

ANÁLISIS VISUAL GUIADO PARA EL ESTUDIO DE DATOS GEOQUÍMICOS

Antonini, Antonella S.^{1,2,3}, **Luque, Leandro E.**^{1,2,3}, **Tanzola, Juan E.**^{4,5}, **Asiaín, Lucía**^{4,5},
Ferracutti, Gabriela R.^{4,5}, **Bjerg, Ernesto A.**^{4,5}, **Castro, Silvia M.**^{1,2,3} y **Ganuzza, M. Luján**^{1,2,3}

¹Laboratorio de I+D en Visualización y Computación Gráfica (VyGLab) (UNS-CIC)

²Dpto. de Cs. e Ing. de la Computación, Universidad Nacional del Sur (DCIC-UNS)

{antonella.antonini, leandro.luque, smc, mlg}@cs.uns.edu.ar

³ICIC, Instituto de Ciencias e Ingeniería de la Computación (UNS-CONICET)

⁴INGEOSUR, Instituto de Geología (UNS-CONICET)

⁵Dpto. de Geología, Universidad Nacional del Sur (DG-UNS)

gferrac@uns.edu.ar, {jetanzola,lasiain,ebjerg}@ingeosur-conicet.gob.ar

RESUMEN

Los datos geoquímicos, multidimensionales por naturaleza, son esenciales para comprender diversos procesos planetarios y medioambientales, entre los que se destacan el origen y la evolución de la Tierra y del Sistema Solar. Por otro lado, los avances tecnológicos en el instrumental analítico, han aumentado espectacularmente el volumen y la variedad de datos geoquímicos disponibles. La exploración de estos conjuntos de datos multidimensionales de naturaleza tan compleja exige técnicas de análisis visual (AV) adecuadas, un alto grado de interactividad y metáforas intuitivas. Dado que existe una correlación natural entre la complejidad de los datos y la complejidad de las herramientas para estudiarlos, los objetivos analíticos son más difíciles de alcanzar. Por lo tanto, tiene sentido considerar métodos que guíen o ayuden a los usuarios en el proceso de análisis visual. En este contexto, nuestro objetivo general es contribuir al diseño y desarrollo de técnicas de analítica visual que contribuyan efectivamente al análisis de datos geoquímicos tanto de rocas como de minerales.

Palabras claves: *Análisis Visual de Datos, Análisis Visual de Datos Geoquímicos,*

Visualización de Datos Multidimensionales, Explicabilidad, Guías de Exploración de Datos

CONTEXTO

Este trabajo se realiza en estrecha colaboración con investigadores del INGEOSUR-CONICET (Instituto Geológico del Sur), del Departamento de Geología de la Universidad Nacional del Sur, y el VyGLab (Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica) del Departamento de Ciencias e Ingeniería de la Computación (DCIC-UNS), instituciones de investigación de reconocido prestigio tanto nacional como internacional.

1. INTRODUCCIÓN

La capacidad actual de generación y recopilación de datos supera con creces la de su procesamiento de una forma adecuada y eficaz y más aún la de una asimilación inteligente de los mismos. En las últimas décadas, la capacidad de aprovechar esos datos no ha aumentado al mismo ritmo que la disponibilidad de los mismos, siendo ésta una constante en la mayoría de las disciplinas. En particular, se puede afirmar que en la actualidad las Geociencias han entrado en la era de los grandes conjuntos de datos.

En este campo, Reichstein *et al.* 2019 [Reichstein et al. 2019] plantean que, para sacar el máximo partido de este crecimiento explosivo de los datos del sistema terrestre y de su diversidad, en los próximos años debemos enfrentarnos a dos grandes tareas: extraer conocimiento de esta avalancha de datos y derivar modelos que aprendan mucho más de los datos de lo que pueden hacerlo los enfoques tradicionales para extraer información, colaborando con el usuario en su exploración y respetando al mismo tiempo la comprensión evolutiva de las leyes de la naturaleza por parte de éste.

En el área de la geoquímica aplicada, la existencia de grandes conjuntos de datos proporciona oportunidades y desafíos en el procesamiento de datos de estudios geoquímicos [Zuo et al. 2021] tanto de rocas como de minerales. Los datos geoquímicos son muy utilizados en la exploración minera, ya que desempeñan un rol destacado tanto en el descubrimiento de nuevos yacimientos minerales como en el estudio de los ya conocidos.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Nuestras líneas de investigación investigación y desarrollo se basan en la utilización de uno de los métodos más ampliamente utilizados hoy en día para el preprocesamiento de datos geoquímicos de exploración, esto es, el análisis de datos composicionales.

Esto permite procesar los datos de exploración geoquímica con el fin, entre otros, de revelar la asociación espacial de patrones geoquímicos, identificar asociaciones elementales e identificar anomalías geoquímicas asociadas a los procesos mineralizantes.

Estas líneas específicas corresponden al diseño y desarrollo de métodos y herramientas

de analítica visual integrando técnicas de análisis visual y de machine learning aplicados a:

Clasificación y diferenciación automática de rocas máficas y ultramáficas y visualización de los métodos aplicados a las tomas de decisiones durante este proceso.

Según Ferracutti et al. [Ferracutti et al. 2017], afloramientos del cuerpo intrusivo Virorco junto con cuerpos intrusivos adyacentes, presentan interés económico [Ferracutti et al. 2005, Ferracutti et al. 2013]. Éstos son cuerpos intrusivos con estratificación magmática primaria. Esta característica se manifiesta como una alternancia de niveles de rocas máficas y ultramáficas. En lo que respecta a la actividad exploratoria de una empresa minera, es fundamental poder discriminar entre rocas máficas y ultramáficas durante la descripción de testigos de perforación.

La clasificación de rocas máficas y ultramáficas, está basada en los porcentajes de determinados minerales que integran estas rocas [Le Maitre et al. 2002]. La identificación de los mismos se lleva a cabo a partir de un análisis petrográfico de muestras de roca, utilizando microscopios de luz transmitida.

En lo que respecta al ámbito de la exploración minera, este proceso plantea una serie de problemas. En primer lugar, tanto la confección de secciones delgadas como el análisis petrográfico, conllevan un determinado tiempo, el cual no es compatible con los tiempos que maneja una empresa minera. En segundo lugar, un análisis petrográfico se realiza sobre muestras de roca puntuales, lo cual resultaría inviable de realizar sobre la totalidad de las rocas de perforaciones que abarcan miles de metros.

En este contexto, resulta de interés el estudio y desarrollo de técnicas de analítica visual que integren modelos de Machine Learning y

de Visualización. El uso de técnicas de Machine Learning nos permitirá automatizar el proceso de análisis de grandes cantidades de datos, considerar y evaluar las proporciones de los distintos minerales acelerando así la clasificación de las rocas máficas y ultramáficas. Por otro lado, en este proceso de automatización se debe explicar a los usuarios, geólogos en este caso, cómo y por qué se tomaron las decisiones que llevaron a dicha clasificación.

Análisis Visual Guiado de datos geoquímicos de espinelos (grupo de minerales). Las técnicas de visualización nos proporcionan herramientas útiles para comprender y explorar datos geoquímicos en minerales. El análisis visual proporciona información valiosa acerca de la información subyacente en estos datos. Los datos geoquímicos son complejos y multidimensionales. Además, contienen valores atípicos, el contenido de los elementos varía en miles de ppm (partes por millón) y los distintos elementos químicos pueden estar correlacionados. Si bien existen diversas herramientas que permiten visualizar datos multidimensionales, estas no proveen técnicas que se ajusten adecuadamente a datos geoquímicos en general y a espinelos en particular. Por otro lado, si bien hay algunas aplicaciones que permiten el análisis visual de estos datos [Ganuzo et al. 2012, Ganuzo et al. 2014, Ganuzo et al. 2017], éstas se restringen a un máximo de 6 dimensiones. Esto motiva el desarrollo de métodos y herramientas adecuadas para analizar datos geoquímicos que sean poderosos. Un efecto adverso de las herramientas complicadas es que los objetivos analíticos son más difíciles de alcanzar; sin embargo, el análisis visual de datos se concibe como una colaboración entre el usuario y el sistema informático con el fin de completar una tarea determinada. Por lo tanto, tiene sentido considerar métodos que guíen o

ayuden a los usuarios en el proceso de análisis exploratorio de estos datos.

La búsqueda de una integración eficaz entre usuario y sistema (método/aplicación) en la que el sistema ayude activamente al usuario a alcanzar su objetivo de análisis, ha sido objeto de investigación durante muchos años [Ceneda et al. 2016, Ceneda et al. 2018, Escarza et al. 2011, May et al. 2012, Sacha et al. 2014, Streit et al. 2012]. Sin embargo, este problema sigue sin resolverse de forma sustancial. Como consecuencia, los usuarios pueden verse abrumados por sistemas de análisis visual poderosos pero complicados, lo que también limita su capacidad para producir resultados relevantes. Así, la guía es prometedora para lograr una colaboración eficaz usuario-sistema que favorezca el análisis visual.

3. RESULTADOS OBTENIDOS Y ESPERADOS

De las líneas de trabajo delineadas se han obtenido resultados parciales.

En lo que respecta a la clasificación automática de rocas máficas y ultramáficas y a la visualización de cómo se toman las decisiones durante este proceso, se han aplicado diversos modelos de aprendizaje automático supervisados y no supervisados para la clasificación de muestras de datos. Los resultados obtenidos de la clasificación automática fueron cotejados con los esperados por los geólogos, quienes afirmaron la utilidad y exactitud de estas nuevas técnicas.

Además, en la clasificación visual que realizan los geólogos se observan ciertos minerales en particular. Mediante el análisis de la importancia de ciertas características, pudimos determinar que al considerar adicionalmente otros minerales en la clasificación de las rocas los resultados podrían mejorarse significativamente.

Con respecto al Análisis Visual de datos geoquímicos de espinelos se ha desarrollado una herramienta de Visualización extensible, *SpinelWEB* [Antonini et al. 2021], para el Análisis Visual de datos geoquímicos. Si bien dicha herramienta está especialmente orientada a datos geoquímicos del grupo de minerales del espinelo, puede utilizarse para el análisis de datos geoquímicos en general. Actualmente se están desarrollando alternativas que permitan visualizar estos datos. Entre éstas, se está trabajando con diversas técnicas como *Paired y Non-Paired General Line Coordinates* [Luque et al. 2021, Antonini et al. 2022] y técnicas utilizadas en las visualización de clusters y conjuntos. Conjuntamente se está trabajando en guías de exploración con el objetivo de sacar ventaja de la sinergia entre los expertos del dominio y método/aplicación. Estas guías de orientación en la exploración, no se han explorado previamente en el contexto de aplicaciones geológicas.

4. FORMACION DE RECURSOS HUMANOS

A continuación, se detallan proyectos de investigación, tesis finalizadas y en desarrollo y becas obtenidas dedicadas a temáticas vinculadas con las líneas de investigación presentadas.

Proyectos Vigentes

-PIBAA - CONICET (2872021010 0824CO)
Análisis Visual de Datos Multidimensionales sin Pérdida de Información. Directora: Dra. M. Luján Ganuza.

-PGI 24//N048 *Análisis Visual de Datos.*
Directora: Silvia M. Castro.

-PICT-2017-1246 *Análisis Visual en Geociencias.* Directora: Silvia M. Castro.

Tesis en Desarrollo

-Análisis Visual de Datos Multidimensionales, tesis de Doctorado en Ciencias de la Computación. Alumna: Antonella S. Antonini. Directora: Dra. Silvia Castro. Codirectora: Dra. M. Luján Ganuza.

Becas

-Antonella S. Antonini. Título del plan propuesto: *Análisis Visual de Datos Multidimensionales.* Beca doctoral 2019 CONICET. Adjudicada a partir de abril de 2019 y por un término de 5 años. Directores: Dra. Silvia Castro - Dr. Ernesto Bjerg.

5. BIBLIOGRAFÍA

[Ceneda et al. 2016] Ceneda, D., Gschwandtner, T., May, T., Miksch, S., Schulz, H.J., Streit, M. and Tominski, C., 2016. *Characterizing guidance in visual analytics.* IEEE TVCG, 23(1), pp.111-120.

[Ceneda et al. 2018] Ceneda, D., Gschwandtner, T., May, T., Miksch, S., Streit, M. and Tominski, C., 2018, June. *Guidance or No Guidance? A Decision Tree Can Help.* In EuroVA@ EuroVis (pp. 19-23).

[Ceneda et al. 2019] Ceneda, D., Gschwandtner, T. and Miksch, S. 2019. *A review of guidance approaches in visual data analysis: A multifocal perspective.* In Computer Graphics Forum (Vol. 38, No. 3, pp. 861-879).

[Ferracutti et al. 2005] Ferracutti, G., Mogessie, A., and Bjerg, E. 2005. Chemical and mineralogical profile of the Las Águilas mafic-ultramafic drill core, San Luis Province, Argentina. *Mitteilungen der Osterreichischen Mineralogischen Gesellschaft* 151: 40 (2005).

[Ferracutti et al. 2013] Ferracutti, G., Bjerg, E. and Mogessie, A. 2013. Petrology, geochemistry and mineralization of the Las Águilas and Virorco mafic-ultramafic bodies, San Luis Province, Argentina. *International Journal of Earth Sciences*, 102(3), 701-720.

[Ferracutti et al. 2017] Ferracutti, G., Bjerg, E., Hauzenberger, C., Mogessie, A., Cacace, F. and Asiain, L. 2017. Meso to Neoproterozoic layered mafic-ultramafic rocks from the Virorco back-arc

intrusion, Argentina. *Journal of South American Earth Sciences*, 79, 489-506.

[Le Maitre et al. 2002] Le Maitre, R. W., Streckeisen, A., Zanettin, B., Le Bas, M. J., Bonin, B., Bateman, P., Bellieni, G., Dudek, A., Efremova, S., Keller, J., Lameyre, J., Sabine, P. A., Schmid, R., Sørensen, H. and Wooley, A.R. 2002. *Igneous Rocks: A Classification and Glossary of Terms*. Cambridge 774 University Press, 236 p., Cambridge.

[Antonini et al. 2021] Antonini, A.S., Ganuza, M.L., Ferracutti, G.R., Gargiulo, M.F., Matković, K., Gröller, E., Bjerg, E.A., Castro, S.M. (2021). *SpinelWeb: an interactive web application for visualizing the chemical composition of spinel group minerals*. *Earth Science Informatics*, 1-8.

[Antonini et al. 2022] Antonini, A.S., Luque, L., Ganuza, M. L., Castro, S.M. (2022). *Towards a Taxonomy for Non-Paired General Line Coordinates: A Comprehensive Survey*. *International Journal of Data Science and Analytics*.

[Escarza et al. 2011] Escarza, S., Larrea, M., Urribarri, D., Martig, S., Castro, S., (2011) *Integrating Semantics in the Visualization Process*, Book "Scientific Visualization: Interactions, Features, Metaphors", pp. 92-102, Series: Dagstuhl FOLLO-UPS Vol.2, Hans Hagen Ed, Publisher, SchlossDagstuhl--Leibniz-Zentrum fuer Informatik. Dagstuhl, Alemania. ISBN 978-3-939897-26-2.

[Ferracutti et al. 2015] Ferracutti, G.R., Gargiulo, M.F., Ganuza, M.L., Bjerg, E.A., Castro, S.M. (2015) *Determination of the spinel group end members based on electron microprobe analyses*. *Mineral. Petrol.* 109, 2, 153-160.

[Ganuza et al. 2012] Ganuza, M. L., Castro, S. M., Ferracutti, G.R., Bjerg, E. A., Martig, S. (2012) *Spinelviz: An interactive 3d application for visualizing spinel group minerals*. *Comput. Geosci.* 48, 50–56.

[Ganuza et al. 2014] Ganuza, M.L., Ferracutti, G., Gargiulo, M.F., Castro, S.M., Bjerg, E.A., Gröller, E., Matkovic, K. (2014) *The spinel explorer - interactive visual analysis of spinel group minerals*. *IEEE TVCG.* 20, 12, 1913–1922.

[Ganuza et al. 2017] Ganuza, M. L., Ferracutti, G., Gargiulo, M. F., Castro, S. M., Bjerg, E. A., Gröller, E., Matkovic, K. (2017) *Interactive visual*

categorization of spinel-group minerals. Proc. of the Spring Conf. on Comput. Graph.

[Luque et al. 2021] Luque, L.E., Ganuza, M.L., Antonini, A.S., Castro, S.M. (2021). *npGLC-Vis Library for Multidimensional Data Visualization*. *En Cloud Computing, Big Data & Emerging Topics: 9th Conference, JCC-BD&ET, La Plata, Argentina, June 22-25, 2021, Proceedings (Vol. 1444, p. 188)*.

[May et. al 2012] May, T., Steiger, M., Davey, J., Kohlhammer, J. *Using signposts for navigation in large graphs*. *Computer Graphics Forum*, 31(3pt2):985–994, 2012.

[Reichstein et al. 2019] Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. (2019) *Deep learning and process understanding for data-driven Earth system science*. *Nature* 566:195–204.

[Sacha et al. 2014] Sacha, D., Stoffel, A., Stoffel, F., Kwon, B.C., Ellis, G. and Keim, D.A., 2014. *Knowledge generation model for visual analytics*. *IEEE TVCG*, 20(12), pp.1604-1613.

[Streit et. al 2012] Streit, M., Schulz, H.-J., Lex, A., Schmalstieg, D., Schumann, H. *Model driven design for the visual analysis of heterogeneous data*. *IEEE TVCG*, 18(6):998–1010.

[Zuo et al. 2021] Zuo R, Wang J, Xiong Y, Wang Z (2021) *The processing methods of geochemical exploration data: past, present, and future*. *Appl Geochem* 132:105072.