



Aplicación de inteligencia artificial para el análisis de facies sedimentarias

Por **Isabel Eugenia Giannoni**, **Mariano Germán Arregui**, **Augusto Nicolas Varela** (Y-Tec y CONICET) y **Javier Vásquez** (YPF).

Este trabajo fue seleccionado en las 3º Jornadas de Revolución Digital para Petróleo y Gas.



El uso de inteligencia artificial y machine learning permite clasificar facies sedimentarias con mayor precisión, reduciendo la subjetividad del análisis. Esta herramienta acelera el estudio de testigos de roca y mejora la comprensión de los ambientes sedimentarios.

Las facies sedimentarias, un concepto fundamental en geología, refiere a las características físicas, químicas y biológicas de una unidad de roca sedimentaria que se formó por un proceso sedimentario específico. Estas características incluyen la textura, composición y estructuras sedimentarias (mecánicas y biogénicas) presentes en la roca (Walker, 2006). La definición y estudio de facies sedimentarias y sus asociaciones resulta especialmente útil en la exploración y explotación de hidrocarburos y otros recursos, donde la comprensión del paleoambiente sedimentario es fundamental para la toma de decisiones, como fue planteado por Miall (1977, 1988).

La Inteligencia Artificial (IA) se ha convertido en la base de numerosas aplicaciones actuales que van desde

el diagnóstico médico personalizado hasta los vehículos autónomos y la personalización de contenidos en línea; transformando así la manera en que interactuamos con la tecnología y abordamos desafíos en diversos campos (Juca-Maldonado, 2024). Dentro de la geología, en campos como la sedimentología y la petrografía, ha permitido realizar análisis más rápidos, precisos y detallados de las rocas (Eftekhari et al., 2024; Gernay et al., 2023; Gonçalves et al., 2017; Holden et al., 2023; Ippolito et al., 2021; Mancini et al., 2020).

El objetivo del presente trabajo es el uso de la IA, específicamente de Machine Learning (ML), a imágenes de alta y baja resolución de muestras de testigos coronas, datos geoquímicos y perfiles de resistencia a la rotura

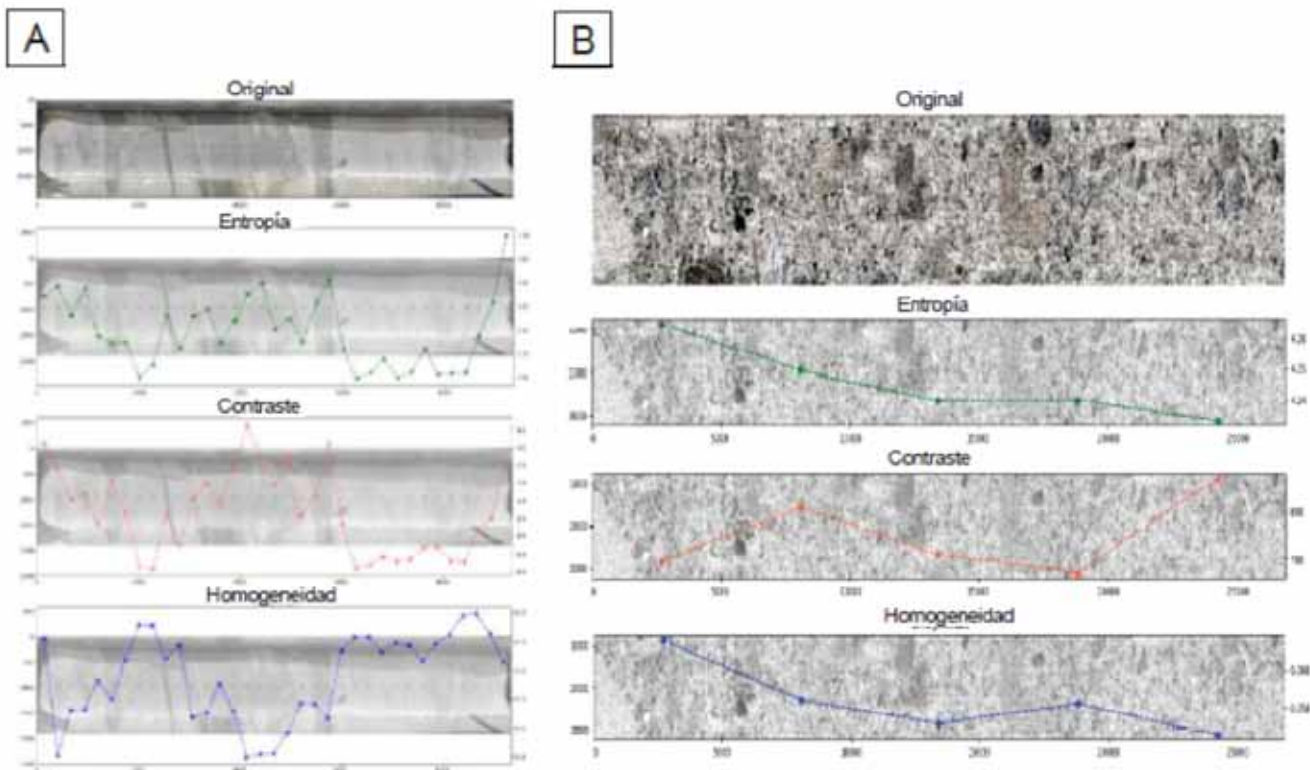


Figura 1. (A) Extracción de texturas de la imagen a partir de imágenes de baja resolución en profundidad. (B) Extracción de texturas de la imagen a partir de imágenes de alta resolución en profundidad.

(scratch test machine) con el fin de clasificar, inferir y predecir patrones de comportamiento asociados a las facies con sus correspondientes procesos y ambientes sedimentarios.

Tradicionalmente, los sedimentólogos han empleado métodos manuales para clasificar y describir los diferentes tipos de sedimentos/sedimentitas presentes en muestras, el cual es un proceso laborioso y muchas veces subjetivo. Sin embargo, mediante el uso de estas técnicas de procesamiento de imágenes comprendidos dentro del subcampo del ML supervisado y no supervisado, los algoritmos pueden analizar imágenes de los testigos de rocas y clasificarlos automáticamente en categorías predefinidas. Esto reduce la subjetividad del observador, mejorando la precisión de las clasificaciones, además de disminuir los tiempos de descripción; permitiendo comparar de manera directa las descripciones de diferentes observadores.

El proceso de análisis consistió en varias etapas, en las que se buscó asegurarse de describir los diferentes tipos de sedimentos/sedimentitas presentes en muestras, el cual es un proceso laborioso y muchas veces subjetivo. Sin embargo, mediante el uso de estas técnicas de procesamiento de imágenes comprendidos dentro del subcampo del ML supervisado y no supervisado, los algoritmos pueden analizar imágenes de los testigos de rocas y cla-

sificarlos automáticamente en categorías predefinidas. Esto reduce la subjetividad del observador, mejorando la precisión de las clasificaciones, además de disminuir los tiempos de descripción; permitiendo comparar de manera directa las descripciones de diferentes observadores. la calidad de la información. Inicialmente se describieron digitalmente y reconstruyeron las 42 diferentes facies sedimentarias de 84 metros de testigos de rocas de un set de datos sintéticos cedido por la empresa EPSLOG S.A en el marco del convenio de colaboración firmado oportunamente con Y-TEC S.A. El set sintético de datos fue recopilado de manera tal que las rocas seleccionadas respondan a los diversos procesos y paleoambientes sedimentarios que se dan en la naturaleza. Esto se logró combinando diferentes metros de diferentes pozos sobre los cuales la compañía EPSLOG S.A realizó servicios.

Para la nomenclatura de dichas facies sedimentarias, se utilizó una denominación basada en Miall (1977, 1988) para rocas clásticas, mientras que para las rocas carbonáticas se utilizó Dunham (1962), ampliada por Embry y Klován (1971) y modificada luego por Lokier y Al Junaibi (2016).

Luego, se realizó el análisis de las imágenes de testigos coronas, tanto de alta como de baja resolución. Para pasar de imágenes RGB, cuya dimensión posee 3 capas a una matriz plana unidimensional se decidió realizar una

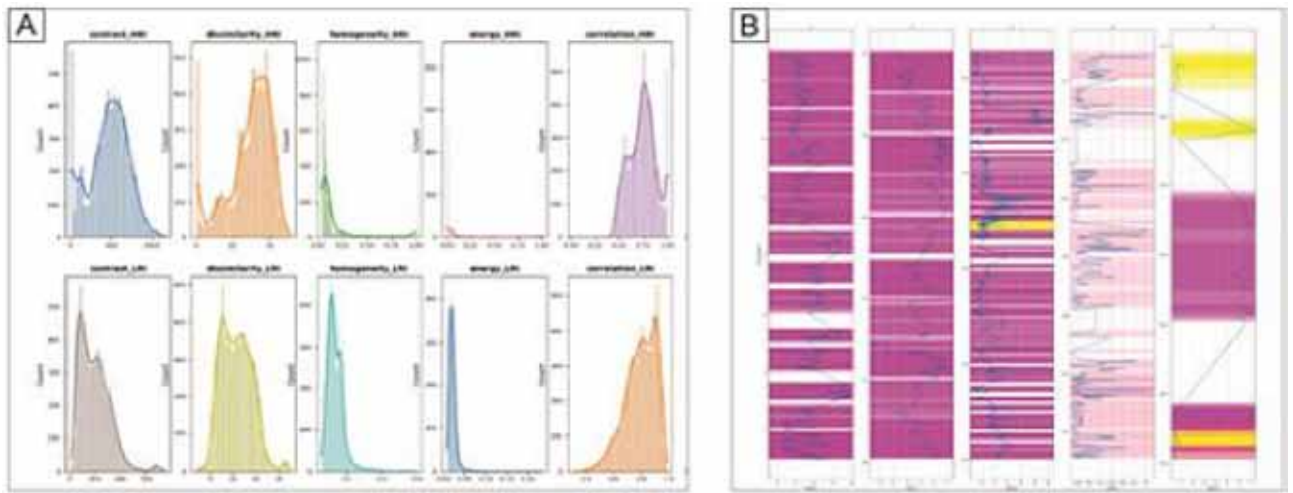


Figura 2. (A) Análisis univariado de la distribución de cada variable; (B) Análisis univariado de cada elemento en profundidad y composición.

segmentación centímetro a centímetro y la obtención de un grupo de propiedades (contraste, homogeneidad, segundo momento angular o ASM, correlación, disimilitud, heterogeneidad, energía, entropía), que surgen del análisis textural de la imagen en función de la profundidad (Fig.1 A y B). Para que los análisis de frecuencia de la matriz fueran representativos, se decidió observar la co-ocurrencia de los valores de los píxeles en todas las direcciones posibles (E-W, NE-SW, N-S, SE-NW) y a diferentes distancias 5, 10, 15 y 20 píxeles (Marceau et al., 1990; Marrón, 2012; Presutti, 2004).

Lo que respecta a los datos químicos, al ser obtenidos mediante fluorescencia de rayos X, fueron sometidos a la transformación necesaria para su utilización eliminando la restricción de la suma constante (Alperin, 2013; Martín-Fernández et al., 2011; Rollinson, 1993). Una vez completada la transformación necesaria para algunas de las variables, se procedió a realizar el análisis estadístico exploratorio del conjunto de datos en su totalidad teniendo en cuenta un enfoque uni, bi y multivariado (Fig.2). Se realizó una evaluación de la distribución de cada variable en profundidad para identificar zonas de interés (Fig.2 A y B), así como la correlación entre pares de elementos para identificar posibles relaciones entre ellos. Una vez obtenidas, homogeneizadas y estandarizadas todas las variables, se procedió a realizar un segundo Análisis de Componentes Principales para realizar la reducción de variables. Como resultado de este procesamiento de datos se obtuvo finalmente un data set compuesto por 17 variables a modelar y 1 variable Target (facies sedimentarias).

En la etapa de modelado, se evaluó la performance predictiva y de clasificación de diferentes modelos de ML supervisados: Extreme Gradient Boosting Machine Classification (XGBM), Logistic Regression (LG) y Random Forest Classification (RF). Como resultado, se obtuvo que el mejor predictor era XGBM con un accuracy de 0.77, a su vez se pudo observar que el segundo momento

angular de las imágenes de baja resolución, en conjunto con el Bario y el Calcio son las variables que más aportan a la construcción de las predicciones (Fig. 3).

Sin embargo, al realizar una validación a posteriori (Fig.4) se observó que, si bien las métricas del modelo eran buenas, al estar frente a muestras nunca vistas la performance no era la esperada. De un total de 120 predicciones, solo el 25 % fueron acertadas (muestras verdes en Fig. 4), siendo el otro 75% desaciertos con distintos niveles de errores. Cabe destacar que de todas las predicciones solo se ha obtenido 1 error en lo que respecta a la composición, siendo el resto de los errores anteriormente mencionados de granulometría y estructura sedimentaria. En este sentido, los errores fueron clasificados en colores dependiendo el criterio geológico: (1) los errores naranjas representan el 2.11 % y son los que poseen error de composición, textura y granulometría. Este error es considerado una predicción totalmente errónea; (2) los errores azules representan el 39.11% del total, corresponden a muestras bien clasificadas composicionalmente pero que poseen discrepancias entre las granulometrías y las estructuras sedimentarias predichas y las

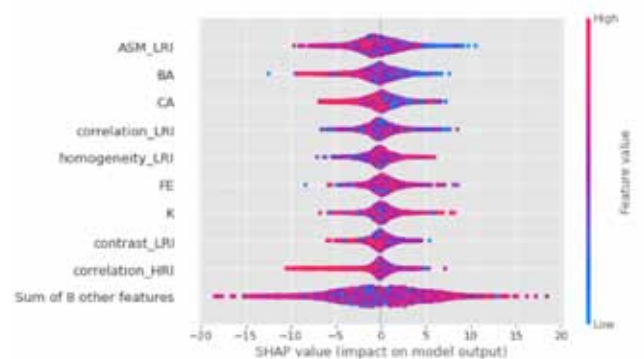


Figura 3. Diagram tipo Shap donde se muestran los aportes de las distintas variables a las predicciones del modelo.

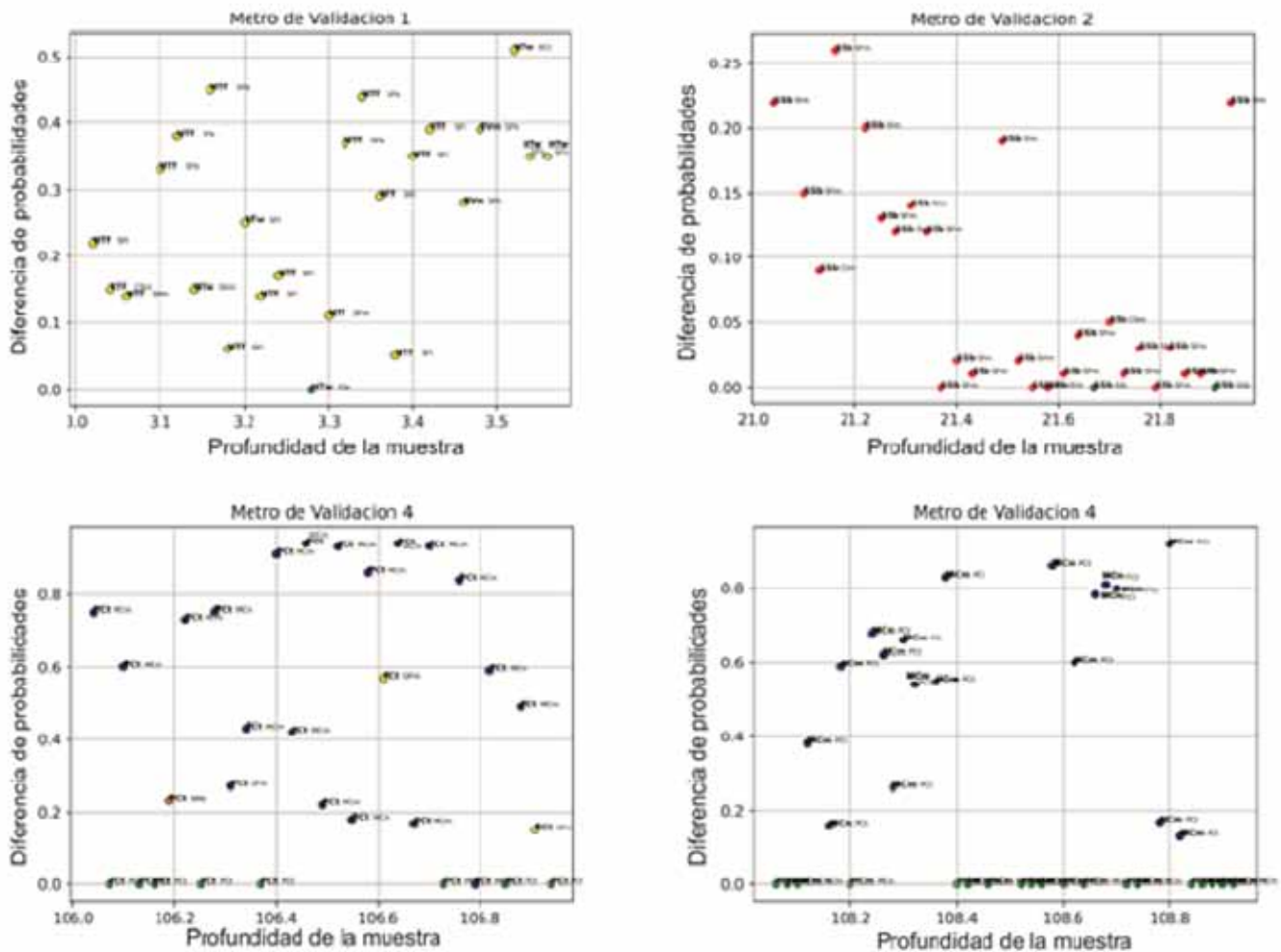


Figura 4: Predicciones del modelo para los 4 metros utilizados. En distintos colores se encuentran los distintos tipos de errores.

reales; (3) los errores amarillos corresponden al 26.11% y se consideran errores moderados, las muestras eran en su mayoría heterolíticas y si bien el modelo erra en su predicción, esta se encuentra dentro de los rangos granulométricos de la heterolítica; y por último (4) los errores rojos, que conforman el 27.11 % restante, corresponden a los considerados no graves ya que en su mayoría son limos clasificados como arenas finas y donde la estructura sedimentaria real era masiva por bioturbación, pero el modelo las considero masivas.

Podemos concluir entonces que: (1) cuando se realizan modelos de predicción de facies y procesos sedimentarios es indispensable contar con una apropiada entrada de los datos para alimentar los modelos. Para conseguir esto es necesario unificar los criterios sedimentológicos, por lo que se confeccionó un código de facies coherente, constante, homogéneo que elimine las posibles subjetividades presentes en las descripciones de distintos observadores. Tal código de facies sigue las nor-

mativas sugeridas por (Dunham, 1962; Embry & Klovan, 1971; Lokier & Al Junaibi, 2016; Miall, 1977, 1988) pero modificado ligeramente para su análisis matemático (2) el modelo entrenado tiene una buena performance para predecir las distintas composiciones a nivel centimétrico de las muestras dadas, pero posee una deficiencia moderada e importante a la hora de predecir granulometrías y estructuras sedimentarias respectivamente. Esto podría deberse que la escala de entrenamiento del modelo es demasiado pequeña para la naturaleza de las escalas de las características geológicas a predecir. Cabe destacar que muchos de estos errores a la hora de predecir facies y agrupamientos de facies muchas veces están presentes también en trabajos realizados por profesionales, sobre todo los errores clasificados como amarillos y rojos, que son los que más se prestan a confusión. Por otro lado, no debe descartarse la posibilidad que los distintos orígenes de los datos (perfiles en el caso de la composición química e imágenes para la granulometría y la estructura

sedimentaria), puedan ser factores preponderantes en la baja performance del modelo.

Agradecimientos

El presente trabajo se realizó en el marco del proyecto DISCO2STORE, quien recibe financiamiento del programa de investigación e innovación Horizont 2020 de la Unión Europea en el marco del Marie Skłodowska-Curie grant agreement No 101007851. Los autores agradecen a EPSLog S.A por el aporte de las muestras. Así también a los tecnólogos y analistas de laboratorio de Y-TEC S.A por el análisis de los mismos.

Bibliografía

- Alperin, M. (2013). Introducción al análisis estadístico de datos geológicos. Editorial de la Universidad Nacional de La Plata (EDULP).
- Eftekhari, S. H., Memariani, M., Maleki, Z., Aleali, M., Kianoush, P., Shirazy, A., Shirazi, A., & Pour, A. B. (2024). Employing Statistical Algorithms and Clustering Techniques to Assess Lithological Facies for Identifying Optimal Reservoir Rocks: A Case Study of the Mansouri Oilfields, SW Iran. *Minerals*, 14(3), 233. <https://doi.org/10.3390/min14030233>
- Embry, A. F., & Klovan, J. E. (1971). A late Devonian reef tract on northeastern Banks Island, NWT. *Bulletin of Canadian petroleum geology*, 19(4), 730-781.
- Germay, C., Lhomme, T., & Perneder, L. (2023). High-resolution core data and machine learning schemes applied to rock facies classification. *Geological Society, London, Special Publications*, 527(1), 121-135. <https://doi.org/10.1144/SP527-2021-193>
- Gonçalves, Í. G., Kumaira, S., & Guadagnin, F. (2017). A machine learning approach to the potential-field method for implicit modeling of geological structures. *Computers & Geosciences*, 103, 173-182. <https://doi.org/10.1016/j.cageo.2017.03.015>
- Holden, T., Kurian, R., Ibrahim, M., Hampson, D., & Downton, J. (2023). Predicting Facies, Rock, and Geomechanical Properties Using Convolutional Neural Networks: A Case Study From an Unconventional Shale Reservoir. *Proceedings of the 11th Unconventional Resources Technology Conference*. Unconventional Resources Technology Conference, Colorado Convention Center, Denver, Colorado, US. <https://doi.org/10.15530/urtec-2023-3862247>
- Ippolito, M., Ferguson, J., & Jenson, F. (2021). Improving facies prediction by combining supervised and unsupervised learning methods. *Journal of Petroleum Science and Engineering*, 200, 108300. <https://doi.org/10.1016/j.petrol.2020.108300>
- Juca-Maldonado, F. (2024). El impacto de la inteligencia artificial en los trabajos académicos y de investigación. *Revista Metropolitana de Ciencias Aplicadas*, 6(Suplemento 1), 289-296. <https://doi.org/10.62452/8nww1k83>
- Lokier, S. W., & Al Junaibi, M. (2016). The petrographic description of carbonate facies: Are we all speaking the same language? *Sedimentology*, 63(7), 1843-1885. <https://doi.org/10.1111/sed.12293>
- Mancini, M., Weindorf, D. C., Monteiro, M. E. C., De Faria, Á. J. G., Dos Santos Teixeira, A. F., De Lima, W., De Lima, F. R. D., Dijair, T. S. B., Marques, F. D., Ribeiro, D., Silva, S. H. G., Chakraborty, S., & Curi, N. (2020). From sensor data to Munsell color system: Machine learning algorithm applied to tropical soil color classification via NixTM Pro sensor. *Geoderma*, 375, 114471. <https://doi.org/10.1016/j.geoderma.2020.114471>
- Marceau, D. J., Howarth, P. J., Dubois, J. M., & Gratton, D. J. (1990). Evaluation Of The Grey-level Co-occurrence Matrix Method For Land-cover Classification Using Spot Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4), 513-519. <https://doi.org/10.1109/TGRS.1990.572937>
- Marrón, B. S. (2012). Co-occurrence Matrix and fractal dimension for image segmentation. *Revista de Matemática: Teoría y Aplicaciones*, 19(1), 49-63. <https://doi.org/10.15517/rmta.v19i1.2104>
- Martín-Fernández, J. A., Palarea-Albaladejo, J., & Olea, R. A. (2011). Dealing with Zeros. En V. Pawlowsky-Glahn & A. Buccianti (Eds.), *Compositional Data Analysis* (1.a ed., pp. 43-58). Wiley. <https://doi.org/10.1002/9781119976462.ch4>
- Miall, A. D. (1977). Lithofacies types and vertical profile models in braided river deposits: A summary. En *Fluvial Sedimentology* (pp. 579-604). Canadian Society of Petroleum Geologists Bulletin.
- Miall, A. D. (1988). Facies Architecture in Clastic Sedimentary Basins. En K. L. Kleinspehn & C. Paola (Eds.), *New Perspectives in Basin Analysis* (pp. 67-81). Springer New York. https://doi.org/10.1007/978-1-4612-3788-4_4
- Presutti, M. (2004, agosto 11). La Matriz de Co-ocurrencia en la clasificación multiespectral: Tutorial para la enseñanza de medidas texturales en cursos de grado universitario. 4a Jornada de Educação em Sensoriamento Remoto no Âmbito do Mercosul.
- Rollinson, H. R. (1993). Using geochemical data: Evaluation, presentation, interpretation. Longman Scientific & Technical ; Copublished in the U.S. with J. Wiley & Sons.
- Walker, R. G. & Geological Association of Canada (Eds.). (2006). *Facies models: Response to sea level change* (8. print). Geological Association of Canada.