



Sociedad Colombiana
de Ingenieros



Sociedad Tolimense
de Ingenieros

**XXIII Seminario
Nacional
DE HIDRÁULICA E HIDROLOGÍA**

XXIII SEMINARIO NACIONAL DE HIDRÁULICA E HIDROLOGÍA IBAGUÉ, 15 AL 17 DE AGOSTO DE 2018

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

Tobón, A., Vergara, S., Cañón, J. E., Álvarez, O. D., Arias, P. A.

Escuela Ambiental, Facultad de Ingeniería, Universidad de Antioquia. Medellín, Colombia.

alejandro.tobon2@udea.edu.co, santiago.vergara@udea.edu.co, julio.canon@udea.edu.co, odavid.alvarez@udea.edu.co,
paola.arias@udea.edu.co

RESUMEN:

En este trabajo se utilizan las capacidades computacionales que ofrece la plataforma Google Earth Engine (GEE) para acceder y procesar diferentes bases de datos geospaciales, con el fin de desarrollar 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 diferentes escalas, tomando como ejemplo algunas ubicaciones sobre el territorio colombiano. Algunas de estas herramientas permiten analizar variaciones espacio-temporales a nivel regional, cuyo principal aporte puede ser 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, y multivariado y análisis de agrupamiento o “clusters”), mediante los cuales se puede hacer un análisis descriptivo de las diferentes variables de interés. Este trabajo presenta algunos resultados obtenidos a partir de la implementación de diferentes análisis estadísticos sobre puntos geográficos de la región colombiana, abarcando diferentes escalas espaciales y temporales, poniendo a prueba las capacidades de la plataforma GEE.

ABSTRACT:

We explore the computational capacities offered by the Google Earth Engine (GEE) platform to access and process different geospatial databases, in order to develop a statistical "toolbox" to characterize different variables of interest and to perform morphological, hydrological and hydraulic analyzes on regional territories at different scales, taking as an example different locations on the colombian territory. Some of these tools allow the analysis of spatio-temporal variations at regional level, useful to characterize the physical behavior of environmental variables. Currently, our toolbox, which adapts code available on the platform and develops its own code, incorporates components of univariate statistical analysis, and multivariate. This paper presents some results obtained from the implementation of several statistical analyzes on geographic points of the colombian region, covering different spatial and temporal scales and testing the capabilities of the GEE platform.

PALABRAS CLAVES: *Google Earth Engine; análisis estadístico; información remota*

Mayor Información

**Comisión Técnica Permanente
de Ingeniería de Recursos Hídricos**

57 (1) 5550520 logistica@sci.org.co



INTRODUCCIÓN

Google Earth Engine (GEE), una plataforma científica desarrollada por Google para facilitar el manejo y procesamiento de enormes volúmenes de información satelital para aplicaciones científicas y académicas. Con GEE se logra el acceso a recursos computacionales de alto desempeño para el procesamiento de gran número de bases de datos con información de 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, con conocimientos intermedios de 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 apuntan a diferentes aplicaciones a escala global.

Con esta plataforma se facilita además la transmisión de conocimiento entre los diferentes usuarios, permitiendo compartir fácilmente los códigos desarrollados. La interfaz de la plataforma permite acceder a diferentes bases de datos que se actualizan regularmente, tales como Landsat, Sentinel, MODIS (Moderate-Resolution Imaging Spectroradiometer), ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer), bases de datos topográficas como SRTM (Shuttle Radar Topography Mission) y climatológicas como el reanálisis de NCEP (United States National Centers for Environmental Prediction), TRMM (Tropical Rainfall Measuring Mission), CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data), PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) e HydroSHEDS (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales). Las bases de datos a las que permite acceder la herramienta hacen que las aplicaciones desarrolladas a partir de esta plataforma sean diversas (Hansen et al., 2013; Global Forest Watch, 2014; Sturrock et al., 2014; Lobell et al., 2015; Pekel et al., 2016; Collect Earth, 2016; Map of Life, 2016).

Por lo tanto, se identifica la oportunidad de explorar GEE 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 diferentes escalas, tomando como ejemplo algunas ubicaciones sobre el territorio colombiano.

DESARROLLO DE HERRAMIENTAS

Los desarrolladores de GEE y sus usuarios 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 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, valores máximos y mínimos a escala de píxel), y multivariado (análisis de componentes principales y análisis de agrupamiento o “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.

Mayor Información



Figura 1.- Interfaz de desarrollo GEE. 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 geostatísticas 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 Magdalena, uno de los corredores fluviales más importantes de Colombia. Este método se aplica al mapa medio multianual, y al mapa medio del año 2008, con el objetivo de identificar visualmente desviaciones del comportamiento medio del río bajo el efecto de fenómenos macroclimáticos como el fenómeno El Niño-Oscilación del Sur (ENSO), muy influyente en el clima colombiano. En esta figura es posible identificar que durante el año 2008 hubo un aumento en promedio del área pixelada definida como “río” a través del método de cluster (color negro), resultado que se asocia con una fase La Niña en 2008, que tuvo mucha influencia sobre el territorio colombiano, con un aumento general en las precipitaciones sobre el área de estudio. Analizando las diferencias entre la media multianual y lo presentado en el año 2008, se puede observar que, en general, el río Magdalena mantuvo su recorrido respecto a lo presentado históricamente, ya que el mapa de diferencias no muestra un cambio en el cluster que abarca lo asociado a la lámina de agua, en este caso, del río Magdalena.

De igual forma, en la Figura 3 se muestra la clasificación de clusters para la precipitación media multianual para 15 años de registros cada cinco días, obtenidos a partir de la base de datos de CHIRPS (Funk et al, 2015). De manera rápida, se obtiene una clasificación que da cuenta de las diferentes regiones hidroclimáticas a nivel colombiano, marcando zonas con condiciones hidrológicas particulares como la Amazonía (sur), la región Pacífico (occidente), la región de la Orinoquía (noreste), y la región Andina (oriente y centro del país). Este resultado es sobresaliente, si se tiene en cuenta que la definición de los clusters se hace únicamente a partir del procesamiento estadístico de las imágenes, sin la intervención de expertos temáticos.

Mayor Información

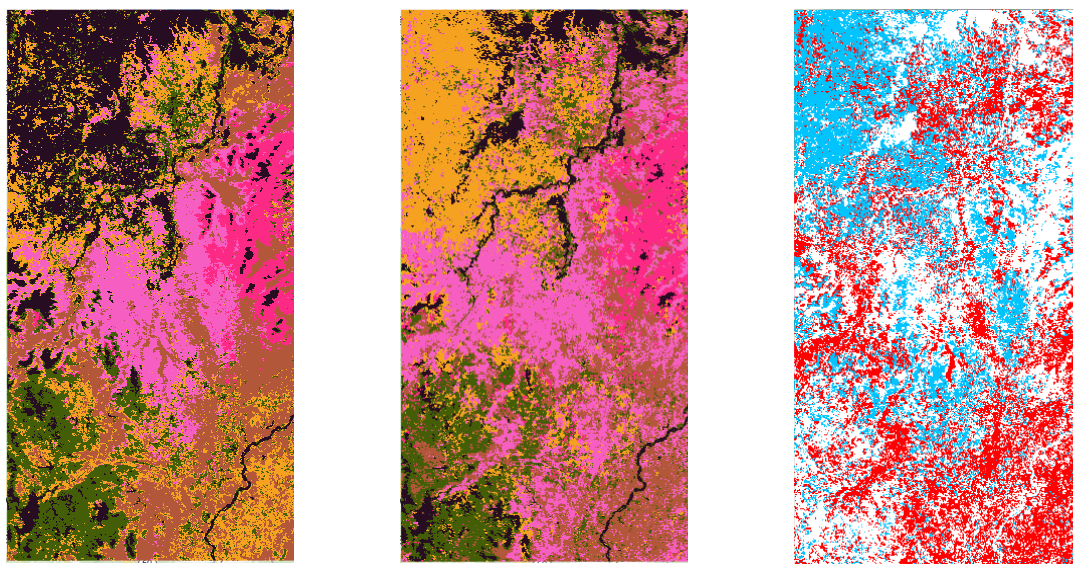


Figura 2.- Clasificación por clusters del NDVI para un tramo del río Magdalena, Colombia. Media multianual (izquierda), media para el año 2008 (medio) y diferencia entre la media multianual y la media del año 2008 (derecha)

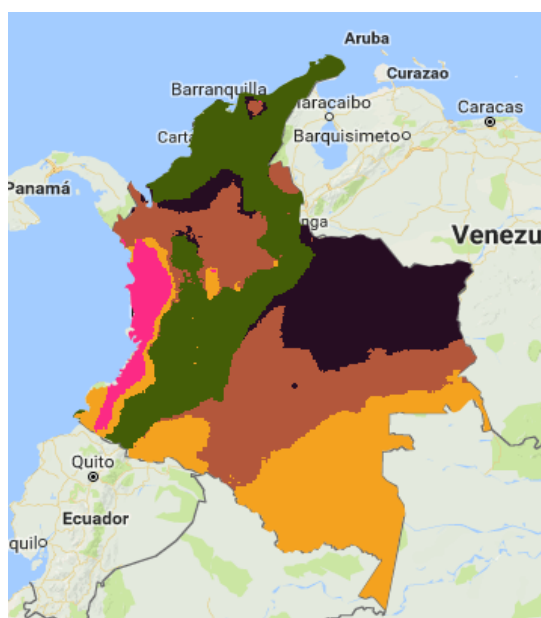


Figura 3.- Clasificación por clusters de la precipitación media sobre Colombia para el periodo 2001-2015

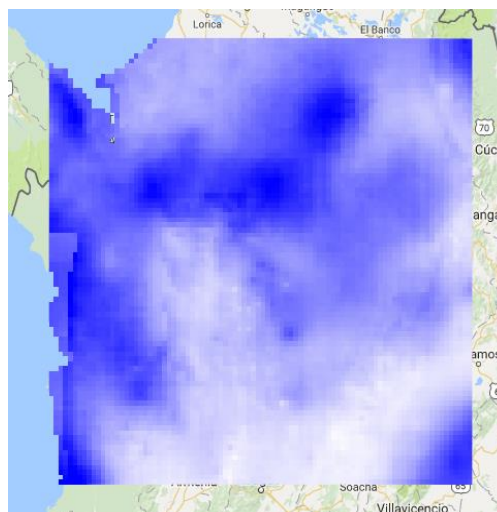


Figura 4.- Primer componente principal de anomalías de precipitación para Colombia (periodo 2001-2002)

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, encontrando 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 más oscuros representan las regiones donde mayor ocurrencia de precipitación, coincidiendo en este caso con las regiones Pacífico y Andina, reconocidas por las altas precipitaciones observadas durante todo el año.

Mayor Información

Las Figuras 5 y 6 muestran parte del código que implementamos para el análisis cluster específico de la Figura 3 y de componentes principales de la Figura 4 respectivamente.

```
Cluster_CHIRPS_Colombia *
1 var util=require('users/alejandrotobon2/default:utilities');
2 var stat=require('users/alejandrotobon2/default:statistics');
3 var collection=ee.ImageCollection('UCSB-CHG/CHIRPS/PENTAD').select('precipitation');
4 print('Original Collection',collection);
5 var region = ee.FeatureCollection('ft:1tdSwUL7Mvp0auSgRzqVTOwdfy17KDbw-1d9omPw')
6   .filter(Metadata('Country', 'equals', 'Colombia')); // Región de trabajo
7 Map.addLayer(region);
8 var scale = 50000; // Escala de trabajo
9 var dis='2001-01-01T00:00:00Z'; // Tiempo de inicio
10 var dfs='2015-12-31T23:59:59Z'; // Tiempo final
11 var delta=1; // Delta de tiempo
12 var unit='month'; // Unidades para el delta de tiempo
13 var fmt='YYYYMM'; // Formato tiempo
14 var converter = 1; // Valor escalar para conversiones de tiempo
15 var mode = 1; // Tipo de agregación (1:sum,2:media)
16 var palette='FEFDFE00,FF000099,FFFF0099,' // Paleta de colores para generación de mapas
17   +'00FF0099,0BCBC299,0000FF99,4C029199';
18   var image = util.ColtoImg(collection, region, dis, dfs, delta, unit, fmt, converter, mode);
19 var means = image.reduce('mean');
20 var training = image.sample({
21   region: region,
22   scale: scale,
23   numPixels: 50000
24 });
25 var clusterer = ee.Clusterer.wekaKMeans(5).train(training)
26 var result = image.cluster(clusterer)
27 Map.addLayer(image.reduce('mean').multiply(ee.Number.constant(12)),{'min': 0, 'max': 6000}, 'Mean')
28 Map.addLayer(result.randomVisualizer(), {}, 'clusters');
```

Figura 5.- Código desarrollado para el análisis cluster a nivel colombiano

```
PCA
1 var util=require('users/alejandrotobon2/default:utilities');
2 var stat=require('users/alejandrotobon2/default:statistics');
3 var collection=ee.ImageCollection('UCSB-CHG/CHIRPS/PENTAD').select('precipitation');
4 var region = ee.Geometry.Rectangle([-77.5, 4.5, -73, 9]); // Región de estudio
5 var scale = 27000; // Escala de trabajo
6 var dis='2001-01-01T00:00:00Z'; // Fecha de inicio
7 var dfs='2002-12-31T23:59:59Z'; // Fecha final
8 var delta=1; // Delta de tiempo
9 var unit='month'; // Unidades para el delta de tiempo
10 var fmt='YYYYMM'; // Formato del tiempo
11 var converter = 1; // Valor escalar para transformaciones del tiempo
12 var mode = 1; // Tipo de agregación (1:sum,2:media)
13 var palette='FEFDFE00,FF000099,FFFF0099,' // Paleta de color
14   +'00FF0099,0BCBC299,0000FF99,4C029199';
15   var image = util.ColtoImg(collection, region, dis, dfs, delta, unit, fmt, converter, mode);
16 print('Original Variable',image);
17 var means = image.reduce('mean');
18 var centered = image.subtract(means);
19 Map.addLayer(centered,{'min': -100, 'max': 100}, 'Anomalias');
20 var pca = stat.PCA(centered, region, scale);
21 var vt = 0.0;
22 var vp = ee.List.repeat(1,pca[2].length().getInfo()[0]);
23 var vt=pca[2].get([0,0]);
24 for(var i=1;pca[2].length().getInfo()[0]-1;i++){vt = vt.add(pca[2].get([i,0]));}
25 for(i=0;pca[2].length().getInfo()[0]-1;i++){var rep=pca[2].get([i,0]).divide(vt).multiply(ee.Number(100));vp=vp.set(i,rep);}
26 print('Principal Components',pca[0]);
27 print('%Variance',vp);
28 print('Eigenvalues',pca[1]);
29 Map.addLayer(image.select('200101'),{'min': 0, 'max': 200, 'palette':palette}, '200101');
30 Map.addLayer(image.reduce('mean').multiply(ee.Number.constant(12)),{'min': 0, 'max': 8000, 'palette':palette}, 'Multianual Mean');
31 Map.addLayer(pca[0],{'min': -1, 'max': 1}, 'PCs');
32 Map.addLayer(pca[0].select('pc4'),{'min': -2, 'max': 2, 'palette':'FF000099,FEFDFE00,0000FF99'}, 'PC 4');
33 Map.addLayer(pca[0].select('pc3'),{'min': -2, 'max': 2, 'palette':'FF000099,FEFDFE00,0000FF99'}, 'PC 3');
34 Map.addLayer(pca[0].select('pc2'),{'min': -2, 'max': 2, 'palette':'FF000099,FEFDFE00,0000FF99'}, 'PC 2');
35 Map.addLayer(pca[0].select('pc1'),{'min': -2, 'max': 2, 'palette':'FF000099,FEFDFE00,0000FF99'}, 'PC 1');
```

Figura 6.- Código desarrollado para el análisis de componentes principales sobre el territorio colombiano

Otra aplicación desarrollada en la plataforma GEE consiste en la generación de análisis estadísticos univariados enfocados en la descripción de regiones a partir de diferentes variables climatológicas. La Figura 7 muestra los valores medios de precipitación agregados para el territorio colombiano, con datos cada cinco días tomados de la base de datos de CHIRPS, para el periodo comprendido entre los años 2000 y 2015.

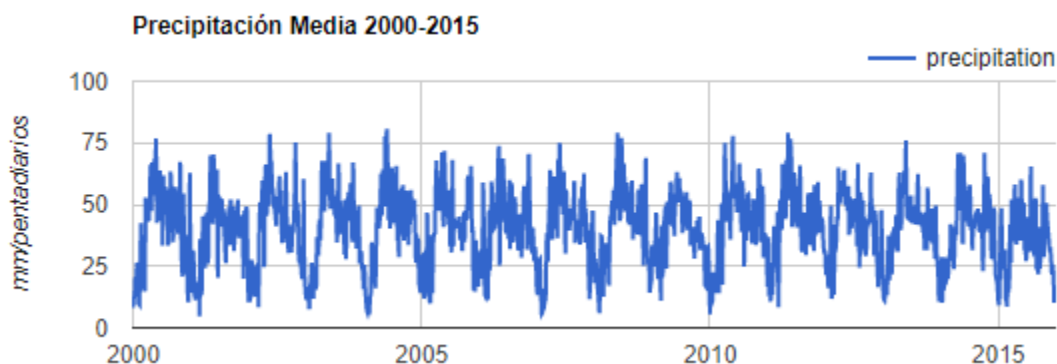


Figura 7.- Precipitación cada cinco días agregada para Colombia entre 2000-2015

Al ser una precipitación agregada sobre todo el territorio, se pierde parcialmente la definición bimodal de algunas regiones, pero se conservan los meses de menor y mayor precipitación. Igualmente, se destaca que no hay evidencia de variaciones significativas ni tendencias en la precipitación multianual promedio en la región para el periodo considerado.

CONSIDERACIONES FINALES

GEE permite el procesamiento de gran cantidad de información en cuestión de pocos segundos, para casos en los que se analizan muchos años de información a resoluciones horarias y diarias. Como no se descargan las imágenes para su procesamiento se facilita el análisis y se elimina la necesidad de tener grandes capacidades computacionales, por lo que se gana en aplicabilidad de las bases de datos satelitales de información ambiental. Entre las limitaciones destacamos la capacidad de la plataforma de integrarse con otros paquetes de análisis, pues no encontramos aún la posibilidad de acoplar herramientas externas a la plataforma. No obstante, los desarrollos realizados son de gran utilidad para los usuarios técnicos en materia ambiental.

En este trabajo resaltamos el uso potencial de la herramienta a escala regional, con aplicaciones que abordan temáticas diversas y complementarias a trabajos que se desarrollan localmente. Los tiempos de procesamiento son cortos, lo que permite analizar rápidamente la variación espacio-temporal de las variables. Sin embargo, en el detalle muy fino, la plataforma presenta limitaciones. La escritura de códigos con estructura basada en píxeles 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 píxeles resulta ideal.

Los códigos de métodos estadísticos de caracterización general de variables ambientales que desarrollamos para este trabajo, aplicables a estudios convencionales de hidráulica e hidrología, se pueden emplear para otras regiones y se pueden modificar según el interés de cada usuario.

Mayor Información



REFERENCIAS

- Collect Earth** (2016). United Nations Food and Agriculture Organization.
<http://www.openforis.org/tools/collect-earth.html>
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... Michaelsen, J.** (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2, 150066. <https://doi.org/10.1038/sdata.2015.66>
- Global Forest Watch**, (2014). World Resources Institute, Washington, DC. <http://www.globalforestwatch.org>
- 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>
- 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.
- Lobell, D., Thau, D., Seifert, C., Engle, E., Little, B.**, (2015). A scalable satellite-based crop yield mapper. *Remote Sens. Environ.* 164, 324–333.
- Map of Life** (2016). <http://www.mol.org>
- Pekel, J.F., Cottam, A., Gorelick, N., Belward, A.S.** (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*.
- 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.

Mayor Información