

## Google Earth Engine (GEE): una poderosa herramienta que vincula el potencial de los datos masivos y la eficacia del procesamiento en la nube

### *Google Earth Engine - GEE: A Powerful Tool Linking the Potential of Massive Data and the Efficiency of Cloud Processing*

Gabriel Alejandro Perilla\* y Jean-François Mas\*\*

Recibido: 13/06/2019. Aprobado: 21/06/2019. Publicado en línea: 11/03/2020.

**Resumen.** Desde hace décadas se están produciendo imágenes satelitales para el mundo entero. Sin embargo, ha habido un rezago tecnológico que ha impedido poder analizar la extensa cantidad de información satelital disponible. Hasta fechas recientes la mayoría de los estudios se enfocaban en regiones relativamente reducidas y periodos cortos. No obstante, el potencial de estas imágenes permite afrontar problemas a escalas globales y abarcar periodos largos. Google Earth Engine es una herramienta innovadora, que permite realizar análisis geoespaciales usando procesamiento y acervos de datos en la nube. De este modo, se convierte en una opción interesante para solucionar el rezago tecnológico, y así poder examinar las grandes bases de imágenes satelitales.

**Palabras clave:** Google Earth Engine, imágenes de satélite, rezago tecnológico, análisis geoespacial

**Abstract.** Earth Observation Satellites have been acquiring images across the entire world for decades. However, a technological lag has prevented analyzing the vast amount of satellite data currently available. Until recently, most studies focused on relatively small regions and short periods of time, but the potential of this imagery allows addressing issues at global scale and spanning long periods of time. Google Earth Engine is an innovative tool that facilitates geospatial analyses using cloud processing and data archives. It thus provides an interesting alternative to solve the technological lag, and hence be able to analyze the large files of satellite data currently available.

**Key words:** Google Earth Engine, satellite images, technological lag, geospatial analyses

---

\* Pontificia Universidad Javeriana, Bogotá. Facultad de Estudios Ambientales y Rurales, Programa de Ecología. Transv. 4° No. 42-00. Edificio J. Rafael Arboleda, S.J. Piso 8 Bogotá, Colombia. ORCID: 0000-0002-1845-2166. Email: gperilla@javeriana.edu.co

\*\* Laboratorio de Análisis Espacial, Centro de Investigaciones en Geografía Ambiental. Universidad Nacional Autónoma de México, campus Morelia. Antigua Carretera a Pátzcuaro No. 8701, col. Ex-Hacienda de San José de La Huerta. C.P. 58190 Morelia Michoacán México. ORCID: 0000-0002-6138-9879. Email: jfmas@ciga.unam.mx. Autor de correspondencia.

## INTRODUCCIÓN

Actualmente existe una enorme cantidad de información satelital obtenida a escala global, recopilada por diferentes sensores que pertenecen a diferentes proyectos o agencias. Sus datos son de diferente naturaleza como, por ejemplo, observaciones atmosféricas, elevación, así como respuestas de la cubierta de la Tierra en varias porciones del espectro electromagnético, registradas con varias resoluciones espaciales y temporales (Yang, John, Zhao, Lu y Knapp, 2016) produce, sustain, and serve satellite climate data records (CDRs). Recolectados durante varios años, o incluso décadas, conforman un vasto repositorio. Hoy, los datos recolectados por diferentes sensores, como Landsat (U.S Geological Survey, 2018), MODIS (MODIS Characterization Support Team, 2017) o Sentinel (European Space Agency, 2015), se distribuyen de forma gratuita.

Es así que los estudios basados en percepción remota se han visto limitados no por la falta de información, sino, al contrario, por la ausencia de capacidad para analizar todos los datos disponibles (Ma, Wang, Liu y Ranjan, 2015; Yang, Huang, Li, Liu y Hu, 2017). Esto ha dado como resultado investigaciones a escalas locales o regionales o de baja resolución espacial o temporal (Gamba, Du, Juergens, & Maktav, 2011; Ma, Wang, Liu y Ranjan, 2015). En ocasiones, los análisis para el monitoreo de los cambios de cubierta del suelo pueden demorar mucho, de modo que los resultados no son útiles para el apoyo a acciones de fiscalización y de respuesta rápida (Gorelick et al., 2017; Shimabukuro et al., 2011).

Google Earth Engine (GEE, <https://earthengine.google.com>) nació a finales del 2010 como una forma para eliminar esa limitante computacional y permitir a los usuarios realizar análisis complejos, involucrando grandes cantidades de datos y expandiendo la frontera de la ciencia en percepción remota. Esta plataforma de análisis permite procesar información geoespacial en la nube sin necesidad de ocupar la memoria de la computadora del usuario. Al mismo tiempo, GEE está ligado directamente a varios programas satelitales que permiten integrar en las bases de datos las imágenes que acaban de

ser tomadas. GEE es, por lo tanto, una solución innovadora para el manejo de los datos masivos ("Big Data"), que permite afrontar problemas globales, dando resultados velozmente. De esta manera, permite proponer y actuar en soluciones para dichos problemas de forma ágil (Gorelick *et al.*, 2017).

## LA PLATAFORMA

GEE está compuesto por cuatro elementos principales. El primero es la infraestructura de Google, la cual pone a disposición del usuario sus servidores, permitiendo así hacer análisis en paralelo con cerca de 10 000 CPUs. Esto agiliza la velocidad de procesamiento, en comparación con una computadora individual. Por ejemplo, el desarrollo del mapa global sin nubes a 15 m de resolución de Google Earth y Google Maps, usando estos servidores, tardó un par de días, mientras que, si se hubiera usado una computadora personal la demora habría sido de 14 años (Moore, 2017).

El segundo elemento es el acervo de datos (*datasets*). Google tiene almacenado todas las imágenes de varios sensores (Landsat, Sentinel, MODIS, entre otros). Estas bases de datos se actualizan a medida que se toman nuevas imágenes (cerca de 6000 nuevas escenas diarias), creando así un enorme catálogo de datos geoespaciales. Estas bases se pueden consultar a través de diferentes criterios (calidad, localización, fechas) sin necesidad de descargar ni solicitar acceso a las imágenes (Gorelick et al., 2017). Igualmente, GEE permite usar información vectorial, pero a diferencia de las imágenes, los vectores no se encuentran organizados en un catálogo y su búsqueda puede resultar tediosa (Google Developers, 2018).

El tercer elemento es la API (*Application Program Interface*), la cual consiste en una serie de comandos y funciones preestablecidas, escritos en lenguaje JAVA, que permite una programación sencilla al desarrollar algoritmos para las investigaciones. Sin embargo, dichas funciones no pueden modificarse ni tampoco se puede acceder a su código, lo cual impide al usuario realizar modificaciones

para lograr análisis alternativos a los preestablecidos. De esta manera, se limita la flexibilidad de programación del usuario, siendo algo restrictivo ya que solo se permite usar las funciones del API. Cabe mencionar que existe también un API para el lenguaje Python, pero debe ser descargado. Esto limita el procesamiento ya que consume memoria de la computadora del usuario. Además, puede incurrir en cobros de uso y carece de documentos de apoyo y ayudas.

Finalmente, el cuarto elemento es el Code Editor, el cual es un entorno de desarrollo integrado en línea, donde se juntan todos los elementos. Aquí es donde el usuario puede, a través de código de trabajo (“*scripts*”), llamar a los datos, procesar y visualizarlos de manera virtual con los servidores de Google, teniendo así sus resultados e información en la nube (Figura 1).

Cabe destacar que, si bien GEE usa sus propios acervos, también permite que el usuario cargue a la plataforma sus propios datos en formato ráster o vectorial. De la misma forma, a pesar de que el procesamiento se haga en la nube de GEE, existe una función para poder descargar la información generada al almacenamiento Google Drive del usuario.

## CASOS DE ESTUDIO

GEE ha sido utilizado para generar datos geográficos, a menudo a escala global, sobre diferentes temas como el monitoreo de vegetación y bosques, el mapeo de aguas superficiales, la detección de islas de calor, de minas, de incendios, etc. (Kumar y Mutanga, 2018). Brevemente explicaremos algunas de estas aplicaciones.

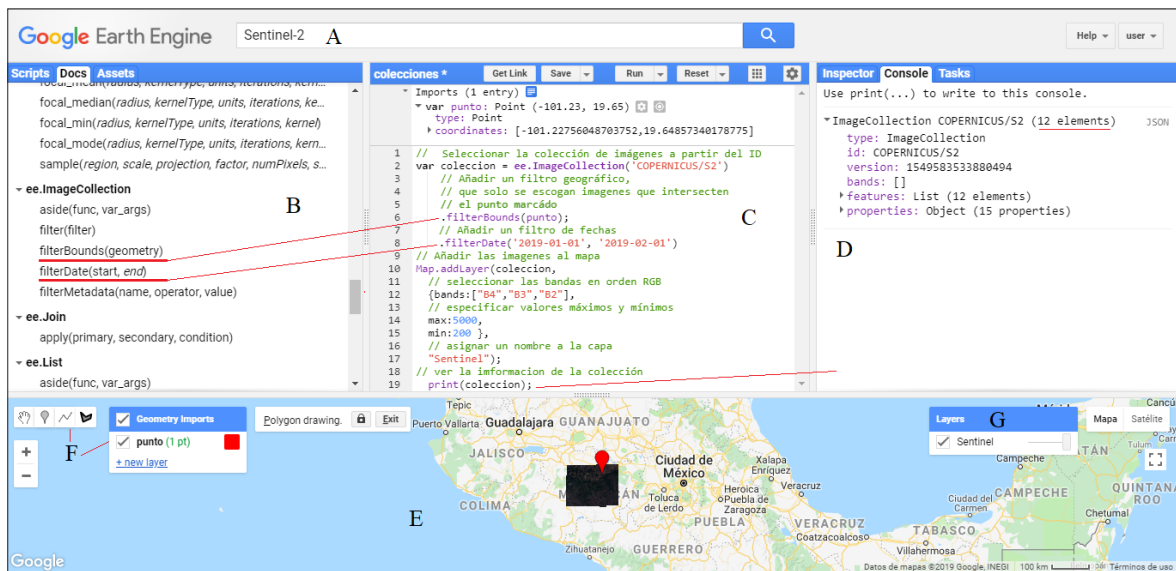


Figura 1. Entorno de trabajo (*code editor*) de GEE y código para obtener imágenes Sentinel-2. A) buscador, donde se puede buscar los datasets requeridos, en este caso se busca el Sentinel-2. B) Panel donde se encuentran las funciones preestablecidas del API. En este caso se señalan dos funciones de filtros que son usadas en el código. C) Panel de programación donde el usuario crea sus algoritmos. En este caso, hay un código para buscar imágenes de Sentinel-2, de enero del 2019 que interceptan con el punto que señala Morelia. Para poder usar dicho código, se usa el ID de Sentinel-2 que se obtiene en el buscador. Además, se especifican los parámetros para visualizar las imágenes. D) Consola donde se muestran solamente los resultados usando la función ‘*print*’ de los análisis que fueron procesados por los servidores de Google. En este caso se muestra la información contenida en la colección del usuario. Se puede ver que para las fechas y zona establecida hay 12 imágenes disponibles. E) Panel de visualización donde se muestran los mapas, ráster o vectores, que el usuario llamó en su código. F) Herramientas de dibujo donde se puede seleccionar una región de interés a través de un punto, una línea o un polígono (filtro geográfico). G) Capas activas en la visualización. Nótese que el nombre asignado a la capa en el código es el nombre que aparece en esta pestaña. Fuente: elaboración propia.

### ***Mapeo de alta resolución de las superficies de aguas globales y sus cambios a largo plazo (Pekel, Cottam, Gorelick y Belward, 2016)***

El Centro Común de Investigación de la Comisión Europea, conocido por sus siglas en inglés como JRC (Joint Research Centre), ha usado GEE para crear mapas de alta resolución de las superficies de aguas en el mundo, lo que permite evidenciar sus cambios, estacionalidad, recurrencia y transiciones (<https://global-surface-water.appspot.com/>).

Para ello se utilizaron más de tres millones de imágenes Landsat tomadas durante un periodo de 30 años (1984-2015) con resolución de 30 metros. Estas fueron procesadas y clasificadas para detectar cuerpos de agua permanentes o estacionales y observar los cambios temporales en dichos cuerpos. Se calcula que una computadora individual habría demorado 1212 años para llevar a cabo este procesamiento, mientras que, con la capacidad de procesamiento paralelo de GEE, se logró en apenas 45 días.

Para alcanzar el objetivo, se utilizaron 64 254 muestras de entrenamiento obtenidas con interpretación visual y repartidas en 9149 imágenes en todo el mundo y para todas las fechas. Estos datos se usaron para entrenar un sistema experto, basado en la información espectral, y variables auxiliares como altimetría para clasificar la totalidad de las imágenes del mundo. Para evaluar el mapa obtenido, se usaron 40 124 puntos de validación repartidos en los 30 años y en todo el mundo, los cuales fueron evaluados visualmente con imágenes de alta resolución. El resultado final arrojó un error de omisión menor a 5% y de comisión menor al 1%.

Estos mapas son de enorme ayuda para poder identificar cuerpos de agua de forma rápida y eficiente y, a su vez, poder entender los cambios de estos en el marco del cambio climático global y diseñar políticas para la seguridad hídrica.

### ***Global Forest Change (Hansen et al., 2013)***

Otra importante base de datos obtenida con GEE es el *Global Forest Change*, <http://earthenginepartners.appspot.com/science-2013-global-forest>, que consiste en una serie de mapas digitales globales, con resolución de 30 m, que para el periodo 2000-2012 indican las áreas forestales, la proporción

de cobertura arbolada y las áreas que presentaron procesos de reforestación (ganancia forestal) y deforestación (pérdida). Recientemente, el periodo de análisis se extendió hasta 2018 para algunos temas (deforestación). Para elaborar la cartografía se analizaron 654 178 imágenes Landsat 7, las cuales fueron remuestreadas, corregidas radiométricamente y filtradas (presencia de nubes) para generar diferentes métricas de series de tiempo, que sirvieron para clasificar las imágenes con un algoritmo de árbol de decisión. De acuerdo con la evaluación realizada por los autores, la base de datos es confiable. Por ejemplo, la clase “pérdida forestal” presenta errores de omisión y comisión de orden de 13%.

### ***Mapbiomas (Mapbiomas Project, 2016)***

A nivel nacional otra iniciativa interesante, basada en GEE, es el proyecto brasileño Mapbiomas (<http://mapbiomas.org/>) que consiste en elaborar cartografía anual de las cubiertas del suelo de los biomas de Brasil para el periodo 1985-2018. La elaboración de los mapas se basa en la clasificación de imágenes Landsat con el algoritmo de árboles de decisión Random Forest. Estas series de tiempo cartográficas permite entender los procesos de cambio a través del análisis de las trayectorias de uso/cubiertas observadas a lo largo del periodo de más de 30 años (Mas et al., 2019).

### ***Mapeo de la agricultura protegida en México (Perilla & Mas, 2019)***

Recientemente, Perilla & Mas (2019) presentaron un mapa de alta resolución de la agricultura protegida en México que emplee plásticos (túneles e invernaderos). Este mapa se obtuvo a través del procesamiento de imágenes de Sentinel-2 y puede usarse para un sistema de monitoreo de la agricultura protegida en México.

## **DISCUSIÓN Y RECOMENDACIONES**

Si bien GEE ofrece una serie de soluciones innovadoras para el análisis de datos masivos, vale destacar que es un proyecto en constante desarrollo (Gorelick et al., 2017). Por ejemplo, el repositorio

de imágenes si bien es muy amplio, no incluye todas las imágenes disponibles. Además, GEE funciona exclusivamente con datos en la proyección geográfica, de modo que no es recomendado hacer, por ejemplo, cálculos de área en GEE. Para calcular medidas espaciales es más fiable usar un sistema de coordenadas métrico. Las funciones existentes son numerosas, pero pueden resultar limitadas para ciertos usuarios. Por ejemplo, no existen funciones para realizar segmentaciones de imágenes o algoritmos de clasificación muy avanzados como los de *deep learning*. En estos casos se tendrán que hacer algunas operaciones fuera de la plataforma.

GEE no consume la memoria de la computadora del usuario, pero es necesario tener una buena conexión a internet para poder cargar el *Code Editor*, las visualizaciones y cualquier información que sea solicitada de los servidores. Adicionalmente, para descargar los resultados de la nube a la computadora del usuario se requerirá espacio disponible suficiente en su Google Drive (la versión gratuita de Google Drive permite hasta 15 gigabytes). En el caso que se pretenda usar GEE con fines lucrativos, o que no estén enfocados en la educación y la investigación, se deberá pagar una licencia (Google Developers, 2018).

A pesar de estas limitaciones, servicios de procesamiento en la red como GEE se utilizan cada vez más. Si bien GEE no es el primer intento de generar una herramienta que permita abordar el Big-Data o intentar procesamiento en la nube (Lin, Chou, Ku, Chung, & Wang, 2018; Yang et al., 2017) si es la primera de este tipo que logra articular efectivamente varios elementos: los acervos de datos, en particular las imágenes de satélite, una infraestructura computacional de gran procesamiento, un libre acceso de los usuarios y una interfaz de uso sencillo. Gracias a esta herramienta se ha incentivado la investigación mundial en percepción remota a escalas globales, que responde a los problemas medioambientales que se enfrenta hoy el mundo. En cierto sentido permitió rebasar una limitación de los datos cartográficos tradicionales que era la disyuntiva entre datos detallados para pequeñas áreas versus grandes extensiones con una escala pequeña de poco detalle. Es ahora posible generar

datos a escala global con una alta resolución (30 m para Landsat). Sin embargo, la calidad de los mapas obtenidos varía sin duda mucho dependiendo de la región, lo cual no se ve reflejado por los ejercicios de evaluación para el conjunto de la base de datos.

En conclusión, por sus novedosas características de procesamiento en la nube, capacidad computacional en paralelo, su amplio catálogo de datos y su interfaz amigable con el usuario, GEE se muestra como una herramienta muy poderosa para afrontar los problemas de Big Data y ampliar los horizontes de la ciencia de datos y la percepción remota. Al mismo tiempo permite que cualquier usuario con conocimientos básicos de programación y acceso a internet pueda realizar investigaciones utilizando grandes cantidades de datos, lo que permite fomentar una comunidad científica y una investigación científica más equitativa y democrática en el mundo (Kumar y Mutanga, 2018).

Así mismo, está revolucionando la elaboración de información geográfica que permite producir datos de alta resolución espacial a escalas globales, algo que hasta hace poco era casi imposible, siendo muy pocos los datos de libre acceso a escala global con una resolución menor a cientos de metros. Estos avances han permitido expandir la frontera hacia una investigación novedosa donde se puede producir información mundial de altísima resolución (<1km), de una forma significativamente más barata, rápida y actualizada que los métodos y prácticas convencionales.

## AGRADECIMIENTOS

Este estudio se inició durante una estancia de investigación en el ámbito del Convenio de Cooperación Internacional entre la Pontificia Universidad Javeriana y el Centro de Investigaciones de Geografía Ambiental (CIGA) de la Universidad Nacional Autónoma de México (UNAM) y se concluyó con el apoyo del proyecto del Programa de Apoyo a Proyectos para la Innovación y Mejoramiento de la Enseñanza (PAPIME) “*Herramientas para la enseñanza de la Geomática con programas de código abierto*” clave PE301919.



## REFERENCIAS

- European Space Agency. (2015). *SENTINEL-2 User Handbook*. Recuperado de <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi>
- Gamba, P., Du, P., Juergens, C., & Maktav, D. (2011). Foreword to the Special Issue on "Human Settlements: A Global Remote Sensing Challenge." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(1), 5–7. <https://doi.org/10.1109/JSTARS.2011.2106332>
- Google Developers. (2018). Get Started with Earth Engine. February 6, 2019, recuperado de <https://developers.google.com/earth-engine/getstarted>. Consultado el 6 de Febrero, 2019.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- MapBiomass Project 2016. The Brazilian Annual Land Use and Land Cover Mapping Project. Recuperado de <http://mapbiomas.org/>.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., & Tyukavina, A. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *SCIENCE*, 342(November), 850–854. <https://doi.org/10.1126/science.1244693>
- Kumar, L., & Mutanga, O. (2018). Google Earth Engine applications since inception: Usage, trends, and potential. *Remote Sensing*, 10(10), 1–15. <https://doi.org/10.3390/rs10101509>
- Lin, F.-C., Chou, T.-Y., Ku, W.-Y., Chung, L.-K., & Wang, C.-J. (2018). Storage and processing of massive remote sensing images using a novel cloud computing platform. *GIScience & Remote Sensing*, 50(3), 322–336. <https://doi.org/10.1080/15481603.2013.810976>
- Ma, Y., Wang, L., Liu, P., & Ranjan, R. (2015). Towards building a data-intensive index for big data computing - A case study of Remote Sensing data processing. *Information Sciences*, 319, 171–188. <https://doi.org/10.1016/j.ins.2014.10.006>
- Mas, J.-F., Nogueira de Vasconcelos, R., Franca-Rocha, W. (2019). Analysis of High Temporal Resolution Land Use/Land Cover Trajectories. *Land*, 8(2), 30. <https://doi.org/10.3390/land8020030>
- MODIS Characterization Support Team. (2017). *MODIS Level 1B Product User's Guide Operation Operation*. Recuperado de <https://mcst.gsfc.nasa.gov/content/11b-documents>
- Moore, R. (Directora de Google Earth, Earth Engine y Earth Outreach). (2017). *Accompanying presentation Earth Engine Users' Summit 2017* (Videoconferencia). Recuperado de <https://youtu.be/5yy1EwtZmhE>
- Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), 418–422. <https://doi.org/10.1038/nature20584>
- Perilla, G.A. & J.-F. Mas (2019). High-resolution mapping of protected agriculture in Mexico, through remote sensing data cloud geoprocessing. *European Journal of Remote Sensing*, 52(1), 532–541. <https://doi.org/10.1080/22797254.2019.1686430>
- Shimabukuro, Y. E., Santos, J., Rudorff, B. F. T., Arai, E., Duarte, V., & Lima, A. (2011). Detección operacional de deforestación y de áreas quemadas en tiempo casi real por medio de imágenes del sensor MODIS. en Mas, J.F. *Aplicaciones de sensor MODIS para el monitoreo del territorio* (pp. 123–143). México: CIGA-INECC. Recuperado de [http://www.ciga.unam.mx/publicaciones/images/abook\\_file/aplicacionesMODIS.pdf](http://www.ciga.unam.mx/publicaciones/images/abook_file/aplicacionesMODIS.pdf)
- U.S Geological Survey. (2018). *LANDSAT 8 (L8) DATA USERS HANDBOOK*. Recuperado de <https://www.usgs.gov/media/files/landsat-8-data-users-handbook>
- Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth*, 10(1), 13–53. <https://doi.org/10.1080/17538947.2016.1239771>
- Yang, W., John, V. O., Zhao, X., Lu, H., & Knapp, K. R. (2016). Satellite climate data records: Development, applications, and societal benefits. *Remote Sensing*, 8(4), 1–24. <https://doi.org/10.3390/rs8040331>