EXPLORACIÓN DE MÉTODOS ESTADÍSTICOS EN GOOGLE EARTH ENGINE PARA DIFERENTES APLICACIONES REGIONALES

Alejandro Tobón, Santiago Vergara, Julio E. Cañón, Paola A. Arias y Óscar D. Álvarez-Villa

Facultad de Ingeniería, Universidad de Antioquia UdeA, Colombia.
E-mail: alejandro.tobon2@udea.edu.co, santiago.vergara@udea.edu.co, julio.canon@udea.edu.co, paola.arias@udea.edu.co, odavid.alvaez@udea.edu.co

Introducción

Atendiendo a la necesidad de romper las barreras existentes entre la información ambiental disponible a nivel mundial y los usuarios que la requieren, múltiples entidades globales se han dado a la tarea de desarrollar herramientas que faciliten el acercamiento a los datos producidos históricamente. Tal es el caso de Google Earth Engine (GEE), una plataforma científica desarrollada por Google para facilitar el manejo y procesamiento de enormes volúmenes de información satelital, tanto para aplicaciones científicas como académicas. Por medio de esta herramienta se logra el acceso a recursos computacionales de alto desempeño para el procesamiento de un gran número de bases de datos provenientes de información obtenida con sensores remotos que están siendo actualizadas constantemente (Gorelick et al., 2017). La herramienta permite a usuarios (principalmente investigadores en diferentes campos de las geociencias) realizar diferentes análisis aún sin poseer grandes habilidades en programación. Una de las bondades de la plataforma es la considerable reducción de los tiempos de procesamiento de información, por lo que los estudios desarrollados a partir de ésta apuntan a diferentes aplicaciones a escala global. Algunos trabajos desarrollados en los últimos años muestran las diferentes disciplinas que interactúan con la herramienta: cambio global de cobertura tipo bosque (Hansen et al., 2013), identificación de zonas con afectaciones por malaria (Sturrock et al., 2014), mapeo de zonas urbanas (Zhang et al., 2015), identificación de zonas de inundación (Coltin et al., 2016), identificación de cambios en el uso del suelo (Collect Earth, 2016), entre otros.

A partir de lo anterior, se identifica la oportunidad de explorar GEE como una herramienta para desarrollar estudios rápidos con información actualizada, incorporando el análisis geoestadístico de las diferentes variables físicas del sistema que se analice y aprovechando todas las bondades computacionales y de acceso a información remota. En este trabajo, presentamos una "caja de herramientas" estadísticas a partir de las cuales es posible caracterizar diferentes variables de interés y realizar análisis morfológicos, hidrológicos e hidráulicos sobre territorios regionales a escala de país, tomando como ejemplo el caso colombiano.

Desarrollo de herramientas

Desde la implementación de la plataforma, un gran número de usuarios se han enfocado en el desarrollo de aplicaciones con fines diversos. De igual forma, los desarrolladores de GEE han implementado herramientas que facilitan el procesamiento de los datos, cuyos códigos se comparten abiertamente. Algunas de estas herramientas permiten analizar variaciones espacio-temporales a nivel regional, cuyo principal aporte puede ser el de caracterizar de manera ágil el comportamiento físico de variables ambientales. En la actualidad, nuestra caja de herramientas, que adapta códigos disponibles en la plataforma y desarrolla código propio, incorpora componentes del análisis estadístico univariado (medias, medianas, desviaciones estándar, máximos y mínimos), y multivariado (análisis de componentes principales (PCA) y análisis de agrupamiento ("clusters"), mediante los cuales se puede hacer un análisis descriptivo de las diferentes variables de interés.

La **Figura 1** muestra la interfaz de desarrollo de la plataforma GEE.

Figura 1.- Interfaz de desarrollo GEE. Fuente: Google. A) Editor de código. B) Consola de visualización de tablas y gráficos. C) Interfaz de visualización de mapas generados y elementos vectoriales.

Aplicaciones y potencialidades para estudios ambientales

Una de las posibles aplicaciones de las herramientas geoesadísticas desarrolladas consiste en la delimitación y detección de cambios particulares sobre la morfología de cauces a partir de variables obtenidas de análisis de imágenes satelitales.

En la **Figura 2** se muestra la clasificación de clusters del Índice de Vegetación de Diferencia Normalizada (NDVI) sobre una región del río Magadalena, uno de los corredores fluviales más importantes de Colombia. Este método se aplica al mapa medio multianual, y al mapa medio del año 2011, con el objetivo de identificar visualmente desviaciones del comportamiento medio del río bajo el efecto de fenómenos macroclimáticos como el ENSO, muy influyente en el clima colombiano. En esta figura es posible identificar que durante el año 2011 hubo un aumento en promedio del área definida como río a través del método de cluster (color negro), resultado que se asocia con una fase Niña prolongada en 2011, que tuvo mucha influencia sobre el territorio Colombiano con un aumento general en las precipitaciones sobre el área de estudio.

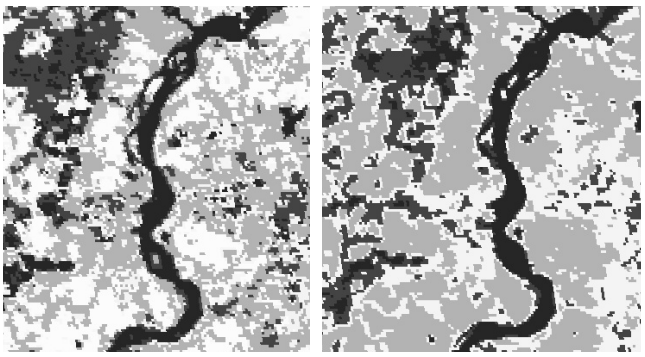

Figura 2.- Clasificación por *clusters* del NDVI para un tramo del río Magdalena, Colombia. Media multianual a la izquierda, media para el 2011 a la derecha.

De igual forma, en la **Figura 3** se muestra la clasificación de clusters para la precipitación media multianual para 15 años de registros pentadiarios, obtenidos a partir de la base de datos de CHIRPS (Climate Hazards Group InfraRed Precipitation with Station Data). De manera rápida se obtiene una clasificación que de cuenta de las diferentes hidroregiones colombianas, marcando zonas con condiciones hidrológicas particulares como la Amazonía (sur), la región Pacífico (occidente), la Guajira (norte), además de la regiónes de la Orinoquía y Andina (oriente y centro del país, respectivamente). También se resalta la identificación de zonas con microclimas particulares, pero de bastante relevancia en términos de precipitación, como la Sierra Nevada de Santa Marta, la cual queda bien delimitada a partir de esta clasificación.

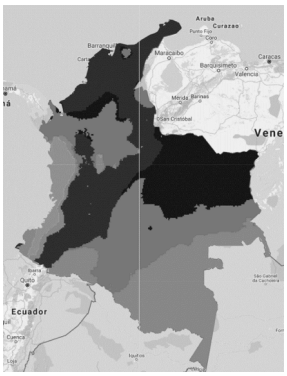

Figura 3.- Clasificación por *clusters* de la precipitación media sobre Colombia para el periodo 2001-2015.

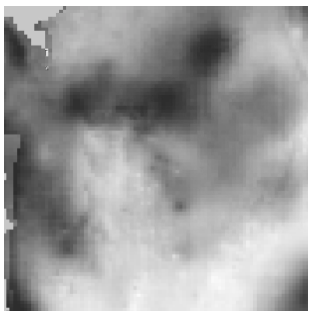

Figura 4.- Primer componente principal de anomalías de precipitación para Colombia (periodo 2001-2002).

La **Figura 4** muestra uno de los resultados asociados a la implementación del análisis de componentes principales. Este desarrollo se implementó para la base de datos de precipitación de CHIRPS, y se encontró que para el periodo 2001–2002, la imagen contiene cerca del 64% de los modos de varianza de la anomalía de precipitación para la región analizada, comprendida entre el Pacífico y el centro de Colombia. Los colores mas oscuros representan las regiones donde mayor ocurrencia de precipitación se registra, coincidiendo en este caso con la regiones Pacífico y Andina, reconocidas por las altas precipitaciones a las que se encuentran sometidas durante todo el año.

Finalmente, otra aplicación desarrollada en la plataforma GEE consiste en la generación análisis estadísticos univariados enfocados en la descripción de regiones a partir de diferentes variables climatológicas. La **Figura 5** muestra los valores medios de precipitación agregados para el territorio colombiano, con datos pentadiarios tomados de la base de datos de CHIRPS, para el periodo comprendido entre los años 2000-2015.

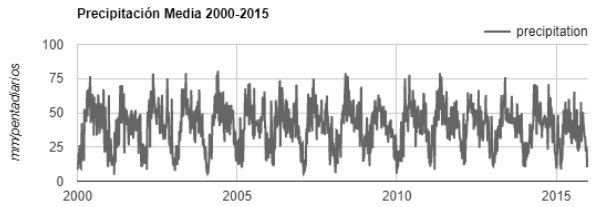

Figura 5.- Precipitación pentadiaria agregada para Colombia entre 2000-2015.

Discusión y conclusión

Los desarrollos presentados dan cuenta de la gran cantidad de información que puede ser procesada en cuestión de pocos segundos para casos en los que se analizan muchos años de información a resoluciones horarias y diarias. El hecho de no tener que descargar la información para su procesamiento facilita la implementación de estos análisis, lo que favorece la manipulación de la información y obvia la necesidad de tener grandes capacidades computacionales, lo que repercute en una ganancia en términos aplicabilidad de las bases de datos satelitales de información ambiental. Entre las limitaciones encontradas destaca la capacidad de la plataforma de integrarse con otros paquetes de análisis, pues entre lo explorado, no se encontró una posibilidad de acoplar herramientas externas a la plataforma. No obstante, se encuentra que los desarrollos realizados dentro de ella son de gran utilidad para los usuarios técnicos en materia ambiental, y para la aplicación de las geociencias a la toma de decisiones. Hemos mostrado en este trabajo el potencial de uso a escala regional, con aplicaciones que abordan temáticas diversas apoyándose en un gran número de bases de datos con información remota abiertas al público, que complementan los trabajos que se desarrollan a escala local. Los tiempos de procesamiento son muy cortos, lo que permite analizar rápidamente la variación espacio-temporal de las variables, sin embargo, al buscar un detalle muy fino, la plataforma presenta limitaciones. La escritura de códigos con estructura basada en pixeles puede resultar útil para diferentes aplicaciones que involucren la distribución espacial de las variables, aunque no necesariamente este tipo de conceptualizaciones sean del interés de todos los desarrollos. Para el caso concreto de los análisis enteramente distribuidos a nivel espacial, una estructura basada en pixeles resulta ideal. Es de resaltar que los productos acá generados consisten en una serie de códigos para la implementación de métodos estadísticos de caracterización general de variables ambientales, y de su aplicación particular en diferentes problemas de la hidráulica y la hidrología convencional, por lo que el usuario final de los mismos estaría habilitado para utilizar las aplicaciones para sus propios intereses, sin importar la región que desee estudiar, e incluso podría utilizar partes de los códigos para generar sus propias aplicaciones. A partir de lo expuesto, se concluye que los productos y las herramientas desarrolladas en GEE dentro del marco de esta investigación son potencialmente utilizables por usuarios técnicos e investigadores, cuyos trabajos se orienten a la producción de conocimiento que permita la gestión integral de los recursos hídricos, especialmente en América Latina.

Referencias

Collect Earth (2016). United Nations Food and Agriculture Organization. http://www.openforis.org/tools/collect-earth.html

Coltin, B., McMichael, S., Smith, T., Fong, T. (2016). "Automatic boosted flood mapping from satellite data". Int. *J. Remote Sens*. 37 (5), 993 –1015

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., et al., (2013). "High-resolution global maps of 21st-century forest cover change". *Science* 342, 850 –853.

Gorelick, N., et al., (2017) "Google Earth Engine: Planetary-scale geospatial analysis for everyone". *Remote Sensing of Environment*. http://dx.doi.org/10.1016/j.rse.2017.06.031

Sturrock, H.J., Cohen, J.M., Keil, P., Tatem, A.J., Le Menach, A., Ntshalintshali, N.E., Hsiang, M.S., Gosling, R.D. (2014). "Fine-scale malaria risk mapping from routine aggregated case data". Malar. J. 13 (1), 1.

Zhang, Q., Li, B., Thau, D., Moore, R. (2015). "Building a better urban picture: combining day and night remote sensing imagery". *Remote Sens*. 7 (9), 11887 –11913.