de grupos discretos, los cuales están definidos en el espacio euclidiano p dimensional (E_p) por medio de sus centroides. La dimensión del espacio euclidiano queda determinada por el número de atributos usados para describir los elementos de la muestra. El centroide de un grupo es el punto en E_p cuyas coordenadas son los valores medios de los atributos de los puntos pertenecientes a ese grupo. El método parte de la posible existencia de un grupo máximo de grupos ($N_{\text{máx}}$). Este número máximo de grupos siempre debe tomarse superior al posible número de grupos que, en términos geológicos, se espera que existan. Además se especifica el número mínimo (N_{min}) hasta el cual se debe tratar de realizar la agrupación. Para comenzar el proceso, deben definirse los centroides de los $N_{\text{máx}}$ grupos de partida. En la práctica por lo general, no se conocen, y el problema puede resolverse de la forma siguiente (Cuadras, 2014).

- a) Se realiza una pre-ubicación de los centroides para el número máximo de grupos, generando sus coordenadas aleatoriamente en el recorrido correspondiente a cada atributo de la muestra.
- b) Se realiza una pre-asignación de todos los elementos i de la muestra en aquellos k grupos para los cuales se obtiene el menor valor de la distancia al cuadrado punto-centroide.
- c) Cuando los n elementos de la muestra se han pre-asignado a los distintos $N_{\text{máx}}$ grupos, se reevalúan las coordenadas de los centroides.

A partir de este punto comienza propiamente el proceso de agrupación euclidiano según el siguiente algoritmo:

- 1) Se calculan los cuadrados de las distancias entre cada punto i y cada uno de los nuevos centroides, moviendo el punto hacia aquel grupo para el cual se obtiene el mínimo valor del cuadrado de la distancia punto – centroide, al mismo tiempo se modifican las coordenadas de cada centroide para tomar en cuenta la reubicación del punto. Este procedimiento se repite hasta que la composición de los $N_{\text{máx}}$ grupos permanezca constante. Cuando se ha alcanzado esta situación se dispone de:
 - Las coordenadas de cada centroide.
 - La distribución de los elementos para cada grupo.
 - La distancia de cada elemento al centroide de su grupo.

- La desviación cuadrática media puntos – centroides.
- 2) El número de grupos N se reduce en uno, a menos que $N=N_{\min}$. Esto se realiza fusionando el par de grupos para los cuales la distancia entre sus centroides sea mínima. Posteriormente se calculan las coordenadas de los centroides del grupo recién formado.
- 3) Una vez realizadas las agrupaciones para $N_{\min} \leq N \leq N_{\max}$, es posible realizar una prueba de hipótesis de Fisher secuencialmente, para definir los modelos de agrupaciones que producen mejoras significativas, en términos estadísticos, a los ajustes puntos – centroides.

Es necesario aclarar, que se debe ser muy cauteloso al seleccionar los modelos de clasificación, pues no siempre los mejores modelos de agrupaciones en términos estadísticos son los que poseen un sentido geológico concreto. Ello implica, que todos los modelos resultantes deben ser analizados y correlacionados con datos geológicos disponibles. Para separar y caracterizar las litologías presentes en los acuíferos de la Cuenca Sur de las provincias Artemisa y Mayabequey de la Cuenca M-IV-2 de la provincia Matanzas, se aplicó el análisis euclidiano de agrupación (AEA) a los datos obtenidos en cada cala a partir de los registros geofísicos de pozo y fue seleccionado el modelo de clasificación que mejor se ajusta a las características geológicas del área de estudio. Para la Cuenca M-IV-2 también se aplicó el Análisis de Componentes Principales y el Análisis de Factores, con el objetivo de investigar la estructura de correlación del sistema de atributos originales y encontrar factores controladores de la variabilidad total del sistema. A continuación se presentan los resultados obtenidos.

Resultados y discusión

La Tabla 1 presenta los centroides del modelo de clasificación que considera tres grupos para describir las litologías presentes en la Cuenca Sur de las provincias Artemisa y Mayabeque.

Tabla 1: Resultados de aplicar el AEA a los registros geofísicos de pozo para la evaluación litológica de la Cuenca Sur de las provincias Artemisa y Mayabeque.

Grupo	N	Ig (μ_r/h)	F
1	350	2,2	9,2
2	63	4,1	70,0
3	198	6,5	10,4

Los Grupos 1 y 3 presentan valores muy semejantes del factor de formación (F) por lo que deben presentar valores muy semejantes de porosidad, Sin embargo, estos grupos difieren en el comportamiento de la radiactividad gamma natural, El análisis de los centroides de cada grupo y las limitadas descripciones de núcleos permite concluir que el Grupo 1 está integrado por rocas carbonatadas carsificadas no arcillosas, mientras que al Grupo 3 se asocian las rocas carbonatadas carsificadas más arcillosas o dolomitizadas,

El Grupo 2 presenta valores más moderados de intensidad gamma natural y los mayores valores del factor de formación, todo lo cual refleja que está integrado por rocas carbonatadas no cavernosas que pueden presentar cierto grado de arcillosidad o dolomitización.

La Figura 3 permite comparar las descripciones de núcleos disponibles en la sala I-II-12K y la clasificación litológica propuesta como resultado del procesamiento realizado. Se puede verificar la efectividad de las técnicas de reconocimiento de patrones empleadas.

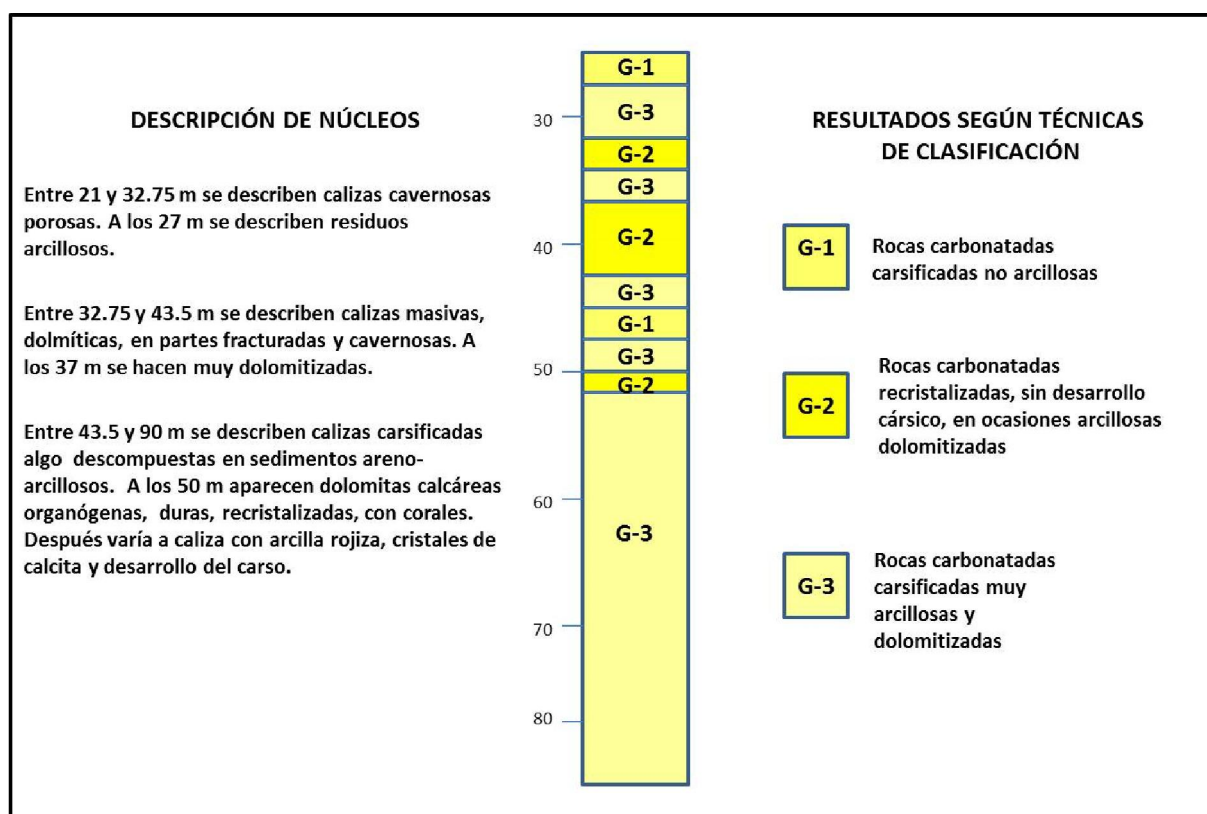


Figura 3. Clasificación litológica propuesta como resultado de aplicar el Análisis Euclidiano de Agrupación y descripciones de núcleos disponibles en la sala I-II-12K de la Cuenca Sur de las provincias Artemisa y Mayabeque.

La Tabla 2 muestra los centroides del modelo de clasificación propuesto para describir las litologías presentes en la Cuenca M-IV-2, obtenido al aplicar el AEA sobre las variables intensidad gamma natural, factor de formación e intensidad neutrónica.

Tabla 2: Resultados de aplicar el AEA a los registros geofísicos de pozo para la evaluación litológica de la Cuenca M-IV-2 de la provincia Matanzas.

Grupo	N	Ig ($\mu\text{r/h}$)	F	Inn(c/seg)
1	73	2,7	6,3	455
2	21	5,3	32,4	889
3	44	8,5	7,1	397

Los Grupos 1 y 3 difieren básicamente por los niveles de radiactividad gamma natural. Ambos presentan bajos valores del factor de formación y de la intensidad neutrónica indicando alta porosidad de estas capas. El Grupo 3 se caracteriza por valores más altos de la radiactividad gamma natural lo cual es reflejo de una mayor arcillosidad de estas rocas.

El Grupo 2 se caracteriza por los mayores valores del factor de formación y de la intensidad neutrónica indicando baja porosidad, mientras que la intensidad gamma natural presenta valores medios, reflejo de la presencia de rocas con cierta arcillosidad.

A partir de las características de estos grupos y analizando su asociación a las descripciones de núcleos queda establecido que el Grupo 1 está compuesto por rocas carbonatadas cavernosas no arcillosas, el Grupo 2 por rocas carbonatadas con escaso o ningún desarrollo cársico, duras, compactas, y a veces arcillosas o dolomitizadas, y el Grupo 3 está integrado por rocas carbonatadas muy cavernosas y arcillosas. La Figura 4 permite comprobar la correspondencia entre esta clasificación litológica y las descripciones de núcleos disponibles en un pozo de este acuífero.

En general ocurre que la valiosa información que brindan las descripciones de núcleos está disponible solo de forma muy limitada, debido a que la perforación de pozos con recuperación de núcleos eleva de manera significativa los costos de las investigaciones. La aplicación de las técnicas de reconocimiento de patrones sobre parámetros geofísicos ha demostrado su efectividad para separar e identificar las litologías presentes en un área de estudio con el apoyo de un número mínimo de descripciones de núcleos.

En los pozos de la Cuenca M-IV-2 también fue aplicado el Análisis de Componentes Principales obteniéndose que dos componentes principales son suficientes para explicar el 88% de la variabilidad total del sistema. La Tabla 3 resume este resultado.

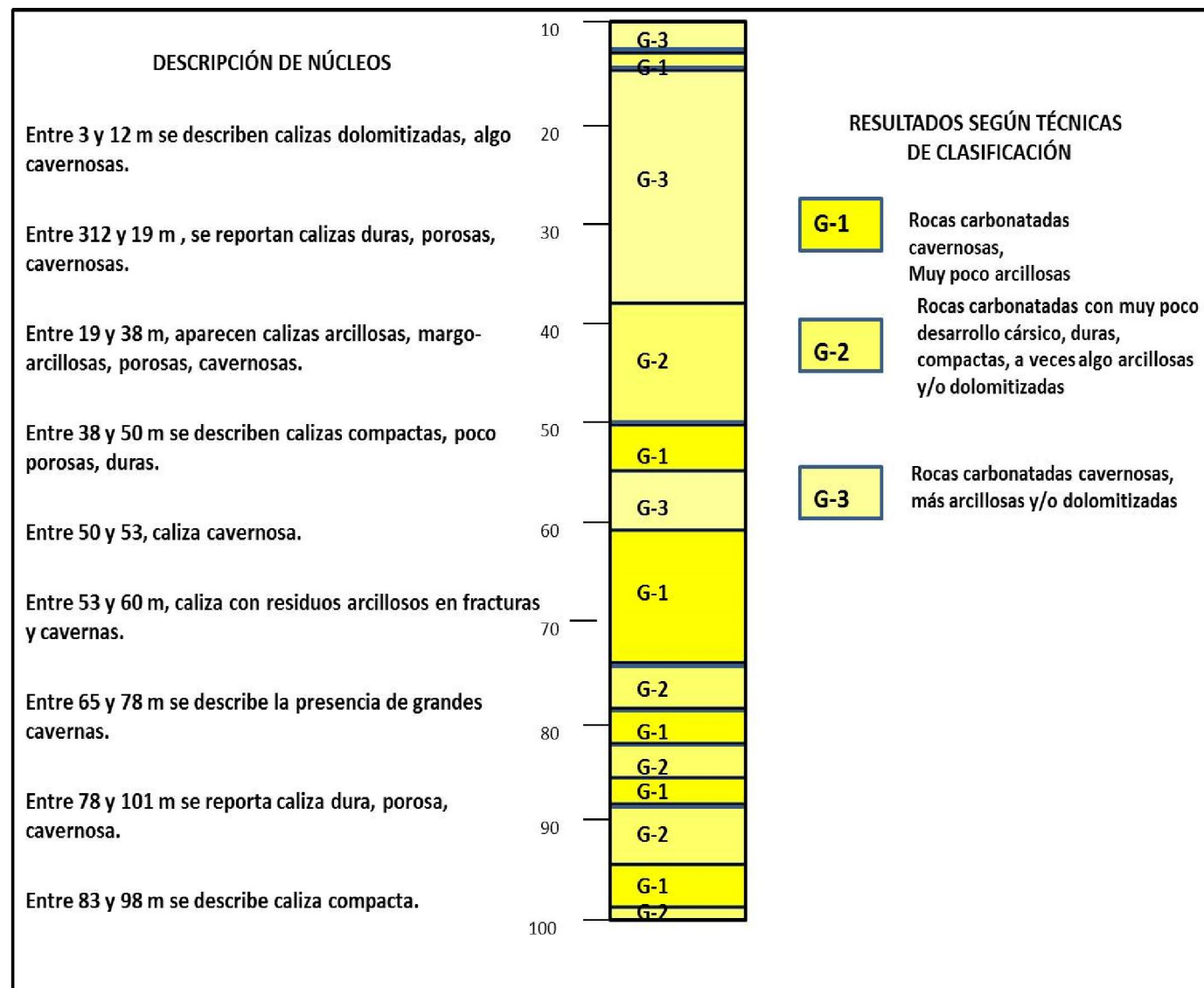


Figura 4. Clasificación litológica propuesta como resultado de aplicar el Análisis Euclidiano de Agrupación y descripciones de núcleos disponibles en la cala N-2 de la Cuenca M-IV-2 de la provincia Matanzas.

La proyección en el plano de las dos primeras Componentes Principales reflejó la separación perfecta de los grupos litológicos definidos demostrando el poder resolutivo de los atributos empleados.

Se aplicó también el Análisis de Factores siguiendo el criterio práctico de retener aquellos factores con varianza mayor o igual a 0,7. Los dos primeros factores presentaron varianza de 1,64 y 1,0 respectivamente, mientras que el tercer factor presentó varianza de solo 0,33.

Tabla 3: Variabilidad que explica cada Componente Principal en la evaluación litológica de la Cuenca M-IV-2 de la provincia Matanzas.

Componente Principal	% que explica	% acumulativo
1	55	55
2	33	88
3	11	100

Después de aplicar rotación oblicua a los ejes de las variables reguladoras, se obtuvo la matriz factorial de estructura simple que se presenta en la Tabla 4.

Tabla 4: Matriz factorial de estructura simple obtenida en la evaluación litológica de la Cuenca M-IV-2 de la provincia Matanzas.

Variables	Factores	
	1	2
Ig	-0,01	0,99
F	0,91	0,07
Inn	0,91	-0,09

Es evidente que el primer factor controla el comportamiento del factor de formación y de la intensidad neutrónica, de forma tal que incrementos de estas variables (que responden a disminución de la porosidad de la roca) provocan incrementos del primer factor. El segundo factor se asocia a las variaciones de la intensidad gamma natural.

Es decir, la dimensión del problema puede reducirse al estudio de dos factores, uno asociado de forma inversamente proporcional al desarrollo cársico, y otro directamente proporcional a la arcillosidad de las rocas.

Conclusiones

1. La aplicación combinada del Análisis Euclidiano de Agrupación sobre los parámetros geofísicos de pozo, con el apoyo de datos de núcleos, permitió definir las litologías presentes en cada pozo.
2. En los acuíferos estudiados fue posible identificar tres grupos litológicos básicos:
 - rocas carbonatadas no cavernosas, duras, recrystalizadas, en ocasiones arcillosas o dolomitizadas.
 - rocas carbonatadas carsificadas muy poco arcillosas.
 - Rocas carbonatadas carsificadas muy arcillosas o dolomitizadas.
3. La proyección en el plano de las dos primeras Componentes Principales demostró el poder clasificatorio del sistema de atributos empleados para separar los grupos litológicos fundamentales.
4. El Análisis de Factores permitió reducir la dimensión del problema estudiado en la Cuenca M-IV-2 al definirse que dos factores son suficientes para explicar el 88% de la variabilidad total del sistema, un primer factor controlado por el factor de formación y la intensidad neutrónica, que se asocia al desarrollo cársico de las rocas, y un segundo factor controlado por la intensidad gamma que varía con la arcillosidad del corte geológico.
5. La metodología aplicada demuestra ser efectiva para caracterizar los principales grupos litológicos del área de estudio con el apoyo de un número mínimo de descripciones de núcleos, lo que contribuye a disminuir los costos de las investigaciones geológicas.

Referencias

- CUADRAS, Carles M. “Nuevos métodos de análisis multivariante”. CMC Editions, Barcelona, España, 2014, 288 pp. ISBN 080-23.
- DÍAZ MONROY, Luis Guillermo; MORALES RIVERA, Mario Alfonso. “Estadística multivariada: inferencia y métodos”. Bogotá Colombia: Editorial Universidad Nacional de Colombia, 2012, 637 pp. ISBN 978-958-701-195-1.
- Kirsch, R., 2006: “Groundwater Geophysics A Tool for Hydrogeology”. ISBN 13 978-3-540-29383-5, Springer Berlín Heidelberg, New York.
- VALCARCE, R. M.; (1998). Evaluación de las propiedades hidrogeológicas de acuíferos cársicos empleando un complejo mínimo de registros geofísicos de pozo. Tesis en opción al Título de Doctor en Ciencias. Departamento de Geofísica, ISPJAE, Ciudad de La Habana. 130 p.