

Reproducibilidad en la ciencia : Open science, reproducibility and transparency in ecology (Powers y Hampton,2019)

Juan Carlos Moral Y. - Programa DCG - Herramientas cuantitativas

Mayo 2020

1. Instrucciones

Leer los artículos de la carpeta Reproducibilidad que se encuentra en la plataforma Teams (en la carpeta Class Materials-¿Recursos Bibliográficos) Escoger uno de los artículos en la carpeta Ta3-Reproducibilidad (Enlaces a un sitio externo.) que se encuentra en las carpetas Class Materials-¿M1-Habilidades del grupo Teams de la asignatura. Cada estudiante debe escoger un artículo distinto. Realice lo siguiente:

- Lea el artículo
- Defina el *Why,How,y What* del artículo
- Explique cómo se podrían implementar las ideas del artículo en su área de investigación
- Entregar este análisis en esta Tarea.

2. *Why*

Según Powers y Hampton (2019) los sistemas ecológicos son dinámicos y heterogéneos, por lo que interactúan numerosos factores que determinan su historia natural. Al depender mucho de las variaciones espacio-temporales este tipo de estudios son muy difíciles de reproducir generando amenazas para el proceso científico. Los problemas de reproducibilidad y transparencia afectan la conservación y gestión de los ecosistemas, como también las percepciones que se tienen del campo en estudio. Por ejemplo, las comunidades o ecosistemas observados dentro del mismo año o día pueden afectar los resultados de los estudios. Si bien existen algunas pautas de investigación reproducible como las implementadas por la Ecological Applications en los últimos tres años, la multitud de enfoques en este tipo de estudios genera poco incentivo en los investigadores para querer replicar y transparentar el trabajo de sus colegas.

3. *How*

Mediante la implementación de la reproducibilidad computacional basada en la capacidad de compartir datos, metadatos y códigos de las investigaciones, a partir de pautas estandarizada que

permitan a la comunidad acceder a una fuente de datos públicos. Aquello se puede lograr utilizando herramientas de código abierto (open source) para tareas como la integración de datos, estadística y modelado. Estas herramientas al estar bajo la supervisión de múltiples usuarios permite que accedan a la visualización y canalización de flujos de trabajo. Algunos de estos ejemplos pueden ser *Kepler*, *noWorkflow* para *Python* y *RData Tracker*.

Se propone el análisis de secuencias de comandos, datos y metadatos a partir de plataformas públicas como por ejemplo GitHub o DataONE o instituciones de educación, que sean capaces de estandarizar los datos y generar nuevos repositorios.

4. *What*

La reproducibilidad computacional permite a la comunidad científica y lectores seguir la toma de decisiones de los autores en revisiones cuidadosas de los flujos de trabajo que están llevando a cabo como también un canal de recepción de datos polémicos. Al adoptar el concepto de Open Science se puede compartir los productos del trabajo generando un menor desperdicio de tiempo para revisores. Aquello también fomenta la revisión entre pares y la integración de datos complejos debido a un enfoque heterogéneo como es observado en la Ecología.

Por su parte, la transparencia de datos y el igual acceso permite eliminar brechas sensibles en la población debido a la falta de información, facilitando así la toma de decisiones tanto académica como en el ámbito político. Un investigador que practica el Open Science permite generar el debate de nuevas ideas y promover la colaboración en un contexto donde la ciencia avanza a pasos agigantados.

5. Aplicación en área de investigación

La aplicación de reproducibilidad y transparencia en las Ciencias de la Tierra nos invitan a adoptar el concepto de Open Science. Autores como promueven el uso de código abierto (*FLOSS* in Geosciences) en nuestro campo, adjuntando multiples plataformas de repositorios y software para llevar a cabo esta tarea (ver en Mader y Schenk (2017)).

En nuestra área de investigación por ejemplo se puede utilizar grandes bases de datos procesados por Machine Learning en la discriminación de diferentes entornos tectónicos utilizando datos geoquímicos e isotópicos (Li *et al.* (2015)). Estos datos se pueden extraer directamente de las bases de datos petrológicas y de acceso abierto como *o*, aplicando metodologías de vectores de soporte útiles a la clasificación según Petrelli y Perugini (2016)(Fig.1). En el desarrollo de la Petrología suelen aparecer una infinidad de nuevos instrumentos que puedan medir la presión y temperatura de manera convencional, sin embargo, muchas de estas ecuaciones no consideran las particularidades tanto para rocas metamórficas como ígneas, y el problema se vuelve más agudo incluso entre los diferentes protolitos existentes en el área del metamorfismo, ya que deben incorporar otros factores como la deformación y la acción de fluidos. Un enfoque de regresión lineal aplicado a las múltiples investigaciones geotermobarométricas empíricas, permitirían al investigador replicar fácilmente las condiciones (Fig.1) y facilitar la integración de datos como los observados en el área de la ecología.

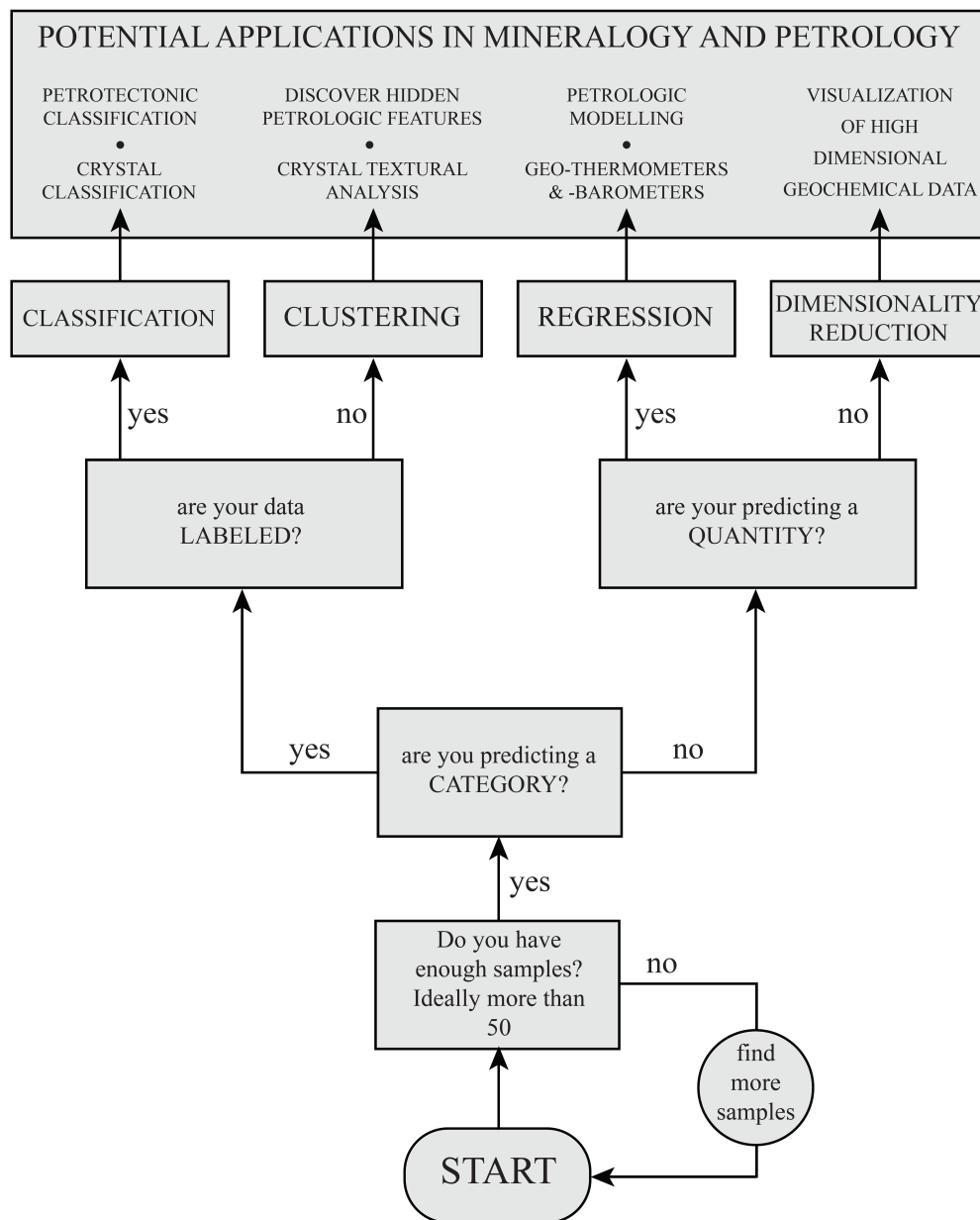


Figura 1: Aplicaciones Machine Learning en Geología. Obtenido de Petrelli y Perugini (2016)

Bajo el concepto de Open Science y reproducibilidad a partir de la publicación de datos, metadatos y códigos, esto podría mejorar. Perfectamente, puede favorecer la aparición de grandes bases de datos que nos permitan diferenciar entre las distintas composiciones químicas y cómo la geotermobarometría convencional varía con respecto a los resultados entregados. Generando una base estandarizada, sería posible la sugerencia a distintos investigadores sobre que tipo de ecuaciones se ajustan mejor a las condiciones de trabajo que ellos presentan, considerando múltiples variables como las mencionadas anteriormente.

En la figura 1 se puede observar las múltiples aplicaciones que se pueden realizar en el campo de la petrología según Petrelli y Perugini (2016). Cabe destacar por ejemplo un análisis de clústeres para texturas de rocas, lo que facilita la determinación de volúmenes de equilibrio o patrones de deformación a través de un tratamiento previo de píxeles de la imagen con software libre como GIMP o ImageJ. Esto puede permitir establecer condiciones estándar para cierto tipo de composiciones de roca, que por lo general quedan a criterio del investigador estableciendo sistemas cerrados y poco representativos de procesos geológicos que ocurren a grandes variaciones espacio-temporales. De igual manera, se puede generar una retroalimentación de los algoritmos utilizados entre los investigadores para la clasificación y engrosar así la base de datos como parte fundamental para integrar datos a futuro y enfrentar problemas complejos de gran variación a escala espacio-temporal.

Referencias

- Li, C., Arndt, N. T., Tang, Q., y Ripley, E. M. (2015). Trace element indiscrimination diagrams. *Lithos*, 232:76–83.
- Mader, D. y Schenk, B. (2017). Using Free / Libre and Open Source Software in the Geological Sciences. *Austrian Journal of Earth Sciences Vienna*, 110(Url 3).
- Petrelli, M. y Perugini, D. (2016). Solving petrological problems through machine learning : the study case of tectonic discrimination using geochemical and isotopic data. *Contributions to Mineralogy and Petrology*, (81):171.
- Powers, S. y Hampton, S. (2019). Open science , reproducibility , and transparency in ecology. *Ecological Applications*, 29(1):1–8.